ASSESSING DATA BREACH FACTORS THROUGH MODERN CRIME THEORY: A STRUCTURAL EQUATION MODELING APPROACH

Pedagani Srivani, Dr. J. Reddeppa Reddy
M.Tech Student, Associate Professor
Department of Computer Science and Engineering
Brilliant Institute of Engineering & Technology, Abdullapur (V), Abdullapurmet (M), Rangareddy (D), Hyderabad - 501 513
reddy.mca08@gmail.com

Abstract Research into data security often emphasizes the need to understand the factors linked to security breaches, aiming to prevent future information security incidents. The advancement of digital technology has made safeguarding an organization's sensitive data more complex. Despite the growth of research in data security, there's currently a shortage of studies that specifically investigate the factors contributing to information security breaches in organizations. Previous studies have primarily examined the security posture of companies and organizations, focusing on the breach type and location. However, few studies have explored external factors that may contribute to organizations' vulnerability to information security breaches. The current study addresses this gap in the literature by integrating modern crime theory (MCT) to investigate the exogenous factors influencing the victimization of public and private organizations to data breach incidents. We use insights from crime theories and information about organizations' technical, organizational, and financial aspects to investigate how attractiveness, visibility, and guardianship affect the likelihood of data breaches. We build a theoretical model to explore the relationship between these factors as independent predictors of data breaches. A covariance-based structural equation modeling (CB-SEM) based framework is developed to conduct a comprehensive examination of the dynamics within the context of cybercrime. Through the examination of collected data from 4,868 organizations, this study demonstrates a good fit of the hypothesized model to the data, supporting the validity of the proposed constructs. The results of this study validate the use of MCT in the study of information security breach, and enable the identification of the major exogenous factors influencing data breaches, such as the attractiveness of valuable data and effectiveness of guardianship measures.

INTRODUCTION

Software effort estimation is a necessary component of software development projects that belong to industrial software systems and digital transformation initiatives. Digital transformation refers to the process of integrating digital technology into various components of a company or organization in order to improve operations, procedures, customer experiences, and overall performance. Industrial software systems are trained software packages designed for use in industrial and manufacturing processes. The project deals with the machine learning based effort estimation in order to create an effective and robust model for predicting effort. The project proposes an Omni-Ensemble Learning (OEL) approach, which is a combination of static ensemble selection along with genetic algorithm and dynamic ensemble selection. The project identifies the impact of software effort estimation in industrial software system, and works on the these attributes to implement a robust ensemble model. The proposed Omni-Ensemble Selection (OES)

provides better overall performance (in terms of evaluation metrics) and on comparing with multiple machine learning models over Finnish and Maxwell datasets.

The primary purpose of this study is to provide a comprehensive classification of social engineering (SE) attacks and countermeasures. By analyzing existing taxonomies and identifying gaps in the literature, this research aims to establish a structured and layered approach to categorizing SE threats. The proposed taxonomy serves as a foundation for further research, helping security professionals and organizations develop effective strategies to detect, prevent, and mitigate social engineering incidents. By organizing SE attacks based on environment, approaches, and mediums, this study provides a clear understanding of how these threats manifest and how they can be systematically addressed

This study encompasses a wide range of social engineering attack techniques, including phishing, pretexting, baiting, and impersonation, among others. It considers different environments where these attacks occur, such as corporate settings, social media, and personal communication channels. Additionally, the research examines countermeasures that span both technical solutions, such as email filtering and authentication protocols, and non-technical strategies, such as user training and policy implementation. By covering multiple dimensions of social engineering, this research ensures a broad and in-depth analysis, making it applicable to organizations across industries.

Need for System

With the increasing sophistication of SE attacks, organizations require a structured and systematic approach to identify and mitigate these threats. Traditional security measures focus primarily on technical vulnerabilities, leaving human factors as a weak link. The proposed taxonomy fills this gap by offering a comprehensive classification system that helps organizations understand the various attack vectors and their countermeasures. This system is essential for developing proactive defense mechanisms, improving employee awareness, and enhancing incident response strategies. By adopting a well-defined taxonomy, organizations can strengthen their security posture and reduce the risk of falling victim to social engineering attacks.

