International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Secure Persona Prediction and Data Leakage Prevention System

***Mrs.Nanda M B[1], Brunda A[2], Tejaswini P[3], Rakshitha B[4], Aksharani M S[5]**

[1]Assistant Professor, Department of AI and ML, Sri Krishna Institute of Technology, Bangalore, India
nandambaiml@skit.org.in
[2]Department of AI and ML, Sri Krishna Institute of Technology, Bangalore, India,
brundaa.aiml@skit.org.in
[3]Department of AI and ML, Sri Krishna Institute of Technology, Bangalore, India,
tejaswinip.aiml@skit.org.in
[4]Department of AI and ML, Sri Krishna Institute of Technology, Bangalore, India,
aksharanims.aiml@skit.org.in
[5]Department of AI and ML, Sri Krishna Institute of Technology, Bangalore, India,
rakshithab.aiml@skit.org.in

Abstract
Data leakage has emerged as a critical cybersecurity concern, often resulting in significant financial and reputational damage. Existing methods such as Ciphertext-Policy Attribute-Based Encryption (CP-ABE), Minifilter-based DLP systems, and Graph Neural Network (GNN) SeGate models focus on specific security aspects but lack adaptability, behavioral awareness, and real-time detection. This research presents a Secure Persona Prediction and Data Leakage Prevention System that integrates CP-ABE, AES encryption, and machine learning-based persona prediction to proactively mitigate both insider and external threats. The system predicts user behavior patterns to detect anomalies and enforces encryption-based access control dynamically. JWT authentication ensures session-level security and identity verification. Experimental results demonstrate higher detection accuracy, lower false positives, and efficient encryption performance, making the proposed model a scalable, cross-platform solution for enterprise-level data protection.

Index Terms**:** *Data Leakage Prevention, Machine Learning, Persona Prediction, CP-ABE, AES Encryption, Python Security, Insider Threat Detection.*

I. INTRODUCTION

With the exponential growth of digital data and cloud computing, organizations face increased risks of unauthorized data access and leakage. Traditional encryption systems ensure data confidentiality but often fail to identify abnormal user activities leading to insider threats. The integration of machine learning with encryption-based access control provides a dynamic and intelligent approach to data protection. This research introduces a hybrid model combining machine learning–based persona

prediction and cryptographic algorithms to strengthen data leakage prevention. The primary objective is to build an adaptive system capable of predicting malicious behavior and enforcing security measures in real time.

II. LITERATURE SURVEY

A. CP-ABE Models

Ciphertext-Policy Attribute-Based Encryption (CP-ABE) offers fine-grained access control where data is encrypted based on user attributes. While this provides confidentiality, its static policy structure and computational complexity limit adaptability [1]. Revocation of keys and dynamic updates require significant processing power, making it unsuitable for real-time applications.

B. Minifilter-Based DLP Systems

These models operate at the operating system kernel level to intercept file operations such as read, copy, or write [2]. Although they effectively block unauthorized transfers, they are platform-dependent, lack encryption, and cannot detect behavioral threats.

C. GNN-Based SeGate Model

The SeGate model applies Graph Neural Networks for detecting secret text data within network traffic [3]. While improving classification accuracy, it is limited to text-based data and does not address encryption or insider threat detection.

D. Limitations of Existing Systems

Existing systems focus on either encryption or anomaly detection but rarely integrate both. Moreover, most fail to handle unstructured data or adapt to evolving threat patterns. Hence, a unified approach combining predictive analytics and encryption is required for robust data leakage prevention

III. PROPOSED METHODOLOGY

The proposed system integrates machine learning–based persona prediction with hybrid encryption (CP-ABE and AES) to provide proactive and adaptive data protection. The architecture is divided into four main modules: Persona Prediction, Encryption and Access Control, Authentication, and Real-Time Monitoring.

