Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/11299
DOI URL: http://dx.doi.org/10.21474/IJAR01/11299

RESEARCH ARTICLE

PRIVACY-PRESERVATION METHODS IN BIG DATA ANALYTICS: A CASE STUDY OF TELECOMMUNICATION COMPANIES IN GHANA

Isaac Lewu Hill[1], Gaoming Yang[1], and Richard Sam Dickson[2]

1. School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China.
2. School of Economics and Management, Anhui University of Science and Technology, Huainan 232001, China.

Manuscript Info

Manuscript History
Received: 10 May 2020
Final Accepted: 15 June 2020
Published: July 2020

Key words:-
Privacy Preservation, Big Data Analytic, Linear Regression Function, Ghana

Abstract

The quantity of data obtained by numerous organizations about individuals is continuously rising. The big dimensional data are often included in data diversity sources.Mostly, data of this kind are stored in a table arrangement and involves data of raw content. Storing confidential data is a necessary task that needs a lot of awareness if the organizations associated with data are to allow the publication of data for different analyses and study purposes. Multiple attacks on privacy are connected with recording and attribute associated attacks. Many researchers and academicians have suggested preservation methods such as l-diversity, t-closeness, K-anonymity, etc. to protect the publication of individual's private data. This paper mainly focused on Preservation Techniques in Big Data Analytics: A Case study of Telecommunication companies in Ghana. The study employed multiple regression methods in analyzing the data, acquired through a closed - ended questionnaire.The result revealed that all various Privacy Preservation Methods used by Telecommunications in Ghana have an important contribution to the Protection of data privacy. The result also showed a strong and positive correlation between privacy-preservation , techniques and user characteristics employed in the study since the multiple regression results explained are over 96%, and the significant level was 0.03788. We recommend that more attention should be put on these techniques to improve the privacy terms and conditions.

Introduction:-

Data has had a vast increase in its amount and quality due to the several utilizations of computers. The accessibility of affordable network technology, storage, and web links has led to this growth. A huge volume of data can include person-specific sensitive information like gender, disease, zip code, religion, etc. that are collected in an unrestricted area. A third-party data examiner is capable of publishing the individual information if a data owner releases the data to him/her to expand more profound insights and to recognize obscure models that mean valuable in making significant decisions that may further promote businesses, produce value-added assistance to customers, prediction, forecasting, and recommendation. Data analytics is a notable application for recommendation systems that are generally employed by many e-commerce sites for proposing products to their customers based on their purchasing

Corresponding Author:- Isaac Lewu Hill
Address:- School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China.

ways and interests. However, publishing the data of individual activities may result in inferential attacks like identifying the gender, zip, etc. of a user (Turkanović et al., 2015; Cai et al., 2018). Privacy Protection problems are identified in numerous application areas like mobile environment, big data mining, and sensor systems. Modern technologies with a higher power to store, process and publish data raises higher concern on the misuse of data stored at various places. The work of Samarati and Sweeney (2017) suggested a publishing pattern called k-anonymity to manage privacy by using generalization and an elimination method. Samarati (2001) introduced a procedure to address the difficulty of publishing microdata. Sweeney (2002) recommended a formal k-anonymity pattern for protection with supplemented database policies. Emam et al., (2009) works suggested a k-anonymity approach for protecting the secrecy of individual data. The proposal of Caballero-Gil et al., (2015) made use of a k-anonymity approach to preserve the secrecy of any location-based settings on a cloud storage server. Basso et al.,(2016) constructed a data perturbation method to decrease the possibilities of inferential attacks. The difficulty faced when dealing with the data perturbation technique is that the query output contains supplementary noises. Hence, only approximate results will be displayed to the requested users without information loss. Data leaking of these types can tend to cause complete identification theft and misuse of sensitive data at several domains in different ways. The researchers from different areas have seen this problem with their perspective and tried to solve it with many techniques. Many research papers are available for Privacy Preservation Techniques. Now at this point, it is necessary to empirically examine these techniques to give a basic idea about each technique and to add to the current discussion on the subject matter.

The main aim of this research is to examine the different Privacy- Preservation Methods use in big data analytics. Specifically, this paper attempts to (1)identify the different Privacy- Preservation Methods employed by telecommunication industries in Ghana (2) analyze each method and (3) recommend the common relevant and most applied method (s) for telecommunication industries.

