Excessive Permissions Investigation with Data-Driven Account Security with Classification Performance

Heri Satria Setiawan
Informatic
University of PGRI Indraprasta
Jakarta, Indonesia
herisatria@unindra.ac.id

Agus Pamuji
Informatic
IAIN Syekh Nurjati Cirebon
Cirebon, Indonesia
agus.pamuji@syekhnurjati.ac.id

Rudi Suparman
Informatic
University of Pelita Bangsa
Cikarang, Indonesia
suparmanj762@gmail.com

Abstract—Many companies lack configuration systems due to the need to protect assets from unauthorized access by individuals or groups. Data mining can help by securing the configuration system to identify accounts in the database. Given the sensitivity of activities on the database system, access permissions are a major concern, especially with unauthorized users. Excessive permissions can compromise database security, making it important to group users into authorized and unauthorized classes. This study uses the decision tree method to extract and investigate factors that affect excessive permissions, and validates the dataset with 10-fold cross-validation to ensure data quality. The final result identifies two classes for user access, showing that the decision tree method performs well with significant values on the AUC curve and the Confusion Matrix.

Keywords—account database, classification, data mining, database security, Excessive Permissions.

I. INTRODUCTION

In developing countries, technology is a major concern in growing the economic sector. Not only on the economy but also security side. The main concern is making technology considered important in government, non-governmental or non-governmental organizations considering the data security activities in information technology in particular. If you look at the security side, the field of computing that discusses security is very broad and very complex [1]. Security in information technology is the same as computer security or cybersecurity. With the concept of security, it is an effort to protect, maintain the resources on hardware and software even on the internet as an extension of a local computer network. There are essentially three main dimensions to information technology security. First, how the resources in this concept in the form of data can be guaranteed to exist in the system in general. Second, building the integrity of resources in the form of data and the dimensions of confidentiality are the demands of every user involved [2]. In this concept, the definition of security must be emphasized, namely on the preliminary side when there is loss or damage to the computing system. The next step is to look at the weaknesses in the system that could potentially be exploited to cause damage. If it is detected, then attacks, threats will arise and must be controlled with an effective strategy.

The security dimension will be used as a reference to secure resources in the form of data packaged in the database system. Data can be considered as a resource that has a high sensitivity rating in addition to targeted attacks on other resources. Although many software developers produce reliable and high-quality products, data must be secured by methods, strategies or even approaches. The goal is to critically identify existing database software as datasets [3]. At the same time, one must understand the potential risks posed as well as weaknesses in these assets. The addition is being able to mitigate the risk in assets. With the implementation of database system security, it is necessary to first understand the requirements plan. The reason is that along with the main task of security is to ensure so that security can measure the security plan against data resources.

The collaboration of data mining with the security sector has been researched and studied for a long time [4]. Data mining is a study in finding knowledge from determined patterns. With a pattern that has been formed based on the data collected. Various methods plus data analysis techniques were carried out to evaluate the performance of each method. If data mining is oriented towards determining patterns and finding knowledge with databases as actual information, then security, especially computing, is a way of protecting data resources. Data mining can improve the performance of information technology security. Methods, techniques and security algorithms, especially in cryptography, have also been widely studied and are quite significant in reducing the number of losses caused by attacks [5]. However, much of the literature mentions the effectiveness of the method, which means that security measures are assumed if there has been the identification of attacks and vulnerabilities [6]. With the data mining concept presented, it will provide improvisation on the advantages of data mining. There are several data mining methods that are used when determining patterns and extracting information. First, perform estimation, prediction, classification, clustering, and also association. Thus, data mining can estimate the hazards in addition to the security risks that arise in the resource. However, data mining must still be considered considering the ability to extract information for early prevention before the occurrence of leaks due to attacks on data. Various kinds of literature state that data mining can detect intrusions on systems with various variants. Given the complex security issues, the discussion of data mining will be narrowed down to the field understudy.

Every database has a user and attached schema. This means that each user on the database system is given an identity [7]. One identity has the same or identical schema. This is an early weakness. Continuity is found that users are given access to all types of access. Granting access rights is also granted to users who are connected to the database system. Excessive permissions are considered as permission attempts against redundant data. This trend can be seen from the activity log which contains anomalous data. In fact, user access rights and permissions must be limited because not all data can be granted permission. Not only data but actions cover the activities of users who are connected to the system. Because security cannot work anonymously, configuration and granting access to data is also a fundamental consideration.

