Journal of Information Security, 2023, 14, 76-91
https://www.scirp.org/journal/jis
ISSN Online: 2153-1242
ISSN Print: 2153-1234

Classification of Big Data Security Based on Ontology Web Language

Alsadig Mohammed Adam Abdallah[1], Amir Mohamed Talib[2]

[1]College of Computer and Information Sciences (CCIS), Riyadh, Kingdom of Saudi Arabia (KSA)
[2]Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Kingdom of Saudi Arabia (KSA)
Email: sadigmo86@gmail.com, ganawa53@yahoo.com

How to cite this paper: Abdallah, A.M.A. and Talib, A.M. (2023) Classification of Big Data Security Based on Ontology Web Language. *Journal of Information Security*, **14**, 76-91.
https://doi.org/10.4236/jis.2023.141006

Received: December 10, 2022
Accepted: January 28, 2023
Published: January 31, 2023

Abstract

A vast amount of data (known as big data) may now be collected and stored from a variety of data sources, including event logs, the internet, smartphones, databases, sensors, cloud computing, and Internet of Things (IoT) devices. The term "big data security" refers to all the safeguards and instruments used to protect both the data and analytics processes against intrusions, theft, and other hostile actions that could endanger or adversely influence them. Beyond being a high-value and desirable target, protecting Big Data has particular difficulties. Big Data security does not fundamentally differ from conventional data security. Big Data security issues are caused by extraneous distinctions rather than fundamental ones. This study meticulously outlines the numerous security difficulties Large Data analytics now faces and encourages additional joint research for reducing both big data security challenges utilizing Ontology Web Language (OWL). Although we focus on the Security Challenges of Big Data in this essay, we will also briefly cover the broader Challenges of Big Data. The proposed classification of Big Data security based on ontology web language resulting from the protégé software has 32 classes and 45 subclasses.

Keywords

Big Data, Big Data Security, Information Security, Data Security, Ontology Web Language, Protégé

1. Introduction

The ability to gather and store massive amounts of data (known as big data) from various data sources, including event logs, the internet, smartphones, databases, sensors, IoT devices, etc., has been made possible by technological ad-

vancements in recent years [1]. These data are gathered, studied, and compared to one another to produce meaningful information that is frequently used in decision-making.

Relational database management systems and desktop visualization/static software are ineffective for handling big data; instead, massively parallel software running on tens of thousands or even millions of servers is needed. The enormous attention big data is receiving is more due to the fact that analysis of just one large set of related data can yield so much more information than analysis of smaller separate sets with the same total amount of data, thus allowing correlations to be found. This allows you to spot business trends, judge the caliber of research, predict disease spread, prevent disease, fight crime, and so much more.

The term "big data security" refers to all the safeguards and instruments used to protect both the data and analytics processes against intrusions, theft, and other hostile actions that could endanger or adversely influence them. Beyond being a high-value and desirable target, protecting Big Data has particular difficulties. Big Data security is not essentially distinct from conventional data security. Big Data security issues are caused by extraneous distinctions rather than fundamental ones.

The typical definition of an ontology is a "explicit specification of a conception" [2]. This indicates that the specification representation offers a formal semantic of the specification and that ontology permits the defining of concepts and interactions between these concepts. Out of all the common models for knowledge representation, ontologies have the highest level of semantic richness [2]. Although none of these models achieves the level of semantic richness that ontologies provide, they are based on models that will be explained in sequence of increasing degree of semantic richness. Out of all the models listed above, a glossary has the least amount of semantic richness. A glossary is a list of words in alphabetical order with their definitions but no explanation of how these words relate to one another. Taxonomy is a model for the nested classification of words at the next level of semantic richness. It uses super- and sub-relations to describe the relationships between words. These relations provide these concepts an ordering of generality. A thesaurus is a taxonomy that has been expanded. A thesaurus explains all possible word relationships. The model that most closely resembles an ontology is a topic map. An abstract model and data format for the creation of knowledge structures is a topic map. The relationships between various themes are described via associations, which are thesaurus-based. Additionally, topic maps can contain external documents that have been embedded via occurrences [3].

Although we focus on the security concerns of big data in this essay, we will also quickly go through the general issues with large data.

1.1. Big Data Characteristics

Volume, Variety, Velocity, Veracity, Value, Variability, Exhaustive, Fine-grained

and uniquely lexical, Relational, Extensional and Scalability are the traits that best represent big data as shown in Figure 1.

1.2. Big Data Technologies

Several tools for evaluating big data exist, including A/B testing, machine learning, and natural language processing. Databases, cloud computing, business information, and visuals like graphs and charts as shown in Figure 2.

