Research on Privacy Protection Based on Joint Learning in Power Industry Big Data Analysis

Feilu Hang*, Linjiang Xie, Zhenhong Zhang,
Wei Guo and Hanruo Li

Information Center of Yunnan Power Grid Co., Ltd, Yunnan 650000, China
E-mail: hangfeilu2021@163.com; xielinjiang2021@163.com; zzh1989@126.com; 2490515@qq.com; luvshinhwa@163.com
*Corresponding Author

Received 10 December 2021; Accepted 20 January 2022;
Publication 24 June 2022

Abstract

In the era of big data, protecting the privacy of smart grid data is critical in ensuring the integrity and confidentiality of that data. Utilizing large amounts of energy data to gain insight into electricity consumers' consumption patterns helps develop power supply strategies. This article presents a Big Data-assisted Joint learning process (BDA-JLP), taking data security issues posed by big data in the electric power industry into consideration for privacy protection using K-anonymity and L-diversity as a foundation. A blockchain with JLP electric utility investigation is being conducted, part of the existing trading model split into phases. To begin, an attribute is chosen to categorize the input database. The comparable class number K and sensitive attribute value category L are limited by the number of original predecessors in the source data table, simplifying the calculation. A mathematical equation is then developed to determine the distance between first cousins multiplied by their combined weight. Linear and clustering with binary K are used to categorize data tables. Cluster and generalize initial data sets, considering

Distributed Generation & Alternative Energy Journal, Vol. 37_5, 1493–1526.
doi: 10.13052/dgaej2156-3306.3759

how the attribute values' internal range changes. The asymmetric encryption method uses two distinct keys for encryption and decryption ensuring that the blockchain system is completely secure. Simulated data show that the BDA-JLP mechanism proposed here has a privacy ratio of 98.3 percent, scalability of 97.0%, improved data management and data protection ratio of 98.2 percent, customer satisfaction ratio of 98.4 percent, and a low energy consumption ratio of 23.9% when compared to other methods currently available.

Keywords: Big data, blockchain joint learning process, power industry, privacy protection.

1 Overview of Privacy Protection Based on Joint Learning in Power Industry Big Data Analysis

In computing, "big data" refers to enormous quantities collected from various sources and available in different formats [1]. Due to their diversity, traditional database designs can't manage the vast amounts of data stored in databases for a long time [2]. Big data is a valuable asset that may be utilized to gain a priceless edge in the marketplace today than a collection of databases in different forms. When businesses talk about big data, this paper refers to investments in business intelligence, profitable intelligence, better perspective, and improved cognitive processing [3, 4]. IoT (Internet of Things) devices like RIFD (Radio-frequency identification), remote sensing, and software logs together with wireless sensor networks are constantly generating more data; thus, the amount of data generated is increasing at an exponential rate. Consequently, conventional applications for producing information no longer meet the needs of the modern world [5, 6]. As the Internet's popularity and capabilities increase, so does the crime committed. Due to our vulnerability to hacking, other parties may get access to information about our activities and exploit it for their ends [7]. The term "Big data" refers to a collection of techniques and structures intended to produce, examine, and analyze large quantities of information from very large amounts of data. In this thought process, big data has the 3V characteristics of volume, velocity, and variety, shown in this definition of the term [8, 9]. The term 'veracity' has been added to IBM's definition of big data. IBM is an abbreviation for International Business Machines and it assists businesses to discover and analyze fresh business insights buried in huge quantities of structured and unstructured data [10]. It is available for free software, together with an IBM server and storage designed for operational analytics, into a single, easy-to-manage solution, including the characteristics of 4V. Every microsecond, ‘Velocity’ generates data in ‘Volume’ from email and Twitter messages [11]. Various sources for generating data include unstructured “Variety” and structured “Variety.” Examples of the former include social media platforms like Twitter, mobile phones, cars, credit cards, M2M sensors, and M2M papers. Furthermore, it is essential to link together power structures, connections, and other data linkages for data to be revealed quickly [12, 13]. You can expect virtually no downtime and complete transparency in your operations with predictive manufacturing. The systematic transformation of data into meaningful knowledge necessitates a large volume of data and sophisticated prediction algorithms. The following are some of the most significant advantages of manufacturing with Big Data applications: Tracking flaws and shortcomings in products. Scalability problems are caused by data volume growth that is too easy to handle; changeability problems cause extensional problems, and relational problems are driven by the fact that the collected data may include the premise that the gathered data may contain collective regions that enable various databases to be integrated or meta-analyze issues cause collective areas that allow meta-analysis problems [14, 15]. Information generation, storage, and processing may be separated into three groups [16]. Access controls and information adulteration procedures are used throughout the information production process to safeguard individual users’ information privacy. Big data application development must take data security and privacy seriously [17]. Therefore, we must balance security and development while maximizing data value. In terms of privacy protection, the K-anonymity paradigm is an old standby. The K-anonymity homogeneity attempt was defeated model could not address by introducing the L-diversity notion [18]. To accomplish data anonymity and privacy protection and others initially suggested the use of clustering technique, whose fundamental concept is using the first-generation and later-test processes to uncover frequently recurring patterns that meet the minimum support criteria [19]. Using both numerical and classification attribute data to calculate the distance between primitives, BDA-JLP results in the subtypes being merged without differentiation during clustering, resulting in increased generalization [20]. Data mining and analysis lose value due to growing data loss. Desensitization strategies must be established and processed for each sensitive piece of data to classify it appropriately in the business system. To fulfil the objective of safeguarding sensitive data, the sensitivity of sensitive data is lowered on the assumption that data availability can be met [21, 22].

Increasing the understanding of energy usage and gathering huge amounts of data on consumers, mostly in two ways: A business's operational data on the system in use includes all types of job order data, real-time energy use by various instruments, and other data information [23, 24]. Date of birth, address (including zip code), phone number (including area code), and other identifying information about the user may be gathered and used. There will be a leak of user data if exported directly from the database. Furthermore, medical systems and numerous large-scale data platforms are vulnerable to privacy leaks. The personal privacy issues raised by big data sharing are addressed, the Big data consumption and development will be limited as the availability and use of important data is constrained [25, 26]. Three categories can classify information generation, storage, and processing. Individual users' information privacy is protected by access controls and anti-adulteration processes employed throughout the information generation process. The development of big data applications must take data security and privacy seriously. Therefore strike a balance between data security and development while also maximizing the value of collected data. Result. The K-anonymity paradigm is a tried-and-true method of maintaining online anonymity.

The main contribution of this paper,

- They suggested a Big Data-assisted Joint learning process (BDA-JLP) because of the data security problems big data technologies face in developing the electric power sector for privacy protection based on K-anonymity and L-diversity.
- The proposed BDA-JLP dealt with the asymmetric encryption method uses two distinct keys for encryption and decryption ensuring that the blockchain system is completely secure.
- The distance between first cousins multiplied by their combined weight is then calculated using a mathematical calculation. Data tables can be classified using linear and binary clustering with K.
- The numerical results have been performed based on the proposed BDA-JLP to achieve a high privacy ratio of 98.3%, scalability ratio of 97.0%, enhanced data management ratio of 98.2%, data protection ratio of 97.5%, customer satisfaction ratio of 98.4%, and low energy consumption ratio of 23.9% when compared to other existing methods.

The structure of the paper is organized as follows: Section 1 explores the Overview of Privacy protection based on joint learning in power Industry big data Analysis. The literature review and discussion are presented in Section 2. Section 3 discusses the BDA-JLP approach that has been suggested. In comparison to an existing approach, Section 4 presents numerical findings in summary form. Section 5 closes the article by drawing on the findings and recommendations presented in Sections 2 and 3.

2 Literature Survey

Yi Liu et al [27]. described the Attention Mechanism-based Convolutional Neural Network-Long Short Term Memory (AMCNN-LSTM) for deep anomaly detection for time-series data in industrial IoT. Anomalies in edge devices may substantially affect Industrial IoT (IIoT) output. Edge device data was a major challenge to current detection techniques since user privacy has improved. Convolutional Neural Networks with Long Short Term Memory based on Attention Mechanisms Could Be Useful for Troubleshooting (AMCNN-LSTM). The AMCNN-LSTM model uses Attention Mechanism-based CNN Units to address memory loss and gradient dispersion issues. Another benefit of utilizing LSTM unit features in this model is that it can make predictions about time-series data. Using the Top-k selection gradient compression method presented in this article, the timeliness of the proposed framework and communication efficiency are both improved. There have been a lot of tests done on real-world datasets that show this method reliably and rapidly identifies abnormalities while decreasing communication costs by half.