Anything related to data, the sensitivity rating is very important. With the security of the database, it must be secured with a data approach that involves an information disclosure strategy [8]. In other words, Data security is made possible by data. This research will analyze the classification of user types in the database system. As for this concept, a decision tree method will be adopted to investigate anomalous data related to the type of user connected to the database system [9]. Similarly, a framework will be proposed that exploits user information on the permissions side.

The strongest reason for the data mining approach will be to improve and increase the performance of securing data in addition to resources. Very little research has led to behavioural data reflected in user activity logs or history. The rest are users who are connected to the database system that is equipped with access. Database security is dominated by the attack pattern approach carried out by the intruder. The following paper discusses it will be organized consisting of an introduction as a background to the problem, problem identification, problem formulation. Second, research related to database security studies using data mining analysis with various findings that have been achieved. Third, the research method proposed on the implementation of the decision tree is continued on the discussion side and ends with the conclusion which includes conclusions and recommendations [12].

II. RESEARCH METHOD

Security in the database is needed to maintain originality when there are users internally modifying, deleting and duplicating without permission. Even without permission, the fact that a user has access rights that exceed the permission limit is still a consideration and is taken seriously [13]. Thus, the proposed data mining concept will predict the existence of anomaly data on access rights in the database system [14].

A. Problem Statement

The permission of file system as the opportunity to interact in the system is essentially critical, the cnfiguration of system had been enabling the complexity and protected digital resource but security of digital asset for protected sensitive data was not completed. Moreover, many collecting of users, as the owner of digital resource such as the data in the system were covered with permissions from the administrator regulation to encompas the owner, a few end-user have full control after the type of acces within the configuration of systems have been delivered as a grant. Securely, the ownership for the file system Both grant and full control types have to be inspected to identify the power of access within digital resource [15]. Monitoring activity users had been being concentrating whether considering acces grant within systems were overseen formally or full control of activity from the reporting of log systems have available [16]. Technically, deletion and modifications as well as reading for behaviour in the fle system was underwritten by full control type [17]. For example, actionable a few end-users in the system were provided in finding general file beside the modification or production of collecting data that conducted by identified user will be indicated by write, although the clear instruction with SQL queries have supported toward the activities [18].

Gradualy, we have bound excessive permission in the stabilized file system to recognise several attributes and also definitive class initially raised. Despite the reporting of observation, the dataset had been provided in order to determine a few variables which were involved to complement the requirement of analysis. Because the main cases have been admiting deeply on the file system, the retrieving of data from the big system that contains complex configuration item,it is clear that the one of technique attempt to classify toward excessive permisions. As a result, we have disseminated possibly the five original attributes which covered Type Permissions, Level, Type of User Account, Type of User, Status, and User Actions and these atrributes definitely have complied with the level of acces for the several users who have legal right about permissions. Moreover, the status of attribute in the file system has riveted to the flag of current status (i.e. Active, Blocked, and Closed) while the type of users such as Beginner, Intermediate and Expert were authorized simply. Either two main type of information for this case, private or public data due to the background of the inspection from specific dataset would be applying for the clear cases after we have hightlighted the advantages from comparability of dataset regarding of availability of digital information securely [19].