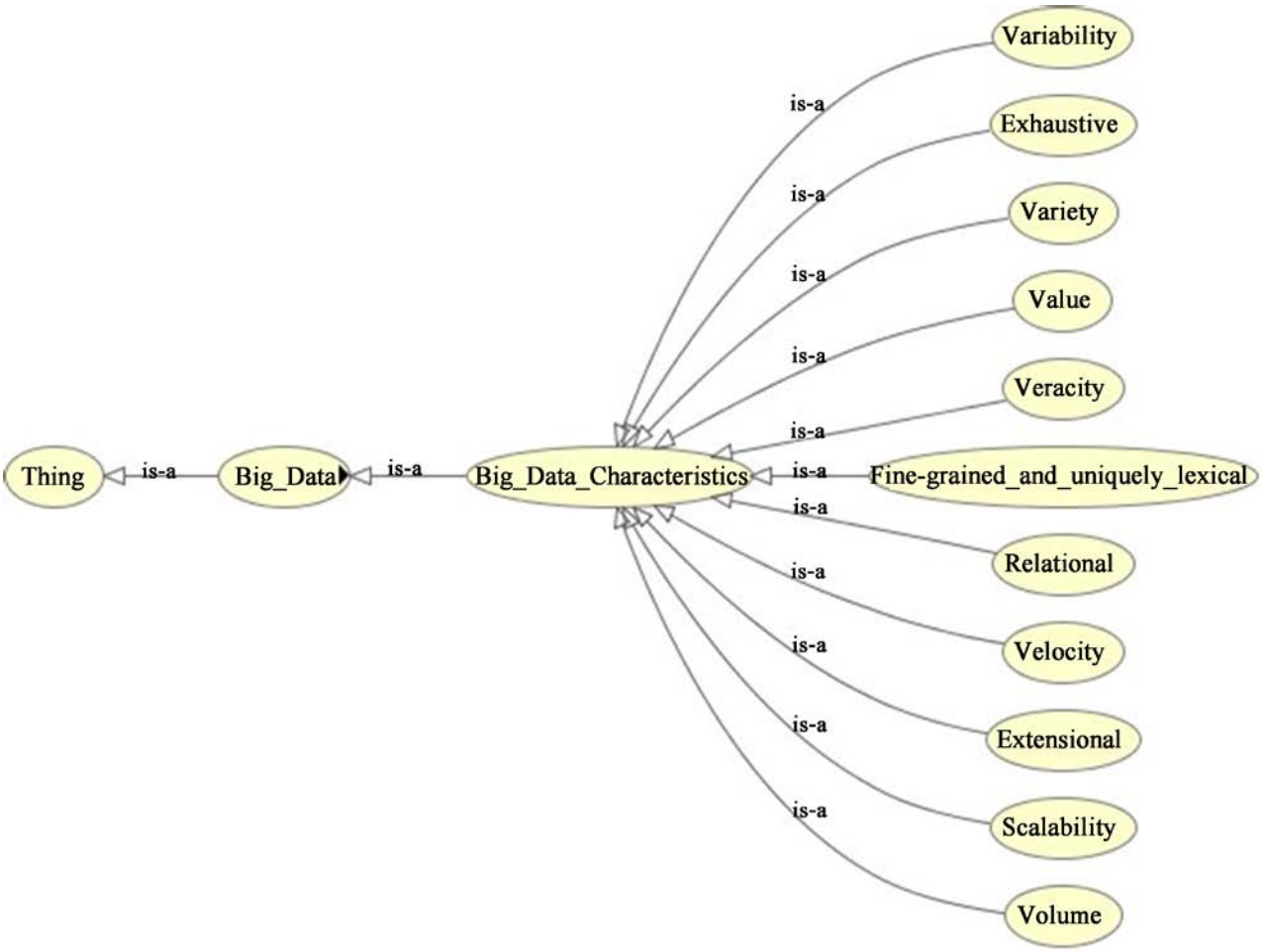

Figure 1. Big data characteristics.

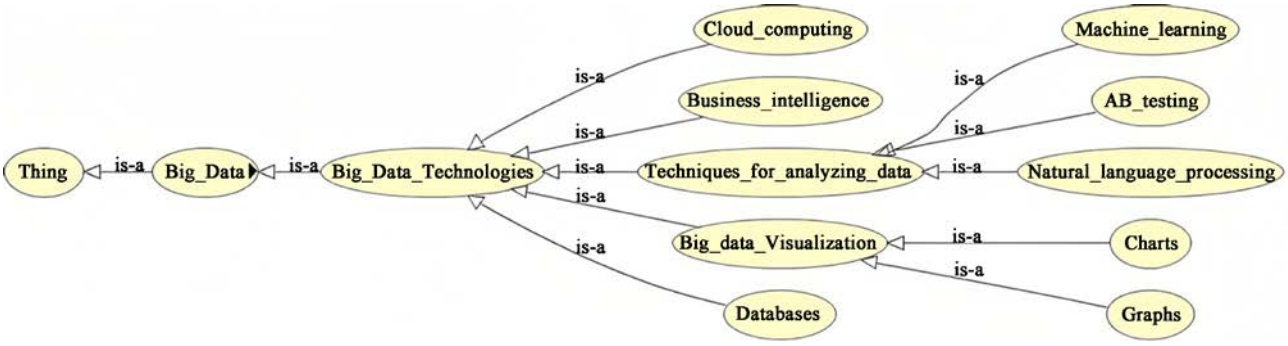

Figure 2. Big data technologies.

1.3. Big Data Applications

Government, international development, healthcare, education, media, insurance, internet of things, and information technology are the industries that use big data the most, as shown in Figure 3.

1.4. Big Data Lifecycle

There are various stages of dealing with the data throughout the big data life cycle. The following are these phases: Data integration, storage and administration, processing, and analysis of the data as shown in Figure 4.

2. Big Data Security Requirements

Managing information security while controlling vast and quick data streams is one difficulty in the Big Data setting. Therefore, security technologies should be

