Human-centric Computing and Information Sciences

June 2025 | Volume 15

KCIA
Korea Computer
Industry Association

www.hcisjournai.com

HCIS Human-centric Computing and Information Sciences

RESEARCH (Research Manuscript) Open Access

Human-centric Computing and Information Sciences (2025) 15:32
DOI: https://doi.org/10.22967/HCIS.2025.15.032
Received: June 14, 2024; Accepted: August 10, 2024; Published: June 15, 2025

A Review of Text-Based Information Security Rating: Fundamental Concepts, Methods, Datasets, Challenges, and Future Works

Yuna Han[1], Juno Lee[1], Jimin Lee[1], and Hangbae Chang[2,*]

Abstract

The growing importance of critical data for national competitiveness and corporate survival underscores the need to classify and protect sensitive documents. Many researchers have introduced information security rating, a data classification method considering document security levels. However, research in this area has faced challenges in methodology development and application due to the lack of a standardized definition and the scarcity of survey papers exploring the latest research trends. To address these issues, this research proposes a standardized term, "information security rating," and establishes a systematic taxonomy of text-based information assets, including domain scope, methodology, and metrics. The primary contribution of this study is to comprehensively review the overall research trends, covering both administrative and technical methodologies, from rule-based methods to deep learning models. It also introduces representative datasets and various evaluation metrics, such as the CIA triad, impact factors, and text classification metrics. Furthermore, this study identifies and proposes five novel limitations from different perspectives, including the challenge of unbalanced confidential data, the need for alternative security evaluation metrics, and convergence approaches. Overall, this study will serve as a fundamental guideline, by providing insights into future research directions.

Keywords

Review, Information Security Rating, Emerging Trend in Security, Open Challenges, Future Work Directions

1. Introduction

In the era of technological competition, possessing critical information is crucial for securing national competitiveness. The number of malicious threats attempting to steal vital information, both externally and internally, has steadily increased [1]. Various information assets manage most critical organizational and corporate information in textual form, underscoring the need to protect the inherent core information within documents. Recent incidents underscore this urgency, such as the legal actions against a former senior executive of Proofpoint in 2021 for allegedly leaking strategic business documents [2] and Yahoo's former employee in 2022 for leaking internal documents containing approximately 570,000

***Corresponding Author:** Hangbae Chang (hbchang@cau.ac.kr)
[1]Department of Convergence Security, Chung-Ang University, Seoul, Korea
[2]Department of Industrial Security, Chung-Ang University, Seoul, Korea

pages of source code and algorithms to a competitor [3]. These cases illustrate the persistent occurrence of data breaches. Moreover, a report by IBM revealed that in 2023, the cost incurred due to data breaches amounted to $4.45 million, emphasizing the essential nature of economic security measures [4]. The leakage of documents containing sensitive research, development, or corporate secrets affects national competitiveness and corporate survival, necessitating effective techniques for preventing such leaks.

Traditional methods such as data loss prevention (DLP) have been critical in securing data at system and network levels [5]. However, DLP systems often require significant resources and suffer from frequent false positives, limiting their effectiveness in identifying and protecting sensitive data [6]. Data classification systems, which categorize information by security level, have been developed to mitigate insider threats. Gartner emphasizes that data policies and classification guidelines are crucial to supporting security solutions such as DLP and governance [7]. For example, Boldon James, a data classification firm, categorizes documents into four levels—confidential, internal, general, and public—to enhance DLP and reduce risks [8]. Similarly, Indiana University provides a system for classifying research and development information into critical, restricted, university-internal, and public categories, highlighting the importance of researchers' responsibility in data management [9].

Administrative and technical approaches divide research efforts in terms of trends. From an administrative perspective, research has focused on evaluating and classifying data based on the CIA triad (confidentiality, integrity, and availability) and security impact factors [10]. However, the CIA-based assessment method for classification management has been criticized for its redundancy in using confidentiality as a cause-and-effect variable, thereby obscuring the criteria. Additionally, assessing data classification based on external attributes such as metadata poses challenges in accurately determining the value of information embedded within documents. A recent study, which is considered our baseline research paper, proposes that information lifecycle-based rating factors—including information value (cost of information creation, timeliness, usability, etc.) and risk assessment for external leakage—should be considered alongside the CIA triad to enable effective economic security activities [11].

Moreover, administrative aspects of information classification employ research on various evaluation metrics for security rating management, such as the CIA triad and security impact factors. However, the subjective nature of administrative approaches necessitates using technical mechanisms to facilitate objective decision-making based on quantitative metrics [12]. Technical mechanisms effectively automate classification management guidelines according to security requirements, employing diverse approaches such as automatic rule-based security rating methods, similarity-based clustering techniques, and machine learning and deep learning models [12–15].

Exploring social, industrial, and research trends has shown that the need for information classification is constantly emerging. However, the concept of information classification still requires additional standards, necessitating a thorough and detailed preliminary investigation to construct an efficient information security rating model. This paper aims to point out these problems, establish a conceptual taxonomy for the standardization of information security rating, and provide a comprehensive introduction to concepts, taxonomies, methodologies, evaluation metrics, current information security rating limitations, and future directions. This review paper's specific objectives, scope, and contributions are introduced in the following subsections 1.1 and 1.2.

1.1 Motivation and Research Objectives

This paper addresses three main problems: (1) the rising need for information security rating due to increasing text-based document leaks, but a lack of research in this area; (2) the fragmented concept of information security rating; and (3) the absence of comprehensive surveys on information security rating research. We aim to standardize terminology and conceptual definitions, as the National Institute of Standards and Technology (NIST) [16] and Gartner [17] use "data classification" for security classification, while other literature refers to "security classification" and "sensitivity classification." This study also seeks to create a taxonomy of information security rating by organizing text-based information assets across research domains, methodologies, and evaluation metrics, focusing on managerial and technical methods. Through a comprehensive review of trends, we discuss limitations in datasets, metrics, and methodologies, and propose future research directions.

1.2 Scope and Contribution of This Review

This paper reviews the following key questions, following a research-question prioritization approach [18]:

Question 1: What is the conceptual definition of information security rating?

Question 2: What research domains (e.g., corporate, R&D) are involved in information security rating?

Question 3: What methodologies are used for performing information security rating, and how can they be applied in the real world?

Question 4: Which representative datasets are used to train information security rating?

Question 5: How are metrics for evaluating information security rating developed from managerial and technical perspectives, and what are the advantages and disadvantages of each?

Question 6: Based on current research trends in information security rating, what are the limitations and future research directions regarding methodologies, datasets, and evaluation metrics?

According to these questions, this paper reviews the definition, research domains, methodologies, applications, datasets, and evaluation of information security rating. It excludes product analysis and focuses on future directions from a broad perspective rather than detailing specific processes or algorithms. Therefore, this paper serves as a guideline for fundamental research in information security rating by systematically reviewing the current research status and analyzing open challenges to provide future research direction. Specifically, the contributions of this paper are as follows:

- Defining the concepts of information security rating.
- Reviewing the concepts of data classification and addressing the difference between topic categorization and security rating.
- Specifically reviewing the information security rating within the scope of corporate and research domains.
- Analyzing and designing taxonomies of text-based information security rating in terms of text-based information assets, methods of information security rating, and evaluation metrics of security rating.
- Reviewing the methods of information security rating in detail regarding management and technical approaches.
- Reviewing and analyzing datasets and evaluation metrics for the information security rating task.
- Addressing the open challenges of information security rating research regarding datasets, methods, metrics, and their application.
- Proposing the open challenges of security rating and future directions.

This introduction sets the stage for systematically exploring information security rating to guide future developments in academia and industry.

2. Related Literature Analysis

As outlined in Section 1, the ultimate purpose of security rating is to protect and utilize data. This section explores existing research related to data protection, mainly focusing on the distinctions between DLP systems and information rights management (IRM) and the necessity for information security rating, emphasizing its novelty through a comparative analysis of surveys across various fields.

2.1 Data Leak Prevention, Data Protection, and Data Classification

With rising incidents of sensitive data breaches, methods such as DLP and IRM have been introduced to prevent leaks, protect data, and classify information. DLP is designed to prevent unauthorized distribution using techniques such as data identification, classification, policy enforcement, and real-time monitoring [19, 20]. It enhances real-time compliance support, incident response, and security awareness by monitoring network traffic, email, and cloud storage [5, 21]. Conversely, IRM, controls access rights based on document sensitivity, ensuring unauthorized users are restricted from confidential data [22]. Both DLP and IRM face challenges, such as complex policy settings and difficulty addressing insider threats, leading to the rise of data classification as a proactive strategy.

Table 1. Summary of critical studies on data leakage prevention, data protection, and data classification

Research field	Methodology	Contribution	Limitation
Data leak prevention [23]	Proposing a fine-grained learning adaptive neighbors (LAN) framework to identify abnormal activity in real-time using activity logs in a graph neural network (GNN) to predict anomaly scores.	Overcame the inability to detect insider threats in real-time; that comes with post-hoc-based insider threat detection methodologies, and addressed data imbalances in a self-supervised manner.	Since labeling abnormal samples requires much effort, an interactive framework for anomaly detection needs to be designed with time efficiency.
Data protection [24]	Designing self-embedding digital watermarking via the Canny operator and DCT compression-based digital image encryption.	Increased resource use and encryption time efficacy with enhanced attack resistance and low image distortion.	Complexity in integrating multiple algorithms; needs adaptability to various multimodal datasets (text, large images, 3D, video).
Data classification [26]	Applying a lightweight DistilBERT model for spam classification to enhance information security.	Adequate to identify spam and non-spam email to ensure information security.	The optimization algorithm is necessary to strengthen the result of the misclassification of non-spam email data.

Through an analysis of key studies in these fields, this paper highlights its unique contribution by exploring their methodologies, findings, and limitations. Table 1 summarizes the results of this analysis.

In summary, studies on DLP, data protection, and data classification focus on detecting insider threats and malicious behavior [5, 23], encryption, watermarking, and text categorization. For instance, a key study on DLP analyzed insider threats, offering recommendations for mitigating risks through controls, education, and policy development, despite its limitations in global perspectives [5]. Data protection research centers on watermarks and encryption, such as blockchain-based frameworks for image forensics in Internet of Things (IoT) environments, though improvements in accuracy remain necessary [24, 25]. Finally, data classification research has evolved into information security rating, categorizing confidential and public information based on its value, addressing contemporary social issues [11, 26–28].

2.2 Related Survey Paper

To evaluate the relevance and originality of this review, we summarized related review papers on data protection and leakage prevention, as shown in Table 2 [29–36]. The analysis suggests that there is a lack of security-centric data classification research in a review of relevant surveys. Thus, this paper's contribution lies in defining security classification, systematically reviewing the current state, and proposing future research directions.

