ISSN 2278-3091
Volume 9, No.3, May - June 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse88932020.pdf
https://doi.org/10.30534/ijatcse/2020/88932020

# The Efficiency of Compliance to Data Privacy Act of 2012 using the Web-based Self-Survey Applying Quality Analysis Technique

**Paraluman Maria Fatima B. Pesito[1], Maria Visitacion N. Gumabay[2], Irma T. Plata[3]**
[1]University of Northern Philippines, Vigan, Ilocos Sur, Philippines
benzonparaluman@gmail.com
[2]Saint Paul University Philippines, Tuguegarao City, Philippines
gumabayvc@gmail.com
[3]IsabelaState University, Echague, Isabela, Philippines
irma.t.plata@isu.edu.ph

**ABSTRACT**

The study focuses on determining the level of efficiency of compliance of the Ilocos Sur provincial government office according to standards of the Data Privacy Act 2012. The survey questionnaire is tailored from the National Data Privacy Commission. There were eight key items like Transparency of Processing, Collection, Technical Security, Organizational Security, Proportionality, Physical Security, Security of Personal Data, and Legitimate Purpose used to create a pattern that would help predict a “High” threat. The determination of the level of "Threat", whether high, medium, and low for the processing of personal data is using the Association rule algorithm and Logistic regression.

Computation of precision and recall is performed to validate the pattern identified. It computes the probability of an event occurrence by employing a data mining technique called the quality analysis technique. Also, all answers/responses are stored in a database for pre-processing and processing using data mining algorithms. The results of the processed data are displayed graphically. The Weka 4.5 data mining tool utilization achieves the results specifically for results metrics. 1,010 participants were involved in the survey using the developed web-based self-survey system, providing an online survey, manages survey results, automatic computations, analysis, and graphical interpretation. The employment of the association rule identifies the relationship between the eight key areas. Results revealed that 'technical security' has an association with a 'low' threat. In contrast, organization security, transparency of processing, technical security, and collection have an association with 'medium' threat, and organization security, transparency of processing, collection, and technical security have a 'high' threat. The logistic regression algorithm results reveal that 'high' level threat is recorded at the Provincial Budget Office and Provincial Human Resource Management office. At the same time, Ilocos Sur Community College has a 'medium' level of threat. Lastly, in terms of the level of efficiency, precision and recall have high scores, which implies positive results (high precision and high recall). Concerning recall, a result of 0.856 was established, while precision resulted in 0.998 and 0.85.

**Key words:** Data privacy, Data mining, Quality Analysis Technique, Web-based system, Weka

## 1. INTRODUCTION

### 1.1 Data Protection Law in the Philippines

The Philippine data protection law imposes a series of requirements and compliance designed to protect individuals against the risks that result from the processing of personal data. The Philippines was rank as the 33rd out of 233 in Kaspersky's list of countries most prone to a data breach [1]. Breaching of personal data and used illegally, it became more disastrous. The data breach, according to security firm Trend Micro in 2017, made Filipino people susceptible to fraud and other risks. With such an issue, which is quite alarming, the National Privacy Commission (NPC) conducts a campaign on personal data privacy. Also, visits to each sector in the government and conducts seminars and training specifically for Data Privacy to be compliant to the Data Privacy Law or the R.A. 10173 [2].

The creation of an independent body called the National Privacy Commission (NPC) as per provision of the RA 10173 to administer, implement, monitor, and ensure compliance in consonance with international standards set for data protection [2]. The rampant use of personal data in social media, access devices, mobile Apps, and delivery of vital services requires the public and community to handle and be aware of how to manage personal information responsibly. This premise was stipulated in the Implementing Rules and Regulations (IRR) particularly, the processing of data by any natural and juridical person in the government or private sector. Due to the volume of population, data acquired, and continuing to acquire in government agencies, it is mandatory to comply with the law [3]. In terms of the data privacy status in the Philippines, the implementation of the Data Privacy Act (DPA) of 2012 has an impact in the business industries, in government agencies, large corporations, and other global companies' affiliation that operate within the boundary.