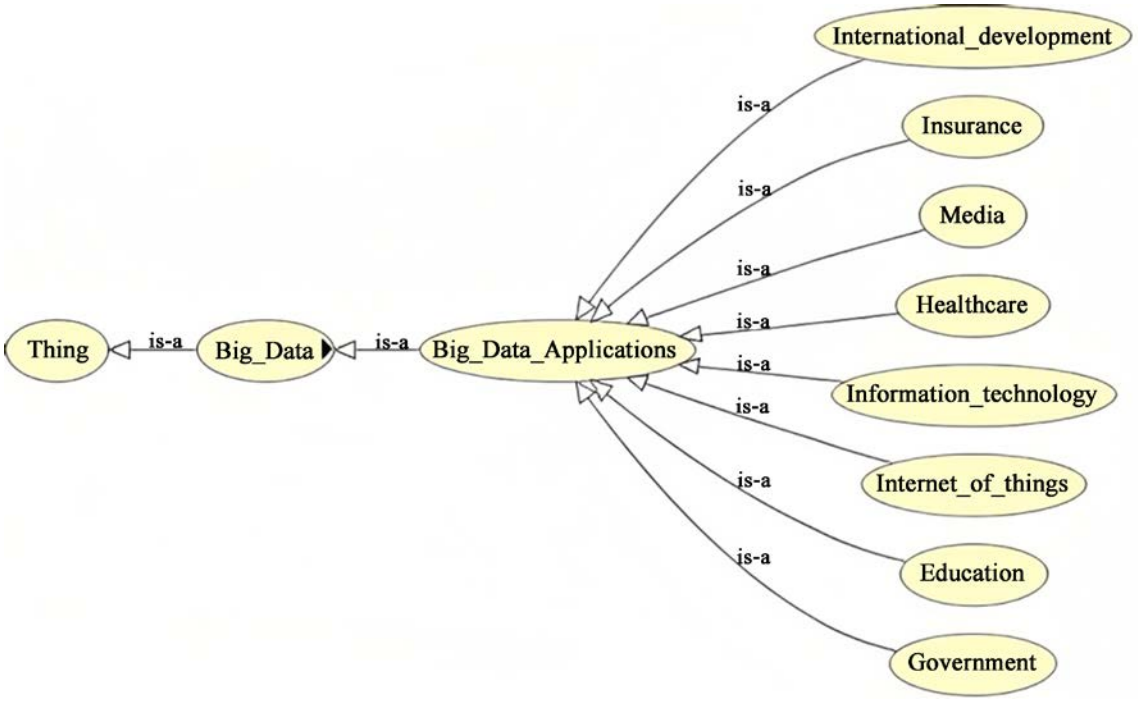

Figure 3. Big data applications.

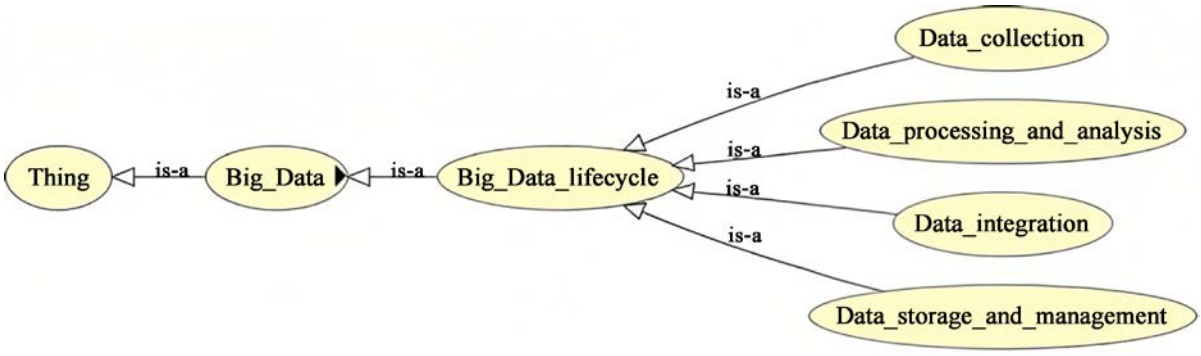

Figure 4. Big data lifecycle.

adaptable and simple to scale in order to facilitate the integration of future technological advancements and handle adjustments in application needs.

Finding a compromise between various security demands, privacy duties, system efficiency, and quick dynamic analysis on various massive data sets is necessary (data in motion or static, private and public, local or shared, etc.).

3. Big Data Security Challenges

According to [3] [4], there are two primary components to security in the context of big data: information security and data security as shown in **Figure 5**.

Big Data security generally seeks to provide real-time monitoring to identify security risks, vulnerabilities, and aberrant behaviors; granular role-based access control; robust protection of personal information; and the creation of security performance indicators. In the event of a security incident, it promotes quick decision-making. The difficulties in achieving these objectives are listed and explained in the following sections.

4. Securing Data of Big Data

Organizations are under pressure to find innovative ways to produce and deliver value to customers through supply chain management because of the world economy's rapid growth and improvements in customers' use of information technology [4]. Businesses will be more successful if they collaborate with other businesses to cut costs, manufacture high-quality goods, and maximize the value added from service for their clients. Additionally, it has been argued that a company will be more competitive if it outsources some of its manufacturing to companies that are unrelated to its main industry.

A service supply chain, according to Arlbjrn *et al.* [5], is a broad notion that includes companies that deal with things like the provision of replacement parts, third-party providers, finance, insurance, retail, and governmental services. The definition of a service supply chain system is a network of suppliers, service providers, consumers, and other supporting units that carry out transactions involving the resources needed to produce services, transform these resources into support and core services, and provide these services to clients [6]. Finance, telephony, internet service, mobile apps, and tourism are among the service industry sectors represented in a service supply chain [7]. A service provider must be creative in developing offerings that set itself apart from its rivals [8]. By scanning the business environment, a corporation can successfully execute a service supply chain by understanding the business processes, supply chain networks, and end-user needs and wants.

Figure 5. Big data security challenges.

In light of the fact that both internal and external elements have an impact on a customer's spending behavior, [9] defines scanning the environment as the process of gathering and applying information about events, trends, and relationships in an organization's internal and external environments. The status quo, education, employment, and religion are a few examples of internal influences. External elements include the economy, environment, family, and friends [10] [11]. A company must occasionally evaluate consumer data that has been categorized into various segments to include both current and potential customers. A corporation must invest a lot of effort to accurately predict client desires because market segmentation changes frequently [12]. In order to achieve lower operational expenses, a service supply chain's performance must be quick and adaptable in response to customer demand [13]. Creating a framework for measuring supply chain performance will have a big impact on identifying success factors, bottlenecks, waste, operational problems, customer needs that must be satisfied, efficient business procedures, factual decisions, tracking progress, and making improvement suggestions [14] [15].