Table 2. Comparison of related review papers

Purpose of paper	Topic	Contribution and limitation
Data leak prevention	Data leakage prevention [29]	Identifies sensitive data during the initial stages of leakage prevention and handling, but lacks specific data classification criteria.
	Insider threat detection [30]	Reviews deep learning approaches for insider threat detection with potential performance gains over machine learning, but lacks proactive methods for insider threat mitigation.
	Anti-phishing [31]	Thoroughly examines the design and effectiveness of anti-phishing programs, but needs more focus on practical implementations and protection of organizational documents.
Data protection	Data protection in blockchain applications [32]	Reviews technologies for sensitive data protection in blockchain and suggests methods for framework design, but is still in the early stages, limiting practical applications.
	AI-based IoT security and privacy [33]	Proposes a new architecture to enhance IoT security and privacy, reviewing AI-based solutions to security issues, but lacks empirical validation and a discussion on the security needs of organizational documents.
Data classification	Topic categorization [34]	Reviews semi-supervised learning (SSL) techniques in text categorization, highlighting trends over the past 5 years, but lacks practical validation and a discussion on information security rating.
	Public security [35]	Conducts a meta-study across 19 public security fields (e.g., cybersecurity, fraud detection), but lacks exploration of text-based information security studies within public security meta-research.
	Evaluation metrics [36]	Discusses various metrics for data classification, with a focus on selection criteria, but focuses primarily on binary classification, lacking validation for multi-class classification problems.
	Information security rating (proposed)	Comprehensively reviews information security rating, analyzing taxonomies and discussing open challenges for future research, but requires rich analysis of text-based research and development documents.

3. Methodology

As shown in Fig. 1, this review follows a two-step approach: research and review to provide a comprehensive analysis of text-based information security ratings, based on a study of state-of-the-art security techniques. This method effectively organizes theoretical backgrounds, analyzes specific areas, and identifies future research directions [32]. The first step defines the domain and categories for review, focusing on corporate and research fields. It examines trends and applies methodologies to case studies. Given the limited studies on security rating, key papers published up to 2024 are included, without restricting the review to a specific timeframe. Keywords guide queries, and abstracts are reviewed for relevance. The second step defines the concept of information security rating, analyzing information assets, methods, and evaluation metrics to develop a rating system. Existing literature is reviewed to identify limitations and suggest future research directions, offering insights into the current and future state of information security rating. This two-step methodology—combining research surveys and trend reviews—serves both research and industry by thoroughly reviewing the underexplored field of information security ratings and discussing practical applications that focus on stability and reliability. Consequently, this paper covers past, present, and future trends in information security rating.

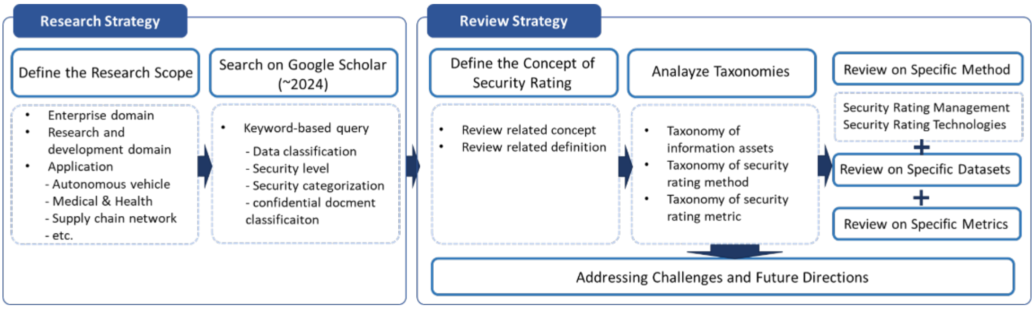

Research Strategy
Define the Research Scope
Search on Google Scholar (~2024)
• Enterprise domain
• Research and development domain
• Application
- Autonomous vehicle
- Medical & Health
- Supply chain network
- etc.
• Keyword-based query
- Data classification
- Security level
- Security categorization
- confidential docment classificaiton
Review Strategy
Define the Concept of Security Rating
Analayze Taxonomies
Review on Specific Method
• Review related concept
• Review related definition
• Taxonomy of information assets
• Taxonomy of security rating method
• Taxonomy of security rating metric
Security Rating Management
Security Rating Technologies
+
Review on Specific Datasets
+
Review on Specific Metrics
Addressing Challenges and Future Directions

Fig. 1. Methodology of this review.

4. Fundamental Concepts

4.1 Definition and Concept of Information Security Rating

As information systems evolve and society shifts toward data-centric operations, the demand for information security rating has increased, leading to numerous publications. However, the fragmented terminology, such as "security classification" [37, 38], "security categorization" [39], "data classification" [16, 17, 40], and "sensitivity rating" [41], underscores the need for a standardized term. For instance, NIST and Gartner provide guidelines for data classification based on data sensitivity [16, 17, 40], while research on text-based information uses terms such as "security classification" [42], "security-level classification" [43], "sensitivity classification" [44], and "text classification" [45]. "Data classification" is often used ambiguously in topic categorization and information security rating, as shown in Fig. 2. This dual usage creates confusion, underscoring the need for clear distinctions. For example, the 1993 United States government report "Security Classification of Information" aimed to protect information by assigning classification levels (e.g., top secret, secret, public) [38].

Synthesizing previous studies, this paper defines information security rating as classifying information assets within organizations, considering economic security factors. This redefinition, grounded in prior research [11], broadens the scope of information security rating and sets a new standard in the field.

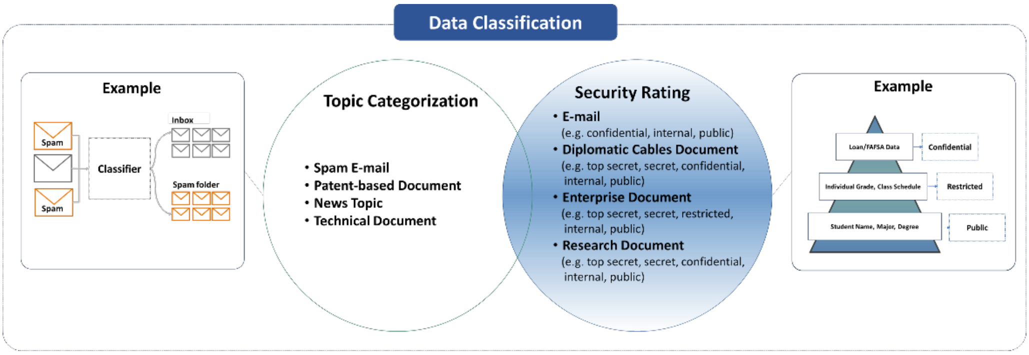

Data Classification
Example
Spam
Spam
Classifier
Inbox
Spam folder
Topic Categorization
• Spam E-mail
• Patent-based Document
• News Topic
• Technical Document
Security Rating
• E-mail
(e.g. confidential, internal, public)
• Diplomatic Cables Document
(e.g. top secret, secret, confidential, internal, public)
• Enterprise Document
(e.g. top secret, secret, restricted, internal, public)
• Research Document
(e.g. top secret, secret, confidential, internal, public)
Example
Loan/FAFSA Data
Confidential
Individual Grade, Class Schedule
Restricted
Student Name, Major, Degree
Public

Fig. 2. Concept of security rating.

4.2 Scope of Information Security Rating

Information security ratings classify organizational assets by security level, using specific guidelines or evaluation models. These assets include both electronic and non-electronic data generated, discovered, or imported within the organization [46]. The scope is divided into corporate information [12] and research and development (R&D) information [47, 48]. Corporate information security rating has evolved alongside digital integration, focusing on mitigating technology leaks and protecting data integrity and confidentiality [49]. Recently, models have been developed to ensure economic security throughout the data lifecycle, including input, use, and output stages [11]. Industry-specific information types demand tailored security ratings [49]. For example, cloud computing has led to studies on e-government security ratings [50], data classification in internet networks [51], and supply chain security ratings [52]. Financial institutions apply systematic protections for personal and document information [53, 54], while the growth of smart healthcare [55] has driven the development of security ratings for mobile devices [56]. R&D institutions, which generate and manage vast amounts of physical and electronic information, require systematic security measures to enhance national competitiveness. Legislative guidelines mandate security ratings [48], though university-level data management strategies remain under-explored [57]. This paper reviews research on text-based security ratings for electronically managed corporate and R&D information, and surveys personal and organizational information security ratings to provide a clearer understanding of current and future trends.

4.3 Taxonomies of Information Security Rating

As previously mentioned, a lack of comprehensive surveys on information security rating suggests a shortage of overarching insights, and standardized concepts or security rating systems. This section refines the information security rating concept introduced in Section 4.1, structuring it across perspectives before examining methodologies and evaluation metrics. The overall taxonomy of text-based information security rating is depicted in Fig. 3, comprising three mainstreams and their sub-branches, such as enterprise and R&D documents classified as information assets.

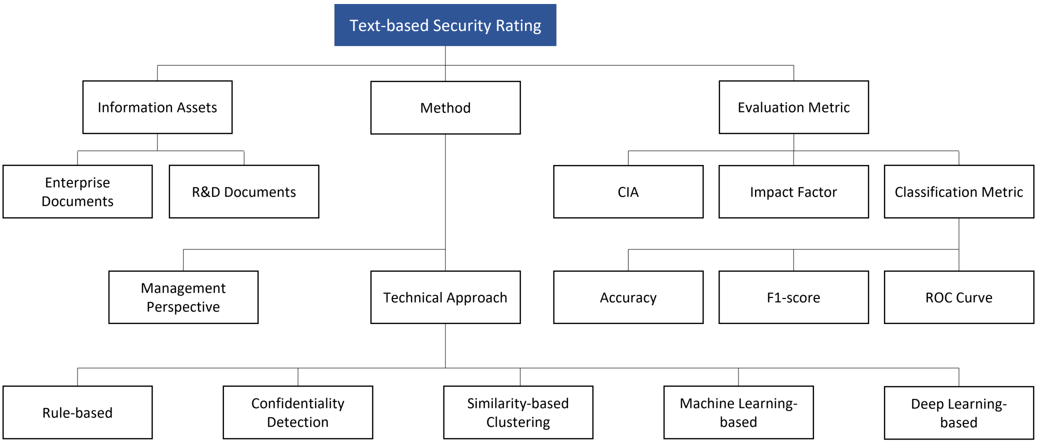


Fig. 3. Text-based information security rating taxonomies.

First, information assets can vary depending on organizational characteristics but typically include electronic information, software, hardware, and personnel data [39, 46]. This paper focuses on documents and text-based information across industries, including corporate and research security domains, as illustrated in Fig. 4 [57]. For example, corporate information requiring trade secret protection includes executive minutes, financial reports, operational manuals, and service agreements [58]. For R&D information, systems categorize units such as R&D outputs and technology transfer data, assigning security ratings to each [48].

