Proceedings of the 3rd International Conference on Business and Policy Studies
DOI: 10.54254/2754-1169/77/20241841

The Application of Big Data and Privacy Impacts Research in Information Technology

Jiahao Yang[1,a,*]

[1]*Tallulafalls School, Tallulah Falls, Georgia, 30573, United States*
a. jiahao.yang@tallulahfalls.org
**corresponding author*

Abstract: The rapid growth of the big data industry has made personal information the basic resource. However, the convenience of obtaining personal information and the commerciality will inevitably lead to the problem of personal information security protection. Based on this background, this paper takes company A as the research case draws out the insufficiency of the protection of customers 'personal information: system risk, human risk, and process management risk in data storage and transmission, and systemic and human factors in data-supported security, and puts forward the countermeasures for the protection of personal information: 1) Clarification of the main body's responsibility; 2) Classification and grading control; 3) Avoiding sensitive data out of the network; and 4) Information use requiring authorization. This study provides relevant suggestions on how to improve the security protection of personal information in the field of big data applications; from the technical aspect, it adopts classification control use authorization and other technologies to reduce network risks; on the other hand, from the management aspect, it clarifies the main safety responsibility, strengthens the professional skills training of personnel, and authenticates the use within the scope of authority, which can help enterprises make future strategies.

Keywords: Personal Information, Big Data, Privacy Protection

1. Introduction

Big data are pervasive in the current technological era. Massive volumes of data have been generated from a number of rich data sources. These big data were characterized by various degrees of authenticity (e.g., exact, imprecise, and uncertain data), social network data, transportation data, and disease reports are a few examples of big data [1-2]. Big data management and mining provide the techniques to cope with this data [3-4]. The use of open data has grown significantly in popularity in recent years. As a result, more large data are freely accessible on open data portals or platforms.

With the rapid growth of the big data industry, personal information has become one of the basic resources in the age of informationization due to its large data samples, authenticity, and collection efficiency. The convenience of obtaining personal information and the commerciality of the use of personal information will not only bring huge economic benefits but also inevitably lead to the problem of personal information security protection. Therefore, the protection of personal information will not only be an individual issue but also an issue that needs to be paid attention to in many fields of society.

DOI: 10.54254/2754-1169/77/20241841

As a telecom operator who has mastered the information of the majority of customers, it is an inevitable trend to integrate big data into enterprise development, and the personal information protection of customers is also imperative. Therefore, this paper aims to actively explore countermeasures for the information security protection of customers and effectively recreate the value of personal information data in an information regulatory environment that ensures security.

2. Case Description

2.1. Case Introduction

A company's big data application of internal business is mainly focused on marketing, network construction, product design, customer service optimization, etc. Taking marketing as an example, through the collection of customer personal information data, the big data is applied to customer portraits [5]. That is, labels distinguish customer characteristics, and customers are specifically positioned by labeling each customer. This can form a unique customer characteristic portrait and can be used as a classification dimension to predict customer needs and preferences according to customer characteristics. And carry out refined marketing to customers.

In practical applications, such as real-time positioning, real-time call, and real-time traffic as the basis of big data for location package push, time-limited package push, through the big data of multi-card slot terminals as the support of new customers in different networks, customer behavior, and consumption preferences for the big data for the 5G era intelligent terminal recommendation, etc. With the development of more flexible and intelligent products, big data marketing based on customer portraits and real-time data will diversify.

2.2. The Analysis of Specific Scenarios

The following is an example of the big data model of early warning of potential off-grid customers in current marketing. Due to the saturation of the number of customers in the regional market, the main direction of Company A's market operation since 2018 has been to conduct in-depth mining through the existing stock customer information data, improve the profit value of the stock customers, and realize the stock operation [6-8]. Taking the retention of off-grid customers in stock operation as an example, since 2018, Company A has actively built a big data model for off-grid early warning of internal customers based on the real-time monitoring of changes in customers' network age, consumption, activity, social connections, and other information, and through the data analysis of the model, it outputs prediction results, deeply excavates existing customers with silent off-grid tendencies, and conducts refined grouping according to the different characteristics of customers, screens different groups of customers, and pushes the most effective retention plans for customer retention and preferential marketing to realize early intervention and active retention of customers off the grid [9-11].