Gawankar *et al.* [16] asserts that further investigation is still required for supply chain performance measurement to advance through study and application. As a result, to compartmentalize market segmentation and demand forecasting via supply chain innovation capabilities, service supply chain management requires trustworthy technologies for data analysis [17]. Service firms have employed Big Data analytics from structured and unstructured data to improve the performance of the service supply chain and leverage firm performance in order to achieve business optimization through innovation. In order to anticipate and satisfy the wants of contemporary clients, big data technology has thus become an essential component of corporate operations and strategy.

Increased flexibility, responsiveness, customer service, and reliability are just a few of the benefits that big data analytics offer service providers [12]. Big Data analytics adoption is crucial for logistics and supply chain operations because better supply chain performance depends on timely, accurate supply chain choices. The bullwhip effect in the supply chain, which results in inefficiencies among the distribution channels, can be resolved through big data analytics. [18] asserts that because Big Data predictive analytics' capabilities is not clearly defined, its potential impact on supply chain performance may be constrained.

There is a need for empirical publications that examine the Big Data analytics that lower standard deviations (demand variance) and add a signaling aspect to forecasting [19]. However, academics and business have started to pay attention to how businesses are using big data analytics, particularly as concerns about data security and privacy have arisen. Service providers have gathered information in the form of video files, status updates, and likes, shares, follows, retweets, and comments from social media feeds that are open sources.

Additionally, big data analytics can make use of information from customer relationship management and enterprise resource planning systems. A signifi-

cant volume of data obtained from many sources comes with privacy and security threats as a result of the availability of data in both structured and unstructured formats. Most businesses have limitations on systematic approaches for ensuring appropriate data access mechanisms, and existing non-Big Data security solutions are frequently not designed to handle the scale, speed, variety, and complexity of Big Data [20]. As a result, service companies lack the analytical tools and techniques to generate useful insights from data to drive strategy and improve service and business performance [21].

Despite the knowledge that Big Data analytics can aid in decision-making and supply chain management processes, many businesses have had trouble implementing the technology. There are various causes for this. One of the reasons why businesses are hesitant to utilize big data is the shortage of workers with the necessary expertise to conduct the analysis. Another is that there hasn't been much investigation into how supply chains may be impacted by Big Data applications (Waller and Fawcett [22]). The literature currently available on this subject is somewhat underemphasized [23], despite the fact that Big Data analytics is helpful in supporting service organizations in generating new products and services.

Big Data is also seen to carry substantial dangers for data security and privacy if service providers are unable to use it properly.

The sparse development of theoretical knowledge on supply chain innovation capacities has been facilitated by the paucity of literature on the relationship between Big Data analytics and data security issues and service supply chain performance. A service organization needs creativity and innovation as competitive weapons to enhance supply chain innovation capabilities. It is challenging to quantify the traditional performance issue of service supply chain operations. This is due to the difficulty in standardizing, visualizing, designing, and delivering services due to their nature [24]. The contribution and effort of the service provider can affect the performance of the service as an independent entity. Some of the best practices for services can be taken from the manufacturing sector, however in some cases, the service business needs particular service performance outcomes [25].

Zhou *et al.* [26] provide a game theory-based secure incentive mechanism and patient-optimized privacy-preserving packet forwarding algorithm that are researched on the mobile health-care system. The patient-optimized privacy (POP) technique of packet transmission uses the marriage idea to ensure the confidentiality of protected health information. In m-health care, the identity (ID) of the patients is employed for optimization.

Big data integrates healthcare data for analytics management, security, and privacy for the well-known decision-making pattern. The author created a methodical decision-making process. Social and economic characteristics are used to identify cases of TB and the human immunodeficiency virus. Yang *et al.* [27] created smart IoT health-care big data storage and self-adaptive access control

for privacy preservation. The paper employs dual access control for both routine and urgent situations. The privilege to safeguard the historical data uses the attribute secret key.

The researcher's focus is drawn to the difficulties in cloud system security by the rise of big data. Discussed are the fundamental ideas of big data architecture and data risk assessment. Big data governance and security controls application mechanisms, however, are not covered.

The cloud service providers face a significant problem in protecting cloud systems from potential dangers. Insider and external attacks are the two categories that are recognized. Insider attack implementation is seen as the cloud security's bottleneck. It is suggested that a cybersecurity architecture be used to support insider attacks on cloud systems. The Hidden Markov model technology is employed in this framework design to predict the behavior of the edge devices, which are divided into four classes based on how they affect cloud security: legal, sensitive, under-attacked, and compromised [28]. This framework moves the edge devices to a virtual honeypot device, a cutting-edge innovation that identifies malicious edge devices and allows for the future tracking, forecasting, and prevention of attacker activity. The preventative measures, however, are never discussed nor introduced.

The cloud security approach introduces the use of encryption tools like Security Information and Event Management SIEM and Data Loss Prevention DLP to safeguard sensitive data from potential hazards. But neither explicit big data categorization techniques nor settings for crucial data indicators are offered, which makes them challenging to put into practice.

Big data security threats from phishing are significant [29]. To demonstrate how phishers might target large data by producing bogus emails to obtain their important information, a case study is used. However, there is no phishing prevention method used for bogus.

There are various levels of big data protection, including communications, processing, authentication, and storage. This article presents a big data integrity and access control technique that safeguards data while it is being processed and stored. Researchers in have explained how to apply a data protection management policy to massive amounts of data without degrading performance. However, the protection mechanism cannot be used because of the high time complexity required. The suggested data protection solutions are also challenging to put into practice because huge data resources vary and are governed by various policies.

