A COMPREHENSIVE APPROACH TO PREVENTING DATA LEAKAGE AND STRENGTHENING CYBERSECURITY

TMP-2023-24-082

PROPOSAL PROJECT REPORT

(UNVEILING PATTERNS AND ANOMALIES TO MITIGATE DATA BREACH RISKS)

S.K. Keshani – IT20299002

B.Sc. (Hons) in Information Technology Specializing in Cyber Security
Department of Computer System and Engineering

Sri Lanka Institute of Information Technology

August 2023

A COMPREHENSIVE APPROACH TO PREVENTING DATA LEAKAGE AND STRENGTHENING CYBERSECURITY

TMP-2023-24-082

PROPOSAL PROJECT REPORT

(UNVEILING PATTERNS AND ANOMALIES TO MITIGATE DATA BREACH RISKS)

S.K. Keshani – IT20299002

Supervisor – Mr. Amila Senarathne
Co – Supervisor -

B.Sc. (Hons) in Information Technology Specializing in Cyber Security
Department of Computer System and Engineering

Sri Lanka Institute of Information Technology

August 2023

DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature	Date
S.K. KESHANI	IT20299002	Keshani	25-08-2023

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

Signature of the Supervisor Date

....

Signature of the Co-Supervisor Date

....

ABSTRACT

In the current digital environment, data breaches are increasingly widespread and provide a danger to companies. Sensitive and confidential data are a requisite for most companies, so protection for this data takes great attention by top management of a company, administrators, and IT managers. Data leakage causes negative impact on companies. The traditional security approaches, such as firewalls, can't protect data from leakage. Data leakage/loss prevention (DLP) systems are solutions that protect sensitive data from being in non-trusted hands. Because of the potential threat of data breaches, it is necessary to implement innovative strategies to strengthen electronic security measures in a world that is becoming more and more digital. This research project examines a comprehensive strategy for minimizing data loss and enhancing cybersecurity by fusing advanced technology and analytical approaches. The project focuses on the area of cybersecurity and IT infrastructure, tackling the essential challenge of safeguarding sensitive data in a complex digital environment.

Four different subcomponents of the research make up a complex entire. The first component, "Protection of Systems by Identifying In habitual User Models," makes use of machine learning algorithms to find abnormalities in user behaviour and access model design. With this proactive strategy, it is possible to identify suspicious activity and take immediate action to prevent unauthorized access or harmful intent. "Utilizing NLP Techniques for Enhanced Data Protection" is the title of the second part, which examines the use of natural language processing (NLP) algorithms for textual data analysis. This technique recognizes sensitive information using entity identification and sentiment analysis, highlighting possible insider risks while enhancing data safety. Utilizing innovative data analysis tools, the third subcomponent, "Unveiling Patterns and Anomalies to Mitigate Data Breach Risks," reveals hidden patterns and anomalies. The potential for violation will be minimized through the use of machine learning algorithms and statistics, allowing organizations to implement proactive security measures. Last but not least, the fourth subcomponent, "Malicious Image Detection and Classification Using Deep Learning Techniques," deals with preventing cyberattacks based on images. Convolutional neural networks are used to generate a deep learning model that can distinguish between legitimate and fake images, making the internet safer.

Expertise and specialized knowledge played a significant role in the organization of this research organization. The project, which places a focus on cyber security, calls for an understanding of the details of anomaly detection, natural language processing algorithms, data analysis methods, and in-depth training methodologies. The importance of the research is also improved by relevant data and real-world cybersecurity scenarios. The combination of these elements results in a comprehensive solution that advances the theoretical foundations of cybersecurity while providing businesses with practical tools to safeguard their critical information. By incorporating modern technologies and analytical frameworks, our study strengthens the resiliency of digital networks against a constantly changing landscape of cyber threats, adding to the continuing conversation on data protection and cybersecurity

TABLE OF CONTENTS

1. INTRODUCTION

The ability to detect patterns and anomalies in large volumes of data can help mitigate the risks of data leakage and cyber-attacks. The data breaches are becoming increasingly common and pose a significant threat to businesses of all sizes. The consequences of a data breach can be devastating, ranging from financial loss to damage to a company's reputation. In this paper, we will explore the importance of identifying patterns and anomalies to mitigate data breach risks. By analyzing data for unusual patterns and trends, we can detect potential security threats before they become major issues. Machine learning algorithms and data visualization techniques can help us identify these patterns and anomalies quickly and accurately [1]. This can lead to early detection and enhanced security, ultimately reducing the risk of data loss and reputational damage for achieving this goal.

This research study aims to dive deeply into the principles of pattern and anomaly detection within this context, lighting the way toward proactive risk reduction. The possibility of identifying security vulnerabilities in their early stages increases by undertaking a thorough review of data for anomalies and unusual trends [2]. This study emphasizes the necessity for an evolution in data protection measures in an environment where growing cyber threats demonstrate unparalleled complexity.

While existing methods for finding patterns and anomalies have some effectiveness, they frequently can't keep up with the constantly changing specifics of modern cyber threats. In this research, the paper proposes a unique strategy supported by modern data analysis tools as a means of overcoming this gap. These include utilizing the prediction skills of machine learning and the contextual understanding provided by natural language processing [3]. The proposed technique aims to provide businesses with a greater capacity to recognize covert patterns and small anomalies that escape conventional security measures by combining these innovative approaches. The main objectives of the project are to develop a robust system for finding patterns and anomalies in data, reducing data breach risks, and improving cybersecurity [4]. The methodology involves implementing neural detection algorithms, pattern analysis, and scalability while ensuring performance, security, reliability, usability, and response time requirements are met.

The complex goals of developing a strong system for pattern and anomaly detection, significantly lowering the risks of data breaches, and improving cybersecurity measures are fundamental to this undertaking. Modern neural detection algorithms, thorough pattern analysis, and a scalable framework serve as the methodology's core. The strategy unwaveringly supports the principles of performance optimization, strict security guidelines, unfailing dependability, ideal user experience, and quick reaction times [5]. This research explores the complex web of data protection in the pages that follow, outlining a proactive strategy for cybersecurity that combines technology advancement with strong analytical skills. In an environment filled with data breaches, it clarifies the necessity of pattern and anomaly detection while simultaneously offering a paradigm-shifting technique that aims to draw the lines of digital protection.