The Philippines ensures its position as a leading IT/BPO outsourcing destination. The IRRs have significant impacts, particularly the I.T. and business process outsourcing industry. Hence, the establishment of the Act is bringing the country as part of international data protection standards. The governing body of the DPA currently has assured different organizations that its priority focuses on educating, guiding, and encouraging these organizations. Furthermore, a possible rise in privacy violations and complaints may be attributed to organizations that are hesitant about the Act's implementation, monitoring, and penal powers. Today, in the Southeast Asian region, out of the ten member states, only three countries, including the Philippines, have successfully implemented comprehensive data protection law. The remaining seven, two, have on-going efforts to draft and pass legislation, three have some relevant policies on privacy and data protection. At the same time, and no data was available for the rest as of this writing. [4]

### 1.2 Data Privacy

Privacy concerns arise wherever personal information is collected, stored, or used. An individual/group's ability to withdraw information about oneself and thereby reveals selectively is called privacy [5]. However, Westin [6] emphasizes that data protection is a complicated matter that has usually been associated with the concept of privacy within the context of personal data processing. Adherence to fundamental principles of recognizing transparency, proportionality, and legitimate purpose in dealing with personal information. Other literature includes the principle of consent as the basis for collecting, fair and lawful, and ensures accurate and quality processing of personal information. Also, the use of consent means implementing adequate safeguards in the processing, transmission, and non-retention of personal information [7].Moreover, the secrecy of personal information called information privacy relates to personal data kept electronically using a computer system. Information related to businesses, medical, financials, criminal records, political records, medical records, and financial data, among others, needs maintenance and security on the privacy of information.

### 1.3 Personal Data and Sensitive Data according to the Data Privacy Act (DPA)

The protection of an individual's personal information is the aim of the Philippine Data Privacy Act of 2012 implementation. Personal information is the recorded identity (*whether recorded in a material form or not*) of an individual, which reasonably and directly determined by the individual holding the information. Also, when put together with other information would positively and directly identify the individual [8]. In this case, protection in the processing of personal information and keeping of identity for privacyis a must.

Moreover, careful processing of sensitive information like a person's race, marital status, age, color, ethnic origin, and affiliations- like political, religious, and philosophical. Equally, consent from the data subject is needed. Anything specific that is given freely, with an informed indication of will whereby the data subject agrees to the collection and processing, this means you are providing the consent of the data subject. Consent may be written, electronic, or recorded means [8].

### 1.4 Data Mining Techniques

Discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems are the processes involved in data mining [9]. Extraction of information from a data set and transforming it into a logical structure is the goal of the data mining process, which can be used for decision making. The data mining algorithm is useful in most applications like Web usage, decision support, mining, and bioinformatics, among others. Item mining is the most problematic phase in data mining. For instance, in a given database, each transaction consists of a set of items, Frequent Itemset Mining (FIM) tries to find itemsets that occur in transactions multiple times than some given occurrences [10]. In recent years, the information industry boomed due to the vast and massive availability of data and the imminent need for turning such data into useful information and knowledge. This is one primary reason why data mining has attracted a great deal of attention. The information and knowledge gained can be used for applications ranging from business management, production control, and market analysis, to engineering design and science exploration.

Researchers treat data mining as an alternative for "Knowledge Discovery in Databases", or KDD [11], while others view data mining as merely a vital step in the process of knowledge discovery in databases [12]. KDD undergoes a mandatory activity called the pre-processing step. The data mining process involves data cleaning, which is the removal of noise or irrelevant data. Next is the data integration, in which multiple data sources may be combined. The data selection, where data relevant to the analysis task, are retrieved from the database is the third step. Data transformation is the last step where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations13], [14].

In the digital era, knowledge discovery is essential to utilize data available on the internet. A vast amount of discoveries was arising due to the availability of data. These were processed using different techniques known since the 1960s. The regression analysis, classification techniques, and association techniques are just some of the techniques used in data mining. The data mining technique is useful in identifying personal preferences and predictions of individuals using the internet. To recognize patterns, the association rule mining, a straight forward technique where experts can make a simple correlation between two or more items often of the same type.

Furthermore, the research community uses regression analysis in identifying and analyzing the relationship among variables. The use of regression analysis is for prediction and

forecasting. It can help in understand the characteristic value of the dependent variable changes if any one of the independent variables is varied. It only means one variable is dependent on another, but it is not vice versa. Construction of association or relation-based data mining technique can be achieved simply with different tools. Tools such as Rapidminer, Weka, and others produced relationship-based results [15].