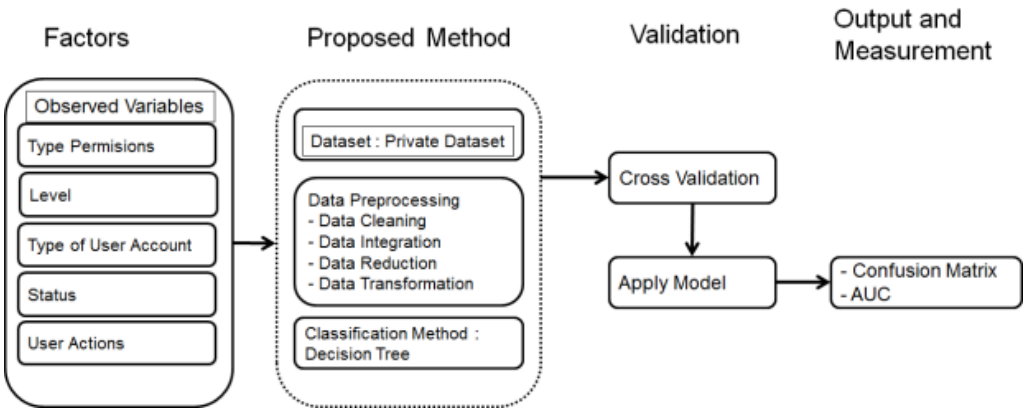


Figure 1. Excessive Permissions Inspection Model framework

According to Figure 1, we present a form of framework related to the investigated case studies. First, the variables observed as factors consist of type permissions, level, type ofuser account, status, and user actions. Second, the method in case studies is permitted on datasets of the private dataset type where data is collected through special institutions and is not publicly available. Next, data preprocessing includes data cleaning, data integration, data reduction, and data transformation [20]. then, the implementation of the model on the decision tree method after the validation process. Finally, the use of confusion matrix and AUC to ensure the model has good performance.

B. Data Preprocessing

The data that has been collected must be managed further which is entered in the data preprocessing stage. Data preprocessing will be time-consuming because it must ensure the data is valid and is almost around 50 to 80%. In addition, there are three most important reasons in data preprocessing, namely, accuracy defines whether the data is true or false, accurate or not. True or false lies in the format of the data or the value of the data itself. The second reason is completeness which will check whether the data in the dataset is recorded even if the data is not available [21]. Third, the consistency that describes data descriptions that can be modified at certain values but while others are not changed.

The power of data mining is quality data and has met the criteria. to meet the requirements, the data preprocessing stage is carried out starting with processing the raw data (from the results of data collection) into data ready to be analyzed. There are 1882 total datasets available with 6 attributes including 1 class [22]. First, the data cleaning stage is carried out to clean data or reduce data that is in abnormal form. In general, datasets have characteristics that are not following the format (Incorrect data) instrument faulty, human or computer error. Among the data cleaning that was overcome were incomplete data describing weak attribute values, lacking certain attributes that were considered interesting, and data collections. examples of incomplete data such as the status attribute containing blank (missing data) as much as 3% of the total dataset. The fix is to fill in the blanks or ignore or even delete data. Found in the dataset there is noise on certain attributes in the form of errors or outliers of almost 2%. The next cleaning data is inconsistent for example, if the status is blocked then the access should be automatically denied (Disallowed) but some are not. In addition, the discrepancy of attribute data (Discrepancy) creates ambiguity around below 2% [23].

The next step in data preprocessing is data reduction where you get data that has been reduced to a smaller size or amount of data. Although it can be reduced, the results still have the same analytical results [24]. The consideration in data reduction is to remember that data is stored in terabytes. Complicated data is also an additional data reduction criterion that results in a long time when executing a complete dataset [25]. The method is

used when there are two data reductions, namely dimensionality reduction and number reduction. The dimensionality reduction method refers to feature selection with a filter approach. The filter approach is carried out by selecting features on the attributes in the dataset, namely status, user action and so on in a subset of feature selection, then applied to the learning algorithm and the final stage is measured on the performance side.

In this case, data transformation is carried out by modifying the dataset into another form with an inverse function or changing the position of attributes in the dataset. The normalization technique can be done by changing the scale to a smaller and more specific range. As a support, the z-score normalization technique is applied. Data transformation is still not sufficient if it has not been discretized by reviewing the types of attributes that exist in the dataset. There are 3 types of attribute types including nominal, ordinal, and numeric. Almost 90% of datasets are private with nominal types. The dataset is divided into several groups using the Binning method for the partition into equal-frequency process.

The refinement stage is by integrating data as part of the data preprocessing stage. Data integration describes combining data from various sources into one form so that it becomes coherent. The main concern with data integration is the integration scheme when combining metadata from multiple sources as well as detecting and resolving data value conflicts. datasets sometimes have ambiguous values and data are indicated as double s must be combined.