Lianyong Qi et al. [28]. Suggested the (LSHT) for using Spatial-Temporal Context and Privacy-aware Data Fusion to Make Predictions in an Industrial Smart City Setting. Transportation, healthcare, business, and social activity-related industrial data are among the many fields where smart cities produce a lot of data as one of the Cyber-Physical-Social Systems (CPSC). The creation and enhancement of different smart city applications may benefit greatly from data from many sources fused and mined effectively and efficiently. As a result, protecting user privacy in smart city data became more important before the data were combined for further mining, analysis, and prediction. This article proposes an innovative privacy-aware data fusion and prediction method based on the traditional Locality-Sensitive Hashing Technique (LSHT) for smart city industrial environments. Finally, a series using real-world datasets as the basis for studies are used to assess this idea. According to the findings of the experiments, this method is better at making predictions than its counterparts. Individuals will have their privacy protected because training will take place locally, and the model will be safely combined among the participants. A joint learning session may lose its ability to aggregate information if only two participants have access to it. The other participant's information is already included in the training's output. As a result, the model will be less accurate if the participants are concerned about protecting their anonymity.

GülçinBüyüközkan et al. [29]. Deliberated the Interval-Valued Intuitionistic Fuzzy AHP (IVIF-AHP) for the new digital service quality model and its strategic analysis in the aviation industry. Digital goods and services are becoming a need for airlines to keep their customers happy. Consequently, traditional customer satisfaction models cannot reflect consumer expectations, and new DSQ models need to be developed. A new and authentic DSQ model will be proposed in this investigation. Customers-centricity and digital tangibles are key components of the concept. There are 35 different criteria associated with each of these components. The IVIF-AHP technique was used to figure out how certain important criteria are. The model's validity was confirmed using a real-world case study of the Turkish airline sector. According to the findings, digital trust has been the most critical factor, with proactive customer service, cybersecurity, and consumer intelligence rounding out the top three.

Yanhui Liu et al. [30]. Expressed the Mixed Linear and Nonlinear Spatiotemporal Chaotic Systems (MLNCML) to preserve privacy in fog computing and on the internet of things. IoT devices are increasingly producing and sending large amounts of data across the network as the IoT sector develops. IoT security is a major concern since the data is sent through an unsecured network channel. There was just one weak link that would be one of the disadvantages of current symmetric encryption and access control methods for data transmission security. This promising distributed ledger technology prevents data from being tampered with maliciously, making it a safe and secure way to store information. This technique encrypts IoT data on an edge node using mixed linear and nonlinear spatiotemporal chaotic systems (MLNCML) before uploading it to the cloud. Access control for Internet of Things data may be dynamic and fine-grained. The proposed method solves the issue of a single point of failure in access control. This method's findings showed that it effectively safeguards IoT data privacy.

According to a literature review, a few existing AMCNN-LSTM, LSHT, IVIF- AHP, and MLNCML techniques need to be modified. Customers are satisfied with high levels of privacy and scalability while using less energy.

As a result, the BDA-JLP architecture presented in this research may assist the electrical industry in securing the privacy of its consumers.

3 BDA-JLP for Privacy Protection Based on Joint Learning in Power Industry

Big data factors play an intelligent data collecting, and processing paradigm for the power industry since different power sectors are rapidly digitizing. Companies can now build and offer strong industrial systems and services using big industrial data. Because of the current uniform privacy protection approach for all sensing data, MCS's crowdsensing services are overly or inadequately protected, and their quality is poor. For this reason, this paper suggests a Big Data-assisted Joint Learning Process (BDA-JLP) based on an evaluation algorithm and data encryption as a way to address the problem at hand. It starts with a privacy protection measuring method that determines private users' data and then applies evolutionary algorithms to develop a logical uploading strategy based on that data. This article proposes a data aggregation method that protects privacy while ensuring data privacy, accuracy, and timeliness are all important considerations. Numerous simulations based on real-world data and theoretical research trajectories show that BDA-JLP is efficient and provides excellent privacy protection in the power industry. Complementarity in joint learning based on distinct models can considerably improve performance compared to merely combining the same learners in a joint learning model (e.g., LSTM with LSTM in this study). In this way, the collaborative learning process benefits from the contributions of each base learner since each learner builds on the strengths of the prior layers of the other learner while avoiding the weaknesses of each learner.

Figure 1 illustrates the system model of the BDA-JLP framework. According to the system model presented in this article for the BDA-JLP framework, there are many sensing users, a semi-trusted sensing platform, and a large number of cloud service providers (CSPs) that are involved in final data transfers. Spatio-temporal sensing tasks (tasks) are carried out by users (abbreviated as users hereafter) who upload Spatio-temporal sensing data in compliance with different privacy protection requirements for the power sector (data). The user receives a payment from the sensing platform to use the cloud service provider platform. CSP is the only place users need to look. As shown in this article, for huge quantities of big data, CSPs are logical in pursuing huge amounts of high-quality customer service data from

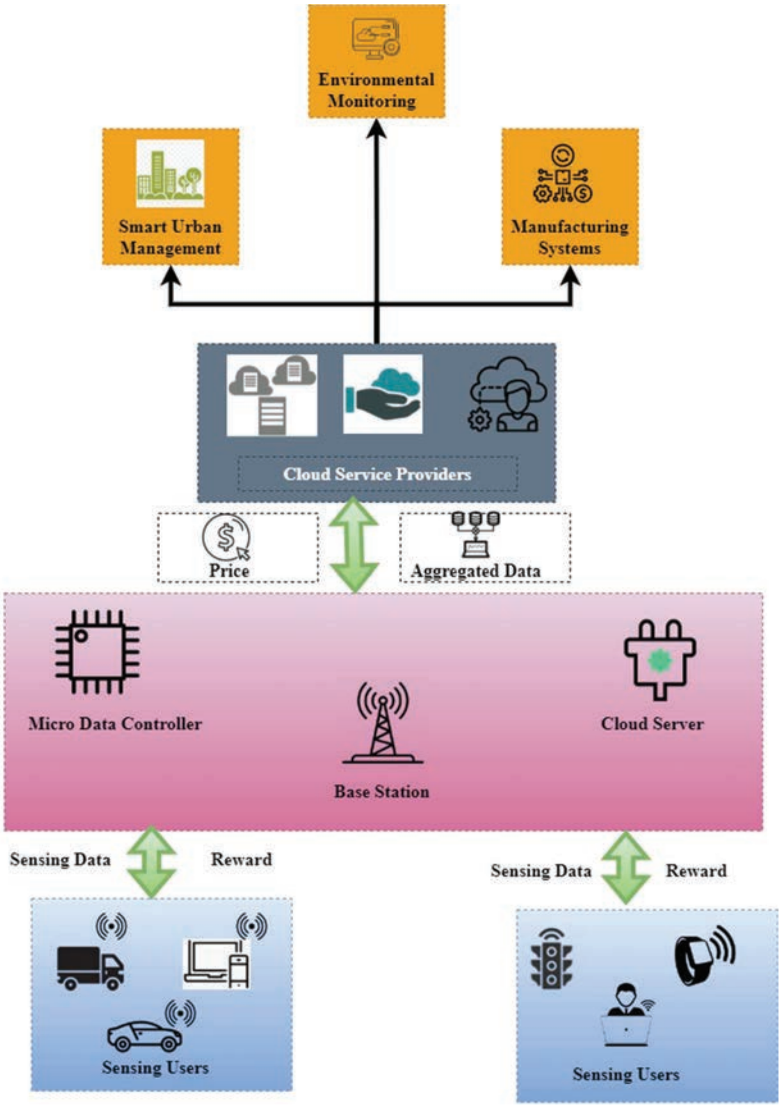

Figure 1 System model of the BDA-JLP framework.

platforms within users' budgets. Assigned responsibilities and interacted with users and the CSP in real-time according to the platform. Currently, it sells the gathered data to CSP and engages a suitable method to repay its (partially) compensated users. To implement the platform, fog nodes or edge nodes may be utilized. Assuming that the CSP and platform are reliable, an attacker will likely be a calculating, profit-driven person who will maintain tabs on any location while gaining access to the MCS's internal state information. The BDA-JLP framework has the following security criteria detailed below: It has been shown that encryption techniques are safe when used in the electricity sector. This implies that no information about a client may leave the platform without the customer's knowledge or permission. The term "data integrity" refers to the process of ensuring that the information received from a sensing user is correct and comprehensive. Using as many intelligent terminals as feasible makes it necessary for the traditional to provide exact privacy protection for the increasingly digitalizing power sectors.