### 1.5 Quality Analysis Technique in Data Mining

The quality analysis technique in data mining includes: checking counts, validation checks, distribution analysis, random sampling of rich subsets, and random sampling [16].The purpose of Checking Counts is to identify the metadata values for existence and accuracy, and the need to perform checking in the record counts can determine the data accurateness. The validation technique assesses how well the data mining technique performed. It is essential to validate mining models by understanding their quality and characteristics before deploying them into a production environment. For some algorithms, it is useful to define the distribution of any continuous columns before processing the model, if the columns are known to contain standard distributions of values. The undefined distributions can result in less accurate mining models. Cross-validation is also essential in the checking of the analysis applied to compare the performances of the modeling procedures. Distribution analysis checks metadata value distributions and looks for oddities. The Random rich subsets the results manually check that extraction was performed correctly. Lastly, the Random sampling manually checks each record from the sample for correct download and extraction from the sample records.

### 1.6Monitoring System Compliance

Monitoring data privacy in the industry has been tedious work for a different organization. Organizations across industries are looking for solutions to develop programs and software that can monitor data privacy compliance. One key measure is to investigate staff to ensure compliance with policies and procedures, contractual agreements, and laws and regulations [17]. Outsourcing services for monitoring are one of the options to provide adequate monitoring in organizations. It is possible to acquire people-support or third-party support since many organizations such as Smart-PLDT, Follosco, Morallos, and Herce Consultancy and other Law firms offering services for DPA. More, the use of document monitoring tools to monitor compliance for International Standard Organizations, DPA, and other regulatory and law.

### 1.7 Goals of the Study

The study focuses on determining the level of efficiency of the compliance of the Province of Ilocos Sur according to standards of the Data Privacy Act 2012. The level of precision and recall were determined by employing a data mining technique called the quality analysis technique. Also, the study includes the development of the Web-based self-survey system as a survey tool.

## 2. PROCEDURES AND METHODS

### 2.1 Data Mining Tools, Techniques, and Algorithm

The Quality Analysis Techniques in determining the level of efficiency in terms of compliance with the Data Privacy Act were applied using the developed Web-based Self-Survey System. The determination of the level of "Threat" for the processing of personal data is using the Association rule algorithm and Logistic regression.
In the association rule algorithm, the patterns are obtained once the definition of the dataset and all associations in the sample are available/found. These results are then validated against the entire datasets. Maximizing the effectiveness of the two algorithm approaches, it uses lowered minimum support on the sample. Since the approach is probabilistic (*i.e., dependent on the sample containing all the relevant associations*), not all the rules may be found in this first pass. The identified sample size of the dataset is very accurate, meaning that association rules can be highly efficiently executed on a sample of this size to obtain a sufficiently accurate result. On the other hand, to validate the pattern identified, computation of precision and recall is highly needed since it computes the probability of an event occurrence. Estimation is done through maximum likelihood and 10-fold cross-validation and computation of precision and recall.

The association rule algorithm's primary purpose is to identify the essential items and their association to the level of "Threats", whether it's Low, Medium, or High. Figure 1 presents the key items and their association with the level of "Threats" (Low, Medium, and High). Based on the study conducted, the use of eight key items to create a pattern that would help predict a "High" threat. The key items are Transparency of Processing, Collection, Technical Security, Organizational Security, Proportionality, Physical Security, Security of Personal Data, and Legitimate Purpose. Also, all answers/responses are stored in a database for pre-processing and processing using the data mining algorithms. The results of the processed data are displayed graphically.

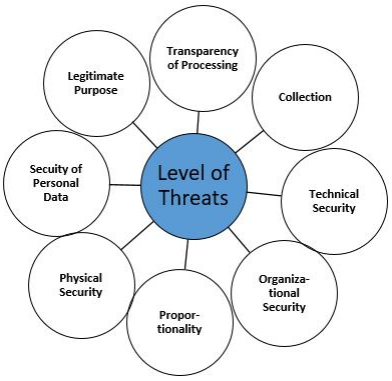


**Figure 1:** Key items and their association with the level of "Threats" (Low, Medium, and High)

In a classification technique, recall is the number of true positives divided by the number of true positives plus the number of false negatives. True positives are data points classified as positive by the positive model (*meaning they are correct*), and false negatives are data points the model identifies as unfavorable that are positive [18]. On the other hand, precision for a class is the number of true positives (i.*e., the number of items correctly labeled as belonging to the positive class*) divided by the total number of elements labeled as belonging to the positive class. Both precision and recall are based on understanding and measure of relevance in data mining techniques. In information retrieval, precision is a measure of result relevancy, while recall is a measure of how many truly relevant results returned. Figure 2 shows the formula for precision and recall [19].