C. Decision Tree

One of the data mining techniques that can classify is a decision tree. The data mining technique can classify objects in the form of datasets . The decision tree diagram becomes the output of data mining technique analysis. In the diagram, some symbols and segments are equally connected. The current that occurs in the decision tree is from the top (root) to the bottom (leaf). In the decision tree, algorithms such as ID3, C4.5 and so on will be used. In this case, ID3 is used to calculate entropy and Information Gain for attribute selection to be a node. The calculation of the value of entropy and information gain can be seen in the equation below.

$$\boldsymbol{Entrophy(S) = \sum_{i=i}^{n} - p_i * \log_2 pi} \quad (1)$$

$$Gain(S, A) = Entrophy(S) \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entrophy(S_i) \quad (2)$$

The steps of the Decision Tree method are as presented in Figure 2 below. First of all, prepare the dataset as the object of the specified case study analysis. The dataset in this case contains log data stored in the database system. Next, the calculation and determination of entropy are adjusted according to the equation. Next, define and create a branch based on the gain value.

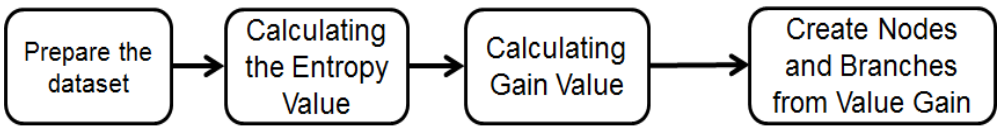


Figure 2. The working stages of the decision tree model

D. Validation and Performance Measurement

Validation is used to ensure data accuracy, both training data and testing data. The technique used is cross-validation through 10 iteration steps. Because it is related to classification, a classification method with a Decision Tree will be applied to ID3 and the output is measured by the confusion matrix and AUC.

The Decision Tree model in this case must be tested for the quality of the model about the dataset. One of them is Accuracy, where the excessive permissions dataset will analyze the level of closeness between the predicted value and the actual value in the dataset. While Precision is used to measure the accuracy between the attributes in the dataset and the response generated from the system, namely the Rapid Miner tool. Then use the F-measure, Recall and so on. Due to using the Confusion Matrix when measuring performance, there are True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$F.1 = 2 . \frac{Precision .Recall}{Precision+Recall} \quad (6)$$

III. RESULTS AND ANALYSIS

A. Data Preprocessing

The power of data mining is quality data and has met the criteria. to meet the requirements, the data preprocessing stage is carried out starting with processing the raw data (from the results of data collection) into data ready to be analyzed. There are 1882 total datasets available with 6 attributes including 1 class. First, the data cleaning stage is carried out to clean data or reduce data that is in abnormal form. In general, datasets have characteristics that are not following the format (Incorrect data) instrument faulty, human or computer error. Among the data cleaning that was overcome were incomplete data describing weak attribute values, lacking certain attributes that were considered interesting, and data collections. examples of incomplete data such as the status attribute containing blank (missing data) as much as 3% of the total dataset. The fix is to fill in the blanks or ignore or even delete data. Found in the dataset there is noise on certain attributes in the form of errors or outliers of almost 2%. The next cleaning data is inconsistent for example, if the status is blocked then the access should be automatically denied (Disallowed) but some are not. In addition, the discrepancy of attribute data (Discrepancy) creates ambiguity around below 2%.

The next step in data preprocessing is data reduction where you get data that has been reduced to a smaller size or amount of data. Although it can be reduced, the results still have the same analytical results. The consideration in data reduction is to remember that data is stored in terabytes. Complicated data is also an additional data reduction criterion that results in a long time when executing a complete dataset. The method is used when there are two data reductions, namely dimensionality reduction and number reduction. The dimensionality reduction method refers to feature selection with a filter approach. The filter approach is carried out by selecting features on the attributes in the dataset, namely status,

user action and so on in a subset of feature selection, then applied to the learning algorithm and the final stage is measured on the performance side.

In this case, data transformation is carried out by modifying the dataset into another form with an inverse function or changing the position of attributes in the dataset. The normalization technique can be done by changing the scale to a smaller and more specific range. As a support, the z-score normalization technique is applied. Data transformation is still not sufficient if it has not been discretized by reviewing the types of attributes that exist in the dataset. There are 3 types of attribute types including nominal, ordinal, and numeric. Almost 90% of datasets are private with nominal types. The dataset is divided into several groups using the Binning method for the partition into equal-frequency process.

The refinement stage is by integrating data as part of the data preprocessing stage. Data integration describes combining data from various sources into one form so that it becomes coherent. The main concern with data integration is the integration scheme when combining metadata from multiple sources as well as detecting and resolving data value conflicts. datasets sometimes have ambiguous values and data are indicated as double so they must be combined.