Scope and Organization of the Study.
The rest of the study is arranged as follows: Section Two covers a revision of theoretical literature on Privacy Preservation Techniques. Section Three covers methodology and estimation procedures such as research design, sampling methods, studying instruments, analytical tools, and techniques. Section Four covers the results and discussion of findings and section Five then presented summary, recommendations, the study's limitations and further studies.

Literature Review:-

Theoretical Review:
Privacy preservation using transformation: The straight forward solution for Privacy- Preservation is the conversion of raw data. The data is restricted so that sensitive information cannot be recovered. The price we pay is a loss of effectiveness in data retrieval and mining and this is usually caused by modifying the primary data. Such techniques for Privacy-Preservation are randomization, k-anonymity, and l-diversity.

Randomization:
According to Chauhan (2013) randomization is the most common technique used as privacy- preservation techniques by communities. It is known that adding noise in the data makes it difficult to find actual data. This concept is employed in the randomization technique. A sufficient large noise is added to make it impossible for the mining of sensitive data. The addition of noise is by masking the attribute value of the record. Mostly this procedure is practiced in public surveys where one can find evasive answer bias. The method of randomization is explained simply in a way that data sets under consideration are sufficiently added to the noise with independent elements so that variance of the noise of original data is not easily found. After the randomization process, the individual records are dissolved into a single distribution having the same behavior of the original data set. The real challenge is that modification of the current algorithm of information mining is done in such a manner that it can work on distribution rather than an individual record. A randomization technique has a primary benefit because it is comparatively easy and does not need the knowledge of other records as well as does not demand a secure server. All data are equally treated on any system that can randomize the data. Attackers can utilize the vulnerability of this technique by recognizing the susceptible data rather than the compact region. Another approach of using randomization is by combining or omitting random objects from the data set.

K-anonymity:
K-anonymity mainly works on the group of data .With anonymization, significant parameters associated with an individual's identification entity (Unique Identification Number) is excluded from the dataset. Nevertheless, an individual's entity is identified by pseudo identifiers such as age, pin code and sex. Ram Mohan Rao et al., (2018) asserted that, the k-anonymity technique can be achieved through a method of inducing or suppressing the pseudo identifiers by setting and modifying the range of data sets. The limitation of employing the suppression method is that the data's effectiveness is reduced. The k-anonymity method limits the granularity representation of data with any produced record mapping onto at least k other than actual data records. K-anonymity is a significant method for privacy de-identification.

l-diversity:
In k-anonymity, the values of sensitive data are set to a generalized form. Thus, the data becomes anonymized leading to data susceptible attacks problems where background information is known to the attacker (Homogeneity attack). The attacker learning the background knowledge can calculate the accurate value of the generalized value by finding a relationship among identifies to narrow down the possible sensitive field. The introduction of the l-diversity method is based on the above flow information of k-anonymity.In l-diversity, both the group of k and the information heterogeneity are maintained. (Ram Mohan Rao et al., 2018; Gkountouna, 2015;Praveena Priyadarsini et al., 2016). The notification here is that, owning "n" separate sets of attributes means having to manage 1-diversity which is a challenging problem. To additionally enhance this model, the t-closeness model is introduced.The t-closeness method requires an equal range between sensitive data and the threshold value. (Patel, 2019; Mathew, 2019).

Data mining modified applications for privacy preservation:
Many times, there exists a possible recovery of information from the result of the application. Therefore, the construction of diverse techniques is to modify the result for privacy preservation. An association rule hiding is a related technique used is the theory of association rule mining (Bux et al., 2018; Rajasekaran & Thanabal, 2019). The achievement of association hiding rule is by using the distortion or blocking method. Changing the value of some bit to binary value in a given transaction is done with the distortion method. The blocking method deals with incomplete entries which are purposely for preventing association rules. Here, there is the possible creation of some unknown or unwanted rules, since there is a loss of both sensitive and non-sensitive rules, causing a reduction in the usage of data and inaccuracy of the results. The primary factor for blocking is to change unknown values preferably than some duplicate value, for example (NULL). Another useful technique is classification. A classifier attacker can have access to raw values. So, the principal purpose here is to decrease the activeness of classifiers by modifying the data.

