International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Increasing Data Anonymity using AES and Privacy Technique

Thamer Khalil Esmeel, Roslina Abd Hamid, Rahmah Mokhtar

Abstract*: Data security and data preserve privacy had been an important area to a huge in recent years. However, rapid developments in collecting, analyzing, and using personal data had made privacy a very important issue. This thesis had addressed the problem the protect user data in the dataset from attacks internal and attacks external by using combination techniques between security technique, and privacy technique and data mining technique. The research objectives were to determine the privacy and security technique in suitable the dataset, and to implement the combination property with chosen and security technique in order to protect user data in the dataset and to validate by comparing result before and after apply privacy techniques in dataset using chosen data mining tool. The research methodology consists of three phases. the analysis phase, combination techniques phase, and results evaluating phase and for every phase has research objective.

***Keywords*: Security techniques, Privacy techniques, Data mining techniques, Weka tools, Weka Explore interface, Weka Experimenter Environment interface**

I. INTRODUCTION

Data security protection and data privacy had been an area of interest in recent years. In this digital world, data play a core role in our life, the data can be classified into two types: public data which are open to everyone, and secret data which accessed only by worked users, encryption is one about the ways for improving information security, where the term 'encryption' refers to changing Understandable information into no Understandable (Abusalim, 2015). Encryption is a way by which data digitals is converted to encoded data couldn't be decoded if the user has the encryption key. It represents one of several techniques that may be applied to protect our research dataset from unauthorized access, encryption is one of the ways means to ensure the security of important information. encryption algorithm run changes on the plaintext such as original messages before encryption can transform it's into cipher-text scrambled messages after encryption, there are much encryption algorithms, are available which can be used in data security. Encryption algorithms are divided into two groups, encryption of symmetric key called (secret key), and encryption of symmetric key called (public key). encryption of the key of symmetrical is the shape of encryption where encryption, and decryption are done by the same key. Encryption of asymmetric is the shape of encryption where encryption, and decryption are executed by using different keys, one (a public key) and the other (a private key), was used by (Abood & Guirguis, 2018).

Nowadays, data privacy is a great important task, especially in a large data set. Privacy means the identity of a person who is not disclosed while disclosing any type of data or using the data for any search. There are many ways to protect privacy such as data concealment and data distortion, to achieve the purpose of privacy protection. These methods are achieved by changing sensitive and insensitive attributes. Sensitive attributes are fields that should not be disclosed or published against the person, which if detected lead to many problems and non-sensitive attributes are fields that if detected do not lead to any problem (Kaur, 2017). Data mining involves searching for certain patterns and facts about the structure of data within large complex datasets (Muhammad, 2016). Data mining can discover important relationships which can improve health, business processes, and many other specializations, mining patterns of hidden and strategic knowledge from big datasets that are stored electronically, and the challenges faced by many organizations. Data mining scope is harmless mining of implicit data that was unknown before and which may be useful from data warehouses. Machine learning uses statistical techniques and visualization to discover information and present it easily understood by humans. The data mining scope is a major recognition of the gathering of the data big amounts and stores it easily across computer systems (Abbas et al., 2015). The propose had been in protecting user data in the dataset from attacks internal and attacks external, preserve privacy to user data in the dataset by using combination techniques among (security technique, and privacy technique, and data mining technique).

II. RELATED WORK

The encryption algorithm and an authentication system to transfer information securely. The algorithm is a variance of AES and implemented between multiple devices, AES uses the only private key (symmetric key) to encrypt data, and the implementation works on one reject multi-border of degree 8. Which was using to calculate double inverse tables and S boxes and invert the S boxes required to work from each layer in the algorithm,

Manuscript published on 30 September 2019
* Correspondence Author
Thamer Khalil Esmeel, (a) Faculty of Computing. Universiti Malaysia Pahang. Malaysia. Email: thamerkhalil29@yahoo.com
(b) Ministry of Higher Education and Scientific Research., University of Mosul, Computer Science, College of Arts, Iraq
Roslina Abd Hamid, Faculty of Computing. Universiti Malaysia Pahang. Malaysia Email: roslina@ump.edu.my
Rahmah Mokhtar, Faculty of Computing. Universiti Malaysia Pahang. Malaysia Email: drrahmah@ump.edu.my

compared with AES, using sixteen one reject multi-border of grade" 8 "instead of one in progress, the key sizes for AES are usually They are 128, and 192, and 256 bits in size,