1.1 BACKGROUND & LITERATURE SURVEY

Researchers and practitioners have both long been interested in the examination of patterns and anomalies in data. Early on, statistical methods were employed to find patterns and anomalies in data. However, the development of machine learning algorithms and advanced data analysis methods has made it feasible to today identify even minor patterns and abnormalities in massive amounts of data.

The amount of research on this issue is extensive and spans a variety of topics, including cybersecurity, banking, healthcare, and more. Rule-based approaches, signature-based detection, and machine learning algorithms are some of the several techniques that researchers have suggested for spotting patterns and abnormalities. When it comes to creating reliable and creative methods for spotting patterns and abnormalities in data, there is still a sizable research gap. Our research fills this gap by using modern data analysis methods and machine learning algorithms to create a reliable system for spotting patterns and abnormalities in data.

- **Why Identify Patterns and Anomalies?**

To discover possible security issues early, it is essential to spot trends and abnormalities in data. Data analysis allows us to spot unusual activity that can point to a breach attempt or illegal access to private data. Additionally, spotting patterns of behaviour enables us to foresee potential security threats and take countermeasures. For example, if we detect a dramatic rise in unsuccessful login attempts, we may investigate the problem and make our authentication processes stronger to guard against a potential data breach. Data loss is the term used to describe a situation in which crucial data is lost to the business, such as during a ransomware attack.

- **Early Detection, Enhanced Security**

Such breaches can have negative consequences, including loss of money and reputational harm. In order to reduce these risks, it is crucial that we find patterns and abnormalities in our data. We may identify possible breaches early on and take proactive steps to strengthen our security by examining our data for anomalous activity and trending behaviour. With this strategy, we can always remain one step ahead of possible dangers and stop them before they can do any damage.

- **Detecting Abnormal Activities**

The capacity to spot abnormal activity in our organization's data is one of the main advantages of spotting patterns and anomalies. By establishing a baseline for typical behavior, any variations may be noted and further studied. By doing this, possible data breaches may be stopped before they happen. It's crucial to keep in mind however that not all anomalous activity is malicious. Genuine users occasionally engage in behavior that differs from the norm for various reasons. To effectively identify actual abnormalities, it is important to have an in-depth understanding of the data and user behavior inside your firm.

- **Identifying Trending Behavior**

A crucial part of reducing the dangers of a data breach is recognizing trending behavior. We can find patterns and trends in massive datasets that provide us with important information about possible weaknesses. This enables us to prevent these risks from becoming problems by making proactive efforts to reduce them.

One illustration of this is seeing patterns in user behavior. We can spot behavioral changes that could be an indication of a possible security problem by tracking user activity over time. One sign of a possible breach can be if a user starts accessing sensitive information unexpectedly outside of usual working hours. We may take precautions to avoid a breach by seeing these patterns early on.

- **Tools for Identifying Patterns and Anomalies**

We can use machine learning algorithms and data visualization approaches to spot patterns and anomalies. On the basis of past data, machine learning algorithms may be trained to quickly identify odd patterns and trends. This can assist businesses in identifying possible security risks before they develop into significant incidents. Techniques for data visualization may also be used to spot patterns and anomalies that aren't always obvious from the data itself.

K-means clustering, SVMs, and neural networks are a few popular machine learning techniques for anomaly identification. These algorithms may be used to spot patterns and trends that are out of the standard. Scattered plots, heat maps, and histograms are examples of visualization techniques that may be used to spot patterns and irregularities in huge datasets. Organizations may better understand their data and proactively reduce the risk of data breaches by using these technologies.

- **Data Breach Prevention Strategies**

User Behavior Analytics (UBA) is one of the best methods for stopping data breaches. UBA is a cybersecurity procedure that examines user activity and finds abnormalities using machine learning techniques. UBA can see possible security issues and prevent them before they do damage by keeping an eye on user activities.Insider threats, which frequently originate from employee carelessness or malicious intent, can be found using UBA. Organizations may immediately spot anomalous activity patterns with UBA and take appropriate action to stop data loss or theft. Additionally, by offering thorough data on user activities, UBA may assist enterprises in meeting regulatory needs.

1.2 RESEARCH GAP

- **Inadequate methods for identifying patterns and anomalies.**

The research gap in identifying patterns and anomalies lies in the inadequacy of current methods. Traditional statistical techniques, rule-based approaches, and basic anomaly detection algorithms are no longer enough to keep up with the ever-evolving landscape of data breaches. Novelty approaches using advanced data analysis techniques such as machine learning, natural language processing, and random forest classifiers have shown promise in detecting abnormalities, but there is still a lack of robust and innovative strategies.

To address this research gap, we aim to develop a system that can quickly and accurately identify meaningful patterns and anomalies in large volumes of data. We will implement advanced data analysis techniques to identify unusual patterns and trends and use machine learning algorithms to detect potential security threats. By doing so, we hope to reduce data breach risks and improve cybersecurity for organizations.

Existing - Traditional Statistical Techniques, Rule-based Approaches, and Basic Anomaly Detection Algorithms.

Novelty – Advanced-Data Analysis Techniques – Machine Learning Algorithms, Natural Language Processing, Random Forest classifiers.

Existing methods for identifying patterns and anomalies in data include traditional statistical techniques, rule-based approaches, and basic anomaly detection algorithms. While these methods have been effective to some extent, they are limited in their ability to handle large volumes of data and complex data structures. To address these limitations, advanced-data analysis techniques have emerged as a novel solution. These techniques include machine learning algorithms, natural language processing (NLP), and random forest classifiers. By leveraging these tools, organizations can better detect abnormal activities and identify trending behaviour, leading to enhanced security and early detection of potential threats.