Query auditing:
A study by Kundalwal et al., (2018) suggested a hybrid method that involved two separate inference control methods, restriction of the size of a query set, and a k-anonymity method to ensure the privacy of individual's data. The restriction of a query collection size is to prevent data sensitivity from inferential attacks, whereas the implementation of k-anonymity is to prevent the exposure of data from a linking attack (Kumar et al., 2015; Bergmann et al.,2016). Both techniques have reached a particular level of privacy with satisfying data utilization. An aggregate query provides public interface access in cases where a common database is not available. In such cases, the exposure of sensitive data is known if attackers learn a series of queries. To prevent such attack, a query(s) is rejected from a series of queries. This is achieved by deciding on which query to reject and which to allow depending on the sensitivity of the data. In Query auditing we deny queries which have covering results, such inquiries can fill in as a potential danger to the security of hidden information. Other than covering, we utilize the disavowal dependent on the size of the inquiry. The query must fulfill the admissible size before execution. In inquiry inspecting, we can think of two potential ways, one which involves the grouping of unknown queries; called online variant and the other situation involving the arrangement of known inquiries; called the disconnected version. A k-anonymity strategy ensures a specific degree of protection. Thus, we can utilize such procedures on the outcome of inquiries for safeguarding purposes. The other technique for protection safeguarding in inquiry evaluating is irregular examination. The result of a query is registered from the arbitrary estimation of informational collections so that attacker cannot produce sets of inquiries. In another technique, we can include noise in the aftereffect of the query so that attacker never gets delicate information. We need to add a little noise to accomplish security. In the modern time of extraordinary large databases and Big Data, data is constantly examined and the above best strategy is applied and depended upon. Generally, information is not solely accessible at one spot. The information is

distributed at different spots and the proprietor of information needs to decide on working collectively with the datasets. It is conceivable to use cryptography methods for these types of applications.

Limitations of protection: the curse of dimensionality.
The above-talked about strategies function admirably when we face a circumstance like the kind of information. The curse of dimensionality says that on the off chance that we have information with different measurements the intensity of any strategy is restricted. Ordinarily, attackers have huge information on data and therefore technique such as l-diversity is proposed. To accomplish a higher degree of protection, numerous credits must be suppressed, leading to loss of information. The curse of dimensionality works with binding impact and it is accomplished by suppressing records in the database. However, the data can be determined by utilizing identifiers in the database. Employing techniques such as l-diversity requires an additional broadening of qualities yet, there is an opportunity for conducive attacks. Thus, the above approaches are repeated which may result in an ineffective database. Therefore, a trade-off decision is applied at certain stages where calculations seem difficult or infeasible to find.

The above works analyze different security dangers, protection safeguarding strategies and models with their constraints and proposes an information lake-based futuristic security conservation procedure to handle the protection safeguarding in an unstructured information. All these examinations are in relation to our study yet contextually not in the situation of telecommunications in Ghana. They are generally done in the western world. This has propelled us to undertake this study to make an addition to the contemporary literature.

Methodology:-

Research Design:
This study employed quantitative and survey-based techniques as the most suitable design for gathering and examining data. Lincoln & Kalleberg (1990) argued that, variables and relationships are the central concepts in quantitative research. Data collected and used in this analysis was from a primary source. The primary data source was by administering questionnaire to the management of telecommunication industries in Ghana.

Target Population:
The population in this research comprised Top Management of all telecommunications in Greater Accra. According to the YEN.COM.GH (2019) website, there are eight (8) telecommunication industries in Ghana.

Sample Size:
We employed a simple random sampling method in this study.

The sample covered 8 telecommunication companies, where 16 respondents (2 Top management from each company) were chosen with a margin of error of 5%.

Data Collection Method;
In the study, we used closed-ended questionnaire for the primary gathering of data to achieve privacy- preservation techniques by using the Five Point Likert Scale. We divided the questionnaire into three sections. Introduction thus, our background information and the purpose of conducting the research. Section **A** contained information of respondents and Section **B** dealt with Dependent variable (privacy preservation) and independent variables (techniques).

Data Analysis:
Analyzing the Techniques Used in Privacy- Preservation in Big Data Analytics :
A quantitative data analysis method was used for gathering and examining the data. Information from the survey was then coded and analysed using the Statistical Package for Social Science (SPSS V 25.0).

Model Specifications:
Our study made use of various inferential statistics by factoring the variables in the simple regression model. The measures of the independent variables, using the Five Points Likert scales were converted to mean values and then to percentages to permit the employment of multiple linear regression model. The regression equation was:
$Y = ß_0 + ß_1 X_1 + ß_2 X_2 + \varepsilon$.

Where Y = Privacy Preservation, X_1 = techniques (Independent Variable), X_2 = User Characteristics and **ε** = Margin of error.