Using a 128-bit key, unlike symmetric encryption, asymmetric encryption uses two keys, private and public keys, public keys are shared between connected parties, while (private keys) are kept secret, keys are supposed to be large in size to maintain their strength, Thus typically ranging from (512 bits) to (2048 bits) or more for asymmetric encryption, connected parties can start with (public keys), and (private keys) and agree on key of shared (session key) used as an AES key, and the (Diffie-Hellman) exchange protocol of key uses multi of algorithms AES and algorithm RSA as the result of reliable encryption systems, the algorithm focuses on creating a new protocol for key creation and agreement, using a hybrid framework, implementing a communication protocol between two devices, with the possibility of working with multiple devices using the central server, and the result was presented to create a new way to encryption and enhance, the algorithm has been used by (Daddala et al., 2017).

An efficient new method of data encryption algorithm using two keys, one provided and created by the system for the individual user, the other from the user as a basic rule for the algorithm, by AES there was the symmetric key using only one key, so the proposed encryption system was symmetrical, using a two-key method, the new algorithm had multiple useful properties, was fast, could be easily implemented in hardware, like DES, when charging one part of the data in an encryption code there will be a lot of bits charged, can vary the length of the data block, you can easily change the encryption mode, and used for the approach proposed by large organizations, governments, individuals, banks, etc., the algorithm has been used by (Abusalim, 2015). Propose to compare and find the best (FPGA) application for DES to secure smart cards, knowing that smart cards occupy a small area of chips, low power consume, so applications require security devices with more space and power constraints and less productivity, These results can be useful in choosing the correct FPGA application for DES application and for a custom application. The results are considered in the framework of FPGA applications, languages, pipeline-related applications, etc. it has been used by (Dichou et al., 2015). Propose a new image encryption system based on virtual optical technology, a digital hologram transformation phase, and the public key RSA exchange. In the system, the encryption keys include the length-wave of the laser beam, the focal- length of the test lens, the focusing distance and the measuring factor. Plain text photo is encoded into encryption using a phase-shifted 3D image function. The decoding process involves calculating the phase using phase-scanning and sequential line algorithms. The keys are exchanging between the receiver and sender using the key RSA exchange algorithm. For determination, the accuracy of the information retrieved by the proposed technique, the average error between the original and the recovered image is calculated, the image encryption system has been used by (Chatterjee et al., 2018).

Propose an algorithm to learn the most expecting decision tree. Expected results that this algorithm produces decision trees that approximate the accuracy of non-specific decision trees, indicating that a nearby neighbour classifier and the vector is theoretically expected to practice anatomical data, as well as original data. Several datasets are publicly available, demonstrating the good performance to the end neighbours and the classification of support carriers by using k-anonymity, the algorithm has been used by (Mancuhan, 2017). The semantic addition algorithm which deals with the specific, discrete and nonordinal nature of the name data, the proposed algorithms Included both the addition of unrelated noise, which was suitable for independent attributes and associated noise, which can deal with multi-variate datasets with dependent, experimental attributes. Proposals provide general and distorted mechanisms for distorting the name data while maintaining better semantics than basic ways based on data distribution, the algorithm has been used by (Rodriguez-Garcia et al., 2017).

Algorithm for xk-anonymity approach. The algorithm has been create to prefer and give the accuracy and maximum range of data hide linked with the suggested approach, highlighting the results from each experiment and how it influenced the creation of an adaptive algorithm, the xk-anonymity model-based on two models of foundational privacy, anonymity and diversity models l , And many data and power in the dataset, the algorithm has been used by (Brown, 2019). Proposing three ways to aggregate (K-anonymous) for datasets that have an almost zero precision loss and are very strong. It was not only possible to replace all feature vectors with the feature vector inside each bag, but also the loss of precision due to differential privacy can be associated with a small number, ensuring low loss of accuracy when training the LLP on a data set (PPDM), and evaluating the PPDM model LLP) on two data sets, one was an adult data set and the other was an elevated dataset Instagram, both of which provide empirical evidence of low-resolution loss after application of the model (PPDM LLP), the application has been used by (Yan, 2018).