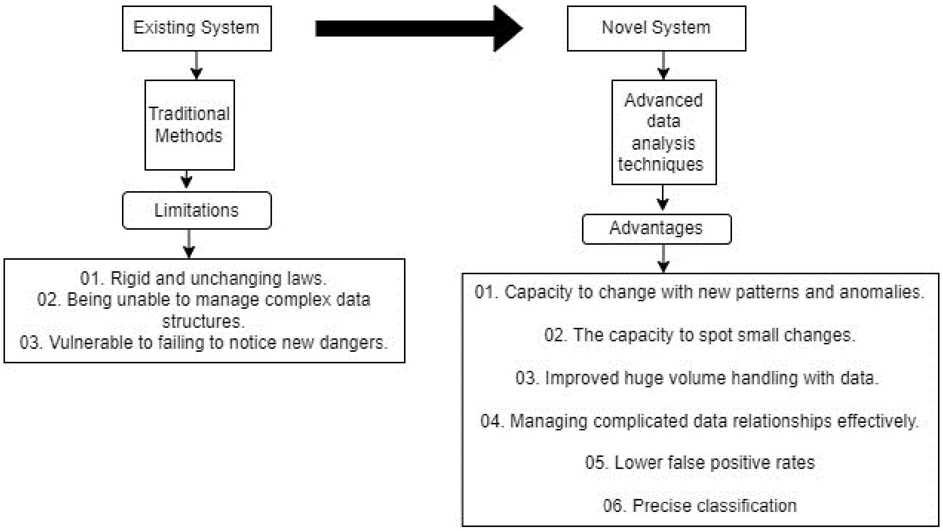

Existing System
Traditional Methods
Limitations
01. Rigid and unchanging laws.
02. Being unable to manage complex data structures.
03. Vulnerable to failing to notice new dangers.
Novel System
Advanced data analysis techniques
Advantages
01. Capacity to change with new patterns and anomalies.
02. The capacity to spot small changes.
03. Improved huge volume handling with data.
04. Managing complicated data relationships effectively.
05. Lower false positive rates
06. Precise classification

Fig 1: Existing System methods vs Novel System techniques.

- **Reliance on outdated approaches and Lack of robust and innovative strategies.**

The current research on identifying patterns and anomalies in data suffers from a significant gap. The existing approaches rely heavily on outdated techniques, which fail to provide robust and innovative strategies to mitigate data breach risks. This gap poses a serious challenge for organizations as they struggle to maintain cybersecurity in the face of an ever-evolving threat landscape.

To address this issue, our research proposes the use of advanced data analysis techniques such as machine learning algorithms and natural language processing (NLP) to identify unusual patterns and trends. By leveraging these techniques, we can detect anomalies and potential security threats with greater accuracy and speed, reducing the risk of data loss and reputational damage. Our approach represents a significant departure from traditional signature-based detection and rule-based systems, which are no longer sufficient in today's complex and dynamic data environments.

Existing - Signature-based Detection and Rule-based Systems

Novelty - Machine Learning and Natural Language Processing, Ensemble Approaches

Existing data breach prevention methods rely heavily on signature-based detection and rule-based systems. While these approaches have been effective in the past, they are becoming increasingly outdated as cybercriminals develop new techniques to evade detection. To stay ahead of the curve, we must embrace novel approaches such as machine learning and natural language processing (NLP), as well as ensemble approaches that combine multiple detection methods.

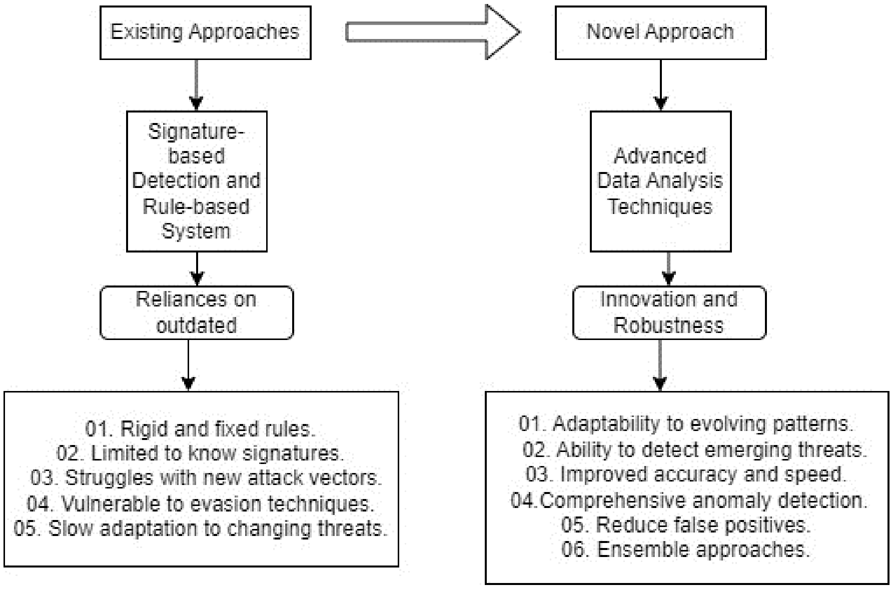

Existing Approaches
Novel Approach
Signature-based Detection and Rule-based System
Advanced Data Analysis Techniques
Reliances on outdated
Innovation and Robustness
01. Rigid and fixed rules.
02. Limited to know signatures.
03. Struggles with new attack vectors.
04. Vulnerable to evasion techniques.
05. Slow adaptation to changing threats.
01. Adaptability to evolving patterns.
02. Ability to detect emerging threats.
03. Improved accuracy and speed.
04.Comprehensive anomaly detection.
05. Reduce false positives.
06. Ensemble approaches.

Fig 2: Existing System Approaches vs Novel System Techniques.

By incorporating machine learning algorithms into our data breach prevention strategies, we can detect anomalies and potential security threats with greater accuracy and speed. NLP can also help us identify unusual patterns and trends in large volumes of data, allowing us to quickly pinpoint potential risks. Ensemble approaches, which combine multiple detection methods, offer an even more robust defines against data breaches and other cybersecurity threats.

1.3 RESEARCH PROBLEM

In today's digital age, the amount of data generated is growing exponentially. With this growth comes the challenge of identifying meaningful patterns and anomalies in large volumes of data. This is particularly important when it comes to cybersecurity and data leakage prevention. The difficulty lies not only in finding these patterns and anomalies but also in doing so quickly and accurately. There is a need for effective tools and strategies that can reduce the risk of data loss and reputational damage.

The research problem we are addressing is the lack of robust and innovative approaches for identifying patterns and anomalies in data. Current methods rely on outdated approaches such as traditional statistical techniques and rule-based systems. While these approaches may have been effective in the past, they are no longer sufficient given the complexity and volume of data being generated. Our objective is to develop a more advanced system that uses machine learning algorithms and natural language processing (NLP) to detect anomalies and potential security threats.

- **Difficulty in identifying meaningful patterns and anomalies in large volumes of data.**

The exponential growth of data has made it increasingly difficult to identify meaningful patterns and anomalies in large volumes of data. With the sheer amount of data being generated every day, traditional methods of data analysis are no longer sufficient. It is important to develop new tools and techniques that can quickly and accurately identify patterns and anomalies in data to mitigate the risk of data breaches.