Manual data collection for all kinds of industrial data is being implemented. Wireless networks for MCS are up to a billion sensors, such as GPS and accelerometers, used in internet-based mobile sensing devices and environmental sensors and gyroscopes. These mobile sensing devices work together to create an interactive and participatory network. Because of this, MCS has been widely accepted as a standard paradigm for intelligent data collection, analysis, and sharing in industry, where data collection, processing, and sharing for industrial sensing operations are outsourced to individuals or groups. As a result, MCS can significantly improve the Industrial Internet of Things (IIoT). Due to this, they are extensively utilized in various creative industrial applications, including Transportation and municipal management systems that are both intelligent and efficient. Health monitoring, manufacturing systems, and social security surveillance are examples of this. They are frequently used in smart city management. MCS provides many advantages for IoT's intelligent transportation system, which primarily comprises of self-Cars, traffic infrastructure, roadside equipment, and GPS services all fall under this category.

Real-time monitoring to keep tabs on road conditions and vehicle status offers real-time traffic monitoring and navigation services. Vehicle crowdsensing may be built with linked cars to provide more customized, coordinated, and safe services to users. Traffic monitoring in real-time allows for real-time driving and vehicle information and traffic conditions from a cloud service provider (CSP), allowing for real-time crowdsensing. MCS's sensing data and its users have many significant privacy and security concerns despite its apparent benefits. Large quantities of personal and private information, including identification, account information, online browsing history, physical activity, and location, may be found in Spatio-temporal sensing data. As a result, users of sensing technology will be reluctant to participate in sensing

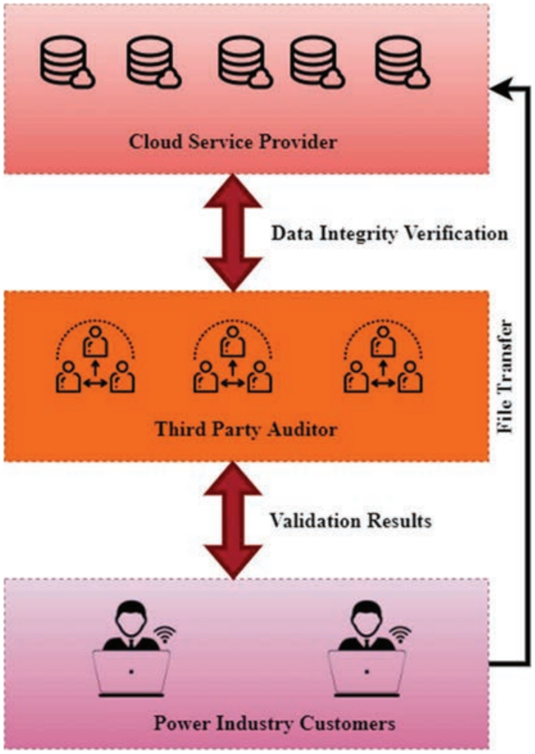

Figure 2 Data possession verification model in the power industry.

activities or share their sensing data. Students and researchers in academia and the power sector have studied ways to safeguard sensing data privacy while maintaining high-quality control and safety (QoS). To a certain degree, this problem may be solved using many techniques such as data encryption, anonymity, and differential privacy due to the absence of appropriate privacy metrics. All big data sensing strategies result in either excessive or insufficient privacy protection. As the sensing platform collects and processes data, it becomes more important to ensure the submitted information's privacy, confidentiality, integrity, and timeliness.

Figure 2 shows the data possession verification model in the power industry. Cloud and large data storage security are getting more sophisticated. More essential as big data links all things like computers. Transportation systems, military, industrial power production, smart homes, and other sectors use cloud and Internet of things integration. Data storage and privacy protection in the power sector provide many security issues. A finite amount of storage and processing power is available on the Internet of Things. However, scalable storage and powerful computer resources may be easily provided via the cloud. By using cloud resources in various ways, big data has broadened its application scope. A lack of control over data volume expansion leads to scaling challenges, whereas increased data volume results in changeability and extensional issues. Meta-analysis problems that lead to collective regions that make meta-analysis problematic or data collection concerns containing collective areas that allow numerous databases to be combined can cause relational complications. Vertical scaling, or acquiring a faster server with more powerful CPUs and more memory, is a part of the process of scaling up. Less network hardware means decreased power consumption, but this may only be a temporary cure for many platforms, especially if more growth is anticipated. Optimizing SQL queries and applying indexing algorithms are the best solutions for most database scalability difficulties. Can drastically reduce the number of queries performed by combining articles and writers into one.

Data is stored by cloud service providers in storage resources (CSP). Cloud computing for data analysis, verification, and storage lowers IoT's compute, storage, and connectivity costs while improving efficiency. It fulfils some of the IoT's real-time needs to some degree. However, because of the centralized nature of the CSP's storage resources, incidents such as faulty hardware and software and malicious malware that cause harm to the system pose a significant danger to the safe preservation of data. Even if the cloud storage system is damaged and the data is lost, IoT devices will be difficult to discover in time. As a result, the Internet of Things (IoT) storage must provide data integrity and availability, and the platform must include data integrity checks.

As the amount of data gathered by sensors grows, a new issue arises in the power industry's storage security: how to effectively perform reduce the computational and communication cost of a cloud storage server while performing a data integrity check (provable data possession, PDP). The cloud service provider CSP's computer, network, and storage capabilities store customer data. Data, the cloud storage server, verifies customers ownership, and it is initiated by the user after user permission and is completed by it. Data is stored by users on a cloud storage server in this vision. Users won't save the original data on their computers because they want to keep storage costs down. To ensure data integrity and availability, the user must execute a data-possession check on the cloud storage server. To verify property in the electricity industry, an auditor TPA is employed in place of the user, and only the verification results are provided back to them. This reduces computing and communication costs for the user. TPA must perform the data possession validation without collecting user information to safeguard big data privacy in the power sector.

Furthermore, the user is unaffected by the verification data's complexity, making it simpler to verify and more efficient. Data is duplicated on the cloud storage services to accommodate numerous users. It's necessary to react to the challenge and provide proof back to the TPA. The electricity industry has implemented a third-party auditor (TPA) to enable general auditing of the data validation process based on user consent and privacy protection based on user permission. The electricity industry has implemented a third-party auditor (TPA). Data possession verification is completed on the cloud storage server rather than on the desktop device through TPA.

Figure 3 shows stages of the big data life cycle. Methods such as access restriction, privacy protection, and information fabrication are used in the

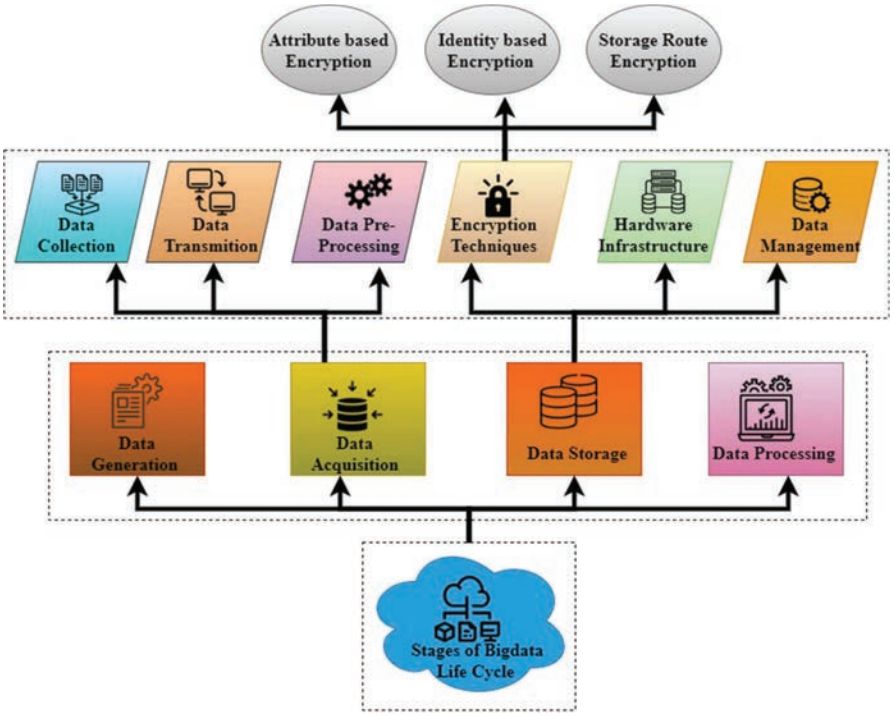

Figure 3 Stages of bigdata life cycle.

data creation phase. Access restriction methods are designed to prevent individuals from gaining unauthorized access to sensitive information. However, data verification procedures alter the data before sharing it with another unreliable group. Generation of information (data) may be generated from different decentralized resources. The amount of data generated by humans and computers has increased steadily during the last several years. The pace at which it is generated tells us a lot about its direction. It describes information production in three phases: organized, semi-organized, and unstructured data, with varying precision. Total data is assigned to an advanced structure for increased capacity and inspection during the information (Data) collection stage.