Data mining methods in Weka tools through Explorer Interfaces and Knowledge Flow Interfaces and Experimenter Interfaces, to validate their approach, using the dengue dataset within 108 cases, but Weka tools used 99 rows, 18 attributes for determining disease prediction and accuracy of their use. Classifications from different algorithms for better performance and data classification, help users mining useful information from data and easy diagnostic an algorithm suitable for its accurate predictive model, from the results, conclude that Naive Bayes, J48 is the best algorithms in performing for rated accuracy because the maximum of Detective accuracy = 100% With 99 correctly categorized cases, maximum ROC = 1, it meant at least absolute error and took minimal time to build this model through the results of Explorer Interfaces and Knowledge Flow Interfaces, the model has been by used (Shakil et al., 2015).

The approach is to perform thirteen workbooks using cross-checking (N-fold) available in the Weka tool for machine learning and obtained an accuracy of better than 96.28% to a detection of malware, which was more than the upper detection accuracy (95.9%), in these upper five classifiers (FT, LMT, J48, Random forest, and NBT), the approach got the accuracy of detection 97.95% by used (Random forest), the approach has been used by (Sharma & Sahay, 2016). The polynomial logistic regression is best rated with the highest accuracy in each binning level for both

SLECR and SLDC followed by a decision tree, the ability to accurately predict rejections with low ranges to predict the lead time period, usually the performance of the works used such as (K-nearest neighbors, multinomial logistic regression, decision tree, support vector machine) are better when bidirectional categorical variables and Naive Bayes work best when categorical variables are converted into ordinal values, and the results will greatly benefit different parties in the supply chain by providing improved Vision and Vision Predictability, the approach has been used by (Hathikal, 2018). The Boosted tree and Random forests gave logical results for analysis, Random Forests generated 100 trees, and Boosted Tree generated 200 trees to give one result on based Bagging learning, each algorithm ranked the variables on based the most important, the best four ranked variables had been selected for further analysis bagging technique used to confirm variables on based the most popular instances, final variables were selected from (data mining, Machine Productivity, Pigment Fastness, and Pile Weight), multiple regression method was applied to predict the equation on based the textile quality score, Before applying linear regression, many algorithms such as artificial (neural network and multivariate adaptive regression) were applied to predict the equation, but these algorithms did not yield good, the approach has been used by (Saad, 2018).

A. Data Mining Techniques in Weka Tools

i. *Naive Bayes algorithm:* is a selective classifier calculates the probability set by calculating the combination and succession of values in the dataset. It is assumed all that variables that contribute to the classification are independent of others. The Bayes naive workbook is on based Bayes theory and total probability theory, the algorithm has been used by (Bhagyashree et al., 2018).

ii. *J48 algorithm:* Optimal enforcement of C4.5 is called. The output by J48 is (Decision Tree). A (Decision Tree) is the same structure tree that contains different nodes, such as the root node, middle node, and the leaf node, each node in the tree has a resolution lead to decision leads to a result. A (Decision Tree) divides the data set entry area into reciprocal spaces, where each region contains a label, value, or procedure to describe or clarify its data points. The partitioning criterion in (Decision Tree) is used to calculate which attribute is best for dividing that part tree of training data that reaches a particular, the algorithm has been used by (Kiranmai & Laxmi, 2018).

iii. *Multilayer Perceptron algorithm (Neural network):* A single-layer perceptron can only classify linear separable problems. For inseparable problems, it is necessary to use more layers. The multilayered network (forward feed) contains more than hidden layer whose nerve cells are called hidden neurons, the algorithm has been used by (Amin & Habib, 2015).