Existing System

A considerable research work has been conducted in the field of information security risk (ISR) literature. Li and Li conducted an insightful analysis using CiteSpace-based visualization techniques to map knowledge structures and illuminate the ISR landscape [4]. Despite their valuable contributions, their analysis does not explicitly pinpoint research gaps. On a related note, Mayer et al. delve into the intention behavior gap, shedding light on motivators and obstacles while providing practical insights for interventions [5]. Within regulatory contexts, Ashraf [6] explore the implications of the Securities and Exchange Commission's (SEC) guidance on cyber risk factor disclosure, examining how peer breaches influence the cyber risk disclosures of non-breached firms. Additionally, another study by Bouveret [7], introduces a quantitative framework for cyber risk assessment, providing a holistic perspective that is applied to cross-country data. In [8], authors examine, from a spatio-temporal perspective, the factors and the context associated to data breaches targeting healthcare sector in the United States.

Disadvantages

- The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to detect Assessing Data Breach Factors Through Modern Crime Theory.

- Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.
- Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

Proposed System

The proposed system implemented Structural Equation Modeling in the investigation of crime theory and data breaches is justified by its capacity to comprehensively examine complex relationships among multiple variables concurrently. CBSEM

facilitates the integration of various constructs such as organizational attractiveness, visibility, and guardianship into a unified framework, aligning with the multifaceted nature of the phenomenon under study. By employing CBSEM, we can rigorously test hypotheses derived from crime theory, validating theoretical propositions and empirically assessing the proposed relationships. Furthermore, CB-SEM enables quantitative analysis, allowing for the quantification of the impact of different factors on the likelihood of a data breach, hypothesis testing, and model comparison. Overall, CB-SEM offers a robust statistical approach to understanding the dynamics of crime theory in the context of data breaches, providing empirical support for theoretical frameworks while capturing the intricate interplay of factors influencing organizational vulnerability to data breaches.

Advantages

• We collect multivariate data about victim and non-victim organizations, using web-scraping techniques.

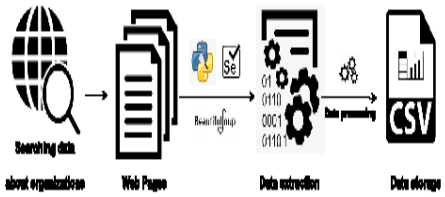
• We design a collection of indicators that capture organizational, financial and technical items.

• We examine the use of crime theories in the study of security breach in cyberspace.

• We propose a theoretical framework that examines the relationship between organization's attractiveness, visibility, and guardianship as independent predictors of data breaches on organizations.

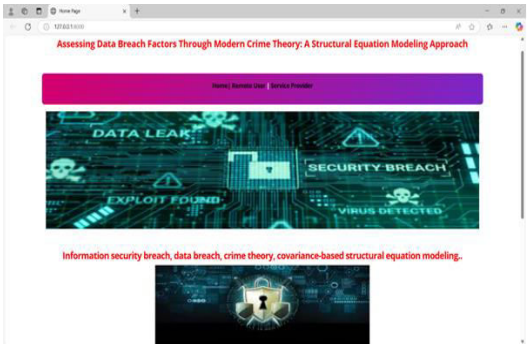
• We examine the use of covariance-based structural equation modeling in the study of factors related to security breach.

• We empirically assess the proposed model's validity and reliability and we highlight the most influential indicators when dealing with information security breaches.