Second, text-based information security rating methodologies can be categorized into administrative and technical approaches. Administrative methods focus on compliance with internal security policies, IT infrastructure, and human resource management [59]. Security managers must understand data requirements and design appropriate rating models [11]. Technical approaches address inefficiencies in administrative methods and challenges in obtaining objective results [60]. These include traditional rule-based ratings, inference, automatic algorithms, clustering, word embedding-based confidentiality detection, and machine learning-based classification [12, 61–64]. Section 5 details specific methodologies.

Lastly, evaluating information security ratings involves metrics such as the CIA triad, assessing confidentiality, integrity, and availability [65]. Recent evaluation methods incorporate impact factors to reduce ambiguity and subjectivity [11, 66], with metrics such as economic impact, accuracy, F1-score, and receiver operating characteristic (ROC) curves. Sections 5 and 7 elaborate on these methodologies and evaluation metrics.

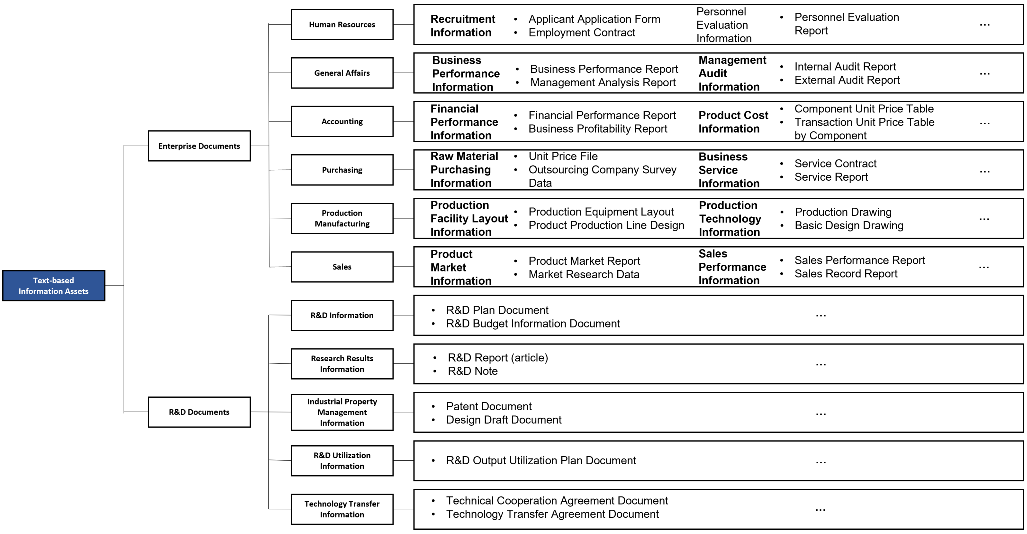

Text-based Information Assets
Enterprise Documents
R&D Documents
Human Resources
Recruitment Information
• Applicant Application Form
• Employment Contract
Personnel Evaluation Information
• Personnel Evaluation Report
...
General Affairs
Business Performance Information
• Business Performance Report
• Management Analysis Report
Management Audit Information
• Internal Audit Report
• External Audit Report
...
Accounting
Financial Performance Information
• Financial Performance Report
• Business Profitability Report
Product Cost Information
• Component Unit Price Table
• Transaction Unit Price Table by Component
...
Purchasing
Raw Material Purchasing Information
• Unit Price File
• Outsourcing Company Survey Data
Business Service Information
• Service Contract
• Service Report
...
Production Manufacturing
Production Facility Layout Information
• Production Equipment Layout
• Product Production Line Design
Production Technology Information
• Production Drawing
• Basic Design Drawing
...
Sales
Product Market Information
• Product Market Report
• Market Research Data
Sales Performance Information
• Sales Performance Report
• Sales Record Report
...
R&D Information
• R&D Plan Document
• R&D Budget Information Document
...
Research Results Information
• R&D Report (article)
• R&D Note
...
Industrial Property Management Information
• Patent Document
• Design Draft Document
...
R&D Utilization Information
• R&D Output Utilization Plan Document
...
Technology Transfer Information
• Technical Cooperation Agreement Document
• Technology Transfer Agreement Document
...

Fig. 4. Text-based information asset taxonomy in enterprise and R&D documentation scope [58].

5. Security Rating Methods

This section reviews methodologies for information security rating from both managerial and technical perspectives. Fig. 5 illustrates the general flow of these strategies, which are detailed in the following subsections.

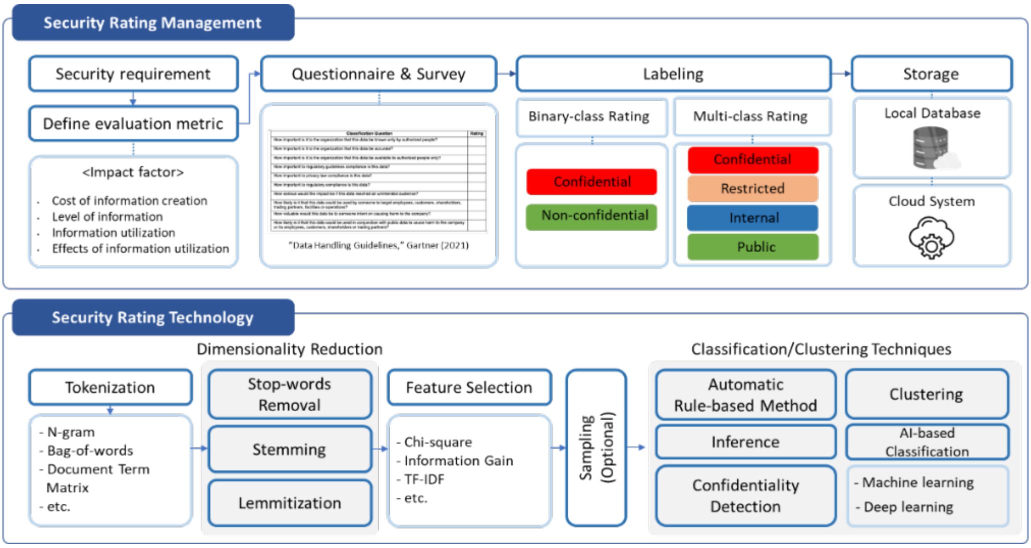


Fig. 5. Information security rating methodologies from management and technical perspectives.

5.1 Security Rating Methods from a Management Perspective

In organizations, security ratings are based on defined information types and categorized under various security levels. A policy-driven framework encompasses data classification policies and handling guidelines [67]. Organizational data is typically classified into levels (e.g., Level 1: public data; Level 5: most sensitive data) [57], and each level has specific handling requirements [68]. After data classification, differentiated measures are required for electronic and physical protection of information assets, user behavior control, data destruction, and data labeling perspectives.

A major limitation of the policy-driven framework is the potential for subjective classification into specific security ratings if no objective criteria for data classification exist. Therefore, 14 factors have been identified for corporate information security rating, such as manpower, time, capital, availability, usability, level of quality, novelty, use frequency, use range, value creation potential, marketability, development maintainability, business continuity, and competitiveness. Fig. 6 illustrates the flowchart of the security rating method regarding managerial view based on impact factor by reconstructing the theoretical method [11] and its practical system [69]. However, managerial guidelines remain institution-specific, limiting the development of universal methodologies.

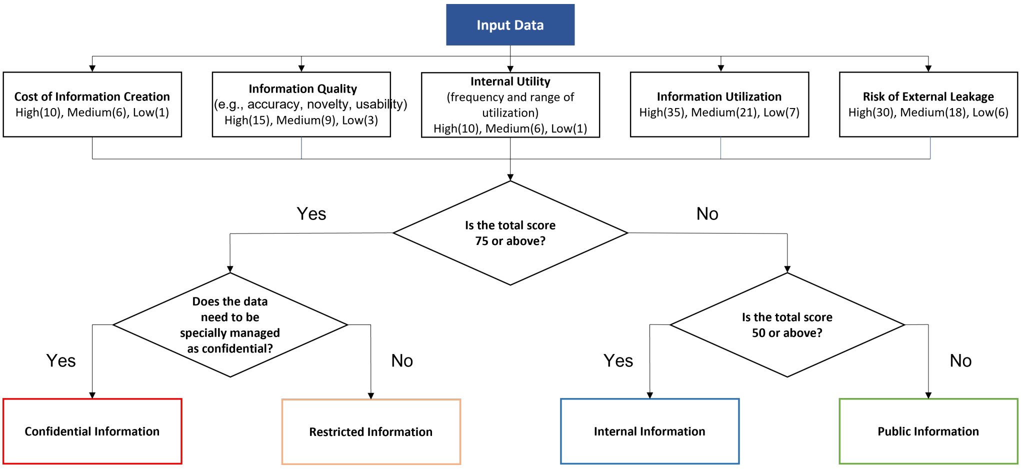

Input Data
Cost of Information Creation
High(10), Medium(6), Low(1)
Information Quality
(e.g., accuracy, novelty, usability)
High(15), Medium(9), Low(3)
Internal Utility
(frequency and range of utilization)
High(10), Medium(6), Low(1)
Information Utilization
High(35), Medium(21), Low(7)
Risk of External Leakage
High(30), Medium(18), Low(6)
Yes
Is the total score 75 or above?
No
Does the data need to be specially managed as confidential?
Yes
No
Is the total score 50 or above?
Yes
No
Confidential Information
Restricted Information
Internal Information
Public Information

Fig. 6. A flowchart of the security rating method by an evaluator based on impact factors was reconstructed using the information security rating methods presented in [11] and [69].

5.2 Security Rating Methods using Technical Approaches

Technical approaches for text-based security rating involve analyzing the structure and meaning of text data, as shown in Fig. 7. The process begins with tokenization, dimensionality reduction, and feature selection, using techniques such as chi-square, information gain [70], TF-IDF [71], and word embedding techniques. These processed data are then classified or clustered based on predefined labels.

Traditional rule-based security rating evaluates resource sensitivity by applying resource and access policies to determine data access. For example, one of the studies proposed a method where a user sends a query to the system, which checks the user's location, working hours, and profile to determine access to data of varying sensitivity levels (high, medium, and low) [12]. Industries still utilize document fingerprinting, which tracks sensitive words through unique patterns, and regular expressions in SQL queries to identify confidential documents. Some methods use manually constructed security keyword dictionaries [72] or IBM’s confidential cue phrases [73]. Another study [74] used data labels as metadata for security levels, while another [52] expanded this to manage creation time, identifiers, and transactions in XML, enabling effective access control [75].