Hypothesis Testing;
The basic hypothesis used for testing the study on a general level was; H0:β_1=0: If the regression parameter estimate associated with the independent variable is zero and H1:$\beta_1 \neq 0$: If the regression parameter estimate associated with the independent variable is significantly different from zero.

For testing the hypothesis, we used the estimated coefficients to deduce the correlation between the variables and the P-values for a valid conclusion. If each P-value is less than the significant value of 0.05 (α= 0.05) the null hypothesis is rejected. Otherwise, the null hypothesis is accepted.

Response Rate:
We obtained a target on each sample of 16 respondents of 8 telecommunications in Ghana and achieved 16 out of 16 responses.

This represented a response rate of 100%. This is a positive response rate for data examination as Kohli, (2011) declared that responses of 50% and higher are sufficient for analyzing data.

Table 1:- Response rate.

	NUMBER	PERCENTAGE
Response	16	100%
No response	0	0%
Total	16	100%

Descriptive Statistics;
Demographics of respondents:
Top Management:
We found out that out of the 16 respondents, 10 (63%) of the responders and 6 (37%) of the responders were male and female respectively as seen in Table 2.

In terms of age 5(31%) of the responders are below the age of 30 years, 7 (44%) of the responders are between the age range of 31-45 years, 4 (25%) of the responders are above 46 years as seen in Table 3. About the highest qualification, only 6 (38%) of the responders had a degree, 3(19%) of the responders had ICA/ACCA, 4(25%) of the responders had Master's Degree, and 3 (19%) of the responders had Ph.D. as seen in Table 4.

Also, 7 (44%) of the responders have been with their respective telecommunication for less than 5 years, 3 (19%) of the respondents have been with their respective telecommunication between 6-10 years, 5 (31%) of the respondents have been with their respective telecommunication between 11-25 years, and only 1 (6%) of the respondents has been with his/her telecommunication for more than 25 years as seen in Table 5.

Table 2:- Gender of respondents.

Gender/Company	Male	Female
MTN GHANA	2	0
GATEWAY	1	1
MOBILE CHOICE	1	1
TELIGENT WIRELESS	0	2
AIRTEL TIGO GHANA	2	0
GLOBACOM GHANA	2	0
EXPRESSO	1	1
VODAFONE	1	1
	10	6
	63%	37%

Table 3:- Ages of Respondents.

Age/Company	Below 30 years	31-45 years	Above 46 years
MTN GHANA	1	1	0
GATEWAY	0	1	1
MOBILE CHOICE	0	2	0
TELIGENT WIRELESS	1	0	1
AIRTEL TIGO GHANA	0	0	2
GLOBACOM GHANA	1	1	0
EXPRESSO	1	1	0
VODAFONE	1	1	0
	5	7	4
	31%	44%	25%

Table 4:- Educational Background of Respondents.

Higher qualification/Company	Degree	ICA/ACCA	Master	PhD
MTN GHANA	1	1	0	0
GATEWAY	1	0	1	0
MOBILE CHOICE	1	0	1	0
TELIGENT WIRELESS	1	0	0	1
AIRTEL TIGO GHANA	1	0	0	1
GLOBACOM GHANA	0	1	0	1
EXPRESSO	0	1	1	0
VODAFONE	1	0	1	0
	6	3	4	3
	38%	19%	25%	19%

Table 5:- Work Experience of Respondents.

Work tenure/Company	Below 5 years	6-10 years	11-25 years	Above 25 years
MTN GHANA	0	1	1	0
GATEWAY	2	0	0	0
MOBILE CHOICE	2	0	0	0
TELIGENT WIRELESS	2	0	0	0
AIRTEL TIGO GHANA	0	0	2	0
GLOBACOM GHANA	0	1	1	0
EXPRESSO	1	1	0	0
VODAFONE	0	0	1	1
	7	3	5	1
	44%	19%	31%	6%

Information on Privacy Preservation Techniques:

The research revealed that all the telecommunications are using two or more privacy preservation techniques in their operation.

Table 6:- Privacy Preservation Techniques.