B. Model Validations

The strength of the data is when used as an analysis can be considered valid. Data that is evaluated to meet the validity, validation method also determines. One way how to test a dataset containing user logs along with information regarding application access permissions refers to the cross-validation method. In general, some researchers do not mention which method is used when discussing data mining. Although validation is considered not mandatory, validation tests can provide certainty to the results of data analysis, especially about certain cases. For example, information security is related to data while security generally deals with physical and software and other scopes. Therefore, database security is not only related to how to protect the data collection stored in the database management system but focuses on accessing the contents of the data. Why is this being ignored, there is very little attention to access which contains an anomaly in the form of excessive permissions.

Several methods can prove how temporary dataset testing is applied to analysis in a data mining approach. The use of the cross-validation technique is a reference for testing the dataset due to its reliability when carried out by researchers or academics. The rest is also applied to the industrial sector and becomes a recommendation for standard units of data measurement. In the cross-validation method, the user dataset is divided into 10 iterations with the results presented in the table below.

Table 1. 10 Fold cross validation test results.

Testing #	Accuracy
1	80,36%
2	87,18%
3	83,91%
4	86,12%
5	77,17%
6	93,63%
7	87,77%
8	78,25%
9	75,22%
10	83,87%

The table above clearly shows the changes of each validation step, especially in the results of the percentage accuracy of the test iteration. The largest value in iteration 6 reached 93.63% which gave confidence to the dataset to be reliable in predicting cases on database security. Next, we will analyze the average accuracy of the test results, the case

at hand as the reference point to determine the success of the Decision Tree method in classifying excessive permissions on the dataset. The calculation is 83.35%, it indicates in this study, that the Decision Tree method includes good classification.

C. Performance Model

The proposed database security is dominated by using encryption techniques when there is a potential attack, as one of the proposed intrusion detection systems. Although encryption techniques can reduce risk, Bharwadj proposes a combination with the concept of data mining. The techniques used are classification and clustering so that the results are better than before. The results of the model performance show the ability of the Decision tree method to have a better accuracy value above the standard so that it meets very well. The same thing with the AUC value is also considered quite good even though the indicators are almost close to superior values such as Recall, F Measure and so on. Thus, the central point is in the dataset when training and testing is carried out. Based on the validation results that have been presented, it is considered significant.

Table 2. Decision Tree Classification Model Performance.

Critera	Value	Standard Deviation
Accuracy	82,7%	2,1%
Classification Error	17,3%	2,1%
AUC	75,1%	5,2%
Precision	82,7%	2,1%
Recall	100,0 %	0,0%
F Measure	90,5%	1,2 %
Sensivity	100,0%	0,0%
Specifity	0,0 %	0,0%

Classification using a decision tree is located at the root where the status attribute becomes the top part by lowering other attributes such as type-user_account, type-permissions and type_permissions. The status attribute value has three criteria such as activated, blocked, and closed. Users who are declared blocked in any type of user will not be given access to any of the database features.

According to Figure 3, the results of the classification model using the decision tree method. Most importantly, the determination of the classification is on the status attribute of the attached account when connected to the database system. There are two statuses, namely active and inactive, that the user is given the opportunity to access if he has an indication of the type of user and the type of user account with a low level.