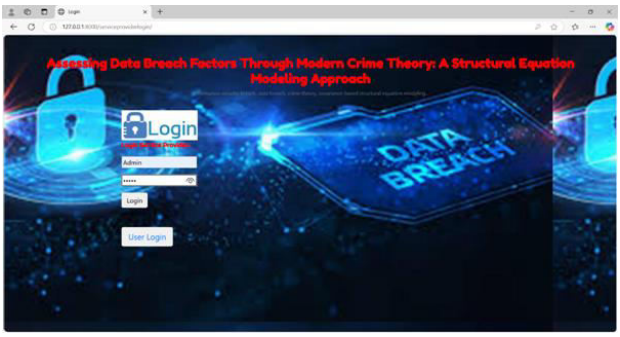


System Architecture

RESULTS AND ANALYSIS

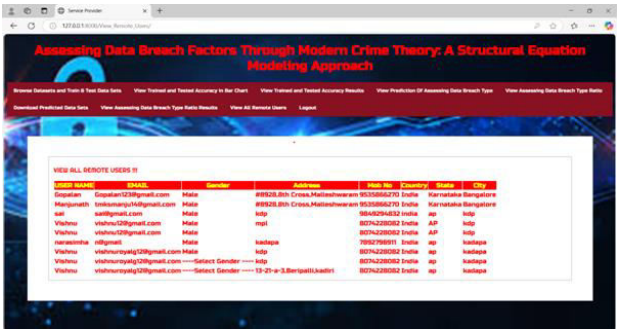


Home page

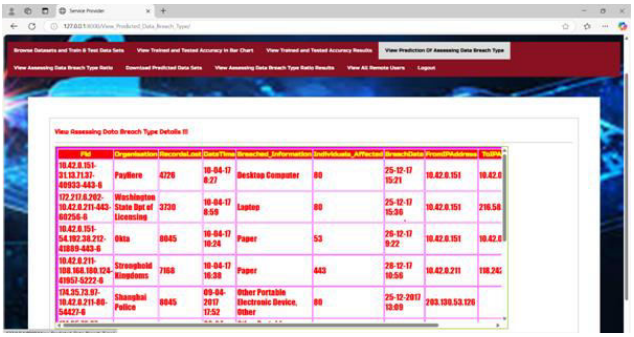


Login page

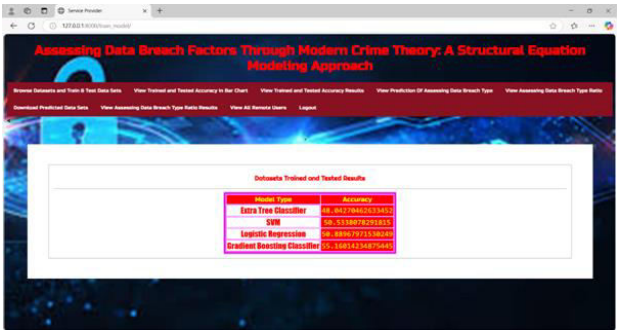
Accuracy Chart in Line Graph



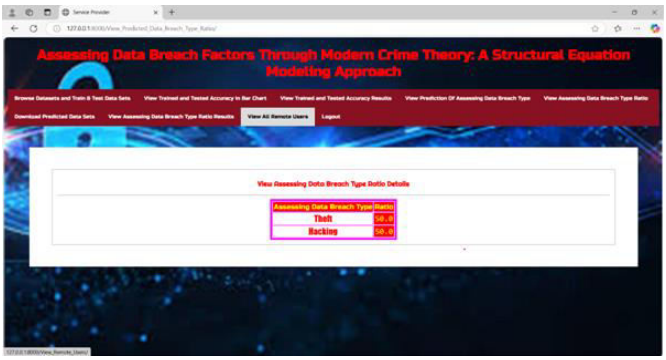
MenuPage



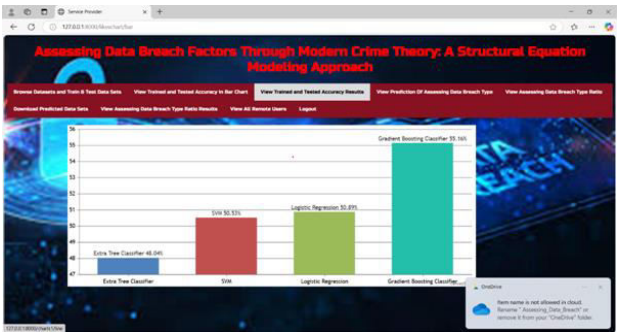
Prediction Value



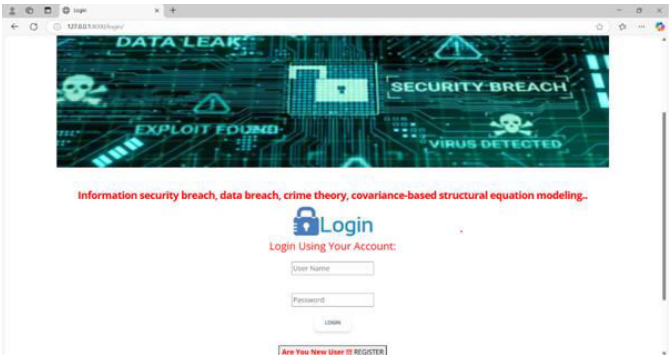
Accuracy Pag



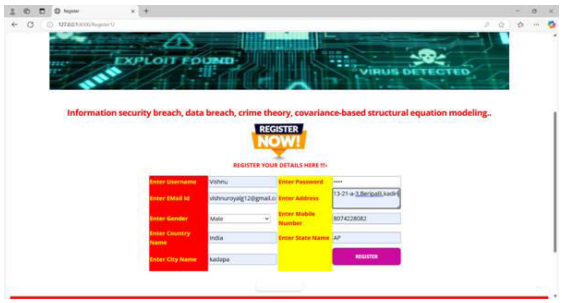
Prediction Ratio



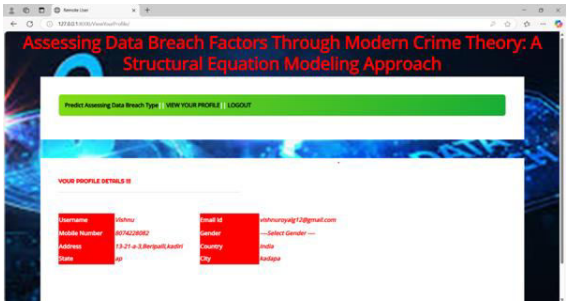
Accuracy Charts in Bar Graph



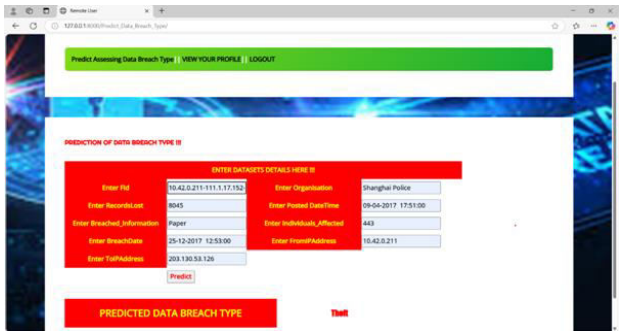
User Login



User Regsiteration



User Home page



User Predclition

CONCLUSION

In this study, we investigated how routine activity theory can help us understand victimization in cyberspace, focusing specifically on data breaches. This theory highlights the importance of factors like target attractiveness, visibility, and guardianship in creating opportunities for cybercrime. We developed a theoretical model to explore how these factors influence the likelihood of data breaches. Using a CBSEM framework, we proposed that these factors positively affect the likelihood of data breaches. CB-SEM allows for a comprehensive analysis of security breaches factors and the underlying constructs of criminal theory, offering valuable insights for developing effective crime prevention strategies. This study seeks to advance the existing literature on cybercrime by incorporating crime theory into the analysis of data breaches. Through this integration, the research aims to enhance our understanding of cyber threats and assist industries in identifying and addressing vulnerabilities within digital systems.

REFERENCES

1. IBM/Ponemon. (2023). Cost of a Data Breach Report. [Online]. Available: https://www.ibm.com/security/data-breach