Techniques/Usage	Using	Not using
k-anonymity	15	1
l-diversity	13	3
t-closeness	11	5
Randomisation	16	0
Data distribution	16	0
Cryptographic techniques	10	6
MDSBA	15	1
	96	16
	86%	14%

Privacy Preservation Techniques:

Table 7:- Descriptive Statistics for Privacy Preservation Techniques.

	k-anonymity	l-diversity	t-closeness	Randomisation	Data distribution	Cryptographic techniques	MDSBA
Mean	1.6	1.6	1.6	1.6	1.6	1.6	1.6
Standard Error	1.029563	1.363818	0.678233	1.16619	0.927362	0.748331	0.748331
Median	0	0	2	0	1	2	2
Mode	0	0	0	0	0	0	0
Standard Deviation	2.302173	3.04959	1.516575	2.607681	2.073644	1.67332	1.67332
Sample Variance	5.3	9.3	2.3	6.8	4.3	2.8	2.8
Kurtosis	-1.00748	4.575095	-3.08129	2.66436	1.930773	-0.61224	-0.61224
Skewness	1.016267	2.129671	-0.31536	1.714392	1.446728	0.512241	0.512241
Range	5	7	3	6	5	4	4
Minimum	0	0	0	0	0	0	0
Maximum	5	7	3	6	5	4	4
Sum	8	8	8	8	8	8	8
Count	5	5	5	5	5	5	5
Mean	1.6	1.6	1.6	1.6	1.6	1.6	1.6

Regression Analysis:

Privacy- Preservation methods in Big Data Analytics :

The results revealed that the regression model of the privacy -preservation methods and user characteristics used by the various telecommunications in Ghana have a strong and positive correspondence with privacy- preservation (regression model sig = 0.03788). The model explained 96.45% of the total variance in Privacy -Preservation in Big Data Analytics (R Square =0.964527).

The regression analysis also revealed that privacy- preservation techniques used by the various telecommunications in Ghana significantly contribute to Privacy Preservation in Big Data Analytics (B= 3.157, P=0.0014≤ 0.05).

Table 8a:- Summary output.

Regression Statistics	
Multiple R	0.982103
R Square	0.964527
Adjusted R Square	1.33333
Standard Error	1.158475
Observations	1
Multiple R	0.982103

Table 8b:- Summary output.

	df	SS	MS	F	Sig
Regression	4	109.4738	27.36845	81.57122	0.03788
Residual	3	4.026193	1.342064		
Total	7	113.5			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%
Intercept							-9E-163
							-9E-163
Intercept	7.571429	0.214321	12.56431	0.00000			
Technique	3.156743	0.328745	10.67432	0.001452	1.564523	2.98354	1.643725

User	2.137564	0.236674	9.031679	0.002866	1.384361	2.890766	1.384361

Observation	Predicted 0	Residuals
1	8.244889	-8.24489

Discussion Of The Correlation Regression Results:

Discussion:-

The result above demonstrated that the different privacy- preservation methods used by the telecommunication companies in Ghana contribute significantly to privacy preservation.

This result revealed that the total variation estimated is over 96% which implies that the variation of the error term is little beneath 4% which is a decent indication for vast model effectiveness. Privacy- preservation technique is statistically significant with an estimate and a P-value of 3.156743 and 0.001452 respectively as shown in Table 8b. Therefore we reject the null hypothesis and conclude that, there are privacy-preservation methods that are utilized in big data analytics by Telecommunications in Ghana.

The result also showed a strong and positive correlation between privacy-preservation and user characteristics employed in the study which is statiscally significant (β= 2.137564, P= 0.002866). Thus user characteristics plays a significant role in privacy-preservation methods employed by telecommunication companies in Ghana.

Conclusion and Recommendations:-

In conclusion, telecommunications in Ghana should invest in the different methods understudied to have the option to improve on privacy-preservation since each method contribute fundamentally to big data analytics.

Our findings above explained that Randomisation and Data Distribution methods (16 counts each out of the total 96 counts of method usage) are used by telecommunications more than that of MDSBA, K- Anonymity, L- Diversity, T -Closeness, and Cryptographic methods (having 15, 15, 13, 11, and 10 usages respectively) as shown in Table 6. Therefore, we suggest telecommunication companies should give more attention and preferences to Randomisation and Data Distribution methods.

We recommend a further study on the extent to which these techniques contribute to privacy preservation and also whether the industry characteristics or external factors affect the impact of the methods in privacy preservation.

This study stands limited due to inadequate funds and time for the researchers to acquire enough primary data.Therefore, data samples were acquired through reliable acquaintances but may contain noise and few anomalies that lead to a lack of viable findings.

Competing interests:

The authors declare that they have no competing interests.