One approach to addressing this challenge is to use machine learning algorithms to detect anomalous behaviour. These algorithms can analyze large volumes of data and identify patterns that may be indicative of a security threat. By using these advanced techniques, organizations can quickly detect potential security threats and take action to prevent data breaches.

- **Lack of effective tools and strategies for quick and accurate identification.**

The lack of effective tools and strategies for quick and accurate identification is a major challenge in mitigating data breach risks. Traditional statistical techniques and rule-based approaches are often inadequate, while basic anomaly detection algorithms can be unreliable. Novelty approaches such as machine learning and natural language processing (NLP) offer more advanced data analysis techniques, but there is still a need for more robust and innovative strategies.

To address this gap, our research aims to develop a system that uses machine learning algorithms to detect anomalies and potential security threats. We will also implement advanced data analysis techniques for identifying unusual patterns and trends. By doing so, we hope to reduce the risk of data loss and reputational damage caused by cyber-attacks.

- **Need for minimizing the risk of data loss and reputational damage.**

In today's digital age, data is a valuable asset that organizations must protect at all costs. The risk of data loss and reputational damage is a real threat that can have serious consequences for businesses and individuals alike.

Data breaches can result in financial losses, legal liabilities, and damage to an organization's reputation. It is crucial that companies take proactive measures to minimize the risk of data loss and protect their sensitive information from cyber threats.

2. OBJECTIVES

The main objective of this project is to develop a robust system for identifying patterns and anomalies in data, reducing the risk of data breaches, and improving cybersecurity. We aim to achieve this by implementing advanced data analysis techniques such as machine learning algorithms and neural detection algorithms to detect anomalies and potential security threats.

To accomplish these goals, we will collect and analyze large volumes of data and use pattern analysis to identify unusual patterns and trends. Our system will also be scalable to handle increasing amounts of data over time. Additionally, we will ensure that our system meets non-functional requirements such as performance, security, reliability, usability, and response time.

2.1 Main Objective

- **Developing a robust system to find patterns and anomalies in data, reducing data breach risks and improving cybersecurity.**

The main objectives of this project are to develop a robust system that can effectively identify patterns and anomalies in data. By doing so, we aim to reduce the risk of data breaches and improve cybersecurity. With the increasing amount of sensitive data being stored and transferred online, it is crucial that we have effective tools and strategies to prevent data loss and reputational damage.

To achieve these objectives, we will be implementing advanced data analysis techniques for identifying unusual patterns and trends. We will also be using machine learning algorithms to detect anomalies and potential security threats. By combining these approaches, we hope to create a powerful system that can quickly and accurately identify any suspicious activity and mitigate the risks of data leakage.

2.2 Sub-Objectives

Implement advanced data analysis techniques for identifying unusual patterns and trends - These techniques could include exploratory data analysis, clustering, time series analysis, and other statistical methods that can assist in finding hidden patterns that might not be apparent via traditional analysis.

Use machine learning algorithms to detect anomalies and potential security threats - Machine learning techniques, consisting of clustering, classification, and anomaly detection algorithms, can learn from historical data to identify deviations from expected behavior. These algorithms can identify abnormal patterns that might indicate a data breach or cybersecurity threat.

Develop strategies to proactively mitigate data breach risks based on detected patterns - This could involve implementing automated response mechanisms, escalating alerts to security personnel, adjusting access controls, or other actions aimed at preventing or minimizing the impact of a potential data breach.

Use Ensemble Anomaly Detection Framework. - Develop an ensemble approach that combines multiple anomaly detection techniques, enhancing detection accuracy and robustness against varying threat scenarios.

Anomaly Explanation and Visualization - Create a visualization interface that provides detailed explanations for detected anomalies, helping security analysts understand the underlying causes and implications.

Unsupervised Pattern Recognition - Implement unsupervised learning techniques to automatically discover and recognize hidden patterns inside complex datasets, enhancing the system's ability to detect novel threats.

To identify unusual patterns and trends, we will be implementing advanced data analysis techniques. These techniques involve the use of machine learning algorithms that can detect anomalies and potential security threats. By analysing large volumes of data, we can uncover hidden patterns that might not be visible through traditional statistical techniques.

Machine learning algorithms are particularly useful in detecting anomalies because they can learn from historical data and adapt to new situations. This means that as new types of attacks emerge, our system will be able to detect them more quickly and accurately. Additionally, by using data visualization techniques, we can make it easier to identify patterns and trends that might otherwise go unnoticed.

3. METHODOLOGY

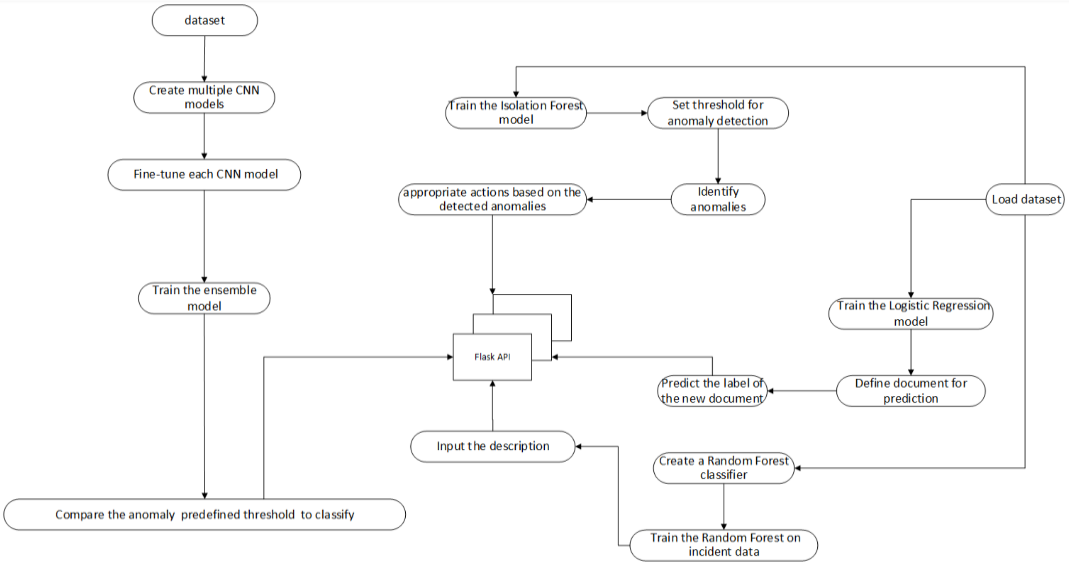
The methodology for the research component "UNVEILING PATTERNS AND ANOMALIES TO MITIGATE DATA BREACH RISKS" involves a systematic approach to enhance cybersecurity measures. It begins with the collection and careful preprocessing of diverse datasets encompassing both normal patterns and potential anomalies. Exploratory data analysis is then conducted to gain insights into data characteristics, followed by the strategic selection of relevant features and attributes to capture intricate relationships within the data.