B. Privacy Techniques

i. *Differential privacy:* Differential privacy distorts data of sensitive by way addition noise with some data or data attributes unchanged. Thus, the processed data still retains certain statistical characteristics for data extraction, the technique has been used by (Yao, 2018).

ii. *k-Anonymity:* Is one of the most adopted methods used to address individual privacy issues while sharing data, while maintaining the usefulness and accuracy of data, and trying to protect individual privacy by adding K-1 records to a dataset to minimize the risk of the original log attributes being redefined, It provides a solution to band the detection data of sensitive, the technique has been used by (Chan et al., 2016).

iii. *Sample-uniqueness:* The masked data are a sample of the original data, which is construed as the population. Thus, if an intruder identifies a unique record in the released data (sample), cannot be sure it was unique in the original data (population), which thwarts re-identification, the more protection.

C. Security Techniques

Advanced Encryption Standard (AES): AES is a modern encryption strategy proposed by NIST to replace DES back in 2001. AES can provide any set of databases. During decryption, the AES process encrypts (10 rounds) of (128-bit) keys. (12 rounds) for (192-bit) keys and (14 rounds) for (256-bit) keys to exit with the last encrypted message. AES allows a (128-bit) length of information to divide into 4 active-active blocks. These segments are treated as a bytes line and a 4 * 4 array is compiled as "state", for encryption and decryption, encryption begins with "Add round master stage". However, shortly before the final round, the output faces 9 basic rounds, up to every 4 conversions.

III. METHODS

There are three main phases in this paper, which are associated with the three research objectives, respectively consists of the analysis phase, combination techniques phase, and results evaluating phase and for every phase has research objective.

A. The Analysis Phase

In this research, the work on data collection to chosen suitable the dataset and analyses it, where the dataset used one of the heart patients files downloaded from one the website in the formula Attribute-Relation File Format (ARFF) And changes its format to Comma-Separated values (CSV) to use it in the research experience shows as Table I.

Table I Dataset File Format (CSV)

	sex	age	chest pain type	blood pressure	cholesterol	Fasting blood sugar > 120
1	Male	60	Asymptomatic	130	206	FALSE
2	Male	49	Abnormal Angina	130	266	FALSE
3	Male	63	Asymptomatic	130	254	FALSE
4	Male	53	Asymptomatic	140	203	TRUE
5	Female	58	Angina	150	283	TRUE
6	Male	58	NoTang	132	224	FALSE
7	Male	63	Angina	145	233	TRUE
8	Male	67	Asymptomatic	160	286	FALSE
9	Female	41	Abnormal Angina	130	204	FALSE
10	Male	56	Abnormal Angina	120	236	FALSE
11	Female	62	Asymptomatic	140	268	FALSE
12	Male	56	NoTang	130	256	TRUE
13	Male	44	Abnormal Angina	120	263	FALSE
14	Female	50	NoTang	120	219	FALSE
15	Male	43	Asymptomatic	150	247	FALSE
16	Female	69	Angina	140	239	FALSE
17	Male	60	Asymptomatic	117	230	TRUE
18	Male	59	Asymptomatic	135	234	FALSE
19	Male	44	NoTang	130	233	FALSE

B. The Combination Techniques Phase

In this research use the comparison properties and combination between all research techniques

i. The comparison property in data mining techniques in Weka tools on the dataset before using privacy techniques and data mining techniques in Weka tools on the new dataset after use privacy techniques.

ii. The combination property of privacy techniques and security techniques on the new dataset.

The apply Combination techniques can be divided into three main areas:

❖ **The first area** is using data mining techniques in Weka tools by (J48, Naive Bayes, and Neural Network). Steps applying the use of data mining in Weka tools on the dataset and get the results:

- Step1: Start.
- Step2: Input the dataset.
- Step3: Comparison input dataset if (Yes) go to use privacy algorithm and progress, else (No) go to data mining in Weka tools.
- Step4: choose (No) go to data mining to apply three algorithms in Weka tools (J48, Naive Bayes, and Neural Network) on the dataset.
- Step5: Compare the different accuracy provided by the dataset with different classification algorithms and identify the significant classification algorithm for a particular dataset.
- Step6: Save prediction.
- Step7: End.