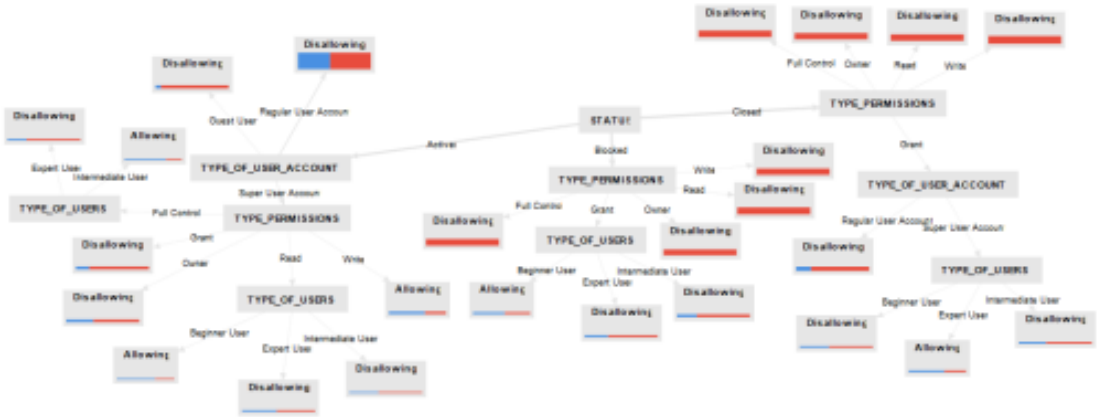
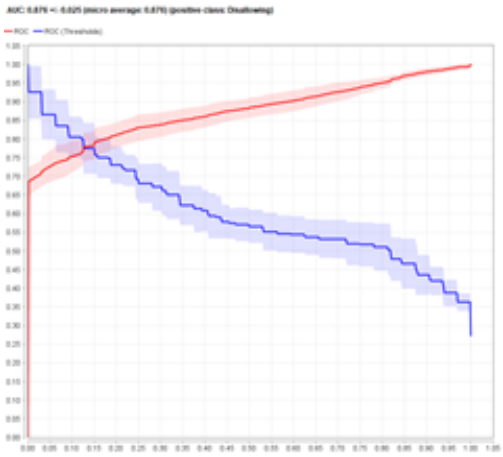


Figure 3. Decision Tree Map Model

The reason for blocking is done by the super admin when the activity log occurs it is considered an anomaly so that it is blocked and access is denied. The following is the result of the model generated by the decision tree method in classifying user datasets related to excessive permissions.

The decision tree measurement diagram and evaluation with a confusion matrix are presented in Figure 4. There are two pictures, the first is the AUC diagram (on the left) which describes the results of the measurement model in the decision tree. Second, the diagram that describes the evaluation of the model presents information on the accuracy of the model with a confusion matrix (on the right).



accuracy: 82.35% +/- 1.94% (micro average: 82.35%)

	true Allowing	true Disallowing	class precision
pred. Allowing	111	120	48.05%
pred. Disallowing	212	1438	87.15%
class recall	34.37%	92.30%	

Figure 4. AUC Diagram of Decision Tree Measurement and Confusion Matrix Evaluations

The findings resulted in a good accuracy value indicated by a value of 82.35%. In addition, the AUC curve shows a value of 0.876 with a good category. The performance of the decision tree method is evidence of its effectiveness in classifying datasets related to excessive permissions. The predicted class is user_action with two attribute values, namely Allowing and Disallowing.

The classification method applied in the study with the dataset of excessive permissions is the identification which has not yet reached how to predict it. The rest of these cases are still being studied and further experiments carried out. This condition is especially in the prediction of the danger of excessive permissions in database security with the concept of data mining. In addition to data mining with classification techniques, there are also associations and clustering and even estimation. Although it can only classify based on attributes in the dataset, the decision tree can describe from root to leaf so that it can be explored more deeply to explore data, especially database security.

Data mining classification performance can be measured again by statistical testing. Analysis of factors that are strongly suspected of excessive permissions must be explained explicitly. To find out and identify the depth of valid data, it can also be measured using the concept of fuzzy logic. Fuzzy logic is applied at the data preprocessing stage even in the type of classification analysis, although fuzzy logic is more suitable with the clustering method.

REFERENCES

[1] W. Chang and J. Wu, *Fog / Edge Computing For Security , Privacy ,*. 2021.

[2] A. A. Z. T. A. F. X. Yi, *SCADA Security Machine Learning Concepts for Intrusion Detection and Prevention*. 2021.

[3] C. Kanth, Y. Vengalarao, S. . Kumar, P. . Chaitanya, and S. Himaja, "A Study On Database Security Issues And Overcome Techniques," *JAC A J. Compos. Theory*, vol. XIV, no. Xi, pp. 1–6, 2021.

[4] A. I. Technology and A. Govardhan, "A FULLY DISTRIBUTED SECURE APPROACH USING NONDETERMINISTIC ENCRYPTION FOR DATABASE SECURITY IN CLOUD," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 7, pp. 2218–2228, 2022.

[5] K. Jamsa, *Introduction to Data Mining and Analytics-Jones & Bartlett Learning LLC (2021)*. 2021.

[6] M. R. Anwar, R. Panjaitan, and R. Supriati, "Implementation Of Database Auditing By Synchronization DBMS," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2 SE-Articles, pp. 197–205, 2021,

[7] D. Binu and B. Rajakumar, *Artificial Intelligence in Data Mining*. London: Oxford University, 2021.