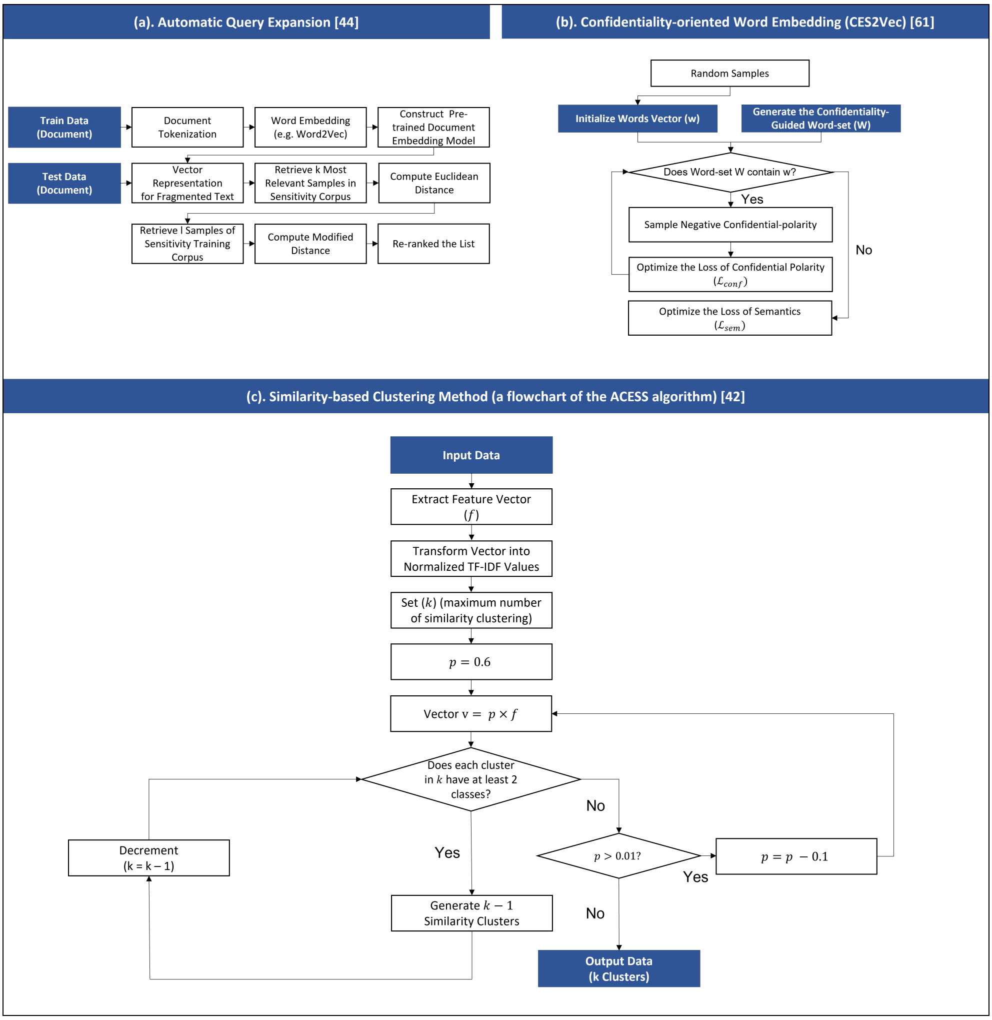

(a). Automatic Query Expansion [44]
Train Data (Document)
Document Tokenization
Word Embedding (e.g. Word2Vec)
Construct Pre-trained Document Embedding Model
Test Data (Document)
Vector Representation for Fragmented Text
Retrieve k Most Relevant Samples in Sensitivity Corpus
Compute Euclidean Distance
Retrieve l Samples of Sensitivity Training Corpus
Compute Modified Distance
Re-ranked the List
(b). Confidentiality-oriented Word Embedding (CES2Vec) [61]
Random Samples
Initialize Words Vector (w)
Generate the Confidentiality-Guided Word-set (W)
Does Word-set W contain w?
Yes
Sample Negative Confidential-polarity
Optimize the Loss of Confidential Polarity ($\mathcal{L}_{conf}$)
No
Optimize the Loss of Semantics ($\mathcal{L}_{sem}$)
(c). Similarity-based Clustering Method (a flowchart of the ACESS algorithm) [42]
Input Data
Extract Feature Vector (f)
Transform Vector into Normalized TF-IDF Values
Set (k) (maximum number of similarity clustering)
$p = 0.6$
Vector v = $p \times f$
Does each cluster in k have at least 2 classes?
No
Decrement (k = k – 1)
Yes
$p > 0.01?$
$p = p - 0.1$
Yes
Generate $k - 1$ Similarity Clusters
No
Output Data (k Clusters)

Fig. 7. A diagram and flowchart of diverse technical approaches for information security rating. (a), (b), and (c) represent automatic query expansion, word embedding, and clustering, respectively.

However, rule-based systems struggle to label data not predefined by keywords. Techniques such as fuzzy logic [76] and fuzzy inference have been proposed to address this uncertainty. Fuzzy logic [76] and fuzzy inference have been proposed to address this issue. A study [49] applied fuzzy techniques based on ISO/IEC 27001 for risk assessments, and another study [77] used association rule mining to identify confidential items with high confidence and support values. Recently, automatic systems have improved efficiency. One study used query extension to re-rank similar data in Twitter document embedding [44] as shown in Fig. 7(a), and another calculated sensitivity weights between substrings to identify confidential information [78]. Although these methods automate tasks, they rely on repetitive rules and may not adapt well to diverse texts, leading to a shift toward learning-based methods.

Word embedding techniques assume that confidential information is located in close vector spaces and is used to detect confidential words in security rating. A notable study [61] proposed CES2Vec, a word embedding that differentiates the confidential polarity of words, based on observing that military terms such as "warplane" differ in security level from commercial "airplane," showing higher accuracy than conventional text-based embedding techniques, including GloVe and Word2Vec. The construction process of the CES2Vec model is shown in Fig. 7(b). Another research [79] demonstrated that clarifying topic boundaries in texts with confidential information improves the accuracy and efficiency of security rating models.

Clustering for information security rating typically means grouping unlabeled text datasets based on the contextual similarity of words or documents, and classifying them into various security levels. The k-means algorithm, a prominent method for clustering in information security rating, has been effectively utilized. For instance, studies on clustering based on distance metrics for paragraphs vectorized through TF-IDF [42, 62] have been conducted, later progressing to techniques that prune impurity topics based on the distribution of security ratings (secret, confidential, and unclassified) and evolving into the automated classification enabled by security similarity (ACESS) method [80] as shown in Fig. 7(c). Further advancements include calculating a confidential score to identify highly sensitive terms [13].

However, setting too many or too few clusters can reduce security rating accuracy, and applying broad category ratings to diverse texts is challenging [62]. For example, even if $k = 100$ achieves high accuracy, dividing data into 100 security levels may lead to misclassification or overfitting, making it impractical in real-world settings [81]. Large k-values can obscure complex document structures, increasing the risk of false positives [13]. Thus, a meta-space classifier has been introduced. This advanced method rebuilds documents using a dual-classification system, estimating the likelihood of documents being public or secret and adjusting misclassified documents into the correct categories [6].

Probabilistic and statistical models calculate the probability of text data belonging to specific security levels [82]. One study used naïve Bayes to estimate prior probabilities and document frequencies, assessing the likelihood of a document's security level [83]. Another study applied support vector machine (SVM) to a proprietary dataset in Turkey, showing better performance than naïve Bayes [84]. A k-nearest neighbor (kNN) model was proposed in [45, 85] assigning weights based on occurrence and identifying neighbors near test samples, offering a simple and efficient solution. However, its limitations in handling large document volumes led to using the T-tree-based TsF-kNN model [86], which improved efficiency and accuracy, particularly in cloud storage applications [87]. In general, ensemble models are known as effective models to handle overfitting issues for large datasets. For instance, one study [88] improved accuracy by parameterizing risk-level probabilities based on the conditional probabilities of a parent node. Another study [89] applied stochastic gradient descent for linear classifiers on complex unstructured data. These machine learning methods focus on vectorizing and learning features from text rather than deeply analyzing long-text contexts. For example, a study [90] used latent Dirichlet allocation for topic modeling to rate document paragraph security and detect embedded confidential information.

Machine learning methods are efficient for resource-limited environments but may struggle with context-rich text classification. Deep learning models use multi-layered neural networks and excel in complex unstructured data classification [91]. For instance, a nonlinear neural network addressed the imbalance in security rating datasets using k-means clustering on under-sampled data [92]. Adaptive neuro-fuzzy inference systems combine fuzzy logic and neural networks for better performance in complex datasets [43, 93]. Convolutional neural networks (CNNs), originally for image classification, have been adapted for text-based security ratings [94] but face issues like sequence truncation, which researchers have addressed by overlapping paragraph sequences to prevent information loss [64]. Depth-wise separable CNNs further balance accuracy and efficiency by separating channel dimensions during training [95]. Additionally, models such as a bidirectional long short-term memory can process up to 1,200 tokens, showing potential for large sequence learning [96]. Recently, keyword-based graph2vec, which builds embeddings based on word relationships, has proven effective for intrinsic document valuation by demonstrating the frequency and relationship of words using nodes, weight, and edges [60]. However, deep learning remains largely confined to supervised learning, highlighting the need for models that can handle unlabeled data across various domains.

5.3 Information Security Rating in Practical Application

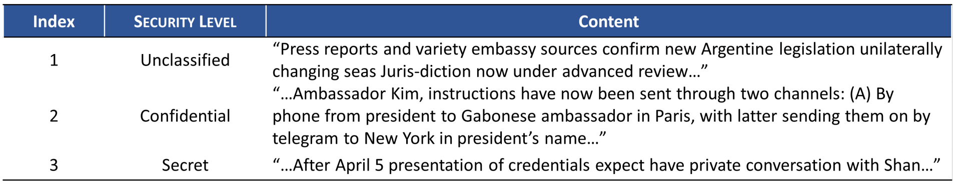
Information security extends theoretical guidelines into practical applications across various industries and organizations [97]. For example, security strategies can classify security levels for electric power systems [98] or define data protection requirements for specific universities or corporations [9, 99]. This section reviews case studies to demonstrate the real-world applicability and effectiveness of information security ratings. Information security rating is widely applied in industries such as display manufacturing, finance, and healthcare. For example, Hong et al. identified critical national information (e.g., electrode wire, planarization film) in South Korea's AMOLED industry using a GNN-based model for patent documents, developing a system to visualize the corporate information value in the display field [60]. Second, Kang and Kim [53] showed an example of calculating the impact on the bank based on the CIA-triad by dividing the document classification system to manage personal information effectively. In healthcare, a sensitive data classification scheme is used to ensure proper control of highly confidential information [55]. In R&D, as shown by Berkeley's classification system based on protection level, availability, and recovery needs, information security rating is essential for protecting vast amounts of confidential data [99]. These studies demonstrate the adaptation of universal frameworks to meet specific organizational needs regarding information security rating.

6. Datasets

Constructing or utilizing datasets is essential for conducting information security ratings using technical strategies. Texts often contain confidential data, typically managed internally or disclosed in a limited manner, making research datasets scarce. Most studies rely on publicly available sources such as WikiLeaks, Reuters, TUBITAK UEKAE, Enron emails, and the selectively accessible Digital National Security Archive (DNSA).