$$Precision = \frac{tp}{tp + fp} \quad Recall = \frac{tp}{tp + fp}$$

**Figure 2:** The precision and recall formula

Figure 3 presents the process flow on how the Weka 4.5 data mining tool utilization achieves the results specifically for results metrics. The conduct of proper cleaning of data complies with the requirements of the data mining tool. The training set of data is determined by analyzing a set of training database instances until a data model was built that describes a predetermined set of classes of concepts. Then, loading to the tool was executed. To achieve the results accurately in the data mining tool for results metrics. In information retrieval, precision was a measure of result relevancy, while recall was a measure of how many truly relevant results were returned.

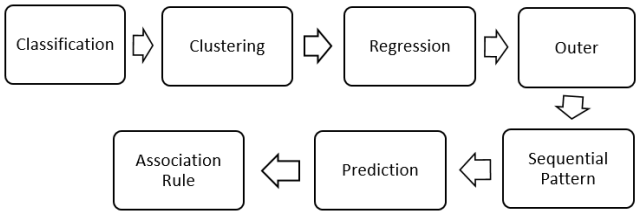


**Figure 3:**Process flow using Weka as the Data mining tool

## 2.2 Software Development Methodology

Figure 4 presents the software development model as a basis in the development of the Web-based Self-Survey System. It involves requirements analysis, design, implementation, and testing performed to build the system successfully [20] [21].

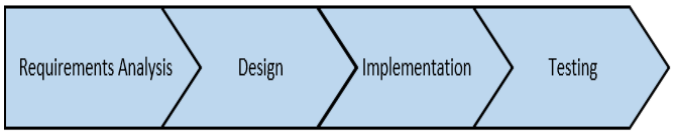


**Figure 4:** The software development model as a basis in the development of the Self-Survey Web-based System

In the Analysis Phase, a comprehensive description of the system is developed. Through interviews, observation, review of documents related to data privacy, and review of related literature and studies, the functional and non-functional requirements based on the users' requirements and needs were identified and analyzed. The users' requirements include the system scope, functions, system attributes, user authorization, and access privileges, specifications, interface requirements, and database requirements.

In the Design Phase, a plan solution to address the user requirements and needs identified was carried out. The design includes an architectural system process, conceptual database schema, logical diagram, data structure definition, and the graphical user interface design.

During the Implementation Phase, the result of the analysis phase and the different designs serve as basis/blueprint in the development of the system. The phase involves the actual code writing and compilation into the operational system and where database and text files are created. The system was developed using the MVC.Net framework and use the following web-based development tools like the Microsoft Visual Studio include the IDE (Integrated Development Environment) that enables to create the web-based system. The bootstrap, a free and open-source front-end web development framework, supports the design and development of the system. The bootstrap supports HTML and CSS-based design templates for typography, forms, buttons, navigation, and other interface components [22]. The XAMPP includes the web server application (Apache), the database (MariaDB), and the scripting language (PHP)[23]. Figure 5 is a sample screenshot of the system.

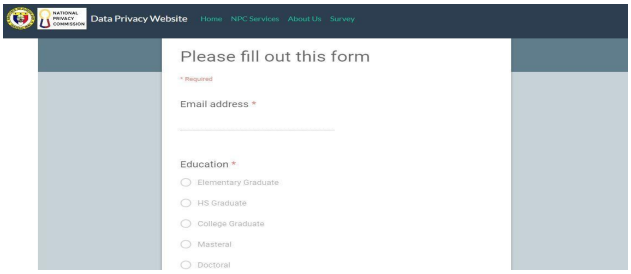


**Figure 5:** The survey form module

During the testing phase, the verification process determines whether the system satisfies the conditions/scope imposed at the analysis phase. The validation process involves the evaluation during and at the end of the development process, determining if it satisfies specified requirements. Also, debugging, in which bugs and system glitches are found, corrected, and refined accordingly, is done in the testing phase.

## 2.3 Process Flow of Applying the Algorithm to determine the efficiency of compliance

Figure 6 is the process flow of applying the algorithm to determine the efficiency of compliance. The first step in

analyzing the results is the selection of attributes needed to be performed based on goals. This step answers what data is needed and is available for consumption. An added difficulty of this step is the integration of data in creating one data set, which included all predictors needed to start the process.Another step to consider in assuring quality results is pre-processing and cleansing of data. Data consistency is enhanced in this stage. It includes data clearing, such as handling missing values and removing outliers. Data Transformation stage is the generation of better data. The data were transformed into a proper format using the data pre-processing technique of a data mining tool.