On the surface, the safeguarding process seems to be composed of three distinct steps: data collection, data transfer, and pre-processing of the data, because it's difficult to tell one way from another how information gets sent and the pre-preparing of information, information pre-handling activities may take place both before and after the transmission of information. There are three tasks to be completed in each phase: Collecting information. "information collection" refers to retrieving bare facts from verifiable objects. The process should be well thought before it is carried out. Off-base information collection would impact the subsequent information inquiry technique and lead to incorrect results. Research methods that collect and analyze information depend on information sources' physical characteristics and the research objectives. This article will collect the raw information and transfer it to a stockpiling framework, often in a server farm, to manage the resultant data. The transmission process may be divided into IP spine transmission and server farm transmission. For different reasons, pre-processing the informative indexes collected may be useful in various ways. There will be expenses involved in moving and keeping raw data.

Interestingly, certain information research methods and applications may impose strict requirements on the quality of the information they rely upon. Large information frameworks should include information-preparation processes to enhance the quality of the information they contain. Encryption methods (ET) are mostly used in the data storage process. Another option is to categorize cryptographic methods according to whether they encrypt data using attributes, identities, or storage routes (SRE). As a result, hybrid clouds are utilized for safeguarding private cloud data that should be kept private. Segmentation, identification, and relationship rule processing may all be utilized to classify these types of processes. Although segmentation and identification split the incoming data into distinct categories, association

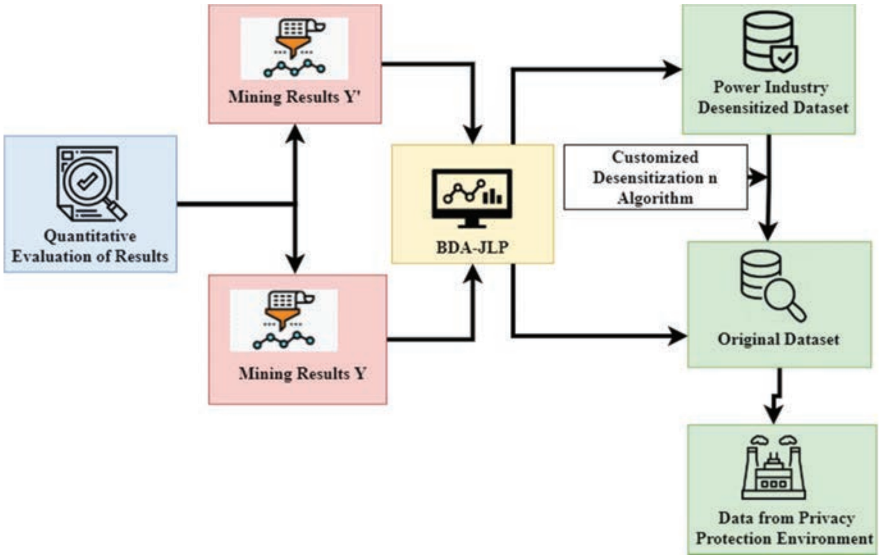
Mining Results Y'
Quantitative Evaluation of Results
Mining Results Y
BDA-JLP
Power Industry Desensitized Dataset
Customized Desensitization n Algorithm
Original Dataset
Data from Privacy Protection Environment

Figure 4 Quantitative evaluation model of desensitization algorithm.

rule processing identifies the important relationships and patterns in the data. A data storage device has two parts: a physical infrastructure and data administration. It may have various APIs for interacting with and evaluating the data contained in the system. Generally, the data processing step is used to acquire, transmit, and immediately pre-process data, collect useful information. Information may be gathered from various places, such as websites with text, pictures, and videos. Big data preservation in the power sector requires privacy protection and knowledge extraction, both parts of data processing. PPPI employs anonymization methods, including generalization and denial, to safeguard the privacy of power sector data. PPPI has several challenges in maintaining the information's usefulness while protecting users' privacy.

Figure 4 explores the Quantitative evaluation model of the desensitization algorithm. More and more businesses are customizing desensitization algorithms. As a result, due to the upcoming use of desensitized data, it is now a top priority for companies, even though data security is still a priority in the meanwhile. Another way, when desensitization improves data availability, it becomes more useful since the data are more closely aligned with the actual production data. Although there is no universal system for evaluating desensitization algorithms, because of the high degree of individualization in these models and the lack of a uniform standard for evaluating the impact of various algorithms, each power sector only has a basic and insufficient internal assessment model. For systematically and comprehensively in practical work, it represents industrial data integrity. For example, and it's objective and desensitized. A quantitative desensitization algorithm assessment model that is accurate and compatible is now required. It's important to note that our quantitative assessment methodology is "black box" focused. Desensitization algorithm testing may be completed with only two data sets: before and after desensitization. There is less danger of leaking customized desensitization algorithms since testers do not comprehend the design process.

The hidden knowledge of the dataset is reflected in our assessment model, which enables us to quantitatively evaluate desensitization impact by comparing changes in association rules before and after desensitization. It's important to note that the association rule mining algorithm is designed to work with Boolean datasets. The initial step in this research is to pre-process those datasets to evaluate them. This paper then uses the same transaction mapping method to turn Transaction data sets, including numerical and categorization characteristics such as an ID number, a name, and any income level. Desensitization and data anonymization privacy protection techniques are widely available in the power sector, and the assessment model's data-oriented features allow us to compare their effectiveness, which is without a doubt the best representation of the equality of the desensitization process. This paper can easily see the effect of desensitization on data sets by familiarizing yourself with the regulations of the mining organization before and after desensitization. The evaluation model for the desensitization method is presented in the following picture. It is a data recorder, a methodology of assessment based on the mining of associations. As a result, the compatibility of various desensitization and anonymization techniques is largely dependent on the pretreatment data that has been tailored for each method. Data preparation requires transactional mapping to convert different qualities into transactional attributes; nevertheless, this is time-consuming. The information does not have a logical meaning for fields such as an ID card number; instead, it is utilized for transactions directly. The implementation of a new transaction system is very difficult. Several other variables, such as the quality of the transactional technique, have an immediate impact on the evaluation result of the desensitization algorithm. As a result, the data preparation portion of the desensitization algorithm assessment model has been carefully constructed to make it general and enhance the capacity of the association rules to represent the total data set's characteristics. Detailed descriptions of data preparation methods are provided in the next section. The next section

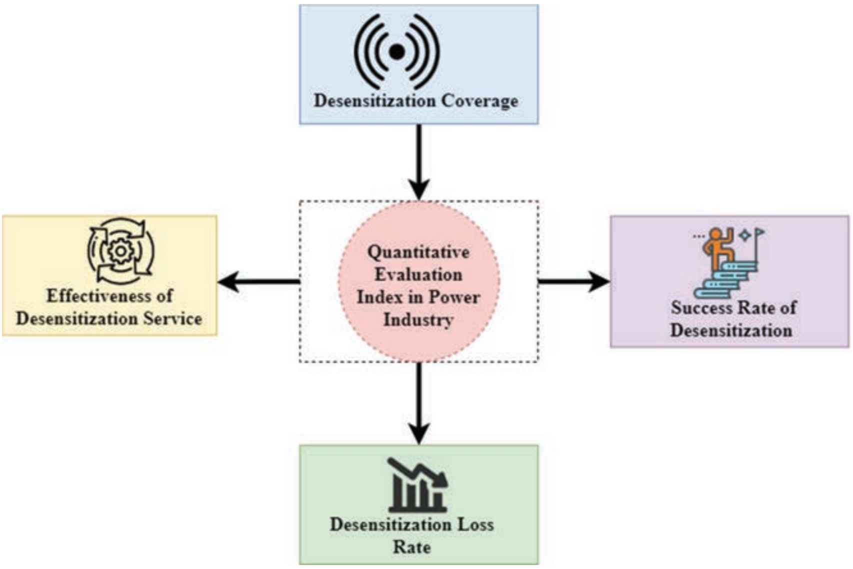
Desensitization Coverage
Effectiveness of Desensitization Service
Quantitative Evaluation Index in Power Industry
Success Rate of Desensitization
Desensitization Loss Rate

Figure 5 Quantitative evaluation index in the Power industry.

will describe a quantitative assessment model for the desensitization method, including data preparation, particular evaluation procedures, and evaluation indicators. This article will next explain the quantitative evaluation model in detail. Data mining, known as knowledge discovery, analyzes source data to uncover potentially valuable information. Association rule learning, one of the data mining techniques, is critical in uncovering hidden relationships and valuable information in large datasets. With association rules, you're looking for patterns or relationships hidden in the data, as implied by the name. Association rules are used in market basket analysis to discover the connections among the elements in a dataset.