The chosen machine learning algorithms for anomaly detection, including isolation forests, one-class SVMs, or autoencoders, are then trained on prepared datasets. During training, model parameters are fine-tuned and optimized to ensure optimal performance in identifying both normal and anomalous instances. Subsequently, the trained models are deployed on validation datasets to identify potential anomalies, guided by established threshold values that demarcate between regular and unusual data patterns.

Evaluation metrics, such as precision, recall, F1-score, and receiver operating characteristic (ROC) curves, are employed to carefully assess the performance of the anomaly detection models. Detected anomalies are scrutinized to understand their characteristics and potential implications, distinguishing genuine threats from false positives. The robustness and generalization ability of the models is validated through extensive testing on new and unseen datasets, ensuring their efficacy in real-world scenarios. Ultimately, this comprehensive methodology culminates in a thorough documentation of the entire process, from data preprocessing to model selection, parameter optimization, and evaluation. The insights gleaned from this research will contribute to an advanced understanding of data breach risk mitigation, providing valuable tools to unveil concealed patterns and anomalies that would potentially compromise data security.

System Diagram

Overall System Diagram


dataset
Create multiple CNN models
Fine-tune each CNN model
Train the ensemble model
Compare the anomaly predefined threshold to classify
Train the Isolation Forest model
Set threshold for anomaly detection
Identify anomalies
appropriate actions based on the detected anomalies
Flask API
Input the description
Load dataset
Train the Logistic Regression model
Define document for prediction
Predict the label of the new document
Create a Random Forest classifier
Train the Random Forest on incident data

System Diagram for Individual Research Component

(UNVEILING PATTERNS AND ANOMALIES TO MITIGATE DATA BREACH RISKS)

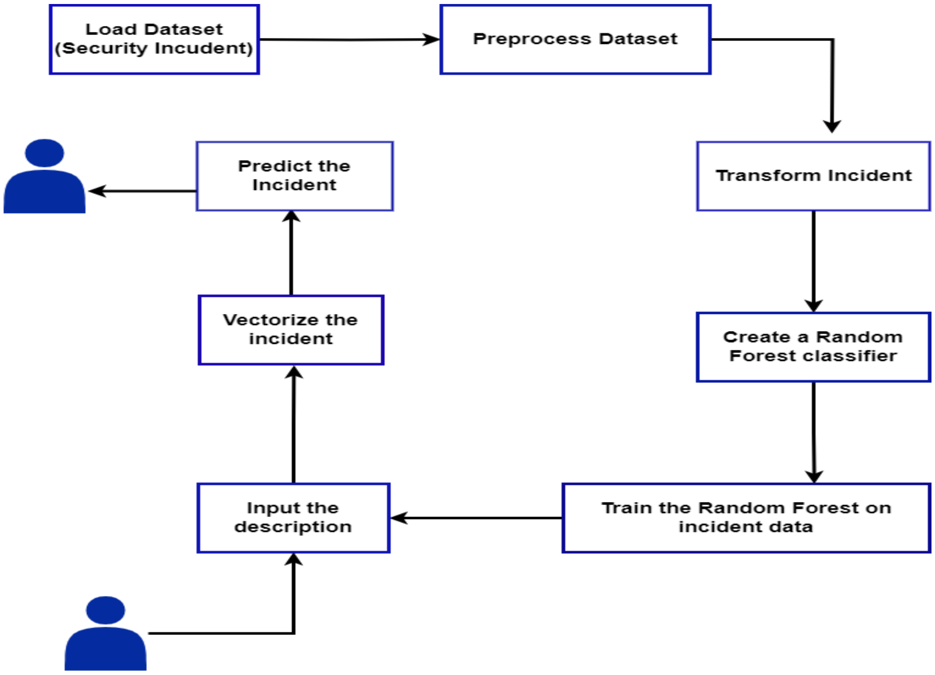

Load Dataset (Security Incudent)
Preprocess Dataset
Transform Incident
Create a Random Forest classifier
Train the Random Forest on incident data
Input the description
Vectorize the incident
Predict the Incident

Figure3: System Diagram for Individual Research Component

Technologies

- Machine Learning
- Neural Network Techniques
- Random forest Classifier model
- Flask API
- Python
- PyCharm

4. PROJECT REQUIREMENTS

The project requirements for unveiling patterns and anomalies to mitigate data breach risks are crucial. The system must have functional requirements such as data collection, neural detection algorithms, pattern analysis, and scalability. Additionally, non-functional requirements like performance, security, reliability, usability, and response time must be met.

To ensure effective data breach prevention and cybersecurity, the system must be able to analyze large volumes of data using advanced analysis techniques and machine learning algorithms. It must also be scalable to accommodate future growth and changing needs. Finally, it must be user-friendly and easy to use for both technical and non-technical users.

4.1 Functional Requirements

- **Data Collection –** Data collection is a crucial functional requirement for identifying patterns and anomalies in data. The system must be able to collect large volumes of data from various sources, including network traffic, user activity logs, and system logs.
- **Nural Detection Algorithms –** Nural detection algorithms are another critical functional requirement for the system. These algorithms use machine learning techniques to analyze data and detect unusual patterns and trends that may indicate potential security threats.
- **Pattern Analysis –** Pattern analysis is also an essential functional requirement. The system must be able to identify meaningful patterns and anomalies in the data, which can help security analysts quickly detect and respond to security incidents.
- **Scalability –** Scalability is the final functional requirement. The system must be able to handle large volumes of data and be scalable to meet the needs of growing organizations

4.2 Non-Functional Requirements

- **Performance:** The system must be able to handle large volumes of data and process them quickly. This requires efficient algorithms and optimized code.
- **Security:** The system must be designed with security in mind, using encryption, access controls, and other measures to protect sensitive data from unauthorized access or theft.
- **Reliability:** The system must be highly reliable, with minimal downtime and the ability to recover quickly from failures. This requires redundancy, fault tolerance, and other measures to ensure continuity of service.
- **Usability:** The system must be easy to use and intuitive, with a user-friendly interface and clear documentation. This requires a focus on user experience and usability testing.
- **Response Time:** The system must respond quickly to user requests, with low latency and high throughput. This requires optimization of network protocols, database queries, and other system components.