As shown in Table 1 below, the off-grid early warning big data model defines the target group, off-grid definition, and variable information and involves customer personal information such as number, network age, call behavior, active behavior, account consumption, and balance, etc., which is modeled and verified according to the above customer information big data, and finally presented as the specific content of the current third version of the off-grid early warning big data model through repeated practice and optimization.

DOI: 10.54254/2754-1169/77/20241841

Table 1: Big data model for early warning of potential off-grid customers

Target group	Definition of off-grid	variable		Modeling samples and verification data selection	Result verification
High-value customers (One year old.) and above, ARPU>=50 or DOU>=1G)	Account closure or consecutive Downtime is more than one months (on monthly downtime, and At the end of the month still Shutdown status)	The magnitude of the change in ARPU	BH_ARPU_J2	Full Sample: Select On-line customers follow the stream The percentage of lost customers is 1:8 modeling	Forecast for N+March Off-grid accuracy 8.1% coverage 23.8%, overall accurate 98.4% accuracy; Forecast for N+2 months Off-grid accuracy 9.3%, coverage 30.8%, overall accurate 98.3% accuracy;
		The magnitude of the change in the times of calls	BH_CALL_CNT_J2		
		The magnitude of the change in active days	BH_HYTS_CNT_J2		
		ARPU	ARPU_J2		
		DOU	DOU_J2		
		Number of calls	CALL_CNT_J2		
		Number of active days	HYTS_CNT_J2		
		The number of top-ups	cz_cnt_j2		
		The base balance is enabled	EBOX_AMT_j2		
		The base balance is not enabled	EBOX_UNAMT_j2		
		Complimentary balances are not enabled	ZS_UNAMT_j2		
		Net age	wl_j2		
		Cornet or not	if_dh		

According to the above big data model, statistical analysis is carried out on the changes in customers' personal information data every month, and the off-grid prediction of all customers can be carried out, which can predict whether customers have off-grid tendencies in advance and judge customer behavior tendencies, so as to carry out customer segmentation and maintenance, and reduce customer off-grid loss [12].

Through the analysis of the above examples, the application of big data within Company A is reflected in the off-grid customer retention work, mainly using the identity information and service content information of the customer's personal information and carrying out the fine marketing of internal business through the collection and statistical analysis of customer service content information, so as to realize the operational value of personal information data.

In this process, in the process of data collection, data output, and data analysis, personal information data requires the participation of internal personnel and the support of the data system,

DOI: 10.54254/2754-1169/77/20241841

and it can be seen that there is a certain risk of sensitive data leakage in both human and system. Based on this, this paper will further study how enterprises can do a good job in the basic work of personal information protection in the process of rationally using user personal information data to carry out the application of internal big data, improve their business operation level, promote enterprise precision marketing, optimize customer service and consumer perception.

3. Analysis on the Problem

From the perspective of enterprise process management and personnel management, the following analyses the personal information protection of telecom operators from two aspects: customer information data storage and transmission and data support. The personal information protection policy of Company A is discussed as an example.

3.1. Analysis of Data Storage and Transmission

In the big data environment, the customer personal information data of telecom operators is mainly collected, stored, and transmitted by multiple relatively independent system platforms, and as the distribution point of customer personal data information, the storage and transmission links involve security risks in systems, personnel, processes, etc., and there are certain loopholes, which cannot fully guarantee the information security of customers [13].

a. System risk: Due to historical reasons, telecom operators have a large number of system platforms for collecting and storing customer information, and the data exchange between platforms is still not fully realized.

b. Human risk: The customer information data of telecom operators is collected through multiple operating systems, such as customer information registration, business handling, and customer interaction, and the data storage is scattered on various open platforms. On the one hand, the permissions of internal staff are different, and the permissions between different platforms are scattered, making personnel management difficult. And there are no very strict specifications for data acquisition of high-authority personnel, which need to be consciously complied with by humans. On the other hand, data collection and storage are complex, the division of security management responsibilities is not clear, and multi-level management not only diversifies risks but also easily causes loopholes and negligence, resulting in data leakage of human factors.

c. Process management risk: Telecom operators involve a lot of sensitive customer data, but in daily work management, the sensitive data that needs to be used cannot be stored in full ciphertext, and data encryption is not in place. In the same way, there is also the problem of insufficient encryption during transmission.