Figure 5 shows the quantitative evaluation index in the Power industry. Following the receipt of the desensitization method will be based on the association rules mining findings assessed for accuracy in terms of privacy protection in the electricity sector by combining the quantitative metrics listed below. Specifically, determining a desensitization method's coverage rate depends on how well the algorithm covers the input data. This algorithm is compatible to handle various data formats in varied situations. According to intuition, this indication indicates if the desensitization algorithm can handle many types of usual information. When it comes to the deregulated electricity market, one of the most pressing issues is determining a global power quality

(PQ) index that can quantify the cost of PQ in contracts between consumers and utility companies. Existing PQ indexes are typically isolated and lack the cost implications of poor PQ. The use of ANNs and fuzzy logic to develop a quantitative global index for PQ evaluation and pricing in the competitive power market is shown here. Companies with desensitization requirements often have big databases, vast data dimensions, and a variety of field kinds to choose from. The desensitization coverage rate is just one of many important indicators that may be used to determine whether or not the desensitization system has completely desensitized the data in the system. Each data point in each field can be appropriately desensitized if the present method for desensitization is capable of doing so. If the desensitization success rate is low, the data have been incorrectly desensitized. If the data have changed after processing, then the data do not comply with certain types of fields with verification regulations. If a series of strange events occur, the data have changed incorrectly after processing, such as garbled code. As part of the desensitization algorithm assessment methodology, this paper computes the success rate of the desensitization algorithm in the following manner: Each record's length is calculated to see whether it meets the length requirement for that particular kind of field. If this is not the case, it is referred to as a desensitization failure. Determine if each data point differs from the data point used for desensitization. If the data after desensitization is the same as the data before desensitization, it is determined that desensitization has failed. For various reasons, traditional relational databases will not be phased out anytime soon. Companies have heavily invested in relational databases and the associated infrastructure. Structured data volumes continue to rise even as transactional workloads remain the most common form. For various reasons, traditional relational databases will not be phased out anytime soon. Companies have heavily invested in relational databases and the associated infrastructure. Structured data volumes continue to rise even as transactional workloads remain the most common form. It is possible to gain a competitive advantage in today's market by utilizing big data. As a whole, it's not merely a collection of databases. Investing in business intelligence, profitable intelligence, improved perspective, and increased cognitive processing are all expressions used by corporations when discussing big data. Sensor networks and IoT devices like RFID (Radio-frequency identification) create an ever-increasing amount of data, as are remote sensing and logs from the Internet of Things. The term "Big Data" refers to a collection of data that is both large in terms of volume and continually expanding. A dataset that is so huge and complicated that typical data management solutions cannot store or process it effectively is the case. Big data is a type of data. However, it has a much larger volume. After desensitization, determine if the verification rules field complies with the verification requirements. The failure of desensitization is determined by whether or not the desensitized data comply with the verification criteria. Each record field must be processed to determine success rate detection the desensitization. There may be a direct correlation between the success of each record's desensitization and the whole record's overall success. The success rate of desensitization may be determined after desensitization. The number of desensitization successes in each area can be calculated by dividing the total number of fields processed by each field's number of desensitization successes. It represents the amount of information lost due to the desensitization process, measured in percentages. Today, As a general goal, most companies want to keep the security of the desensitization algorithm while minimizing the impact on knowledge models represented by original data sets. This way, a desensitized data set can retain as many of the useful characteristics of the original data sets as possible (including statistical characteristics), and the data can play a proper role in subsequent recommendations and recommendations. The desensitization loss rate may be the most important quantitative indication. It is possible to say that the rate of desensitization loss directly reflects the quality of the desensitization algorithm used. This article focuses on the critical aspect of how association rule mining may reveal the relationship between frequently occurring item sets. According to the input data specification for the desensitization algorithm evaluation model, the desensitization algorithm evaluation model is used to transform the two evaluation data sets into transactional data sets before and after desensitization. The two datasets are input into a desensitization assessment model to learn association rules. The findings of the two datasets are then compared to generate the association rule mining results.

(i) Calculate the privacy protection distance in the power industry

Collect and arrange the following data in each database in the power industry data from the user's sector, street address, power usage in real-time, account balance, name, age, and other personal information the identifier may be combined to produce a source of data for further analysis and application. The last step in the finishing process is the following steps: Choose an identifier for each data record in the first information B_H preliminary division categorization characteristic that is $T = \{R_1, R_2, R_j, \ldots R_J\}(R_1{*}R_2 * \ldots *$

$R_j * R_J)$. Cluster the data items in R_j based on the primitives' distance formula to discover the generalization class. Consider the case when there are two datasets k_i, k_n including numerical information $\{S_1, \ldots, S_{m1}\}$ data that has been encrypted $\{B_1, \ldots, B_{m2}\}$. In terms of a numerical value S_a, there is a difference between two data values of $H_{S_a}(k_i, k_n)$ its meaning is as follows

$$H_{S_a}(k_i, k_n) = \frac{1}{\gamma|\overline{S_a}|} \cdot \frac{|k_i - k_n|}{|E_{R_j}|} \tag{1}$$

Where $\overline{S_a}$ is depending on the average,γ is a wide range of qualities, $|E_{R_j}|$ is the attribute's domain value's length S_a in the first classification set $R_j \cdot \gamma|\overline{S_a}|$. The larger γ is, the more distinct this attribute's data. It is the same as putting a numerical value on data. To successfully minimize the quantity of data loss, the generalization distance between records must be increased while the significance of distance measurement must be reduced. In addition, data loss delays production and can cause you to lose clients if a security incident follows it. When sensitive data is stolen or compromised, your organization must notify your customers, resulting in a loss of confidence and respect.

When it comes to the classified property B_t, let E occupy the space of categorization and T_E to be the tree of generality on^E. In the case of two different classifications $k_i, k_n \in E$ the length.

$H_{B_t}(k_i, k_n)$ between k_i, k_n is explained in the following way,

$$H_{B_j}(k_i, k_n) = X(\Delta(k_i, k_n))|X(T_E) \tag{2}$$

Where $\Delta(k_i, k_n)$ is the division whose root has the fewest shared predecessors with all other sections i, n and $X(T)$ is the total of the generalization tree's hierarchical distances T_E. To make it possible to standardize distance measurements.