A security hardening technique [29] based on attribute relation graphs was presented, focusing less on data protection and more on the value of data and how to extract valuable information from data. To safeguard sensitive data, the Computing on Masked Data (CMD) tool combines RND and DET cryptographic techniques. CMD approaches are intended to enhance the secrecy, integrity, and availability of the cloud system. However, CMD approaches were unable to

account for the overhead during the time-consuming masking procedure, and the management of keys calls for effective solutions.

Sahafizadeh and Nematbakhsh [29] discuss a technique for deciding which qualities should be safeguarded in order to secure huge data. Big data is seen as a single item with unique features that are prioritized in accordance with their significance. However, no ranking strategy can account for the characteristics of unstructured huge data.

The difficulty of managing huge data across a range of scientific disciplines is described in. Knowledge of the Map Reduce High Performance Computer Cluster (HPCC).

Data sets are defined, and the necessary information is immediately extracted, using data discovery and data mining techniques. In any case, huge data cannot benefit by data mining techniques, a group of large data security traits.

To gather and stabilize various data kinds, as well as handle massive amounts of rapidly changing data in real-time applications for security analysis, some automated techniques are used. However, no approach for overcoming security concerns and privacy violations is presented, and there is no explanation of how to safeguard massive data from many sources against potential dangers. In order to attain data privacy, which has extra needs that could be strengthened by encryption and concealed data access control systems, data confidentiality is thought to be crucial [30].

But huge data cannot use confidentiality measures because they are too difficult to implement, such as access control and encryption. An intelligently powered monitoring security model. This model uses a dynamic, self-adaptive questionable user log system, as well as self-assuring software that includes a package that incorporates the suspicious behavior keywords entered by users, to increase security. The fishy user is labeled as a crucial user when their number of deviant behaviors exceeds a certain threshold. To determine if the activity is a dubious action or simply motivated by curiosity, a model analysis is used. The self-assuring architecture also includes a library for identifying and archiving the keywords of unusual user activities and associated crucial logs. Although there is no definition of typical conduct and no security against data loss and data leaking.

In order to process large volumes of data from many sources, high performance cryptography, data provenance, security visualization, and qualified employees are all suggested as novel threats detection techniques [31]. Threats to information security are discussed in detail, as is the SIEM system. The proposed approaches and tools, however, were unable to protect huge data and could not adhere to standard practices in information security.

Big data security was separated into two categories by the big data research [32]: access control and information security. The suggested security hardening methodology made use of attribute relation graphs and put more emphasis on the value of the data than the actual data. However, neither a definition nor a description of the link between the protected attributes in the selection method

is provided.

When real-time applications leverage the security of big data analytics techniques to provide useful insight and important meanings from data streams, a great requirement for cybersecurity setups arises. On local network-related data sets, the suggested security analytics technique is evaluated with the goal of lowering the false positive rate of a prediction model for fraudulent transactions. However, local network-related data sets are insufficient to evaluate the prediction model against large data that is produced from numerous sources and contains a variety of data kinds. The Hadoop Ecosystem's privacy and security vulnerabilities are described in [33]. Multiple encryption mechanisms must be used since the Name Node and the Data Node have complete authority over the data. This is necessary to stop unauthorized individuals from accessing the data. Anyhow, the high time complexity and cost of this security method render it useless.

The capacity of the suggested integrated classification and security method in this study to address the confidentiality level of the file contents sets it apart from competing methods. In fact, our suggested approach includes a data security algorithm that was created particularly to reduce the dangers associated with the aggregation and movement of enormous amounts of sensitive data.

Many academics have employed different cryptographic techniques to secure data in cloud storage. To safeguard data security from a nosy cloud and achieve secure data exchange, Xu *et al.* devised a generic hybrid proxy re-encryption (PRE) technique [34]. Then, in order to accomplish fine-grained data sharing delegation, Zeng *et al.* devised a conditional PRE scheme.

To implement fine-grained access control over outsourced encrypted data, attribute-based encryption was added to cloud computing [35]. Significant attention has also been paid to the challenge of ensuring the secrecy of the encrypted data while allowing a semi-trusted cloud to compute between ciphertexts. Homomorphic encryption was consequently introduced [36].

Another popular problem is how to securely search through encrypted data. Searchable encryption has been suggested to achieve this goal. The searchable symmetric encryption (SSE) system was first introduced by Song *et al.* Using dynamic searchable symmetric encryption, Kamara *et al.* made SSE more useful (DSSE). The usability and security of DSSE were then improved by Xu *et al.* A new structure for forward and backward private DSSE was recently developed by Ghareh Chamani *et al.* [37], which lowers SSE leakage.

For a single keyword search, Boneh was the first to introduce the public-key encryption with keyword search (PEKS) approach. Following the initial PEKS study, some researchers worked hard to make PEKS adaptable. For instance, Boneh and Waters provided a PEKS system that supports range, subset, and conjunctive searches, while Shi offered a multi-dimensional range query strategy on ciphertexts. However, the low retrieval efficiency with linear retrieval complexity presented a challenge for the PEKS methods. The structured PEKS approach was first introduced in 2015 by Xu *et al.* [38], who also achieved sub-linear retrieval

complexity. The use of PEKS in IoT scenarios was then further investigated by Xu *et al.* [38], who also suggested a lightweight PEKS method for cloud-assisted wireless sensor networks and a parallel keyword search scheme for the IoT. For secure outsourcing storage, remote data integrity is a consideration in addition to secrecy. In order to confirm the integrity of the cloud data, proof of data possession and proof of data retrievability [38] were proposed.

Numerous security-related topics, including trusted devices, access control, network security, and intrusion detection, are being researched. In an effort to impose security isolation, Pettersen *et al.* built a prototype using secure enclave technology on edge devices.