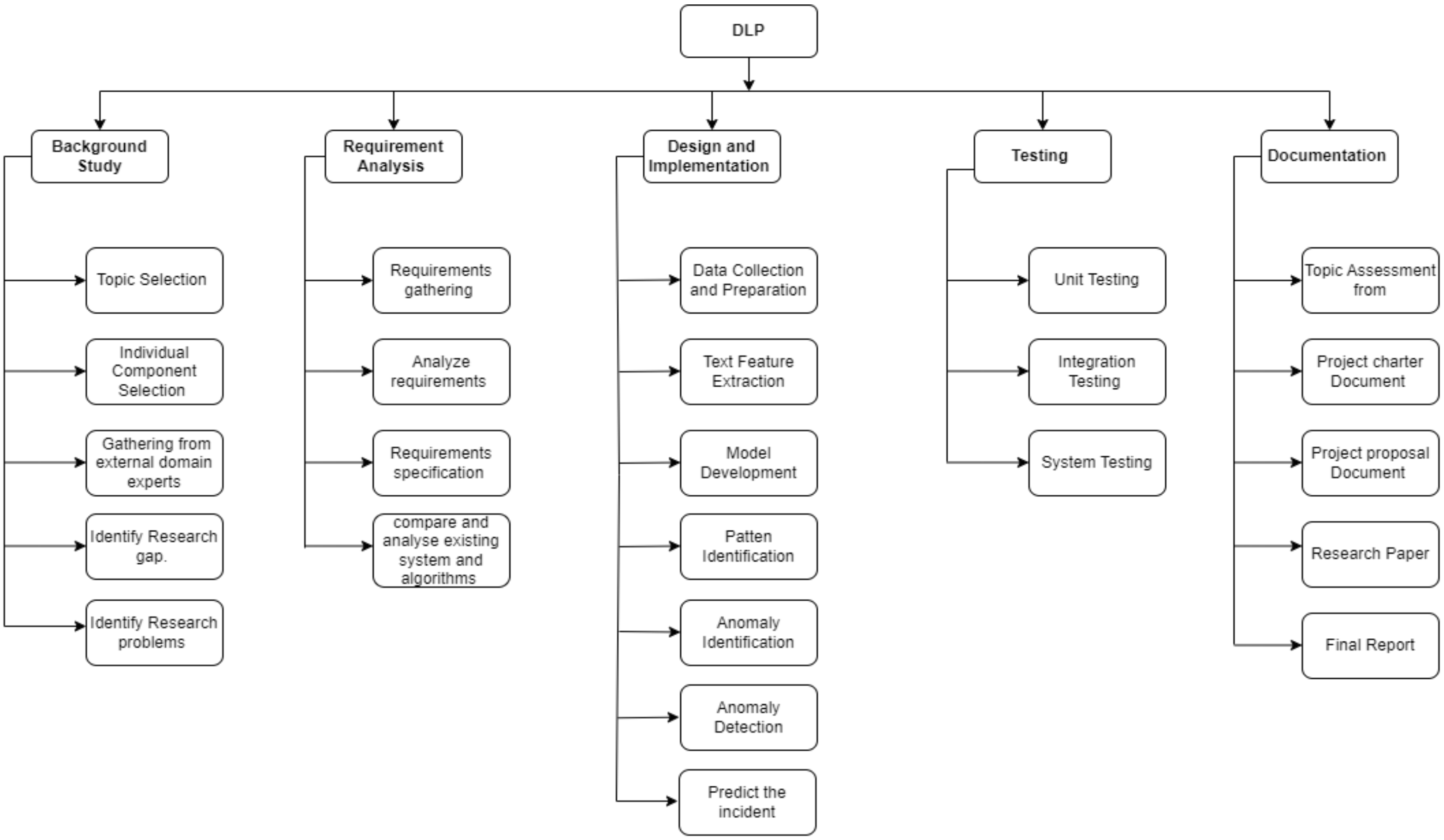
4.3 Software Requirements

- PyCharm

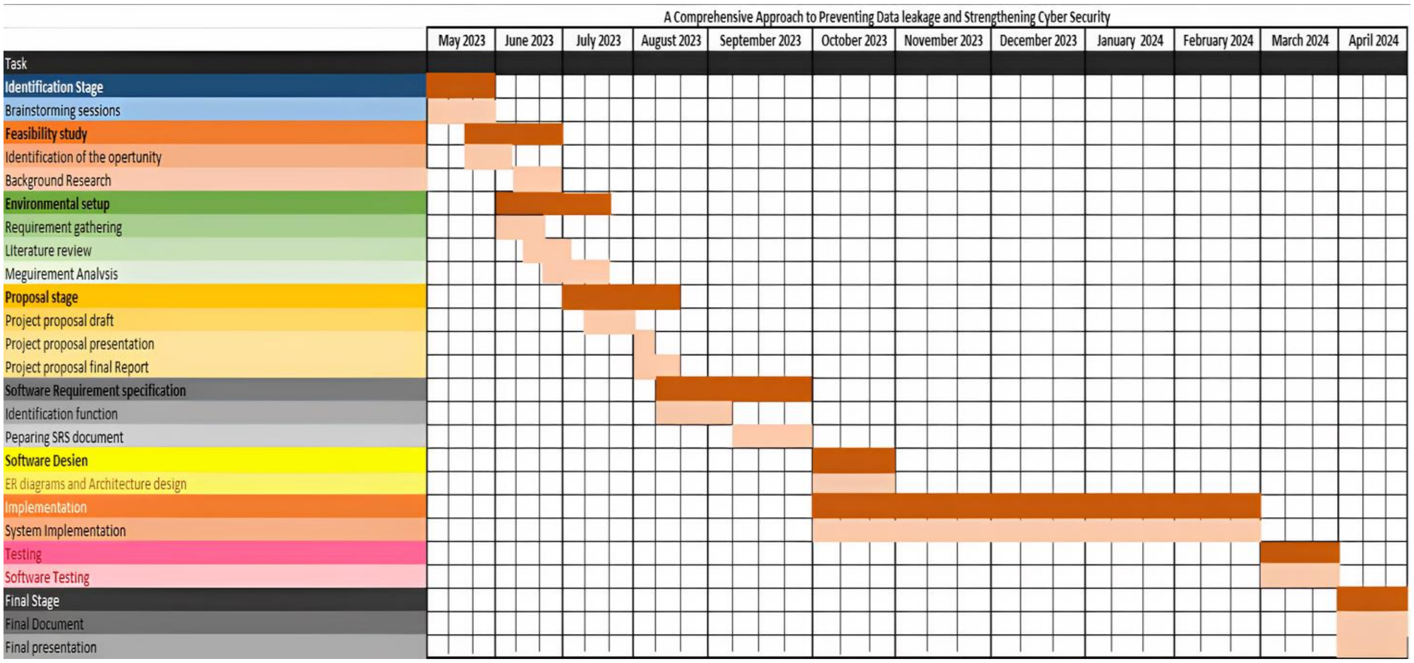
we have discussed the importance of identifying patterns and anomalies as a means of mitigating data breach risks. We have explored various tools and strategies for detecting abnormal activities and identifying trending behavior, including machine learning algorithms and data visualization techniques. Additionally, we have highlighted the need for more robust and innovative approaches to data breach prevention, particularly in the face of reliance on outdated techniques and lack of effective tools. Our main objective is to develop a system that can find patterns and anomalies in data, reducing data breach risks and improving cybersecurity.