[8] A. J. Moreno-Guerrero, G. Gómez-García, J. López-Belmonte, and C. Rodríguez-Jiménez, "Internet addiction in the web of science database: A review of the literature with scientific mapping," *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, pp. 1–16, 2020, doi: 10.3390/ijerph17082753.

[9] C. M. J. Ryngaert and N. A. N. M. Van Eijk, "International cooperation by (European) security and intelligence services: Reviewing the creation of a joint database in light of data protection guarantees," *Int. Data Priv. Law*, vol. 9, no. 1, pp. 61–73, 2019, doi: 10.1093/idpl/ipz001.

[10] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[11] S. Baloch and M. S. Muhammad, "An Intelligent Data Mining-Based Fault Detection and Classification Strategy for Microgrid," *IEEE Access*, vol. 9, no. 1, pp. 22470–22479, 2021, doi: 10.1109/ACCESS.2021.3056534.

[12] S. Sahoo, A. Subudhi, M. Dash, and S. Sabut, "Automatic Classification of Cardiac Arrhythmias Based on Hybrid Features and Decision Tree Algorithm," *Int. J. Autom. Comput.*, vol. 17, no. 4, pp. 551–561, 2020, doi: 10.1007/s11633-019-1219-2.

[13] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," *Procedia Manuf.*, vol. 35, pp. 698–703, 2019, doi: 10.1016/j.promfg.2019.06.011.

[14] H. Li and S. Li, *Two-Echelon Location Routing Problem with Multi-fuzzy and Pick-Delivery Model and Algorithm*, vol. 928. 2020.

[15] D. Iordache, "Database – Web Interface Vulnerabilities," *Strateg. XXI - Secur. Def. Fac.*, vol. 17, no. 1, pp. 279–287, 2021, doi: 10.53477/2668-2001-21-35.

[16] G. Zhou, *Data Mining for Co-Location Patterns*, First Edit. Boca Raton: Taylor & Francis Group, LLC, 2022.

[17] H. Sulistiani and A. A. Aldino, "Decision Tree C4.5 Algorithm for Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia)," *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 40–50, 2020, doi: 10.21107/edutic.v7i1.8849.

[18] A. Husejinović, "Credit card fraud detection using naive Bayesian and c4.5 decision tree classifiers," *Period. Eng. Nat. Sci.*, vol. 8, no. 1, pp. 1–5, 2020.

[19] A. Arabameri *et al.*, "Novel credal decision tree-based ensemble approaches for

predicting the landslide susceptibility," *Remote Sens.*, vol. 12, no. 20, pp. 1–27, 2020, doi: 10.3390/rs12203389.

[20] J. Liang, Z. Qin, S. Xiao, L. Ou, and X. Lin, "Efficient and Secure Decision Tree Classification for Cloud-Assisted Online Diagnosis Services," *IEEE Trans. Dependable Secur. Comput.*, vol. 18, no. 4, pp. 1632–1644, 2021, doi: 10.1109/TDSC.2019.2922958.

[21] M. K. Anam, B. N. Pikir, and M. B. Firdaus, "Penerapan Na¨ive Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemeritah," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 139–150, 2021, doi: 10.30812/matrik.v21i1.1092.

[22] N. Yuvaraj *et al.*, "Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification," *Comput. Electr. Eng.*, vol. 92, no. June 2020, p. 107186, 2021, doi: 10.1016/j.compeleceng.2021.107186.

[23] K. Gajowniczek and T. Ząbkowski, "Interactive decision tree learning and decision rule extraction based on the imbtreeentropy and imbtreeauc packages," *Processes*, vol. 9, no. 7, pp. 1–16, 2021, doi: 10.3390/pr9071107.

[24] S. Xue, "'Face Database Security Information Verification Based on Recognition Technology,' IJ Netw," *Int. J. Netw. Secur.*, vol. 21, no. 4, p. 4, 2019, doi: 10.6633/IJNS.201907.

[25] C. K. Wee and R. Nayak, "A novel machine learning approach for database exploitation detection and privilege control," *J. Inf. Telecommun.*, vol. 3, no. 3, pp. 308–325, 2019, doi: 10.1080/24751839.2019.1570454