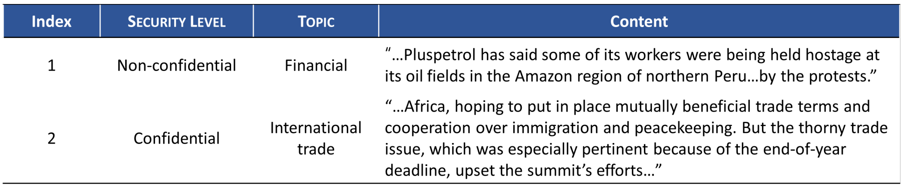
The WikiLeaks dataset consists of classified diplomatic cables from the United States embassies and consulates worldwide from 2003 to February 2010. After removing HTML tags, 10,706 documents are categorized paragraph-by-paragraph as "Unclassified," "Confidential," or "Secret." The dataset includes documents from four embassies: Baghdad, London, Berlin, and Damascus, classified by the highest security level within each document [42, 61, 72, 90, 95]. These documents are publicly accessible on the WikiLeaks website [100], as shown in Fig. 8.

The Reuters dataset contains 21,578 news articles published by Reuters since 1987, which are commonly used for text categorization research. Stored in SGML format across 22 files, it covers topics such as "Earn," "Acquisitions," and "Crude." Each article includes a title, body, and topic label, primarily related to economics, finance, and industry. A sample is shown in Fig. 9, and the dataset is publicly accessible via the Natural Language Toolkit library [13, 62].



Index	Security Level	Content
1	Unclassified	"Press reports and variety embassy sources confirm new Argentine legislation unilaterally changing seas Juris-diction now under advanced review…"
2	Confidential	"…Ambassador Kim, instructions have now been sent through two channels: (A) By phone from president to Gabonese ambassador in Paris, with latter sending them on by telegram to New York in president's name…"
3	Secret	"…After April 5 presentation of credentials expect have private conversation with Shan…"

Fig. 8. A sample of WikiLeaks [100].



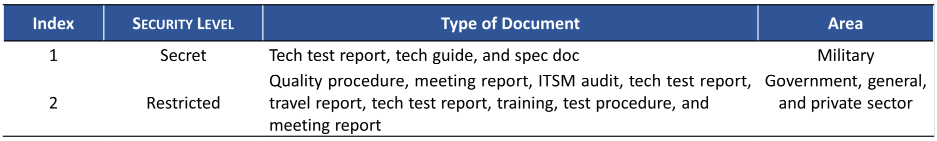
Index	Security Level	Topic	Content
1	Non-confidential	Financial	"…Pluspetrol has said some of its workers were being held hostage at its oil fields in the Amazon region of northern Peru…by the protests."
2	Confidential	International trade	"…Africa, hoping to put in place mutually beneficial trade terms and cooperation over immigration and peacekeeping. But the thorny trade issue, which was especially pertinent because of the end-of-year deadline, upset the summit's efforts…"

Fig. 9. A sample of Reuters [62, 101].

The TUBITAK UEKAE dataset, used for information security and NLP research, contains 222 Turkish documents provided by Turkey's National Research Institute of Electronics and Cryptology. Available in CSV, JSON, and XML formats, the documents are categorized into "Secret" (30), "Restricted" (165), and "Unclassified" (27). Access is public or private, depending on the project [43, 84, 93]. Fig. 10 details document types, accessible via TUBITAK UEKAE's official website [102].

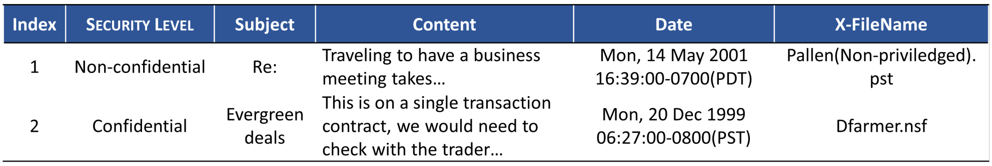
The Enron Email dataset contains about 500,000 emails released by the United States Federal Energy Regulatory Commission in 2001 after Enron's bankruptcy [103]. Collected from 158 employees, it includes email metadata and content, with 64,304 emails categorized as "Confidential" or "Non-Confidential." A sample is shown in Fig. 11, and the dataset is publicly available on Kaggle and the Carnegie Mellon University website [44, 62, 77].

The DNSA dataset includes over 5,000 United States government documents related to national security since World War II. Documents are classified as "Confidential," "Secret," "Top Secret," or "Unclassified," focusing on topics such as Afghanistan, China, and the Philippines. It is available by subscription for academic research and education. Fig. 12 shows a sample of document-level extracted keywords [63, 81, 82].



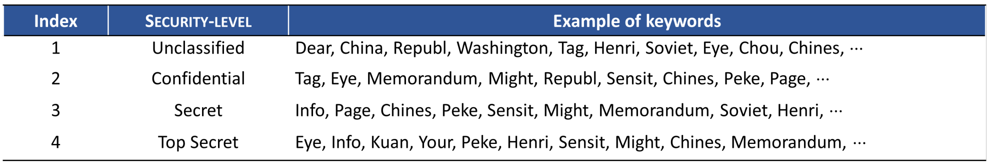
Index	Security Level	Type of Document	Area
1	Secret	Tech test report, tech guide, and spec doc	Military
2	Restricted	Quality procedure, meeting report, ITSM audit, tech test report, travel report, tech test report, training, test procedure, and meeting report	Government, general, and private sector

Fig. 10. Type of document of TUBITAK UEKAE [43].



Index	Security Level	Subject	Content	Date	X-FileName
1	Non-confidential	Re:	Traveling to have a business meeting takes…	Mon, 14 May 2001 16:39:00-0700(PDT)	Pallen(Non-priviledged).pst
2	Confidential	Evergreen deals	This is on a single transaction contract, we would need to check with the trader…	Mon, 20 Dec 1999 06:27:00-0800(PST)	Dfarmer.nsf

Fig. 11. A sample of Enron Email [103].



Index	Security-level	Example of keywords
1	Unclassified	Dear, China, Republ, Washington, Tag, Henri, Soviet, Eye, Chou, Chines, ···
2	Confidential	Tag, Eye, Memorandum, Might, Republ, Sensit, Chines, Peke, Page, ···
3	Secret	Info, Page, Chines, Peke, Sensit, Might, Memorandum, Soviet, Henri, ···
4	Top Secret	Eye, Info, Kuan, Your, Peke, Henri, Sensit, Might, Chines, Memorandum, ···

Fig. 12. A sample of DNSA keywords of each document level [63].

7. Evaluation Metric

7.1 CIA

The CIA Evaluation Metric is based on the three pillars of information security: confidentiality, integrity, and availability. Security ratings are determined by assessing a breach's impact—low, moderate, or high—on each aspect (C, I, A) [46]. The overall impact is then derived by integrating these levels, allowing for the definition of security grades [53], as shown in Equations (1) and (2):

$$\text{CIA} = \{(\text{C}, \text{impact}), (\text{I}, \text{impact}), (\text{A}, \text{impact})\}, \quad (1)$$

$$\text{Total Impact}_{CIA} = \sum_{i \in \{low, med, high\}} (C_i, I_i, A_i). \quad (2)$$

Some studies extend beyond the CIA to include factors such as communication partner authenticity, obligation acceptance, content authenticity, and goal-conform usage [37]. Another CIA application is shown in Equations (3), where the inverse relationship between confidentiality and availability labels total impact as public (1–3), private (4–6), or secret (7–10) [10]:

$$\text{Total Impact}_{Inverse\ CIA} = \frac{\sum_{i=1}^{n}\{(C_i + \mathrm{k}/A_i) / 2 + I_i\}/2\}}{n}. \tag{3}$$

An inherent limitation of CIA-based metrics is the evaluator's subjectivity in assessing the impact on confidentiality, integrity, and availability after a security incident. Additionally, using confidentiality as a dependent and independent variable to determine data sensitivity introduces ambiguity in the assessment.

7.2 Impact Factor

The impact factor evaluation method extends beyond the CIA to assess security ratings based on various factors. For instance, additional criteria such as absolute and relative monetary value, regulatory compliance, stock prices, revenue loss, and customer loss can be applied in fields requiring security rating [66]. Moreover, 14 impact factors were identified for corporate security rating, including manpower, time, capital, availability, usability, quality, novelty, frequency of use, value creation, marketability, and competitiveness [11]. Exploratory factor analysis grouped these into five categories: cost of information creation, information level, utilization, internal effect, and external leakage risk. This method balances relative and absolute standards, reducing evaluator bias and considering multiple business-related factors, ultimately enhancing economic security through security rating.

7.3 Accuracy, F1-Score, and ROC Curve

Typical quantitative metrics used in information security rating include accuracy, F1-score, and ROC curve. As shown in Equation (4), accuracy is the ratio of correct predictions among total predictions. TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{4}$$

From a security rating perspective, TP indicates correctly identified high-security subjects, TN indicates correctly identified low-security subjects, FP occurs when low-rated subjects maintain high-security levels, and FN occurs when high-rated subjects maintain low levels. The F1-score, the harmonic mean of precision and recall, is valuable for evaluating model performance on imbalanced datasets, as shown in Equation (5):

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{\frac{TP}{TP + FP} \times \frac{TP}{TP + FN}}{\frac{TP}{TP + FP} + \frac{TP}{TP + FN}}. \tag{5}$$

The F1-score reflects a model's ability to detect security threats with accuracy while minimizing errors. The ROC curve visualizes the relationship between false positive and true positive rates, with the area under the curve indicating performance—values closer to 1 suggest better accuracy. Although the ROC curve effectively assesses how well models distinguish security threats, it lacks indicators for document importance or confidentiality, limiting its application in security rating. Furthermore, existing managerial and technical metrics still have strengths and weaknesses, as shown in Table 3.

Table 3. Strengths and weakness of information security rating metrics

Metric	Strength	Weakness
CIA	Provides comprehensive coverage of information security, as it considers confidentiality, integrity, and availability, and is able to assess overall security attributes.	Subjectivity by evaluators may be involved, and ambiguity exists in the assessment items as confidentiality is used for both independent and dependent variables.
Inverse CIA	Mathematically expresses the tradeoff between availability and confidentiality in traditional CIA assessments, allowing for practical information security evaluation.	As with traditional CIA, there are issues with subjectivity and ambiguity, which are issues that can arise in assessing confidentiality, integrity, and availability.
Impact factor	Useful to calculate the value and utility of information in the CIA assessment methodology that can vary across the data lifecycle, reducing assessor subjectivity and enhancing economic security.	As an administrative method, information security rating requires human intervention, leading to variability based on evaluator skill, potential errors, and being time-consuming.
Accuracy	Technically useful for evaluating an information security rating model's performance, with higher accuracy leading to better rating predictions.	Limited to environments where information is graded in advance and can be misleading for imbalanced sensitivity levels where one class has more.
F1-score	Useful for measuring unbalanced sensitivity level-based datasets through the harmonic mean method.	It focuses on balancing precision and recall, making it less suitable for situations where predefined labels are not provided.