When it comes to power industry privacy protection, the distance is calculated as follows,

$$E(k_i, k_n) = \sum_{t=1}^{m_1} H_{B_t}(k_i, k_n) + \sum_{a=1}^{m_2} H_{S_a}(k_i, k_n) \tag{3}$$

(ii). Privacy Measurement in the power industry

The location's unified functional state is referred to as its accessible attribute. Customers and investors in the same power industry will have access to an industry's fundamental corporate characteristics, which are the same for everyone. The customer's attributes of a location f_i to the terminal, j is a good example of $G_{j,i}$. This paper can determine the average daily activity time and see a clear power-law distribution by analyzing the users' travel paths. It is possible to decide the average user access time on each website. Derived from past route records across time, reflecting the degree of dependence of the user on the locations, the time it takes to get access $E_{j,i}$ of j to $f_i$from $t_1\ to\ t_2$.a term used to describe the following thing:

$$E_{j,i} = \int_{t_j}^{t_2} \frac{m(t)H_t}{t_1 - t_2} \tag{4}$$

Where t_1 and t_2 indicate the beginning and end of the access record. It would be possible to dynamically update users' access duration and movement routes in real-time. Only one table entry can be called a "record." Data is entered sequentially from left to right in a table, and each row can only contain one entry. Double-click the table that wants to add records to and start typing. In the Table Window, our table has no data yet. The path is taken by j with time t_1, t_2 is $F_j = (f_1, f_i, \ldots, f_m)$ where $1 \leq i \leq m, f_i \in F_j$. Using this new measure, this paper can see how often a certain site is accessed relative to the overall number of times that location has been accessed during a given time in the power sector. The frequency in which information may be accessed $Q_{j,i}$ of j to f_i is outlined in the following

$$Q_{j,i} = \frac{m(f_i)}{\sum f \varepsilon F_j m(f)} \tag{5}$$

Where $m(f_i)$ the frequency distribution is represented by j travelling f_i and $m(f)$ gives customers access to all places at the same time simultaneously frequency. Access regularity is used to represent how often a user goes to a certain place and to calculate the variation in separation cycles $K_{j,i}$ of j travelling f_i is defined as follows

$$K_{j,i} = \frac{\sum_f (y_f - a_{j,i})^2}{number_{j,i}} \tag{6}$$

Where y_f reflects the duration of the spending time together $a_{j,i}$ shows how the strong connection between two things j and f_i and $number_{j,i}$. Indicates the distance between two points.

For measurement of individual privacy for each individual, it is important to evaluate the privacy level of both accessible and customized characteristics before making any uploading decisions FQ_j of j at f_i are given,

$$\begin{aligned} FQ_j &= \alpha Q_{j,i} + \beta E_{j,i} + \varphi G_{j,i} + \omega K_{j,i} \\ &= \alpha \frac{m(f_i)}{\sum f \varepsilon F_j \, m(f)} + \beta \int_{t_j}^{t_2} \frac{m(t) H_t}{t_1 - t_2} + \varphi G_{j,i} + \omega \frac{\sum_f \left(y_f - a_{j,i}\right)^2}{number_{j,i}} \end{aligned} \tag{7}$$

Where $\alpha, \beta, \varphi, \omega$ characteristics of the system to modify the weight of every indication and $\alpha + \beta + \varphi + \omega = 1$. Even if there is a loss of the protection of personal privacy as a consequence of j being able to save information on a computer B_j the degree of individual privacy that j it's possible to say,

$$FQ_j(a_j) = \begin{Bmatrix} FQ_j - B_j, & a_j \rightarrow Y \\ FQ_j, & a_j \rightarrow M \end{Bmatrix} \tag{8}$$

Where $a_j = Y, M$ is an uploading strategy set, FQ_j Privacy level of user j and B_j is uploaded data.

(iii). Privacy-preserving Data Aggregation Scheme in the power industry

Considering users may select whether or not to submit data based on their own privacy needs, the suggested approach is focused on them. When users submit data to the sites, data leakage and privacy are still concerns even with privacy security. This paper developed an additively homomorphic encryption data aggregation method to combat this problem that successfully secures data secrecy, integrity, and timelessness. To decide whether or not to upload data, a user computes a privacy loss threshold B_{max}. Based on the location service's minimal requirements. For example, if a user understands that other users at that place have different privacy levels or have their privacy violated in some way, $(B_1 \leq B_2 \leq \cdots \leq B_m)$ then $B_{max} \leftarrow B_U$ is readily available. A more realistic assessment of B_U Worth is required in the typical power sector when privacy protection is concerned; it is believed that the amount of privacy and the degree of privacy loss are equivalent $FQ_j = \frac{\gamma}{B_j}$. This shows that the more privacy protection the user has, the less privacy is lost when sending sensing data. Customers may get the formula (9) from $I(\sigma_j)$. Meeting the privacy needs of the user.

$$\frac{U}{V} = \int_{\sigma'}^{1} I(\sigma_j) H \sigma_j \tag{9}$$

Where V identifies the overall number of individuals who visit the site. Following the formula (9), $\sigma' = \frac{\gamma}{I^{-1(1-\frac{U}{V})}}$ was calculated to get a reasonable estimate for B_U.

$$\widehat{B_U} = \frac{\gamma}{\sigma'} = \frac{\gamma}{I^{-1(1-\frac{U}{V})}} \tag{10}$$

The gathering of information is referred to as each user gathers data when they receive task requests from the platform g_j. The job of reasonable uploading will be part of my participation.

$$B_j = \text{encryption}(g_j) = g_j + Kt_j \times key_j(G) \tag{11}$$

Where g is the big number that's kept in the system's memory, information is encrypted by the user g_j. Data privacy is ensured while uploading by utilizing the Segmentation algorithm, as stated in Formula (11).

To get better results, it's better to include like students in a collaborative learning model and use complementary models rather than just merging them all (e.g., LSTM with LSTM in this study). Rather than relying on each learner's flaws, the collaborative learning process utilizes each learner's strengths while minimizing the deficiencies of those who came before. Building a big data teaching support platform expands the network space for ideological and political theory course teaching activities, providing students with plenty of precise and real-time data and creating an online learning platform. Using big data as a foundation, a new teaching model for political and ideological theory courses is developed, with significant adjustments to the course's core concepts, methodologies, and procedures.

4 Results and Discussion

BDA-JLP integration has spread like wildfire. This article proposed a new data integrity checking method that combined signatures to address the power industry's privacy protection issue with large amounts of big data. A Safe and Efficient Approach using big data analysis applications with huge aggregated data benefit greatly from the approach since it completely accounts for security, scalability, and privacy protection. However, this paper has significant drawbacks compared to the BDA-JLP in most current power sector settings. Data integrity verification in multiple replica settings, for example, is not covered by this strategy. To that end, this investigates a data integrity verification method that is more real-time and suited to settings with many copies of a file. Further enhancing privacy security in the power industry by using a verification method based on a random oracle.

4.1 Privacy Ratio (%)

Data collection, processing, and transmission may be compromised if consumer privacy is unprotected when gadgets analyze and send massive amounts of data. Maintaining privacy is a difficult task that requires both technological and legal solutions. Big data is rapidly being used by industrial power activities, such as manufacturing facilities and applications for protective control. Security and confidentiality are difficult to provide for in power industrial big data systems because of the systems' complexity. The difficulty of defining and executing secure requirements in a way that can be proven.

The privacy ratio (%) is shown in Table 1. If devices analyze and communicate enormous amounts of data without protecting it, consumer privacy may be jeopardized. Many technological and legal solutions are needed to ensure privacy. Industrial power operations, such as manufacturing facilities and protective control applications, quickly utilize big data. Analyses of many privacy methods and solutions have shown several security flaws and risks. It is possible to enhance privacy protection by collecting and processing data to preserve individual privacy. If everything else fails, devices may be asked to authenticate their identification without disclosing personal information one final time. Various privacy procedures and solutions have been thoroughly analyzed to discover privacy weaknesses and threats. The findings offer new directions for future studies to solve some privacy concerns raised by big data analysis in the electricity sector.

Table 1 Privacy ratio (%)

Number of Customers	AMCNN-LSTM	LSHT	IVIF-AHP	MLNCML	BDA-JLP
10	53.2	60.1	70.9	80.1	90.7
20	51.2	61.2	78.0	82.9	97.0
30	56.0	63.0	76.0	84.0	94.9
40	52.0	64.3	75.0	85.0	98.0
50	54.0	65.0	75.0	86.5	97.4
60	53.8	66.0	76.0	87.0	98.0
70	57.7	62.0	77.9	83.9	95.0
80	53.8	65.7	74.0	81.0	96.9
90	59.0	70.1	73.9	82.0	93.0
100	58.7	69.0	76.8	86.8	98.3

4.2 Scalability Ratio (%) and Data Management Ratio (%)

Integrating proper privacy protection is critical to preventing the emergence of security flaws in systems. Big data technology has issues with dependability, scaling, and energy usage. New facilities may be introduced at any moment thanks to internet capabilities in the power sector big data analysis. A larger database will need more storage capacity to keep up with the growing volume of big data devices, which may significantly impact the cost of production and storage. As consumers and transactions increase, so does big data scalability. As transaction sizes increase, the time it takes to place a transaction and authenticate, it becomes shorter transactions do not need the oversight of a third party, since it allows for more automation, scalability, and reduced transfer costs, all of which may reduce the risk of data exploitation. To increase scalability, the authors of this article recommend keeping distinct databases for different regions. Operational data about a business's system in use include task order data, here real-time energy use by various instruments, and other data for processing. Raising customer awareness of their energy usage and amassing enormous amounts of data about them are the primary goals of these efforts. It is possible to reduce large data breaches by setting access criteria, limiting data collection, usage, or storage to support your business need solely, and implementing technical safeguards to secure data from attackers. In the same way, firms can use data about their suppliers and customers to detect those in financial difficulties, allowing them to act swiftly to minimize their exposure to any potential defaults.