3.2. Analysis of Data-Supported Problems

Data support is mainly to realize access to customers' personal information and data platform and provide data management, data analysis, and data application operations. There are also systemic and human factors in information security risks [14].

a. In the development of big data systems of telecom operators, the O&M personnel of third-party vendors support the big data O&M management system, and there is a risk that third-party O&M personnel steal high-value sensitive data through system backdoors or vulnerabilities in the process of system development and platform maintenance.

b. Telecom operators collect many types of data on customers' personal information. There is a lack of effective identification and separation technology for sensitive and non-sensitive data. There is a possibility that non-sensitive data may be matched and combined to generate sensitive data. As a result, in the actual operation process, customer information data can be generated through the

DOI: 10.54254/2754-1169/77/20241841

association and matching of basic information data, making it difficult to find the risk of data application the first time obtaining data, and the regulatory utility cannot be fully covered.

c. Data support and analysis need to be in contact with first-hand customer information data already encrypted in the current operating technology.

4. Suggestions

Based on the premise of promoting the development of big data applications and addressing the issues in customer information protection for telecommunications operators, this article explores specific strategies for personal information protection in telecommunications operator big data applications from three perspectives: enterprise management, user rights, and resource management. The specific strategies are as follows:

4.1. Clarification of the Main Body's Responsibility

As a telecom operator that holds a large amount of customers' personal information, it should take responsibility for the security of the full amount of customer information held within the enterprise. The staff shall, in accordance with the principle of "who is in charge of who is responsible, who operates who is responsible, who uses who is responsible, who accesses who is responsible", clearly define the division of responsibilities and be responsible for the customer information under their respective management and use, and the logs of the whole process of collecting, transmitting, and using the customer information shall be recorded completely and accurately to ensure that all the operations can be traced back to specific operators and operational bases. Ensure that all operations can be traced back to the specific operator and the basis of operation. In the event of information leakage, it is necessary to pursue responsibility from the source to each link one by one to improve the staff's vigilance and seriousness in managing and using customer information.

4.2. Classification and Grading Control

In the application of big data, there are various types of personal information for different customers; personal information should be classified and graded, graded according to different degrees of sensitivity, according to different degrees of sensitivity involved the information to take appropriate and information security risk-adapted management measures and technical means, in the application of big data, the scope of use of different degrees of sensitivity of the use of data, the use of the extent of the degree of are specifically differentiated to protect the security of customers' personal information in use.

4.3. Avoiding Sensitive Data Out of the Network

In big data applications, for sensitive level customer information data, without explicit authorization from the customer, sensitive data that has not been desensitized shall not be taken out of the internal network and computing environment of the enterprise, and important and sensitive information can only be used in encrypted terminals or in accordance with the specified steps. In order to ensure the safety of personal information, it is not recommended to sell or use personal sensitive information commercially before it is desensitized to reduce the risk of third-party leakage.

4.4. Information Use Requiring Authorization

In the application of big data, in addition to process control, personnel control is also an important link. The use of customer personal information within the enterprise personnel, according to the use of data at different levels of sensitivity, the need for separate authentication management [15-16].

DOI: 10.54254/2754-1169/77/20241841

On the one hand, customer information data should be within the scope of their rights, and all privileged operations are strictly prohibited. On the one hand, all staff should use customer information and data within the scope of their own authority, all privileged operations are strictly prohibited, and the system operations should be approved and recorded in the operation process; on the other hand, centralized management of high-authority work numbers involved in the management of sensitive data.