❖ **The second area**: is using three privacy techniques are (Differential privacy technique, k-Anonymity technique, Sample-uniqueness technique). Steps to apply use privacy techniques:

- Step1: Start.
- Step2: Input the dataset.
- Step3: Comparison privacy implemented on dataset if (Yes) go to use privacy techniques and progress, else (No) go to use data mining in Weka tools.
- Step4: choose (Yes) go to apply three privacy techniques (Differential privacy technique, k-Anonymity technique, Sample-uniqueness technique) on the dataset.
- Step6: Save the dataset in a new name file.
- Steps to apply the use of data mining in Weka tools on a new dataset and get the new results:
- Step7: Input the dataset.
- Step8: Use data mining to apply three algorithms in Weka tools (J48, Naive Bayes, and Neural Network) on the dataset.
- Step9: Compare the different accuracy provided by the dataset with different classification algorithms and identify the significant classification algorithm for a particular dataset.
- Step10: Save prediction.

❖ **The third area**: is using the security technique by four protection methods are the Encryption key by use AES algorithm and the Decryption key by AES algorithm, in Encryption: split file (.csv) into Five files(.csv) and upload them to five servers, in Decryption: the Five files download from the Five servers and Merge the Five files (.csv) into one file.

Steps to applying security techniques to encryption the file:

- Step1: Start.
- Step2: Input the dataset.
- Step3: Comparison privacy implemented on dataset if (Yes) go to use privacy techniques and progress, else (No) go to use data mining in Weka tools.
- Step4: chose (Yes) go to apply three privacy techniques (Differential privacy technique, k-Anonymity technique, Sample-uniqueness technique) on the dataset.
- Step6: Save the dataset in a new file(.csv).
- Step7: Comparison security implemented in a new file(.csv) if (Yes) go to use security techniques and progress, else (No) go to (End).
- Step8: By using AES generated the private key to encrypt the new file(.csv).
- Step9: Split the new file(.csv) into 5 files.
- Step10: Encryption each file by dividing the public key into five keys so that each key of the five keys attaches to one file of the five files.
- Step11: Uploading the five files to the five servers.
- Step12: End.

IV. RESULTS

A. Comparison Property:

The comparison property results, between privacy techniques and data mining techniques in Weka tools.