Figure 6 explores the Scalability Ratio (%) and Data Management Ratio (%). Various data security concerns and problems emerged as it grew simpler to get data. Management can approve a data management document

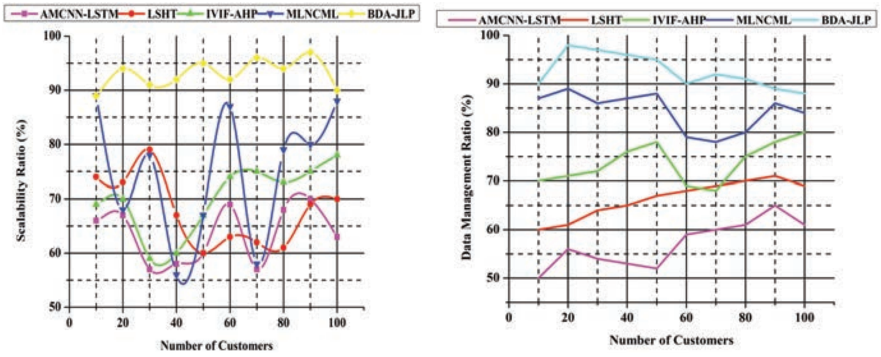

Figure 6 Scalability ratio (%) and data management ratio (%).

disseminated to all workers and the power industry to offer management assistance and advice for data security following the power industry's standards and applicable regulations. If customers use this document, the power industry may successfully prepare for steps to guarantee data security while maximizing the efficiency and advantages of technology. Protecting private telecommunications industry information from unwanted access is one of three criteria for data security management. To guarantee authenticity, communications infrastructure may be strictly controlled. They are establishing management duties and procedures to promptly and efficiently address data security incidents. Stealing, leakage, and modification may all be reported, corrected, and kept from damaging the industry by utilizing the proper management channels. A strong system of surveillance and supervision may help measure customer confidence in the power sectors.

4.3 Data Protection Ratio (%) and Customer Satisfaction Rate (%)

Associated goods, products, and solutions big data gradually emphasize people's everyday lives. As a consequence, the security of private information is becoming essential. It is possible to repurpose devices and power industries in various settings. Finding the right personal data protection solutions may be critical. Having privacy-protected data is critical since if the data are leaked, it may offer a gateway for criminals to access consumer and industry data. The power industry requirements to be served may be considered while searching for the best data security solution. Personal data protection systems may have a major impact on customers and the industry. System solutions are designed to discover ways to keep personal data safe while using big data services. Many power industry factors may impact a solution's efficacy. Finally, a legal authority may be granted to the electricity business and its consumers. Personal privacy data protection may alter existing power businesses, as can legal industries that evaluate privacy.

Figure 7 illustrates the Data Protection Ratio (%) and Customer Satisfaction Rate (%). A better service data network has been given to customers, implying that the system's performance has increased. This means that customer satisfaction is calculated by considering all service quality factors. Most customers are pleased with the various big data technologies. Customer satisfaction is critical in industrial power applications, and the design offers safe storage and job allocation. This measure affects the efficiency of quality service parameters in industrial power applications. It lowers the time

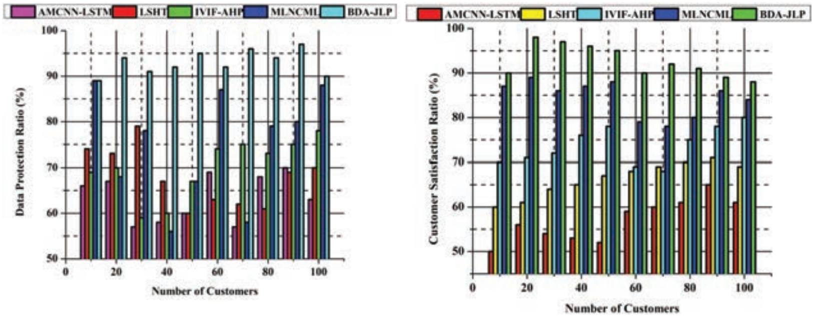

Figure 7 Data protection ratio (%) and customer satisfaction rate (%).

required to complete activities while ensuring complete and happy customers; the suggested method has gained favour with customers. When cloud servers are under-managed, customer happiness is suffering. Big data customers need data storage to manage services and keep data organized. Customer data may be obtained improperly in certain cases since the process takes a lengthy time. Client information is moved to a data centre, which is processed faster while improving the overall quality of service by the suggested method.

4.4 Energy Consumption Ratio (%)

The use of power in big data devices is a significant issue. While big data analysis and technologies in the power sector are energy-constrained, certain processes need enormous energy to function fully. Clustering, an efficient energy activity, may be suggested to solve this problem. Clustering reduces the demand for processing by transmitting data to nearby cloud servers, allowing big data devices to use less energy.

Table 2 shows the energy consumption ratio (%). The table shows that using the suggested approach reduced energy usage. Data centres use cloud networking to check data deduplication at random. The transmission of edge devices getting data requests from other devices necessitates large calculations, which raises the energy consumption of big data. These articles use the internet's maximum possible layer of recognition and convolution to distribute labour securely and decrease industrial energy usage.

This approach is being suggested that displays measures of security, privacy, and scalability; data management; customer happiness; and energy ratio when compared to attention mechanism based convolutional neural network-long short term memory (AMCNN-LSTM), locality sensitive hashing technique(LSHT), interval-valued intuitionistic fuzzy AHP

Table 2 Energy consumption ratio

Number of Customers	AMCNN-LSTM	LSHT	IVIF-AHP	MLNCML	BDA-JLP
10	90.8	70.9	60.8	56.8	35.0
20	97.9	78.0	67.9	55.9	36.9
30	98.0	74.8	68.0	56.0	33.0
40	90.0	74.9	69.0	57.9	32.0
50	96.5	72.0	67.5	54.8	31.0
60	92.0	78.9	62.0	56.0	28.0
70	91.0	76.0	61.0	57.0	26.0
80	94.8	74.9	62.8	58.3	25.8
90	92.0	80.1	63.0	52.0	23.9
100	93.0	79.9	65.0	57.0	26.8

(IVIF AHP) and mixed linear and nonlinear spatiotemporal chaotic systems (MLNCML).

5 Conclusion

Using the data desensitization classification method in typical the ability to harness massive amounts of data application scenarios, a data desensitization method for privacy protection is presented in this study in response to the difficulties of big data in the growing power sector. A K-anonymity-based approach and L-diversity is proposed in this paper by using Big Data to aid in the Joint Learning Process (BDA-JLP). To begin, the information will be divided into categories according to a categorization feature. The data will next be grouped based on the distance between data sets determined by the proposed equation. The clustering data items will be generalized after the process. The generalization results should be free of K-anonymity and L-diversity. A collaborative learning model that incorporates students who are similar and complementary to one another can achieve more effective outcomes than just integrating all of the previously mentioned models. Instead of relying on the weaknesses of each learner, the collaborative learning process uses the capabilities of each learner while reducing the shortcomings of those who have come before. An online platform for ideological and political theory courses can be built from the ground up using large datasets to provide students with precise and real-time information. With the help of big data, a new teaching model for political and ideological theory courses is designed, with major changes to the course's key concepts, techniques, and procedures. This is how the BDA-JLP works in the simulation. It completely utilizes data value while protecting user privacy and safeguarding their information, reducing computing costs for electricity companies. These extensive paper simulations show this proposed BDA-JLP mechanism to achieve a high privacy ratio of 98.3%, scalability ratio of 97.0%, enhanced data management ratio of 98.2%, data protection ratio of 97.5%, customer satisfaction ratio of 98.4%, and low energy consumption ratio of 23.9% when compared to other existing methods.

References

[1] Gao, J., Wang, H., and Shen, H. (2020, May). Smartly handling renewable energy instability in supporting a cloud datacenter. In 2020 IEEE international parallel and distributed processing symposium (IPDPS) (pp. 769–778). IEEE.