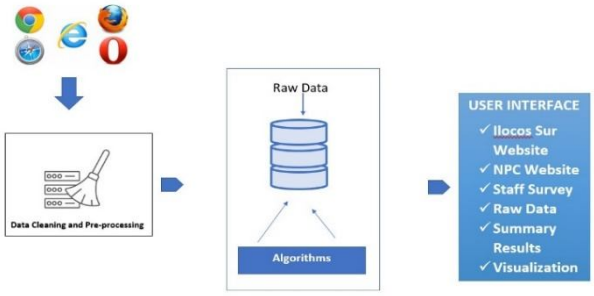


**Figure 6:** Process Flow of Applying the Algorithm to determine the efficiency of compliance

### 2.4 Research Participants

Table 1 shows the participants' frequency and equivalent percentage distribution. The participants were the personnel/workers of the Province of Ilocos Sur with a total of 1,010.

**Table 1**: Distribution of Participants

| Participants | f | Percentage |
|---|---|---|
| Governor's Office | 101 | 10.00 |
| Provincial Administrator | 68 | 6.73 |
| Human Resource Mngt | 17 | 1.68 |
| Vice Governor&Sanggunian | 143 | 14.16 |
| Sanggunian Panlalawigan Secretariat | 27 | 2.67 |
| Provincial Accounting | 19 | 1.88 |
| Provincial Agriculture | 54 | 5.35 |
| Provincial Assessor's Office | 17 | 1.68 |
| Provincial Budget Office | 11 | 1.09 |
| Provincial Engineer's Office | 75 | 7.43 |
| Provincial Legal Office | 11 | 1.09 |
| Provincial Planning and Development Office | 17 | 1.68 |
| Social Welfare & Development | 80 | 7.92 |
| Provincial Population Office | 5 | 0.50 |
| Provincial Treasurer | 27 | 2.67 |
| Provincial Veterinary Office | 54 | 5.35 |
| Envi. Natural Resources Mngt | 39 | 3.86 |
| General Services Office | 155 | 15.35 |
| Ilocos Sur Community College | 80 | 7.92 |
| IT Experts | 10 | 0.99 |
| **TOTAL** | **1,010** | **100%** |

### 2.5 Instrumentation

The survey questionnaire is tailored from the National Data Privacy Commission. Part one of the questionnaire includes demographic profiling in terms of Age, Sex, Civil Status, Educational Attainment, and Length of Service. Part two is the extent of compliance with the Data Privacy Act of 2012 in terms of the eight key items referring to Transparency of Processing, Collection, Technical Security, Organizational Security, Proportionality, Physical Security, Security of Personal Data, and Legitimate Purpose. Moreover, the records produce data sets of 144,000, which is already valid. According to the literature [24], a rule of thumb for data mining is never to have less than 50-100 rows of data for the simplest model's types and scenarios. In the application of the classification algorithm, the data set of a reasonable size is advisable. Focusing on data quality is an essential factor rather than adding more and more data. Above all, the statistically valid patterns have been found, and adding more data does not improve their validity. On the other hand, adding more data, sometimes, you can introduce accidental correlations [25].

A total of 1,002 instances were extracted from the cleaning process. This instance was used to feed in the data mining tool to identify its relationship or association with the identified "Threat Level" such as Low, Medium, and High. These "Threat Level" was also part of the questionnaire, to find out the level in terms of the items or Key Areas defined. The relationship of item "Technical Security" has an association with "Threat" (Low), with 40 instances out of 891 instances of Technical Security.

### 2.6 Data Gathering Procedures

In the conduct of the study, the researchers seek permission from the Office of the provincial governor of Ilocos Sur. Hence, after soliciting the approval and providing a schedule of visits, several interviews, meetings, and observation with the Management Information System (MIS) office and the Human Resource Office were conducted. The goal is to gather pertinent requirements such as the process flow of implementing the DPA as a basis in the development of the Web-based Self-Survey system and identifying the number of personnel from the different offices serving as respondents. Similarly, a consent letter was provided among participants to properly inform them of the study's purpose, results, and effects to the Office. A purposive sampling technique was used in determining specific participants involved in the survey procedure.