References:-

1. Basso, T., Matsunaga, R., Moraes, R., Antunes, N., 2016. Challenges on anonymity, privacy, and big data, in: Proceedings - 7th Latin-American Symposium on Dependable Computing, LADC 2016. https://doi.org/10.1109/LADC.2016.34
2. Bergmann, G., Debreceni, C., Ráth, I., Varró, D. (2016)Query-based access control for secure collaborative modeling using bidirectional transformations. Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems, ACM pp. 351-361
3. Bux, N.K., Lu, M., Wang, J., Hussain, S., Aljeroudi, Y., 2018. Efficient association rules hiding using genetic algorithms. Symmetry (Basel). https://doi.org/10.3390/sym10110576
4. Cai, Z., He, Z., Guan, X., Li, Y., 2018. Collective Data-Sanitization for Preventing Sensitive Information Inference Attacks in Social Networks. IEEE Trans. Dependable Secur. Comput. https://doi.org/10.1109/TDSC.2016.2613521
5. Caballero-Gil, C., Molina-Gil, J., Hernández-Serrano, J., León, O., Soriano-Ibañez, M., 2016. Providing k-anonymity and revocation in ubiquitous VANETs. Ad Hoc Networks. https://doi.org/10.1016/j.adhoc.2015.05.016

6. Chauhan, G., 2013. A review of privacy preservation techniques. Int. J. Eng. Resea & Tech. Vol 2 (12)
7. El Emam, K., Dankar, F.K., Issa, R., Jonker, E., Amyot, D., Cogo, E., Corriveau, J.P., Walker, M., Chowdhury, S., Vaillancourt, R., Roffey, T., Bottomley, J., 2009. A Globally Optimal k-Anonymity Method for the De-Identification of Health Data. J. Am. Med. Informatics Assoc. https://doi.org/10.1197/jamia.M3144
8. Gkountouna, O., 2015. A survey on privacy preservation methods.
9. Kundalwal, M.K., Chatterjee, K., Singh, A. (2108) An improved privacy preservation technique in health-cloud, ICT Express https://doi.org/10.1016/j.icte.2018.10.002
10. Kohli, S. (2011). The usefulness of the quantitative method of statistical analysis (SPSS) as a mode of data analysis for research purposes, Munich, GRIN Verlag, https://www.grin.com/document/170289
11. Kumar, A., Ligatti, J., Tu, Y.C., 2015. Query monitoring and analysis for database privacy - a security automata model approach, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-319-26187-4_42
12. Lincoln, J. R., & Kalleberg, A. L. (1990). Culture, control, commitment: A study of work organization and work attitudes in the United States and Japan. Cambridge University Press, Cambridge, UK. York.
13. Mathew, C., 2019. A Survey on Privacy Preserving Data Mining Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH &TECHNOLOGY (IJERT) NCACCT (Volume 7 – Issue 05),
14. Patel, K., 2019. A study on T-Closeness over K-anonymization Technique for Privacy Preserving in Big Data, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 05
15. Praveena Priyadarsini R., Sivakumari S., Amudha P. (2016) Enhanced ℓ – Diversity Algorithm for Privacy Preserving Data Mining. In: Subramanian S., Nadarajan R., Rao S., Sheen S. (eds) Digital Connectivity – Social Impact. CSI 2016. Communications in Computer and Information Science, vol 679. Springer, Singapore
16. Ram Mohan Rao, P., Murali Krishna, S., Siva Kumar, A.P., 2018. Privacy preservation techniques in big data analytics: a survey. J. Big Data. https://doi.org/10.1186/s40537-018-0141-8
17. Rajasekaran, M., Thanabal, M.S. 2019, A Survey on Sensitive Association Rule Hiding for Privacy Evaluation of Methods and Metrics, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 10
18. Samarati, P., Sweeney, L., 2017. Protecting Privacy when Disclosing Information: k-Anonymity and its Enforcement Through Generalization and Suppresion. Proc IEEE Symp. Res. Secur. Priv.
19. Samarati, P., 2001. Protecting respondents' identities in microdata release. IEEE Trans. Knowl. Data Eng. https://doi.org/10.1109/69.971193
20. Sweeney, L., 2002. K-anonymity: A model for protecting privacy. Int. J. Uncertainty, Fuzziness Knowlege-Based Syst. https://doi.org/10.1142/S0218488502001648
21. Turkanović, M., Družovec, T.W., Hölbl, M., 2015. Inference Attacks and Control on Database Structures. TEM J.
22. https://yen.com.gh/108474-list-telecommunication-companies-ghana-2019.html.