8. Open Challenges and Future Works

8.1 Data Scarcity and Imbalanced Confidential Data

The publicly available datasets for information security rating are insufficient for handling the corporate and R&D information needed by industry. While some datasets contain limited corporate data, their quantity must increase for educational purposes. As suggested in this review, designing a security rating model based on data value remains challenging. It is time to create corporate and R&D datasets, which could be managed privately within organizations or made partially public, enabling the use of algorithms such as few-shot and semi-supervised learning with small sample sizes.

Furthermore, confidential information typically makes up a smaller portion of publicly available datasets compared to public information. In industry, most documents are often classified as confidential, leading to inefficient security investments and frequent use of imbalanced data. Thus, sampling or synthetic data generation can be used to augment confidential data. Therefore, information rating system and reorganization of structures are needed for more economically efficient security investments.

8.2 Difficulty Training Specialized Technical Terms

R&D companies and organizations manage documents containing technical information, posing a challenge: rating models based solely on external attributes often overlook the value of embedded information, making it difficult to protect data that could cause significant harm if leaked, such as national critical technologies. In particular, the evaluation of technical terms can vary greatly depending on the skill of the evaluator during human review. A technical approach to identifying these terms can significantly improve accuracy and performance. Thus, each organization should develop and train a domain-specific dictionary of technical terms.

8.3 Need for Specific Label Annotation Guidelines

The current administrative approach to information security rating relies on organizational security requirements and directives, leading to varying guidelines and security rating policies across organizations. This variation complicates standardization and results in inefficiencies in labeling. Most guidelines are based on the ambiguous criteria of the CIA triad. Recent studies have shifted focus toward real-world applications, highlighting the need for standardized security label annotation methods that align with industry guidelines [104]. Future research could benefit from adjusting internal guidelines to standard directives and tailoring them to the specific needs of each organization.

8.4 Need for Advanced Managerial and Technical Methodologies and Limitations of Applying General Security Rating Models

There are two key reasons for advancing information classification methodologies. First, current methods are task-specific, necessitating artificial general intelligence techniques that span multiple industries; therefore, versatile AI model that can handle both corporate and R&D information by using knowledge distillation and fine-tuning of pre-trained LLMs could help to train task-specific model and surrogate training labels, respectively [41]. Second, administrative and technical approaches are fragmented, limiting their applicability and objectivity; therefore, integration of management and technological approaches in information security rating is essential. For instance, a deep learning-based security rating model following distinction criteria from managerial guidelines could enhance verification and applicability (e.g., small datasets and converged settings).

Furthermore, developing separate security rating models for each domain could increase system construction and maintenance costs and the risk of misclassification when testing untrained documents in real industries. Thus, knowledge distillation and fine-tuning pre-trained LLMs should be considered for applying a general security rating system in real-world applications.

8.5 Need to Improve Evaluation Metrics

The survey found that current evaluation metrics for information security rating models using technical approaches rely on text-classification metrics such as accuracy, F1-score, and ROC curves, commonly used in topic categorization. However, these metrics are insufficient for assessing the significance, novelty, or usability of text-based information from a security perspective. Despite recent efforts to incorporate administrative-level security criteria to address the limitations of the traditional CIA triad, more research is needed to apply impact factors at a technical level. For example, data quality can be evaluated by consistency, conciseness, and interpretability, while novelty can be assessed by aligning keyword trends with information creation timing [105]. Therefore, future research should develop mathematical and statistical methods that integrate management guidelines into technical models, enabling accurate and reliable automated security rating models that incorporate diverse impact factors.

9. Discussion and Conclusion

Advanced technologies have shifted from economic tools to critical national sustainability and security factors, symbolizing power in an era of technological hegemony. Nations now either monopolize these technologies for weaponization or attempt to steal them. As a result, document leaks containing critical information are increasingly common. In 2023, data breach costs hit an all-time high, underscoring the need for greater economic investment in security and the development of security rating models. Leaked trade secrets and R&D technologies are easily replicated, posing significant risks. This paper introduces the “security rating” concept by organizing data classification based on security levels for economic protection. It reviews information security ratings in industrial and R&D sectors, presenting a classification system to refine the concept. The paper explores text-based information assets, security rating methods, and evaluation metrics, and discusses challenges and future research directions related to datasets, methodologies, and applications. This review provides a comprehensive understanding of information security ratings as proactive measures for data protection. As the first review on this topic, the paper focuses on conceptual clarification and broad understanding. However, research security remains underexplored, limiting quantitative trend analysis. Future studies should develop methods to analyze information security ratings in R&D environments and include industrial case studies. Ultimately, practical new information security rating models will emerge from future research.

Author's Contributions

Conceptualization, YH, HC; Methodology, YH; Investigation, YH; Writing of the original draft, YH, JL, JL; Writing of review & editing, YH, HC; Visualization, YH, JL, JL; Supervision, HC; Project administration, HC; Funding acquisition, HC.

Funding

This work was supported by the MSIT (Ministry of Science and ICT), Korea, under the Convergence security core talent training business support program (IITP-2025-RS-2023-00266605) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Competing Interests

The authors declare that they have no competing interests.

References

[1] V. Falkevych and A. Lisnyak, "Internal and external threats in cyber security and methods for their prevention," in *Proceedings of 2023 13th International Conference on Advanced Computer Information Technologies (ACIT)*, Wroclaw, Poland, 2023, pp. 414-419. https://doi.org/10.1109/ACIT58437.2023.10275516

[2] K. Barnett, "Yahoo lawsuit alleges employee stole trade secrets upon receiving Trade Desk job offer," 2022 [Online]. Available: https://www.thedrum.com/news/2022/05/19/yahoo-lawsuit-alleges-employee-stole-trade-secrets-upon-receiving-trade-desk-job.

[3] M. Novinson, "Proofpoint alleges ex-exec took trade secrets to abnormal security," 2022 [Online]. Available: https://www.crn.com/news/security/proofpoint-alleges-ex-exec-took-trade-secrets-to-abnormal-security.

[4] IBM Security, "Cost of a Data Breach Report 2023," 2023 [Online]. Available: https://github.com/jacobdjwilson/awesome-annual-security-reports/blob/main/Annual%20Security%20Reports/2023/IBM-Cost-of-a-Data-Breach-Report-2023.pdf.

[5] A. Al-Harrasi, A. K. Shaikh, and A. Al-Badi, "Towards protecting organisations' data by preventing data theft by malicious insiders," *International Journal of Organizational Analysis*, vol. 31, no. 3, pp. 875-888, 2023. https://doi.org/10.1108/IJOA-01-2021-2598

[6] M. Hart, P. Manadhata, and R. Johnson, "Text classification for data loss prevention," in *Privacy Enhancing Technologies Symposium*. Heidelberg, Germany: Springer, 2011, pp. 18-37. https://doi.org/10.1007/978-3-642-22263-4_2

[7] Gartner Inc., "Building effective data classification and handling documents," 2021 [Online]. Available: https://www.gartner.com/en/documents/4000054.

[8] FORTRA Solution Brief, "Boldon James user-applied classification and Symantec data loss prevention improve user acceptance and risk reduction," 2024 [Online]. Available: https://static.fortra.com/boldonjames/pdfs/solution-briefs/bd-user-applied-classification-and-symantec-data-loss-prevention-sb.pdf.

[9] Indiana University, "IU data management," 2024 [Online]. Available: https://datamanagement.iu.edu/tools/matrix.html.

[10] K. P. Singh, V. Rishiwal, and P. Kumar, “Classification of data to enhance data security in cloud computing,” in *Proceedings of 2018 3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, Bhimtal, India, 2018, pp. 1-5. https://doi.org/10.1109/IoT-SIU.2018.8519934

[11] O. Na, L. W. Park, H. Yu, Y. Kim, and H. Chang, “The rating model of corporate information for economic security activities,” *Security Journal*, vol. 32, pp. 435-456, 2019. https://doi.org/10.1057/s41284-019-00171-z

[12] E. Nwafor, P. Chowdhary, and A. Chandra, “A policy-driven framework for document classification and enterprise security,” in *Proceedings of 2016 International IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/Smart-World)*, Toulouse, France, 2016, pp. 949-953. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0149

[13] P. Subhashini and B. P. Rani, “Confidential terms detection using language modeling technique in data leakage prevention,” in *Proceedings of the Second International Conference on Computer and Communication Technologies*. New Delhi, India: Springer, 2016, pp. 271-279. https://doi.org/10.1007/978-81-322-2526-3_29

[14] F. Wang, D. Jiang, H. Wen, and H. Song, “Adaboost-based security level classification of mobile intelligent terminals,” *The Journal of Supercomputing*, vol. 75, no. 11, pp. 7460-7478, 2019. https://doi.org/10.1007/s11227-019-02954-y

[15] P. Wu, T. He, H. Feng, J. Zhao, F. Qi, W. Zhu, and F. Liu, “Rapid classification method of document security levels based on decision tree,” in *Proceedings of 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*, Changchun, China, 2022, pp. 511-515). https://doi.org/10.1109/ICETCI55101.2022.9832183

[16] W. Newhouse, M. Souppaya, J. Kent, K. Sandlin, and K. Scarfone, “Implementing Data Classification Practices (NIST SP 1800-39A),” 2023 [Online]. Available: https://www.nccoe.nist.gov/publications/practice-guide/implementing-data-classification-practices-nist-sp-1800-39-practice.

[17] R. Chugh, B. Willemsen, and N. Henein, “Gartner Report: How to succeed with data classification using modern approaches,” 2022 [Online]. Available: https://www.proofpoint.com/us/resources/analyst-reports/gartner-report-how-to-succeed-with-data-classification.

[18] A. Zeb, F. Din, M. Fayaz, G. Mehmood, and K. Z. Zamli, “A systematic literature review on robust swarm intelligence algorithms in search-based software engineering,” *Complexity*, vol. 2023, article no. 4577581, 2023. https://doi.org/10.1155/2023/4577581

[19] G. Lopez, N. Richardson, and J. Carvajal, “Methodology for data loss prevention technology evaluation for protecting sensitive information,” *Revista Politécnica*, vol. 36, no. 3, pp. 1-10, 2016.

[20] W. Stallings, “Data loss prevention as a privacy-enhancing technology,” *Journal of Data Protection & Privacy*, vol. 3, no. 3, pp. 323-333, 2020. https://doi.org/10.69554/gxro7494

[21] V. Bandari, “Enterprise data security measures: a comparative review of effectiveness and risks across different industries and organization types,” *International Journal of Business Intelligence and Big Data Analytics*, vol. 6, no. 1, pp. 1-11, 2023.