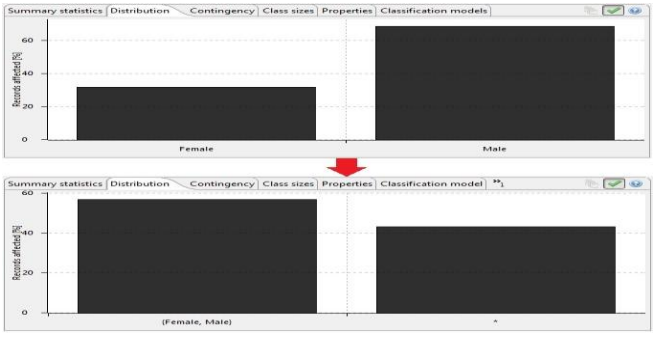

Fig.1. Privacy techniques by analysing Distribution

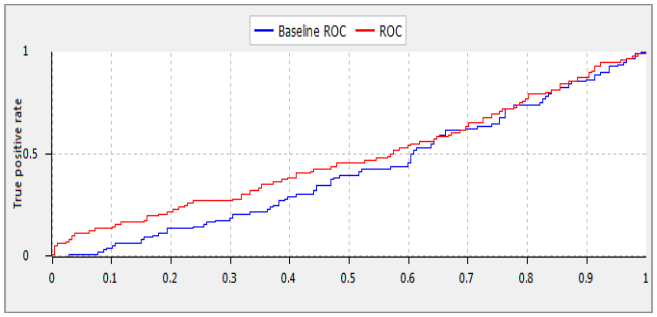

Fig.2. Privacy techniques by analyzing Contingency

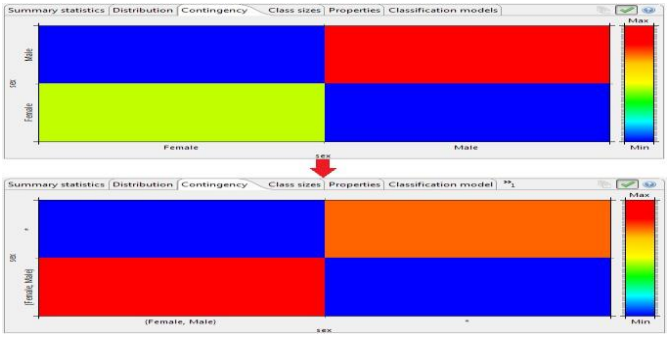

Fig. 3. Privacy techniques by analysing ROC curves

Table II Comparison of statistic prediction results in Weka tools Explorer

Dataset	Naive Bayes	J48	MultilayerPerceptron (Neural Network)
Heart disease.scv	82.1782%	75.9076%	81.8482%
Heart disease privacy.scv	80.0766%	79.3103%	73.9464%

Table III Comparison of statistic prediction results in Weka tools Experimenter Environment

Dataset	Naive Bayes	J48	MultilayerPerceptron (Neural Network)
Heart disease.scv	82.32%	76.79%	77.74%
Heart disease privacy.scv	80.78%	79.93%	74.89%

B. Combination Property

The combination property results, between privacy techniques and security techniques.

Table IV Dataset after Applying three Privacy Techniques

	sex	age	chest pain type	blood pressure	cholesterol	Fasting blood sugar > 120
1	{Female, Male}	[40, 90]	Asymptomatic	130	2**	FALSE
2	{Female, Male}	[40, 90]	Abnormal Angina	130	2**	FALSE
3	{Female, Male}	[40, 90]	Asymptomatic	130	2**	FALSE
4	{Female, Male}	[40, 90]	Asymptomatic	140	2**	TRUE
5	{Female, Male}	[40, 90]	Angina	150	2**	TRUE
6	{Female, Male}	[40, 90]	NoTang	132	2**	FALSE
7	{Female, Male}	[40, 90]	Angina	145	2**	TRUE
8	{Female, Male}	[40, 90]	Asymptomatic	160	2**	FALSE
9	{Female, Male}	[40, 90]	Abnormal Angina	130	2**	FALSE
10	{Female, Male}	[40, 90]	Abnormal Angina	120	2**	FALSE
11	{Female, Male}	[40, 90]	Asymptomatic	140	2**	FALSE
12	{Female, Male}	[40, 90]	NoTang	130	2**	TRUE
13	{Female, Male}	[40, 90]	Abnormal Angina	120	2**	FALSE
14	{Female, Male}	[40, 90]	NoTang	120	2**	FALSE
15	{Female, Male}	[40, 90]	Asymptomatic	150	2**	FALSE
16	{Female, Male}	[40, 90]	Angina	140	2**	FALSE
17	{Female, Male}	[40, 90]	Asymptomatic	117	2**	TRUE
18	{Female, Male}	[40, 90]	Asymptomatic	135	2**	FALSE
19	{Female, Male}	[40, 90]	NoTang	130	2**	FALSE

Table V Dataset after Applying Security techniques by use Encryption by AES

gAAAAABdReKX3dBSKIdUpIyZ2ewmXrq7cQdVFH3Ge1OABrUhSjP
t9tdh-plr0hkXe6r7AEr1UPhNoiBypXXVc3EUfgNxNMhDadJ1KM
KITbJeUbFId7FR-9RkLB1jV_l1ilf7tkvHcmv3_TzcPW5emdERr
_Sl08fdaU-FCf1rdCWSavlDcMA7wxGENFg5E1D2lN-xqh3jMVLH
i2beeVnvEx1qpVMDC5OzQT28e00yoYVYNnbR4_ZUcp265HXc58m
YJ85yRDq8P-O2BIFm9tCiwZWNj3S5KQqLBB_lBGzOkJ4-vMckyq
XNwlir1ZPWgjSoUIG27QDc8_gwGq7tJZ2kp1K60BnmK3hciP5s7
2E6fKqks45aJzBEUdogiSPoM4OEBh-G4kXP14DsBQhGPgvEerg4
qpIAr99MqqnBo72yiTOp41FDowHXgwQexy7m-cN1Dc0fomZlqa5
HCInB7oTuHMxU7EKI9mPkn3NSC_49GVKzDer8e64ql-nSEZDVK0
rGmM0Aw1NQF8jIR2JFoCPRKeaFf-Xp7Lju7qclEJjC_F0YvT26D
_JIZI0k4iZUUxQ-RSRV1q6M3X2z3RC5oEK5pFMTBc-NJb0ep1du
4XM_T0E2zLAJQtzFtSmjjrlDOi8lyJn0lmbzsYkO1m9Ixr4Gbmf