2. (2024). Thales Data Threat Report. [Online]. Available: https://go.thalesesecurity.com/rs/480-LWA-970/images/2024-DTR-Global-A4-Web-ar.pdf

3. J. Straub, “Evaluating the use of technology readiness levels (TRLs) for cybersecurity systems,” in Proc. IEEE Int. Syst. Conf., Apr. 2021, pp. 1–6.

4. X. Li and H. Li, “A visual analysis of research on information security risk by using CiteSpace,” IEEE Access, vol. 6, pp. 63243–63257, 2018.

5. P. Mayer, Y. Zou, B. M. Lowens, H. A. Dyer, K. Le, F. Schaub, and A. J. Aviv, “Awareness, intention, (In)action: Individuals’ reactions to data breaches,” ACM Trans. Comput.-Hum. Interact., vol. 30, no. 5, pp. 1–53, Oct. 2023.

6. M. Ashraf, Potentially Unintended Consequences of the Sec Restricting Managerial Discretion: Evidence From Peer

Data Breaches and Cyber Risk Factors, document SSRN 3807487, 2021.

7. A. Bouveret, “Cyber risk for the financial sector: A framework for quantitative assessment,” IMF Work. Papers, vol. 18, no. 143, p. 1, 2018.

8. N. Nejjari, K. Zkik, and H. Benbrahim, “The breach is dead, long live the breach: A spatial temporal study of healthcare data breaches,” in Proc. Int. Conf. Sci., Eng. Manag. Inf. Technol., 2022, pp. 287–303.

9. M. Barati and B. Yankson, “Predicting the occurrence of a data breach,” Int. J. Inf. Manage. Data Insights, vol. 2, no. 2, Nov. 2022, Art. no. 100128.

10. Z. Fang, M. Xu, S. Xu, and T. Hu, “A framework for predicting data breach risk: Leveraging dependence to cope with sparsity,” IEEE Trans. Inf. Forensics Security, vol. 16, pp. 2186–2201, 2021.

11. H. Sun, M. Xu, and P. Zhao, “Modeling malicious hacking data breach risks,” North Amer. Actuarial J., vol. 25, no. 4, pp. 484–502, Oct. 2021.

12. X. Zhang and X. Chen, “Research on breach prediction for big data through hybrid ensemble learning and logistic regression,” J. Phys. Conf. Ser., vol. 1982, no. 1, Jul. 2021, Art. no. 012049.