In order to deploy policy enforcement components in mobile edge computing, Vassilakis used a formal methodology. Pimentel presented a secure communication protocol for federated content networks to enable secure communication in edge environments. A deep-learning-based model for mobile edge computing was presented by Chen *et al.* [39] to identify harmful applications at the edge of cellular networks.

The security of either cloud storage or edge storage is the only factor taken into account in any of the aforementioned works. Therefore, there is hardly any research on the security of collaborative cloud-edge storage. One of the closest works to safeguard the data privacy of outsourced storage in the cloud-and-edge-assisted IoT was proposed by Mollah *et al.* [40]. This work showed that it was possible to exchange and search outsourced data in the cloud-edge-collaborative model while maintaining privacy by using searchable encryption (SE) and another cryptographic technique. The system is insufficiently secure, though, as edges can get all mobile object private keys. Since the data-search secret key is shared by all edges, it is easily possible for any edge server to be compromised and subsequently used to compromise the security of the entire system.

5. Big Data Security Information

Readers today choose electronic literature resources over paper ones in the big data era. The National Library of China reports that its electronic literary resources are accessed more than 7 million times annually whereas only 200,000 to 80,000 people borrow its print resources [41]. Big data applications in libraries face a challenging problem: how to manage and make use of literature resources.

Integration is crucial for literature sources used in papers. In order to integrate and arrange these paper literature resources in a location that is simple for readers to find, libraries can use sensor data to forecast which paper literature resources will be most popular with readers. Other paper literature resources can be removed from the bookcases or stacked compactly [41]. Libraries can also compute the use rate from the borrow rate in order to incorporate the paper literature resource. To satisfy the needs of the readers, paper literary resources are integrated. Digitization is crucial for resources for electronic literature. In addition to combining physical and digital libraries, print and electronic literature

resources, and paper and electronic literature resources, the digitization of literature resources must encourage the sharing of such resources.

Numerous industries, including insurance, telecommunication, social communities, and others have looked into the issue of churn. To date, a variety of methods have been put out to deal with the churn prediction issue. The primary methods include decision trees 5, logistic regression 3, and support vector machines (SVM) 4. In order to predict the churner in prepaid mobile telephony, Archaux *et al.* [41] used SVM, and the effectiveness of SVM and ANN (artificial neural networks)6 was also evaluated. To solve the churn prediction problem, Auet suggested a data mining system based on evolutionary learning. To predict churners for telecom carriers, Idris *et al.* [42] used random forest, rotation forest, RotBoost, and adorn ensembles.

In their study of three data mining methods for predicting churn in newspaper services, Coussement and van den Poel found that the random forest method outperformed logistic regression and SVM. These researches are all concerned with employing data mining techniques to enhance the accuracy of prediction models, but none of them take into account how social factors affect user turnover.

Another method of predicting user attrition is through social network analysis (SNA) 7. Social network analysis can help to improve the current churn models by looking at the user's communication patterns. As an illustration, Ngonmang created a reliable statistical model to calculate the likelihood that a user will quit the social network based on the properties of the graph. In order to forecast possible churners, Dasgupta *et al.* [43] evaluated the likelihood of user churn based on the neighbors who had already left the system. A form of SNA that can also be considered is information propagation.

To improve the churn prediction performance, Phadke and Zhang *et al.* [44] adopted the receiver-centric propagation model and the sender-centric propagation model, respectively. Kusuma demonstrated, however, that the SNA technique is not usually applicable and that churn prediction in European prepaid users cannot be effectively improved. Additionally, the studies in-depth examined the influence and spread of information. Myers *et al.* [45] investigated how information reaches the nodes in the social network and quantified the external affects over time, whereas Gomez-Rodriguez focused on the issue of tracking channels of diffusion and influence through networks. In contrast to other research, our work presents a thorough paradigm for churn analysis that takes into account both subscriber demographics and social influence. Our study can not only identify subscribers with strong negative influences and important characteristics connected to subscriber turnover, but it can also predict the likelihood of subscriber churn. Our analysis is grounded in actual telecom big data, and the outcomes are more thorough and convincing.

6. Conclusions

Big Data has established itself. By properly analyzing both streaming and static

massive data sets, we can progress many fields of science and medicine and increase the profitability of many businesses. It's nearly impossible to envision the next application without data consumption, data creation, and data-driven algorithms. Security, access control, compression, encryption, and compliance present issues that need to be handled methodically as computer environments become more affordable, application environments become networked, and system and analytics environments are shared over the cloud. In order to make Big Data processing and computing infrastructure much safer, the most pressing Big Data security issues have been outlined in this study. There are 32 classes and 45 subclasses in the suggested classification of Big Data security based on ontology web language produced by the protégé program. This report will encourage the research and development community to focus together on the challenges preventing more security in Big Data platforms and upcoming projects.

In the future, we are going to extend the research by integrating the privacy of Big Data to remark the ontology more comprehensive of Big Data era.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Laudon, K. and Laudon, J. (2020) Management Information Systems: Managing the Digital Firm. Pearson, Rotherham.

[2] Gruber, T.R. (1995) Toward Principles for the Design of Ontologies Used for Knowledge Sharing? *International Journal of Human-Computer Studies*, **43**, 907-928. https://doi.org/10.1006/ijhc.1995.1081

[3] Hitzler, P. (2021) A Review of the Semantic Web Field. *Communications of the ACM*, **64**, 76-83. https://doi.org/10.1145/3397512

[4] Seth, N., Deshmukh, S.G. and Vrat, P. (2006) A Framework for Measurement of Quality of Service in Supply Chains. *Supply Chain Management: An International Journal*, **11**, 82-94. https://doi.org/10.1108/13598540610642501

[5] Arlbjrn, J.S., Freytag, P.V. and De Haas, H. (2011) Service Supply Chain Management: A Survey of Lean Application in the Municipal Sector. *International Journal of Physical Distribution & Logistics Management*, **41**, 277-295.