Table VI Dataset after Applying Security techniques by use Decryption by AES

	sex	age	chest pain type	blood pressure	cholesterol	Fasting blood sugar > 120
1	Male	60	Asymptomatic	130	206	FALSE
2	Male	49	Abnormal Angina	130	266	FALSE
3	Male	63	Asymptomatic	130	254	FALSE
4	Male	53	Asymptomatic	140	203	TRUE
5	Female	58	Angina	150	283	TRUE
6	Male	58	NoTang	132	224	FALSE
7	Male	63	Angina	145	233	TRUE
8	Male	67	Asymptomatic	160	286	FALSE
9	Female	41	Abnormal Angina	130	204	FALSE
10	Male	56	Abnormal Angina	120	236	FALSE
11	Female	62	Asymptomatic	140	268	FALSE
12	Male	56	NoTang	130	256	TRUE
13	Male	44	Abnormal Angina	120	263	FALSE
14	Female	50	NoTang	120	219	FALSE
15	Male	43	Asymptomatic	150	247	FALSE
16	Female	69	Angina	140	239	FALSE
17	Male	60	Asymptomatic	117	230	TRUE
18	Male	59	Asymptomatic	135	234	FALSE
19	Male	44	NoTang	130	233	FALSE

V. DISCUSSION

In this study, the steps of data mining techniques in Weka tools on the dataset were carried out and explained it, and the steps of privacy techniques on the dataset were carried out and explained it, and the steps of security techniques on the dataset were carried out and explained it.

- Data mining technique in Weka tools on the dataset by three algorithms (Naive Bayes, J48, Neural Network) was succeeded in order to the best prediction statistic results before and after apply privacy techniques on the dataset.
- Privacy technique on the dataset by three privacy techniques (Differential privacy, k-Anonymity, Sample-uniqueness) was succeeded in order to preserve privacy the user data in the dataset.
- Security technique on the dataset by one security technique AES was succeeded in order to protect user data in the dataset.

VI. CONCLUSION

There are three different phases in methodology they are so called phase one, phase two and phase three. Phase one is linked with the analysis phase, phase two is linked with the combination techniques phase, and phase three is linked with the results evaluation phase.

- Phase one focuses on the analysis phase which involves conducting the literature review and exploring the existing research works related to these techniques, the analysis in the first phase provides commonly used techniques as listed in this thesis. In this phase, the suitable dataset and determining the suitable techniques and understanding framework in the research phase were determining.

- Phase two emphasized on combination techniques which consist of two properties which are a comparison property and a combination property and implementing into the proposed techniques (data mining techniques in Weka tools, privacy techniques, security techniques).

- Phase three focused on evaluated the results of comparison property and a combination property and analyzed, and evaluating the results of predicting in the comparison property between data mining techniques in Weka tools before and after application the privacy techniques on the dataset, which was satisfying, and evaluated results of combine the privacy techniques and security techniques in the combination property.