[22] I. Raine, “Information rights management: the next step in information governance and security?,” *Legal Management*, vol. 37, no. 9, pp. 11-13, 2018.

[23] X. Cai, Y. Wang, S. Xu, H. Li, Y. Zhang, Z. Liu, and X. Yuan, “LAN: learning adaptive neighbors for real-time insider threat detection,” 2024 [Online]. Available: https://arxiv.org/abs/2403.09209.

[24] L. Wang, S. Banerjee, Y. Cao, J. Mou, and B. Sun, “A new self-embedding digital watermarking encryption scheme,” *Nonlinear Dynamics*, vol. 112, pp. 8637-8652, 2024. https://doi.org/10.1007/s11071-024-09521-y

[25] Q. Yao, K. Xu, T. Li, Y. Zhou, and M. Wang, “A secure image evidence management framework using multi-bits framework and blockchain in IoT environments,” *Wireless Networks*, vol. 30, no. 6, pp. 5157-5169, 2024. https://doi.org/10.1007/s11276-023-03229-4

[26] T. Liu, S. Li, Y. Dong, Y. Mo, and S. He, “Spam detection and classification based on DistilBERT deep learning algorithm,” *Applied Science and Engineering Journal for Advanced Research*, vol. 3, no. 3, pp. 6-10, 2024. https://doi.org/10.5281/zenodo.11180575

[27] X. Ding, B. Liu, Z. Jiang, Q. Wang, and L. Xin, “Spear phishing emails detection based on machine learning,” in *Proceedings of the 2021 IEEE 24th International Conference on Computer Supported*

Cooperative Work in Design (CSCWD), Dalian, China, 2021, pp. 354-359. https://doi.org/10.1109/CSCWD49262.2021.9437758

[28] Q. Wang and Q. Qian, "Malicious code classification based on opcode sequences and textCNN network," *Journal of Information Security and Applications*, vol. 67, article no. 103151, 2022. https://doi.org/10.1016/j.jisa.2022.103151

[29] S. K. Nayak and A. C. Ojha, "Data leakage detection and prevention: review and research directions," in *Machine Learning and Information Processing*. Singapore: Springer, 2020, pp. 203-212. https://doi.org/10.1007/978-981-15-1884-3_19

[30] S. Yuan and X. Wu, "Deep learning for insider threat detection: review, challenges and opportunities," *Computers & Security*, vol. 104, article no. 102221, 2021. https://doi.org/10.1016/j.cose.2021.102221

[31] D. Jampen, G. Gur, T. Sutter, and B. Tellenbach, "Don't click: towards an effective anti-phishing training: a comparative literature review," *Human-centric Computing and Information Sciences*, vol. 10, article no. 33, 2020. https://doi.org/10.1186/s13673-020-00237-7

[32] S. Khanum and K. Mustafa, "A systematic literature review on sensitive data protection in blockchain applications," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 1, article no. e7422, 2023. https://doi.org/10.1002/cpe.7422

[33] O. E. L. Castro, X. Deng, and J. H. Park, "Comprehensive survey on AI-based technologies for enhancing IoT privacy and security: trends, challenges, and solutions," *Human-Centric Computing and Information Sciences*, vol. 13, article no. 39, 2023. https://doi.org/10.22967/HCIS.2023.13.039

[34] J. M. Duarte and L. Berton, "A review of semi-supervised learning for text classification," *Artificial Intelligence Review*, vol. 56, no. 9, pp. 9401-9469, 2023. https://doi.org/10.1007/s10462-023-10393-8

[35] V. D. H. De Carvalho, R. J. R. Dos Santos, T. C. C. Nepomuceno, and T. Poleto, "Enabling public security text-based analytics: a survey to outline research directions," 2024 [Online]. Available: https://doi.org/10.20944/preprints202403.0064.v1.

[36] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 1-11, 2015. https://doi.org/10.5121/ijdkp.2015.5201

[37] J. H. P. Eloff, R. Holbein, and S. Teufel, "Security classification for documents," *Computers & Security*, vol. 15, no. 1, pp. 55-71, 1996. https://doi.org/10.1016/0167-4048(95)00023-2

[38] A. S. Quist, "Security classification of information: Volume 2. Principles for classification of information (No. K/CG-1077/V2)," Oak Ridge National Laboratory, 1993 [Online]. Available: https://sgp.fas.org/library/quist2/index.html.

[39] Gartner Inc., "Data Management Solutions Primer for 2021," 2021 [Online]. Available: https://www.gartner.com/en/documents/3995936.

[40] W. Newhouse, M. Souppaya, J. Kent, K. Sandlin, and K. Scarfone, "Data classification concepts and considerations for improving data protection (NIST Interagency Report 8496)," 2023 [Online]. Available: https://doi.org/10.6028/NIST.IR.8496.ipd.

[41] N. Pangakis and S. Wolken, "Knowledge distillation in automated annotation: supervised text classification with LLM-generated training labels," 2024 [Online]. Available: https://arxiv.org/abs/2406.176330.

[42] K. Alzhrani, E. M. Rudd, T. E. Boult, and C. E. Chow, "Automated big text security classification," in *Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, Tucson, AZ, USA, 2016, pp. 103-108. https://doi.org/10.1109/ISI.2016.7745451

[43] E. Alparslan, A. Karahoca, and H. Bahsi, "Security-level classification for confidential documents by using adaptive neuro-fuzzy inference systems," *Expert Systems*, vol. 30, no. 3, pp. 233-242, 2013. https://doi.org/10.1111/j.1468-0394.2012.00634.x

[44] L. Q. Trieu, T. N. Tran, M. K. Tran, and M. T. Tran, "Document sensitivity classification for data leakage prevention with twitter-based document embedding and query expansion," in *Proceedings of the 2017 13th International Conference on Computational Intelligence and Security (CIS)*, Hong Kong, 2017, pp. 537-542. https://doi.org/10.1109/CIS.2017.00125

[45] L. Tan, J. Yi, and F. Yang, "Improving performance of massive text real-time classification for document confidentiality management," *Applied Sciences*, vol. 14, no. 4, article no. 1565, 2024. https://doi.org/10.3390/app14041565

[46] National Institute of Standards and Technology, “Standards for security categorization of federal information and information systems (NIST FIPS 199),” 2004 [Online]. Available: https://doi.org/10.6028/NIST.FIPS.199.

[47] L. Perrier, E. Blondal, A. P. Ayala, D. Dearborn, T. Kenny, D. Lightfoot, et al., “Research data management in academic institutions: a scoping review,” *PLOS One*, vol. 12, no. 5, article no. e0178261, 2017. https://doi.org/10.1371/journal.pone.0178261

[48] S. Y. Han, “Design of a grade classification system for research and development (R&D) information,” Ph.D. dissertation, Department of Security Convergence, Chung-Ang University, Seoul, 2024.

[49] C. Alonge, K. Adesemowo, A. M. Mustapha, O. T. Arogundade, F. T. Ibharalu, and O. J. Adeniran, “A fuzzy based classification and labelling framework for effective information assets security risk assessment,” 2023 [Online]. Available: https://doi.org/10.21203/rs.3.rs-3358946/v1.

[50] I. K. Ibrahim, S. A. Elmorsy, N. M. Kashef, and M. M. M. Al-Borai, “Securing e-governance services based on two level classification algorithms,” *Mathematical Modelling of Engineering Problems*, vol. 10, no. 2, pp. 442-450, 2023. https://doi.org/10.18280/mmep.100208

[51] K. I. Jones and S. Suchithra, “Information security: a coordinated strategy to guarantee data security in cloud computing,” *International Journal of Data Informatics and Intelligent Computing*, vol. 2, no. 1, pp. 11-31, 2023. https://doi.org/10.59461/ijdiic.v2i1.34

[52] K. L. Chen, M. L. Shing, H. Lee, and C. C. Shing, “Modeling in confidentiality and integrity for a supply chain network,” *Communications of the IIMA*, vol. 7, no. 1, article no. 4, 2007. https://doi.org/10.58729/1941-6687.1021

[53] B. I. Kang and S. J. Kim, “Study on security grade classification of financial company documents,” *Journal of the Korea Institute of Information Security and Cryptology*, vol. 24, no. 6, pp. 1319-1328, 2014. https://doi.org/10.13089/JKIISC.2014.24.6.1319

[54] G. Y. Jang and I. Kim, “Establishing security level standards and case studies for safe electronic financial transactions,” *Journal of the Korea Institute of Information Security and Cryptology*, vol. 28, no. 3, pp. 729-741, 2018. https://doi.org/10.13089/JKIISC.2018.28.3.729

[55] U. Islam, G. Mehmood, A. A. Al-Atawi, F. Khan, H. S. Alwageed, and L. Cascone, “NeuroHealth guardian: a novel hybrid approach for precision brain stroke prediction and healthcare analytics,” *Journal of Neuroscience Methods*, vol. 409, article no. 110210, 2024. https://doi.org/10.1016/j.jneumeth.2024.110210

[56] M. Katarahweire, E. Bainomugisha, and K. A. Mughal, “Data classification for secure mobile health data collection systems,” *Development Engineering*, vol. 5, article no. 100054, 2020. https://doi.org/10.1016/j.deveng.2020.100054

[57] Y. Nugraha and A. Martin, “Towards the classification of confidentiality capabilities in trustworthy service level agreements,” in *Proceedings of the 2017 IEEE International Conference on Cloud Engineering (IC2E)*, Vancouver, Canada, 2017, pp. 304-310. https://doi.org/10.1109/IC2E.2017.48

[58] Trade Secret Protection Center, “Trade secret level self-verification service,” 2024 [Online]. Available: https://www.tradesecret.or.kr/info/secret/search_list.do.

[59] Z. A. Soomro, M. H. Shah, and J. Ahmed, “Information security management needs more holistic approach: a literature review,” *International Journal of Information Management*, vol. 36, no. 2, pp. 215-225, 2016. https://doi.org/10.1016/j.ijinfomgt.2015.11.009

[60] G. Hong, Y. Han, and W. Yoon, “A study on the classification of display technical documents using graph embedding,” *Korean Journal of Industrial Security*, vol. 13, no. 1, pp. 1-25, 2023. https://www.earticle.net/Article/A425381

[61] J. Jiang, Y. Lu, M. Yu, G. Li, C. Liu, S. An, and W. Huang, “CES2Vec: a confidentiality-oriented word embedding for confidential information detection,” in *Proceedings of the 2020 IEEE Symposium on Computers and Communications (ISCC)*, Rennes, France, 2020, pp. 1-7. https://doi.org/10.1109/ISCC50000.2020.9219702

[62] G. Katz, Y. Elovici, and B. Shapira, “CoBAn: a context based model for data leakage prevention,” *Information Sciences*, vol. 262, pp. 137-158, 2014. https://doi.org/10.1016/j.ins.2013.10.005

[63] P. E. Engelstad, H. Hammer, A. Yazidi, and A. Bai, “Advanced classification lists (dirty word lists) for automatic security classification,” in *Proceedings of the 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, Xi’an, China, 2015, pp. 44-53. https://doi.org/10.1109/CyberC.2015.103

[64] K. Alzhrani, F. S. Alrasheedi, F. A. Kateb, and T. E.Boult, “CNN with paragraph to multi-sequence learning for sensitive text detection,” in *Proceedings of the 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, 2019, pp. 1-6. https://doi.org/10.1109/CAIS.2019.8769490

[65] Gartner Inc., “Toolkit: Creating data classification schemes,” 2011 [Online]. Available: https://www.gartner.com/en/documents/1877520.