[6] Pourmorshed, S. and Durst, S. (2022) The Usefulness of the Digitalization Integration Framework for Developing Digital Supply Chains in SMEs. *Sustainability*, **14**, 14352. https://doi.org/10.3390/su142114352

[7] Wang, G., Lu, R. and Guan, Y.L. (2018) Enabling Efficient and Privacy-Preserving Health Query over Outsourced Cloud. *IEEE Access*, **6**, 70831-70842. https://doi.org/10.1109/ACCESS.2018.2880220

[8] Yang, T. and Zhao, Y. (2018) Application of Cloud Computing in Biomedicine Big Data Analysis Cloud Computing in Big Data. 2017 *International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies* (*ICAMMAET*), Chennai, 16-18 February 2017, 1-3. https://doi.org/10.1109/ICAMMAET.2017.8186626

[9] Fawcett, S.E. and Waller, M.A. (2014) Can We Stay Ahead of the Obsolescence

Curve? On Inflection Points, Proactive Preemption, and the Future of Supply Chain Management. *Journal of Business Logistics*, **35**, 17-22. https://doi.org/10.1111/jbl.12041

[10] Fernando, Y., Chidambaram, R.R.M. and Wahyuni-Td, I.S. (2018) The Impact of Big Data Analytics and Data Security Practices on Service Supply Chain Performance. *Benchmarking*: *An International Journal*, **25**, 4009-4034. https://doi.org/10.1108/BIJ-07-2017-0194

[11] Wang, G., Gunasekaran, A., Ngai, E.W.T. and Papadopoulos, T. (2016) Big Data Analytics in Logistics and Supply Chain Management: Certain Investigations for Research and Applications. *International Journal of Production Economics*, **176**, 98-110. https://doi.org/10.1016/j.ijpe.2016.03.014

[12] Cho, D.W., Lee, Y.H., Ahn, S.H. and Hwang, M.K. (2012) A Framework for Measuring the Performance of Service Supply Chain Management. *Computers & Industrial Engineering*, **62**, 801-818. https://doi.org/10.1016/j.cie.2011.11.014

[13] Thatte, A.A., Rao, S.S. and Ragu-Nathan, T.S. (2013) Impact of SCM Practices of a Firm on Supply Chain Responsiveness and Competitive Advantage of a Firm. *Journal of Applied Business Research* (*JABR*), **29**, 499-530. https://doi.org/10.19030/jabr.v29i2.7653

[14] Gunasekaran, A. and Kobu, B. (2007) Performance Measures and Metrics in Logistics and Supply Chain Management: A Review of Recent Literature (1995-2004) for Research and Applications. *International Journal of Production Research*, **45**, 2819-2840. https://doi.org/10.1080/00207540600806513

[15] Balfaqih, H., Nopiah, Z.M., Saibani, N. and Al-Nory, M.T. (2016) Review of Supply Chain Performance Measurement Systems: 1998-2015. *Computers in Industry*, **82**, 135-150. https://doi.org/10.1016/j.compind.2016.07.002

[16] Gawankar, S.A., Kamble, S. and Raut, R. (2017) An Investigation of the Relationship between Supply Chain Management Practices (SCMP) on Supply Chain Performance Measurement (SCPM) of Indian Retail Chain Using SEM. *Benchmarking*: *An International Journal*, **24**, 257-295. https://doi.org/10.1108/BIJ-12-2015-0123

[17] Acar, A.Z. and Uzunlar, M.B. (2014) The Effects of Process Development and Information Technology on Time-Based Supply Chain Performance. *Procedia—Social and Behavioral Sciences*, **150**, 744-753. https://doi.org/10.1016/j.sbspro.2014.09.044

[18] Gunasekaran, A., Papadopoulos, T., Dubey, R., *et al.* (2017) Big Data and Predictive Analytics for Supply Chain and Organizational Performance. *Journal of Business Research*, **70**, 308-317. https://doi.org/10.1016/j.jbusres.2016.08.004

[19] Hofmann, E. (2017) Big Data and Supply Chain Decisions: The Impact of Volume, Variety and Velocity Properties on the Bullwhip Effect. *International Journal of Production Research*, **55**, 5108-5126. https://doi.org/10.1080/00207543.2015.1061222

[20] Kshetri, N. (2014) Big Data's Impact on Privacy, Security and Consumer Welfare. *Telecommunications Policy*, **38**, 1134-1145. https://doi.org/10.1016/j.telpol.2014.10.002

[21] Redden, J. (2012) Big Data as System of Knowledge: Investigating Canadian Governance. In: Elmer, G., Langlois, G. and Redden, J., Eds., *Compromised Data*: *From Social Media to Big Data*, Bloomsbury, London, 17-39.

[22] Rodriguez, L. and Da Cunha, C. (2013) Impacts of Big Data Analytics and Absorptive Capacity on Sustainable Supply Chain Innovation: A Conceptual Framework. *LogForum*, **14**, 151-161. https://doi.org/10.17270/J.LOG.267 https://doi.org/10.17270/J.LOG.267

[23] Cheng, S., Shi, Y., Qin, Q. and Bai, R. (2013) Swarm Intelligence in Big Data Analytics. *International Conference on Intelligent Data Engineering and Automated Learning*, Hefei, 20-23 October 2013, 417-426. https://doi.org/10.1007/978-3-642-41278-3_51

[24] Yap, L.L. and Tan, C.L. (2012) The Effect of Service Supply Chain Management Practices on the Public Healthcare Organizational Performance. *International Journal of Business and Social Science*, **3**, 216-224.