[2] Saravanan, V., Alagan, A., and Woungang, I. (2018). Big data in massive parallel processing: A multi-core processors perspective. In Handbook of Research on Big Data Storage and Visualization Techniques (pp. 276–302). IGI Global.

[3] Manogaran, G., and Lopez, D. (2018). Spatial cumulative sum algorithm with big data analytics for climate change detection. Computers & Electrical Engineering, 65, 207–221.

[4] Nguyen, T. N., Liu, B. H., Nguyen, N. P., and Chou, J. T. (2020, June). Cyber security of smart grid: attacks and defenses. In ICC 2020-2020 IEEE International Conference on Communications (ICC) (pp. 1–6). IEEE.

[5] Chen, J., Ramanathan, L., and Alazab, M. (2021). Holistic big data integrated artificial intelligent modeling to improve privacy and security in data management of smart cities. Microprocessors and Microsystems, 81, 103722.

[6] Manogaran, G., and Lopez, D. (2018). Spatial cumulative sum algorithm with big data analytics for climate change detection. Computers & Electrical Engineering, 65, 207–221.

[7] Babu, D. V., Saravanan, V., Kumar, P., and Singh, S. (2015). Automated robotic receptionist with embedded touch screen. Journal of Chemical and Pharmaceutical Sciences, 415–417.

[8] Nguyen, T. G., Phan, T. V., Hoang, D. T., Nguyen, T. N., and So-In, C. (2020, December). Efficient SDN-Based Traffic Monitoring in IoT Networks with Double Deep Q-Network. In International Conference on Computational Data and Social Networks (pp. 26–38). Springer, Cham.

[9] Gao, J., Wang, H., and Shen, H. (2020). Task failure prediction in cloud data centers using deep learning. IEEE Transactions on Services Computing.

[10] Amudha, G., and Narayanasamy, P. (2018). Distributed location and trust based replica detection in wireless sensor networks. Wireless Personal Communications, 102(4), 3303–3321.

[11] Nguyen, T. N., Liu, B. H., Nguyen, N. P., and Chou, J. T. (2020, June). Cyber security of smart grid: attacks and defenses. In ICC 2020–2020 IEEE International Conference on Communications (ICC) (pp. 1–6). IEEE.

[12] Shakeel, P. M., Baskar, S., Fouad, H., Manogaran, G., Saravanan, V., and Xin, Q. (2020). Creating Collision-Free Communication in IoT with 6G Using Multiple Machine Access Learning Collision Avoidance Protocol. Mobile Networks and Applications, 1–12.

[13] Fenil, E., Manogaran, G., Vivekananda, G. N., Thanjaivadivel, T., Jeeva, S., and Ahilan, A. (2019). Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM. *Computer Networks*, *151*, 191–200.

[14] Amudha, G., Jayasri, T., Saipriya, K., Shivani, A., and Praneetha, C. H. Behavioural Based Online Comment Spammers in Social Media.

[15] Manogaran, G., Baskar, S., Hsu, C. H., Kadry, S. N., Sundarasekar, R., Kumar, P. M., and Muthu, B. A. (2020). FDM: Fuzzy-optimized Data Management Technique for Improving Big Data Analytics. IEEE Transactions on Fuzzy Systems.

[16] Shakeel, P. M., Baskar, S., Fouad, H., Manogaran, G., Saravanan, V., and Xin, Q. (2020). Creating Collision-Free Communication in IoT with 6G Using Multiple Machine Access Learning Collision Avoidance Protocol. Mobile Networks and Applications, 1–12.

[17] Manogaran, G., Baskar, S., Hsu, C. H., Kadry, S. N., Sundarasekar, R., Kumar, P. M., and Muthu, B. A. (2020). FDM: Fuzzy-optimized Data Management Technique for Improving Big Data Analytics. IEEE Transactions on Fuzzy Systems.

[18] Manimuthu, A., Dharshini, V., Zografopoulos, I., Priyan, M. K., and Konstantinou, C. (2021). Contactless Technologies for Smart Cities: Big Data, IoT, and Cloud Infrastructures. SN Computer Science, 2(4), 1–24.

[19] Abiad, M., Kadry, S., Ionescu, S., and Niculescu, A. (2019). Customers' Perception of Telecommunication Services. FAIMA Business & Management Journal, 7(2), 51–62.

[20] Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 571–588.

[21] Hou, J., Li, Q., Cui, S., Meng, S., Zhang, S., Ni, Z., and Tian, Y. (2020). Low-cohesion differential privacy protection for industrial internet. *The Journal of Supercomputing*, *76*(11), 8450–8472.

[22] Abowd, J. M., and Schmutte, I. M. (2019). An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review*, *109*(1), 171–202.

[23] Bäck, I., and Kohtamäki, M. (2016). Joint learning in innovative R&D collaboration. *Industry and innovation*, *23*(1), 62–86.

[24] Tariq, N., Asim, M., Al-Obeidat, F., Zubair Farooqi, M., Baker, T., Hammoudeh, M., and Ghafir, I. (2019). The security of big data in fog-enabled IoT applications including blockchain: A survey. *Sensors*, *19*(8), 1788.

[25] Fang, W., Wen, X. Z., Zheng, Y., and Zhou, M. (2017). A survey of big data security and privacy preserving. *IETE Technical Review*, *34*(5), 544–560.

[26] Shin, D. H., and Choi, M. J. (2015). Ecological views of big data: Perspectives and issues. *Telematics and Informatics*, *32*(2), 311–320.

[27] Liu, Y., Garg, S., Nie, J., Zhang, Y., Xiong, Z., Kang, J., and Hossain, M. S. (2020). Deep anomaly detection for time-series data in industrial iot: a communication-efficient on-device federated learning approach. IEEE Internet of Things Journal, 8(8), 6348–6358.

[28] Qi, L., Hu, C., Zhang, X., Khosravi, M. R., Sharma, S., Pang, S., and Wang, T. (2020). Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment. IEEE Transactions on Industrial Informatics, 17(6), 4159–4167.

[29] Büyüközkan, G., Havle, C. A., and Feyzioğlu, O. (2020). A new digital service quality model and its strategic analysis in aviation industry using interval-valued intuitionistic fuzzy AHP. Journal of Air Transport Management, 86, 101817.

[30] Liu, Y., Zhang, J., and Zhan, J. (2021). Privacy protection for fog computing and the internet of things data based on blockchain. Cluster Computing, 24(2), 1331–1345.

[31] Wang, F., Yang, N., Shakeel, P. M., and Saravanan, V. (2021). Machine learning for mobile network payment security evaluation system. Transactions on Emerging Telecommunications Technologies, e4226.

[32] Xue, M., Xiu, G., Saravanan, V., and Montenegro-Marin, C. E. (2020). Cloud computing with AI for banking and e-commerce applications. The Electronic Library.

[33] Saravanan, V., Nuneviller, M., Pillai, A. S., and Anpalagan, A. (2020). Foundation of Big Data and Internet of Things: Applications and Case Study. Securing IoT and Big Data, 1–14.

Biographies

Feilu Hang was born in 1984 in Zhaotong, Yunnan province, China. Graduated from Yunnan University with a master's degree in system Analysis and integration. At present, I am working in equipment management Department of information center of Yunnan Power Grid Co., LTD. His research interest covers network and information security.

Linjiang Xie was born in 1985 in Qujing, Yunnan Province, China. Graduated from Yunnan University with a bachelor's degree in information security. At present, I am working in equipment management Department of information center of Yunnan Power Grid Co., LTD. His research interest is network security operation.

Zhenhong Zhang, born in Qujing city, Yunnan Province, China in 1989, graduated from Beijing University of Posts and Telecommunications with a master's degree in computer technology. Currently, he is working in the equipment management department of information Center of Yunnan Power Grid Co., LTD., and his research direction is information system operation and maintenance.

Wei Guo, born in 1986 in Kunming, Yunnan Province, China, graduated from Chongqing University of Posts and Telecommunications with a bachelor's degree in Information Management and Information System. Currently, he is working in the equipment management department of information Center of Yunnan Power Grid Co., LTD. His research direction is network and network security operation and peacekeeping management.

Hanruo Li born in 1991 in Zhaotong, Yunnan, China, graduated from Fuzhou University with a bachelor's degree in network engineering. Currently, he is working in the equipment management Department of the information Center of Yunnan Power Grid Co., LTD. His research direction is network security operation and maintenance.