As we move forward in our efforts to prevent data loss and reputational damage, it is crucial that we remain vigilant and proactive. The stakes are high, and the consequences of inaction are dire. By implementing advanced data analysis techniques and using machine learning algorithms to detect potential security threats, we can stay one step ahead of those who seek to exploit our vulnerabilities. Let us work together to make the digital world a safer place for all.

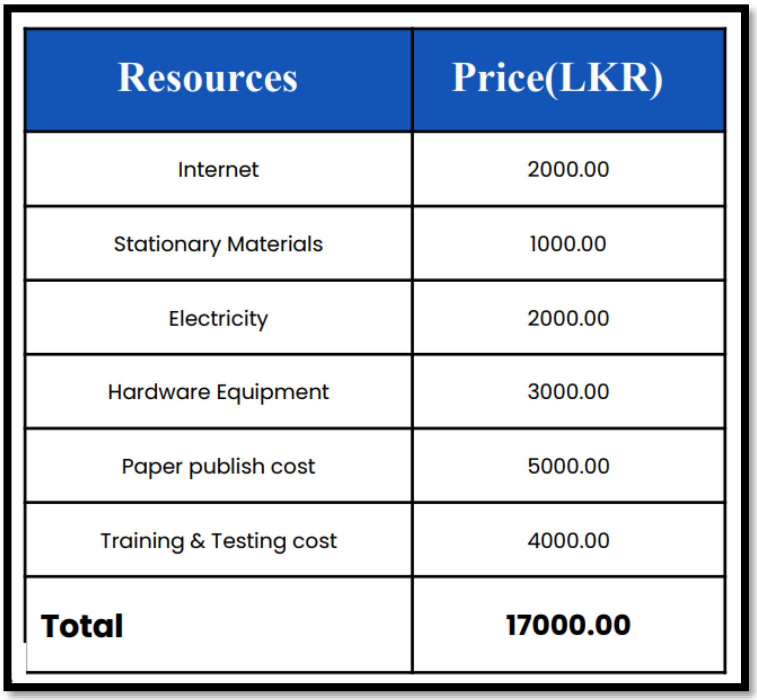
5. WORK BREAKDOWN STRUCTURE *(WBS)*



6. GANNT CHART



7. BUDGET JUSTIFICATION



Resources	Price(LKR)
Internet	2000.00
Stationary Materials	1000.00
Electricity	2000.00
Hardware Equipment	3000.00
Paper publish cost	5000.00
Training & Testing cost	4000.00
Total	**17000.00**

Table 1- Estimated budget details

8. COMMERCIALIZATION

- **Create a sales plan for the industry.**

Industry analysis: Conduct thorough market research to identify industries with the highest demand for comprehensive cybersecurity solutions. Understand their pain points, regulatory requirements, and specific challenges associated with data leakage and cybersecurity.
Tailored fee Proposition: Craft a compelling value proposition that directly addresses the unique needs of each targeted industry. Highlight how our system's capabilities align with its requirements, showcasing its effectiveness in preventing data leakage and strengthening overall cybersecurity.
Strategic Partnerships: Forge partnerships with influential industry associations, organizations, and thought leaders. Collaborate to co-host webinars, workshops, or events that position our system as an innovative solution, gaining credibility and expanding our reach within the industry.

- **Design customer subscription plans.**

Tiered Plans: Develop various subscription levels, each catering to different business sizes and needs. Offer options like basic, standard, and premium plans, each with a distinct set of features and capabilities.
Scalability: Ensure that your subscription plans are designed to accommodate the growth and changing requirements of businesses. Provide flexibility for clients to upgrade or adjust their plans as their needs evolve.
Customization Flexibility: Integrate customization options within subscription plans, allowing clients to tailor features based on their specific data protection needs. This ensures that they only pay for the functionalities they require.

- **Provide excellent customer support.**

Dedicated support team: Assign a team of knowledgeable support representatives to promptly address inquiries and issues.

Multi-Channel assist: Provide support through various channels such as email, live chat, and phone.

24/7 Availability: offer round-the-clock assistance for critical concerns and urgent inquiries.

Knowledge Base: Develop an online resource with FAQs, tutorials, and troubleshooting guides.

Continuous training: Offer training sessions to help clients maximize the benefits of the system.