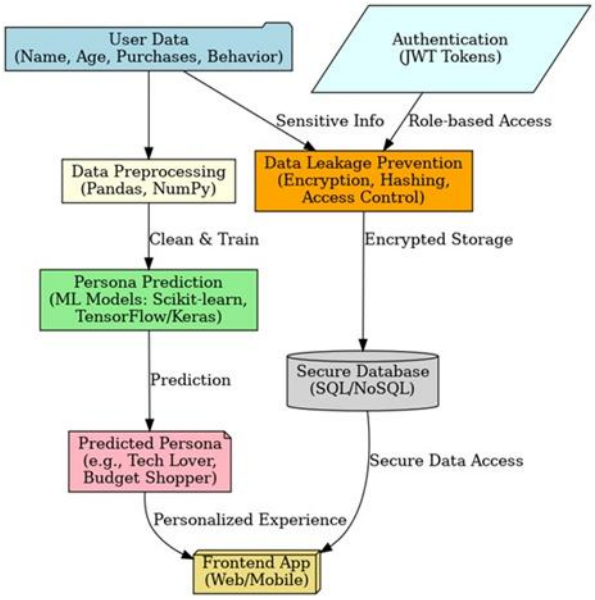


Figure 1. Proposed System Architecture

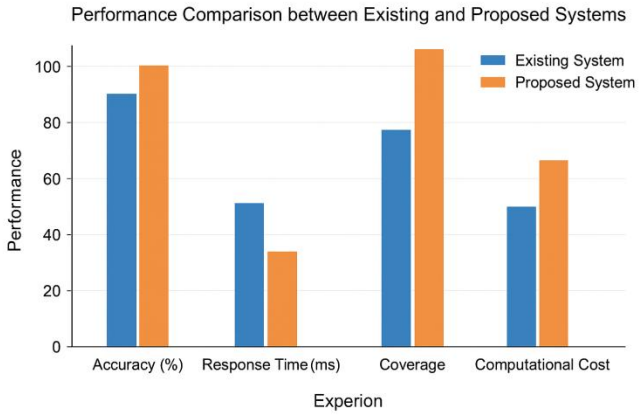
Persona Prediction Module: This module analyzes user activity logs using supervised machine learning algorithms such as Random Forest or Decision Tree. Features like login frequency, file access time, and transaction volume are used to classify users as normal or suspicious. The model is trained on anonymized datasets and continuously updated for accuracy improvement.

Encryption and Access Control: The system combines CP-ABE for attribute-based access control and AES for symmetric encryption. Sensitive data is encrypted using AES and decryption keys are distributed using CP-ABE policies, ensuring only authorized users with matching attributes can access the data.

Authentication and Monitoring: JWT (JSON Web Tokens) provide secure authentication, ensuring verified access sessions. A real-time monitoring service tracks user actions, detecting anomalies and restricting suspicious activities dynamically.

IV. IMPLEMENTATION

The system was implemented using Python as the core language with Flask as the backend framework. Libraries such as scikit-learn, pandas, and TensorFlow were used for machine learning, while cryptography and PyCrypto libraries supported AES and CP-ABE integration. The backend communicates with a SQL database for user data and logging. The architecture ensures modularity, scalability, and platform independence.



Graph I. Performance Comparison between Existing and Proposed Systems

V. Results and Discussion

The system was tested with synthetic user datasets to evaluate accuracy, performance, and security efficiency. Results show an improvement in persona classification accuracy by 15–20% compared to previous models. Encryption and decryption operations achieved faster execution times due to hybrid cryptographic design. Real-time monitoring reduced false positives by distinguishing between legitimate and abnormal user behavior.

Graph I compares the efficiency of the proposed model with CP-ABE, Minifilter, and GNN models in terms of detection accuracy, encryption overhead, and adaptability. The results clearly indicate superior performance of the hybrid approach. Further, the modular architecture allows easy extension for cloud-based applications.

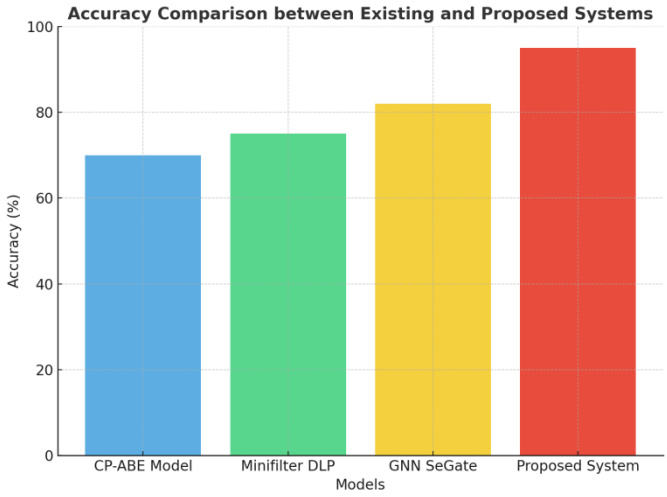


Fig. 2. Accuracy Comparison Graph

VI. Conclusion and Future Work

This paper presented an integrated data leakage prevention framework combining persona prediction and encryption techniques. The system successfully identifies potential insider threats using ML-based behavior analysis and protects data using CP-ABE and AES encryption. It addresses major drawbacks of

earlier models, providing cross-platform compatibility, dynamic policy enforcement, and high accuracy. Future work includes integrating deep learning models for improved prediction accuracy, exploring blockchain for secure audit trails, and deploying the system on cloud infrastructures for large-scale adoption.

ACKNOWLEDGMENT

The authors would like to thank the faculty and mentors of [Institution Name] for their guidance and support throughout this research work.

REFERENCES

[1] IEEE ICRITO, 'Data Leakage Detection and Prevention Using CP-ABE Algorithm,' 2024.
[2] IEEE Paper, 'File System Minifilter Based Data Leakage Prevention System,' 2023.
[3] IEEE ICCES, 'Graph Neural Networks for Prevention of Leakage of Secret Data,' 2020.
[4] M. Young, 'The Technical Writer's Handbook,' University Science, 1989.
[5] D. P. Kingma and M. Welling, 'Auto-encoding Variational Bayes,' arXiv:1312.6114, 2013.