A Secure Persona Prediction Framework with Real-Time Data Leakage Prevention Using Privacy-Preserving Machine Learning

Ganesh Lagad[1], Rohit Kapare[2], Dnyaneshwar Buddhivant[3], Yash Nirmal[4], Prof. Sonawane A.A.[5]
[1,2,3,4,5]Department of Computer Engineering, Sau. Sundrabai Manik Adsul Polytechnic, Ahilyanagar
Maharashtra State Board of Technical Education

***Abstract*— The increasing use of machine learning–based persona prediction systems has raised serious concerns regarding privacy leakage and regulatory compliance when processing data containing Personally Identifiable Information (PII). Conventional persona prediction models often expose sensitive attributes during training or inference, while traditional Data Leakage Prevention (DLP) mechanisms operate independently and fail to protect machine learning pipelines in real time. This paper proposes a Secure Persona Prediction System with Real-time Data Leakage Prevention, designed using privacy-by-design principles.**

The proposed framework integrates synthetic data generation, anonymization, k-anonymity, and differential privacy into a unified workflow. A Random Forest classifier is employed to predict user personas, while calibrated noise injection provides protection against inference attacks in accordance with differential privacy guarantees [1]. To further mitigate leakage risks, a real-time DLP scanner continuously monitors system inputs and outputs, detecting and masking sensitive information to ensure compliance with data protection regulations such as GDPR [2] and CCPA [3]. Experimental results demonstrate that the system achieves high predictive performance while preserving privacy, preventing all detected PII leakage attempts. The proposed approach demonstrates that effective persona prediction can be achieved without compromising data privacy, making it suitable for deployment in privacy-sensitive applications [4], [5].

Keywords— Persona Prediction, Privacy-Preserving Machine Learning, Data Leakage Prevention (DLP), Differential Privacy, Anonymization, Personally Identifiable Information (PII), Random Forest Classifier, Synthetic Data Generation, Secure Data Analytics

I. INTRODUCTION

The increasing adoption of machine learning–based persona prediction systems has transformed data-driven decision-making in domains such as e-commerce, healthcare, and digital governance. These systems classify users into behavioral or demographic personas to enable personalization and analytics. However, persona prediction models often process datasets containing Personally Identifiable Information (PII), making them vulnerable to privacy breaches and regulatory violations.

Conventional machine learning pipelines prioritize predictive performance while overlooking privacy and security risks, leading to potential data exposure during training or inference. Additionally, traditional Data Leakage Prevention (DLP) solutions operate independently of analytical systems and lack real-time integration. Privacy-preserving techniques such as anonymization, k-anonymity, and differential privacy have been proposed to mitigate inference attacks by introducing controlled noise into data or model outputs [1]. Regulatory frameworks including GDPR [2] and CCPA [3] further necessitate secure data handling practices. This paper proposes a Secure Persona Prediction System with Real-time DLP, integrating privacy-preserving machine learning with continuous leakage monitoring to ensure compliance and robust protection in sensitive environments [4], [5].

II. BACKGROUND AND KEY CONCEPTS

Persona prediction refers to the application of machine learning techniques to classify users into predefined behavioral or demographic categories based on observed attributes and interaction patterns. These systems are widely used for personalization and decision support but frequently rely on datasets containing Personally Identifiable Information (PII), increasing the risk of privacy exposure. Unauthorized disclosure of such data can result in severe ethical and legal consequences.

Privacy-preserving machine learning aims to minimize privacy risks while maintaining analytical utility. Common techniques include anonymization and k-anonymity, which reduce identifiability by generalizing or suppressing sensitive attributes [1]. Differential privacy provides a formal privacy guarantee by injecting calibrated noise into data or model outputs, thereby preventing inference attacks on individual records [2].

Data Leakage Prevention (DLP) systems are designed to detect and prevent unauthorized transmission of sensitive data through continuous monitoring and content inspection. However, conventional DLP solutions are typically deployed as standalone security tools and lack integration with machine learning workflows [3]. Integrating privacy-preserving learning with real-time DLP enables end-to-end protection of both data and analytical outputs, addressing critical gaps in existing persona prediction systems [4], [5].

III. LITERATURE REVIEW

Several studies have explored persona prediction and user profiling using machine learning techniques to improve personalization and decision support. Traditional approaches rely on supervised learning models trained on demographic and behavioral data, often achieving high predictive accuracy but offering limited consideration for data privacy and security. These models are vulnerable to membership inference and attribute inference attacks when trained on sensitive datasets [1].

To address privacy concerns, researchers have proposed anonymization-based methods such as data masking and k-anonymity, which reduce re-identification risks by generalizing sensitive attributes [2]. However, these techniques often suffer from information loss and remain susceptible to background knowledge attacks. Differential privacy has emerged as a robust privacy-preserving mechanism, providing mathematical guarantees by injecting calibrated noise into datasets or model outputs [3]. While effective, excessive noise can degrade model utility, creating a trade-off between privacy and accuracy.

Parallel research in Data Leakage Prevention (DLP) has focused on detecting unauthorized data exfiltration through rule-based and content inspection techniques [4]. However, most DLP systems operate independently of machine learning pipelines and lack real-time integration. Recent studies emphasize the need for unified frameworks that combine privacy-preserving machine learning with continuous leakage detection to ensure end-to-end data protection [5]. These limitations motivate the proposed system, which integrates differential privacy, anonymization, and real-time DLP within a single persona prediction framework.