[25] Tseng, M.-L., Tan, R.R., Chiu, A.S.F., Chien, C.-F. and Kuo, T.C. (2018) Circular Economy Meets Industry 4.0: Can Big Data Drive Industrial Symbiosis? *Resources, Conservation and Recycling*, **131**, 146-147. https://doi.org/10.1016/j.resconrec.2017.12.028

[26] Yip, W., Fu, H., Chen, A.T., *et al.* (2019) 10 Years of Health-Care Reform in China: Progress and Gaps in Universal Health Coverage. *The Lancet*, **394**, 1192-1204. https://doi.org/10.1016/S0140-6736(19)32136-1

[27] Yang, Y., Zheng, X., Guo, W., *et al.* (2019) Privacy-Preserving Smart IoT-Based Healthcare Big Data Storage and Self-Adaptive Access Control System. *Information Sciences*, **479**, 567-592. https://doi.org/10.1016/j.ins.2018.02.005

[28] Poolsappasit, N., Dewri, R. and Ray, I. (2011) Dynamic Security Risk Management Using Bayesian Attack Graphs. *IEEE Transactions on Dependable and Secure Computing*, **9**, 61-74. https://doi.org/10.1109/TDSC.2011.34

[29] Sahafizadeh, E. and Nematbakhsh, M.A. (2015) A Survey on Security Issues in Big Data and NoSQL. *Advances in Computer Science: An International Journal*, **4**, 68-72.

[30] Hababeh, I., Gharaibeh, A., Nofal, S. and Khalil, I. (2018) An Integrated Methodology for Big Data Classification and Security for Improving Cloud Systems Data Mobility. *IEEE Access*, **7**, 9153-9163. https://doi.org/10.1109/ACCESS.2018.2890099

[31] Kumari, A., Tanwar, S., Tyagi, S. and Kumar, N. (2019) Verification and Validation Techniques for Streaming Big Data Analytics in Internet of Things Environment. *IET Networks*, **8**, 155-163. https://doi.org/10.1049/iet-net.2018.5187

[32] Bhathal, G.S. and Singh, A. (2019) Big Data: Hadoop Framework Vulnerabilities, Security Issues and Attacks. *Array*, **1**, Article ID: 100002. https://doi.org/10.1016/j.array.2019.100002

[33] Tao, Y., Xu, P. and Jin, H. (2019) Secure Data Sharing and Search for Cloud-Edge-Collaborative Storage. *IEEE Access*, **8**, 15963-15972. https://doi.org/10.1109/ACCESS.2019.2962600

[34] Babu, B.R. and Gopisetty, D.G.K.D. (2019) Private Data Destitution and New Publicize in Cloud Storage Suppliers.

[35] Xu, C., Chen, J., Wu, W. and Feng, Y. (2016) Homomorphically Encrypted Arithmetic Operations over the Integer Ring. *International Conference on Information Security Practice and Experience*, Zhangjiajie, 16-18 November 2016, 167-181. https://doi.org/10.1007/978-3-319-49151-6_12

[36] Xu, P., He, S., Wang, W., *et al.* (2017) Lightweight Searchable Public-Key Encryption for Cloud-Assisted Wireless Sensor Networks. *IEEE Transactions on Industrial Informatics*, **14**, 3712-3723. https://doi.org/10.1109/TII.2017.2784395

[37] Worku, S.G., Xu, C., Zhao, J. and He, X. (2014) Secure and Efficient Privacy-Preserving Public Auditing Scheme for Cloud Storage. *Computers & Electrical Engineering*, **40**, 1703-1713. https://doi.org/10.1016/j.compeleceng.2013.10.004

[38] Erhan, L., Ndubuaku, M., Di Mauro, M., *et al.* (2021) Smart Anomaly Detection in Sensor Systems: A Multi-Perspective Review. *Information Fusion*, **67**, 64-79. https://doi.org/10.1016/j.inffus.2020.10.001

[39] Chen, Y., Zhang, Y., *et al.* (2017) Deep Learning for Secure Mobile Edge Computing. https://doi.org/10.48550/arXiv.1709.08025

[40] Mollah, M.B., Azad, M.A.K., *et al.* (2017) Secure Data Sharing and Searching at the Edge of Cloud-Assisted Internet of Things. *IEEE Cloud Computing*, **4**, 34-42.

[41] Archaux, C., Martin, A. and Khenchaf, A. (2004) An SVM Based Churn Detector in Prepaid Mobile Telephony. *Proceedings of* 2004 *International Conference on Information and Communication Technologies*: *From Theory to Applications*, Damascus, 23 April 2004, 459-460. https://doi.org/10.1109/ICTTA.2004.1307830

[42] Idris, A., Rizwan, M. and Khan, A. (2012) Churn Prediction in Telecom Using Random Forest and PSO Based Data Balancing in Combination with Various Feature Selection Strategies. *Computers & Electrical Engineering*, **38**, 1808-1819. https://doi.org/10.1016/j.compeleceng.2012.09.001

[43] Dasgupta, K., Singh, R., Viswanathan, B., *et al.* (2008) Social Ties and Their Relevance to Churn in Mobile Telecom Networks. *Proceedings of the* 11*th International Conference on Extending Database Technology*: *Advances in Database Technology*, Nantes, 25-29 March 2008, 668-677. https://doi.org/10.1145/1353343.1353424

[44] Phadke, C., Uzunalioglu, H., Mendiratta, V.B., Kushnir, D. and Doran, D. (2013) Prediction of Subscriber Churn Using Social Network Analysis. *Bell Labs Technical Journal*, **17**, 63-76. https://doi.org/10.1002/bltj.21575

[45] Myers, S.A., Zhu, C. and Leskovec, J. (2012) Information Diffusion and External Influence in Networks. *Proceedings of the* 18*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2012, 33-41. https://doi.org/10.1145/2339530.2339540