[66] Gartner Inc., “Toolkit: Creating Data Classification Schemes,” 2011 [Online]. Available: https://www.gartner.com/en/documents/1877520.

[67] Gartner Inc., “Data Classification Policy,” 2018 [Online]. Available: https://www.gartner.com/en/documents/3895169.

[68] Gartner Inc., “Building an effective sensitive data classification and handling policy,” 2010 [Online]. Available: https://www.gartner.com/en/documents/1306014.

[69] Trade Secret Protection Center, “Trade secret classification guide for corporate,” 2021 [Online]. Available: https://www.tradesecret.or.kr/bbs/precedentView.do?gb=211.

[70] M. Shakir, A. Abubakar, O. Yousoff, M. Waseem, and M. Al-Emran, “Model of security level classification for data in hybrid cloud computing,” *Journal of Theoretical and Applied Information Technology*, vol. 94, no. 1, pp. 133-141, 2016. https://www.jatit.org/volumes/Vol94No1/13Vol94No1.pdf

[71] I. Gupta, S. Mittal, A. Tiwari, P. Agarwal, and A. K. Singh, “TIDF-DLPM: term and inverse document frequency based data leakage prevention model,” 2022 [Online]. Available: https://arxiv.org/abs/2203.05367.

[72] T. D. Oyetoyan and P. Morrison, “An improved text classification modelling approach to identify security messages in heterogeneous projects,” *Software Quality Journal*, vol. 29, pp. 509-553, 2021. https://doi.org/10.1007/s11219-020-09546-7

[73] Y. Park, “A text mining approach to confidential document detection for data loss prevention,” IBM Research Technical Report RC25055, 2010. https://dominoweb.draco.res.ibm.com/reports/rc25055.pdf

[74] A. J. Blazic and S. Saljic, “Confidentiality labeling using structured data types,” in *Proceedings of 2010 4th International Conference on Digital Society*, Sint Maarten, Netherlands Antilles, 2010, pp. 182-187. https://doi.org/10.1109/ICDS.2010.70

[75] S. Oudkerk, I. Bryant, A. Eggen, and R. Haakseth, “A proposal for an XML confidentiality label and related binding of metadata to data objects (RTO-MP-IST-091),” 2009 [Online]. Available: https://apps.dtic.mil/sti/tr/pdf/ADA584044.pdf.

[76] W. Cai and H. Yao, “Research on information security risk assessment method based on fuzzy rule set,” *Wireless Communications and Mobile Computing*, vol. 2021, article no. 9663520, 2021. https://doi.org/10.1155/2021/9663520

[77] R. Chow, P. Golle, and J. Staddon, “Detecting privacy leaks using corpus-based association rules,” in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, NV, USA, 2008, pp. 893-901. https://doi.org/10.1145/1401890.1401997

[78] S. J. M. Escobar, J. R. Shulcloper, C. J. Landin, J. S. Ruiz-Castilla, and O. A. P. Garcia, “STClass: a method for determining the sensitivity of documents,” in *Advances in Soft Computing*. Cham, Switzerland: Springer, 2021, pp. 140-152. https://doi.org/10.1007/978-3-030-89820-5_11

[79] J. Jiang, Y. Lu, M. Yu, G. Li, Y. Jia, J. Guo, C. Liu, and W. Huang, “CIDetector: semi-supervised method for multi-topic confidential information detection,” *Frontiers in Artificial Intelligence and Applications*, vol. 325, pp. 1834-1841, 2020. https://doi.org/10.3233/FAIA200299

[80] K. Alzhrani, E. M. Rudd, C. E. Chow, and T. E. Boult, “Automated us diplomatic cables security classification: topic model pruning vs. classification based on clusters,” in *Proceedings of the 2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, Waltham, MA, USA, 2017, pp. 1-6. https://doi.org/10.1109/THS.2017.7943471

[81] P. E. Engelstad, “On the usability of clustering for topic-oriented multi-level security models,” in *Proceedings of the 2015 IEEE European Modelling Symposium (EMS)*, Madrid, Spain, 2015, pp. 14-20. https://doi.org/10.1109/EMS.2015.13

[82] J. D. Brown and D. Charlebois, “Security classification using automated learning (SCALE): optimizing statistical natural language processing techniques to assign security labels to unstructured text,” Defence R&D Canada, 2010 [Online]. Available: https://apps.dtic.mil/sti/tr/pdf/ADA551452.pdf.

[83] K. Akuthota, A. Ganesh, B. R. Avula, and S. Depuru, “Machine learning models for classification of sensitive financial documents,” in *Proceedings of the 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, Hamburg, Germany, 2023, pp. 334-340. https://doi.org/10.1109/ICCCMLA58983.2023.10346685

[84] E. Alparslan and H. Bahsi, “Security level classification of confidential documents written in Turkish,” in *User Centric Media*. Heidelberg, Germany: Springer, 2010, pp. 329-334. https://doi.org/10.1007/978-3-642-12630-7_41

[85] W. Xing and Y. Bei, “Medical health big data classification based on KNN classification algorithm,” *IEEE Access*, vol. 8, pp. 28808-28819, 2019. https://doi.org/10.1109/ACCESS.2019.2955754

[86] M. A. Zardari and L. T. Jung, “Data security rules/regulations based classification of file data using TsF-kNN algorithm,” *Cluster Computing*, vol. 19, pp. 349-368, 2016. https://doi.org/10.1007/s10586-016-0539-z

[87] A. Inani, C. Verma, and S. Jain, “A machine learning algorithm TsF K-NN based on automated data classification for securing mobile cloud computing model,” in *Proceedings of the 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, Singapore, 2019, pp. 9-13. https://doi.org/10.1109/CCOMS.2019.8821756

[88] B. Sheehan, F. Murphy, M. Mullins, and C. Ryan, “Connected and autonomous vehicles: a cyber-risk classification framework,” *Transportation Research Part A: Policy and Practice*, vol. 124, pp. 523-536, 2019. https://doi.org/10.1016/j.tra.2018.06.033

[89] A. Koltays, A. Konev, and A. Shelupanov, “Mathematical model for choosing counterparty when assessing information security risks,” *Risks*, vol. 9, no. 7, article no. 133, 2021. https://doi.org/10.3390/risks9070133

[90] K. Alzhrani, E. M. Rudd, C. E. Chow, and T. E. Boult, “Automated big security text pruning and classification,” in *Proceedings of the 2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2016, pp. 3629-3637. https://doi.org/10.1109/BigData.2016.7841028

[91] S. F. Ahmed, M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A. B. M. S. Ali, and A. H. Gandomi, “Deep learning modelling techniques: current progress, applications, advantages, and challenges,” *Artificial Intelligence Review*, vol. 56, pp. 13521-13617, 2023. https://doi.org/10.1007/s10462-023-10466-8

[92] J. W. Huang, C. W. Chiang, and J. W. Chang, “Email security level classification of imbalanced data using artificial neural network: the real case in a world-leading enterprise,” *Engineering Applications of Artificial Intelligence*, vol. 75, pp. 11-21, 2018. https://doi.org/10.1016/j.engappai.2018.07.010

[93] E. Alparslan, A. Karahoca, and H. Bahsi, “Classification of confidential documents by using adaptive neurofuzzy inference systems,” *Procedia Computer Science*, vol. 3, pp. 1412-1417, 2011. https://doi.org/10.1016/j.procs.2011.01.023

[94] D. Patil, R. Lokare, and S. Patil, “An overview of text representation techniques in text classification using deep learning models,” in *Proceedings of 2022 3rd International Conference for Emerging Technology (INCET)*, Belgaum, India, 2022, pp. 1-4. https://doi.org/10.1109/INCET54531.2022.9825389

[95] Y. Lu, J. Jiang, M. Yu, C. Liu, C. Liu, W. Huang, and Z. Lv, “Depthwise separable convolutional neural network for confidential information analysis,” in *Knowledge Science, Engineering and Management*. Cham, Switzerland: Springer, 2020, pp. 450-461. https://doi.org/10.1007/978-3-030-55393-7_40

[96] S. Achar, N. Faruqui, A. Bodepudi, and M. Reddy, “Confimizer: a novel algorithm to optimize cloud resource by confidentiality-cost trade-off using BiLSTM network,” *IEEE Access*, vol. 11, pp. 89205-89217, 2023. https://doi.org/10.1109/ACCESS.2023.3305506

[97] P. Costa, R. Montenegro, T. Pereira, and P. Pinto, “The security challenges emerging from the technological developments: a practical case study of organizational awareness to the security risks,” *Mobile Networks and Applications*, vol. 24, pp. 2032-2037, 2019. https://doi.org/10.1007/s11036-018-01208-0

[98] Z. Lu, L. He, D. Zhang, B. Zhao, J. Zhang, and H. Zhao, “A security level classification method for power systems under N-1 contingency,” *Energies*, vol. 10, no. 12, article no. 2055, 2017. https://doi.org/10.3390/en10122055

[99] Berkeley Information Security Office, “Data and IT Resource Classification Standard,” 2024 [Online]. Available: https://security.berkeley.edu/data-classification-standard.

[100] X. Yu, Z. Tian, J. Qiu, and F. Jiang, “A data leakage prevention method based on the reduction of confidential and context terms for smart mobile devices,” *Wireless Communications and Mobile Computing*, vol. 2018, article no. 5823439, 2018. https://doi.org/10.1155/2018/5823439

[101] T. D. H. Nguyen, “Financial News Dataset from Reuters,” 2016 [Online]. Available: https://github.com/duynht/financial-news-dataset.

[102] TUBITAK UEKAE [Online]. Available: https://en.bilgem.tubitak.gov.tr/en/.

[103] W. W. Cohen, “Enron dataset,” 2015 [Online]. Available: https://www.cs.cmu.edu/~enron/.

[104] G. Feng, T. Yan, and J. Yang, “Attribute-consistency reversible pedestrian de-identification in intelligent transportation,” *IEEE Transactions on Vehicular Technology*, vol. 74, no. 2, pp. 2187 - 2197, 2024. https://doi.org/10.1109/TVT.2024.3391834

[105] D. Sonntag, “Assessing the quality of natural language text data,” 2004 [Online]. Available: https://dl.gi.de/items/98df0122-7ccf-4071-a442-a170aeebfb09.