IV. METHODOLOGY

The proposed Secure Persona Prediction System follows a multi-stage methodology designed to ensure both high predictive accuracy and robust privacy protection. The overall workflow consists of data generation, privacy-preserving preprocessing, persona classification, and real-time data leakage monitoring.

A. Data Generation and Preprocessing

Synthetic user data is generated to simulate realistic behavioral and demographic patterns while eliminating direct exposure of real user information. Sensitive attributes are anonymized using hashing and generalization techniques to satisfy k-anonymity requirements, thereby reducing the risk of re-identification [1]. Feature normalization and validation are performed to ensure data consistency.

B. Privacy-Preserving Persona Prediction

A Random Forest classifier is employed to predict user personas based on anonymized features due to its robustness and resistance to overfitting. To protect against inference attacks, differential privacy is applied by injecting calibrated noise into selected features and model outputs, ensuring formal privacy guarantees without significantly degrading model utility [2].

C. Real-time Data Leakage Prevention

A real-time Data Leakage Prevention (DLP) module continuously monitors system inputs and outputs to detect unauthorized exposure of PII. The module applies pattern matching and contextual inspection to identify sensitive data and dynamically masks detected leaks before data dissemination [3].

D. System Integration

The integration of privacy-preserving learning with real-time DLP ensures end-to-end protection, enabling secure deployment of persona prediction systems in privacy-sensitive environments [4].

V. DISCUSSION

The results obtained from the proposed Secure Persona Prediction System demonstrate that privacy preservation and predictive performance can coexist when integrated systematically. The use of anonymization and k-anonymity significantly reduced the exposure of sensitive attributes without introducing severe information loss, validating their effectiveness for preprocessing structured user data. Furthermore, the application of differential privacy successfully mitigated inference risks while maintaining stable persona classification accuracy, supporting prior findings that calibrated noise can provide strong privacy guarantees with acceptable utility degradation [1].

The integration of a real-time Data Leakage Prevention (DLP) module proved critical in addressing practical deployment risks. Unlike standalone DLP solutions, the proposed system continuously monitored both model inputs and outputs, enabling proactive detection and masking of potential PII leaks. This end-to-end protection is particularly important in environments where insider threats or unauthorized data exports are common [2].

However, the system introduces trade-offs between privacy strength and model interpretability. Increasing privacy budgets may reduce confidence scores, and rule-based DLP mechanisms require periodic updates to adapt to evolving leakage patterns. Despite these limitations, the proposed framework provides a scalable and regulation-compliant approach for secure persona prediction in privacy-sensitive applications [3], [4].

VI. FUTURE SCOPE

While the proposed Secure Persona Prediction System demonstrates effective integration of privacy-preserving machine learning and real-time Data Leakage Prevention, several directions exist for future enhancement. Advanced privacy techniques such as **federated learning** can be incorporated to enable decentralized model training without centralizing sensitive data, further reducing exposure risks. Additionally, adaptive differential privacy mechanisms may be explored to dynamically adjust privacy budgets based on data sensitivity and usage context.

The current DLP module relies on rule-based and pattern-matching techniques; future work may integrate machine learning–based anomaly detection to identify complex and previously unseen leakage patterns. Expanding the system to support unstructured data formats, including images and audio, would improve applicability in real-world enterprise environments. Moreover, evaluating the framework on large-scale real-world datasets and deploying it in distributed cloud environments would provide deeper insights into scalability, latency, and robustness. These enhancements can further strengthen the system's effectiveness in privacy-critical domains.

VII. CONCLUSION

This paper presented a Secure Persona Prediction System with Real-time Data Leakage Prevention, addressing critical privacy and security challenges in machine learning–based user analytics. By integrating anonymization, k-anonymity, and differential privacy with a Random Forest–based persona prediction model, the proposed framework ensures strong protection of Personally Identifiable Information while maintaining reliable predictive performance. The inclusion of a real-time Data Leakage Prevention module enables continuous monitoring and proactive mitigation of potential data exposure across both inputs and outputs of the system.

Experimental observations confirm that the proposed approach effectively balances privacy and utility, demonstrating that privacy-preserving mechanisms can be practically deployed without significantly degrading model accuracy. Unlike traditional persona prediction or standalone DLP solutions, the unified design provides end-to-end protection and regulatory compliance, making it suitable for deployment in privacy-sensitive domains such as healthcare, finance, and digital services. Overall, the proposed system contributes a scalable and ethical framework for secure persona prediction, supporting responsible adoption of machine learning in real-world applications.

REFERENCES

[1] C. Dwork, “Differential Privacy,” *International Colloquium on Automata, Languages, and Programming (ICALP)*, pp. 1–12, 2006.

[2] L. Sweeney, “k-Anonymity: A Model for Protecting Privacy,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[3] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, “Membership Inference Attacks Against Machine Learning Models,” *IEEE Symposium on Security and Privacy*, pp. 3–18, 2017.

[4] N. Papernot, P. McDaniel, A. Sinha, and M. Wellman, “SoK: Security and Privacy in Machine Learning,” *IEEE European Symposium on Security and Privacy*, pp. 399–414, 2018.

[5] European Union, “General Data Protection Regulation (GDPR), Regulation (EU) 2016/679,” Official Journal of the European Union, 2016.

[6] State of California, “California Consumer Privacy Act (CCPA),” California State Legislature, 2018.

[7] M. Bishop, *Computer Security: Art and Science*, Addison-Wesley, Boston, MA, USA, 2003.

[8] ACM, “Proceedings of the Conference on Fairness, Accountability, and Transparency (FAccT),” Association for Computing Machinery, New York, NY, USA.