On the other hand, the centralized management of the use of high-privilege work numbers involved in the management of sensitive data, use of sensitive data should ensure that the customer authorizes all operations and retains authorization records, the use of data according to the vault management mode, the user and the authorizer should be separated from each other.

5. Conclusion

This paper takes the big data application as the background, takes company A as the research case, studies the development of the big data application of company A, draws out the status quo and insufficiency of the protection of customers' personal information in the field of big data application, and puts forward the countermeasures for the protection of personal information in the process of big data application of company A.

On the one hand, from the technical aspect, it adopts classification control use authorization and other technologies to reduce network risks; on the other hand, from the management aspect, it clarifies the main safety responsibility, strengthens the professional skills training of personnel, and authenticates the use within the scope of authority, especially for the third party to provide big data applications, big data information to summarize the analysis presented to minimize the use of sensitive data visualization.

The results of this paper can make relevant suggestions on how to improve the security protection of personal information in the field of big data applications, etc., which can provide a reference for enterprises' future strategies. However, the data of the telecom operator company A in this paper is secondary data. This will reduce the credibility of the data to some extent, as future research should collect the primary data of company A by survey for further analysis.

References

[1] *Alim, et al. (2019). Uncertainty-aware opinion inference under adversarial attacks. in IEEE BigDat, pp. 6-15.*

[2] *F. Jiang, C.K. Leung. (2015). A data analytic algorithm for managing, querying, and processing uncertain big data in cloud environments. Algorithms 8(4), pp. 1175-1194.*

[3] *K. Leung, et al. (2014). Fast algorithms for frequent itemset mining from uncertain data. in IEEE ICDM, pp. 893-898.*

[4] *He, et al. (2019). Finding mutual X at WeChat-scale social network in ten minitues. in IEEE BigData, pp. 288-297.*

[5] *Leung, K. et al. (2019). Personalized DeepInf: enhanced social influence prediction with deep learning and transfer learning. in IEEE BigData, pp. 2871-2880.*

[6] *Leung, K., Jiang, F. (2015). Big data analytics of social networks for the discovery of "following" patterns. in DaWaK, pp. 123-135.*

[7] *Castrogiovanni, P., Fadda, E., Perboli, G., Rizzo, A. (2020). Smartphone Data Classification Technique for Detecting the Usage of Public or Private Transportation Modes. IEEE Access 8, p. 58377–58391.*

[8] *Balbin, P. F. F. et al. (2020). Predictive analytics on open big data for supporting smart transportation services. Procedia Computer Science 176, pp. 3009-3018.*

[9] *Leung, K., Elias, J. D., Minuk, S. M., Roy, A., de Jesus, R., Cuzzocrea, A. (2020). An innovative fuzzy logic-based machine learning algorithm for supporting predictive analytics on big transportation data. in FUZZ-IEEE, pp. 1-8.*

[10] *Leung, K. et al. (2020). Data mining on open public transit data for transportation analytics during pre-COVID-19 era and COVID-19 era. in INCoS, pp. 133-144.*

[11] *Leung, K. et al. (2019). Urban analytics of big transportation data for supporting smart cities. in DaWaK, pp. 24-33.*

DOI: 10.54254/2754-1169/77/20241841

[12] *Gupta, P. et al. (2021). Vertical data mining from relational data and its application to COVID-19 data. Big Data Analyses, Services, and Smart Data, pp. 106-116.*
[13] *Souza, J. et al. (2020). An innovative big data predictive analytics framework over hybrid big data sources with an application for disease analytics. in AINA, pp. 669-680.*
[14] *Tsumoto, S. et al. (2019). Estimation of disease code from electronic patient records, in IEEE BigData, pp. 2698-2707.*
[15] *Cuzzocrea, A. (2016). Big data provenance: state-of-the-art analysis and emerging research challenges. in EDBT/ICDT Workshops, pp. 37:1-37:4.*
[16] *Hassani, M., Spaus, P., Cuzzocrea, A., Seidl, T. (2015). Adaptive stream clustering using incremental graph maintenance. in BigMine, pp. 49-64.*