9. REFERENCES

[1] Craig, D. (2023, July 26). How to develop an effective cyber security strategy. *Riskxchange.Co*. https://riskxchange.co/1007123/how-to-develop-effective-cyber-security-strategy/

[2] *Data loss prevention vs. Data leak prevention*. (n.d.). Blackberry.com. Retrieved August 25, 2023, from https://www.blackberry.com/us/en/solutions/endpoint-security/data-loss-prevention/data-loss-prevention-vs-data-leak-prevention

[3] De Groot, J. (n.d.). *What is data loss prevention (DLP)? Definition, types & tips*. Digital Guardian. Retrieved August 25, 2023, from https://www.digitalguardian.com/blog/what-data-loss-prevention-dlp-definition-data-loss-prevention

[4] Emenike, S. U., & Nohlberg, M. (n.d.). *Data loss prevention in a remote work environment*. Diva-portal.org. Retrieved August 25, 2023, from https://www.diva-portal.org/smash/get/diva2:1578629/FULLTEXT01.pdf

[5] Noor, U., Anwar, Z., Malik, A. W., Khan, S., & Saleem, S. (2019). A machine learning framework for investigating data breaches based on semantic analysis of adversary's attack patterns in threat intelligence repositories. *Future Generations Computer Systems: FGCS*, *95*, 467–487. https://doi.org/10.1016/j.future.2019.01.022

[6] Paul, M., & Medhe, K. (2019). Using machine learning to detect anomalies in internet browsing pattern of users. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3511054

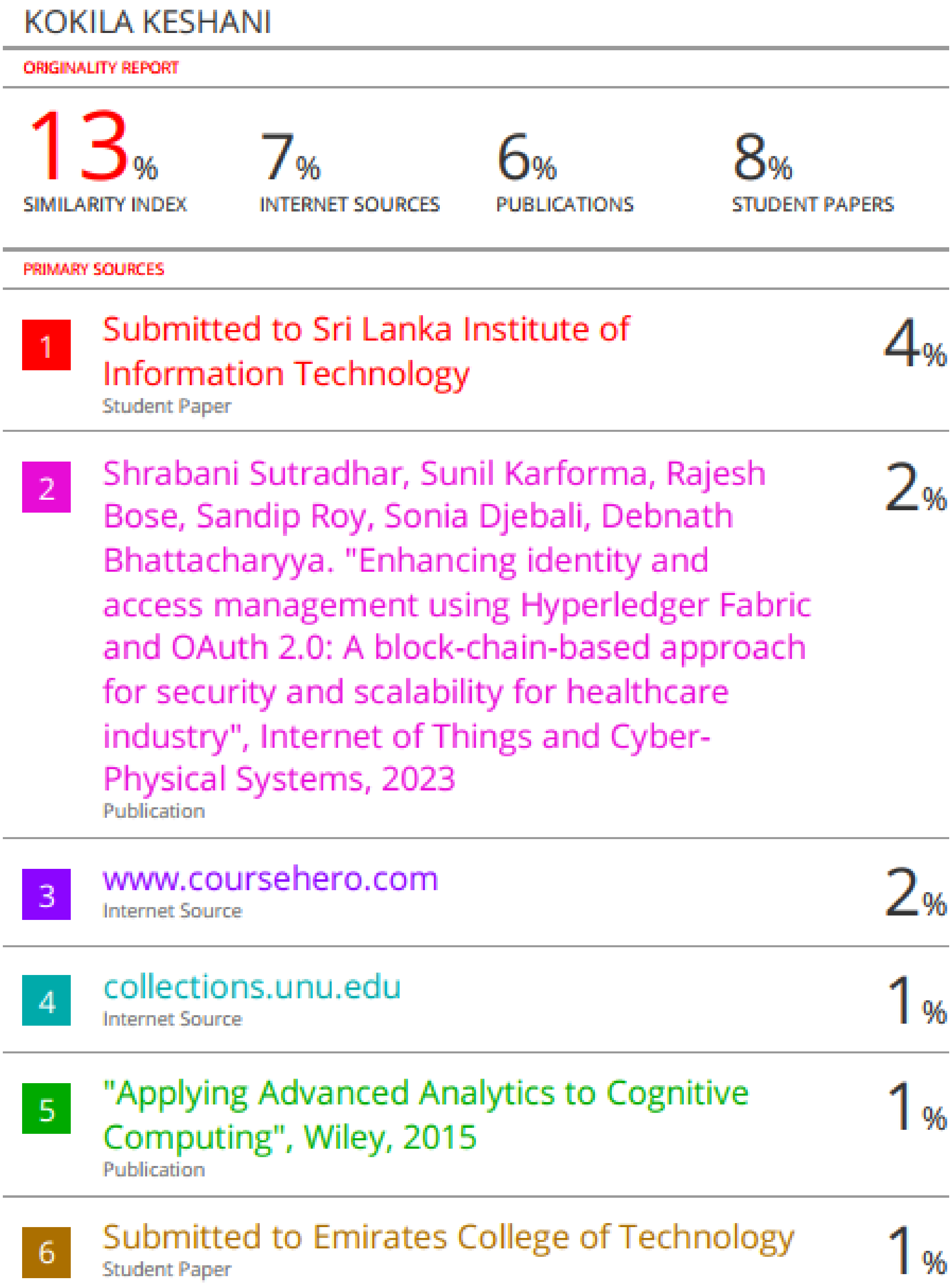
[7] *What is data loss prevention (DLP)?* (n.d.). Microsoft.com. Retrieved August 25, 2023, from https://www.microsoft.com/en-us/security/business/security-101/what-is-data-loss-prevention-dlp

[8] Willbanks, H. (2023, May 31). *Unveiling anomalies — strengthening bank security with behavioral analytics*. Exabeam. https://www.exabeam.com/information-security/unveiling-anomalies-strengthening-bank-security-with-behavioral-analytics/

[9] (N.d.-a). Researchgate.net. Retrieved August 25, 2023, from https://www.researchgate.net/publication/266617827_Data_LeakageLoss_Prevention_Systems_DLP

[10] (N.d.-b). Researchgate.net. Retrieved August 25, 2023, from https://www.researchgate.net/publication/335336220_Data_Loss_Prevention

APPENDICES



KOKILA KESHANI

ORIGINALITY REPORT

13%	7%	6%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sri Lanka Institute of Information Technology Student Paper	4%
2	Shrabani Sutradhar, Sunil Karforma, Rajesh Bose, Sandip Roy, Sonia Djebali, Debnath Bhattacharyya. "Enhancing identity and access management using Hyperledger Fabric and OAuth 2.0: A block-chain-based approach for security and scalability for healthcare industry", Internet of Things and Cyber-Physical Systems, 2023 Publication	2%
3	www.coursehero.com Internet Source	2%
4	collections.unu.edu Internet Source	1%
5	"Applying Advanced Analytics to Cognitive Computing", Wiley, 2015 Publication	1%
6	Submitted to Emirates College of Technology Student Paper	1%