REFERENCES

1. O. Abbas, M.E., Mustafa, S.B. Ibrahim, "The Role of Data Mining in Information Security". Int, J. of Computer (IJC),, 17(1), 2015, 1-20.
2. O.G. Abood, S.K. Guirguis, "A Survey on Cryptography Algorithms" Int. J. of Sci.Res,. Pub. (IJSRP), 2018, 8(7).
3. B. Y. Abusalim, "An Efficient Approach For Data Encryption Using Two Keys". An Efficient Approach for Data Encryption Using Two Keys. 2015
4. M. N. Amin, M. A. Habib, "Comparison of different classification techniques using WEKA for hematological data". American J. of Eng. Res. 4(3), M. 2015, 55-61.
5. S. I. R. Bhagyashree, K. Nagaraj, M. Prince, C.H. Fall, M. Krishna "Diagnosis of Dementia by Machine learning methods in Epidemiological studies: a pilot exploratory study from south India". Social psychiatry and psychiatric epidemiology, 53(1), 2018, 77-86.
6. E. E. Brown, . Adaptable Privacy-preserving Model , 2019.
7. M. Chan, H. Elsherbini, X. Zhang, X. "User density and spatial cloaking algorithm selection: Improving privacy protection of mobile users. Paper presented at the 2016 IEEE 37th Sarnoff Symposium.
8. A. Chatterjee, J. Dhanotia, V. Bhatia, S. Prakash, "Virtual optical encryption using phase shifted digital holography and RSA algorithm". Paper presented at the 2018 3rd International Conference on Microwave and Photonics (ICMAP).
9. B.Daddala, H. Wang, A.Y. Javaid, "Design and implementation of a customized encryption algorithm for authentication and secure communication between devices". Paper presented at the 2017 IEEE National Aerospace and Electronics Conference (NAECON).
10. K. Dichou, V. Tourtchine, F. Rahmoune, "Finding the best FPGA implementation of the DES algorithm to secure smart cards". Paper presented at the 2015 4th International Conference on Electrical Engineering (ICEE).
11. S. Hathikal, "Prediction of Ocean Import Shipment Lead Time for Freight Forwarder Using Machine Learning Techniques". ProQuest Dissertations and Theses, 2018, 74-74.
12. A. Kaur, "A hybrid approach of privacy preserving data mining using suppression and perturbation techniques. Paper presented at the 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). 2017
13. S. A. Kiranmai, A.J. Laxmi, "Data mining for classification of power quality problems using WEKA and the effect of attributes on classification accuracy". Protection and Control of Modern Power Systems, 3(1), 2018, 29.
14. K. Mancuhan, "Data Classification for l-diversity". Purdue University. 2017
15. A. M. Muhammad, "Advances in Clustering based on Inter-Cluster Mapping". Western Sydney University (Australia). 2016.
16. M. Rodriguez-Garcia, M., Batet, D. Sánchez, "A semantic framework for noise addition with nominal data". Knowledge-Based Systems, 122, 2017, 103-118.
17. H. Saad, " An Integrated Framework of Data Mining and Process Mining to Characterize Quality and Production Processes". State University of New York at Binghamton. 2018
18. K. A. Shakil, S. Anis, M. Alam, "Dengue disease prediction using weka data mining tool." arXiv preprint arXiv:1502.05167. 2015
19. A. Sharma, S.K Sahay, "An effective approach for classification of advanced malware with high accuracy). arXiv preprint arXiv:1606.06897. 2016
20. X. Yan, X. "Privacy Preserving Bag Preparation for Learning from Label Proportion". Illinois Institute of Technology. 2018
21. S. Yao, "An Improved Differential Privacy K-Means Algorithm Based on MapReduce". Paper presented at the 2018 11th International Symposium on Computational Intelligence and Design.

AUTHORS PROFILE

Mr Thamer Khalil Esmeel is an aspiring research student at the Faculty of Computing, Universiti Malaysia Pahang, Malaysia. He works at the Ministry of Higher Education and Scientific Research., University of Mosul, Computer Science, College of Arts, Iraq. His area of specialization is in software Engineering. He has great quest to learn and collaborates with great minds.

Dr Roslina Abd Hamid is a lecturer at the Faculty of Computing, Universiti Malaysia Pahang, Malaysia. Her area of specialization is Software Engineering. She has authored many papers locally and internationally with many conference proceedings to her credit.

Dr Rahmah Mokhtar is a lecturer at the Faculty of Computing, Universiti Malaysia Pahang, Malaysia. Her area of specialization is Information System. He is an erudite scholar who has publish in many journals and attended conferences.