## 3. RESULTS AND DISCUSSIONS

### 3.1The Developed Web-based Self-Survey System Functionalities and Features

The functionalities and features of the system describe what it does and the services it provides to the users. Here are some significant functionalities:

- provides dashboard module
- manages Participants Profile
- provides Online Survey which includes questions/survey, survey contents, a user to select the questions to configure the content using the admin account, and the system enables a user to add one or more components to the configuration of the content of the survey
- provides visualization through graphical representation of results and analysis
- manages data collected for reporting, analysis, and interpretation purposes
- authenticates user credentials to view profiles

To describe briefly, the Web-based Self Survey System is composed of modules such as NPC Services, About Us, and the Survey. The National Privacy Commission (NPC) official site displays complete information of the Data privacy law and the services that NPC offers. The website also offers quick links to all the official forms and guidelines of NPC. Figure 7 presents the NPC Services home page, and Figure 8 presents the participants' survey results.

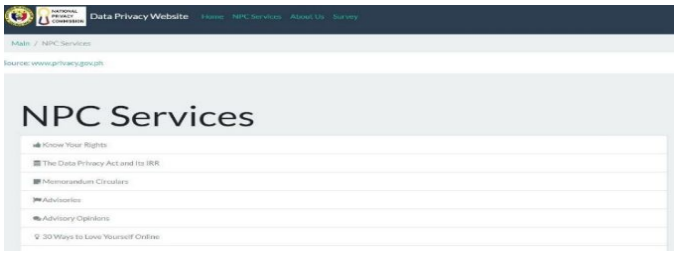


**Figure 7**: presents the Survey module, where Survey results are presented.


**Figure 8:** Participants survey results

Figure 9a and Figure 9b show sample results presented in the Dashboard module. The creation of the Dashboard facilitates the control for data visualizations while the dashboard icon provides users with a summary of the survey for Data Privacy Compliance through Visualization.A graphical presentation for each category uses to understand the results of the survey quickly. The developed system provides an accessible way to see patterns and trends. Each of the categories provides a representation of the results from the questions answered by "Yes" or "No".

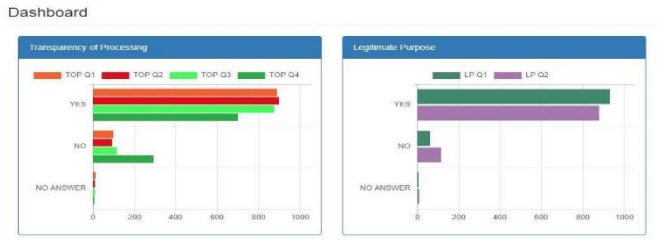


**Figure 9a:** Sample Graphical dashboard module

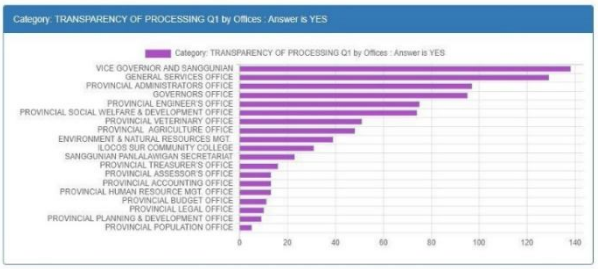


**Figure 9b**: Sample screenshot: transparency of processing by Office who answered "YES"

### 3.2 Applying Algorithms to monitor and determine the efficiency of the compliance to the Data Privacy Act of 2012

The self-survey is a set of text answerable by a "YES" or a "NO". The data mining tool analyzes the text in preparation for the next stage. Regular expressions afforded by Notepad++ serves as editor to remove these tags and other noise data present in the corpus (*e.g., unwanted white spaces*). The data mining techniques, association rule mining, and logistic regression algorithm determine the accuracy levels of each algorithm on datasets.Data Cleaning and Processing. The cleaning process involves removing white spaces in the text model before feeding it to the tool. The said process was executed using Notepad++ through regular expression like \r, (_), \n, and many others aimed to convert the data to a model acceptable by the data mining tool. The first activity involved is data classification. The training sets of data were determined by analyzing a set of training database instances until a data model was built that describes a predetermined set of classes or concepts. Then, the generated model was applied to test the data, which was not part of the training data to determine the classification rate of the model.

### 3.3 Association Rule Mining Algorithm Results

The rule-based approach applies association rule discovery algorithms to find an association between items and then generates item recommendations based on the strength of the association between items. The association rule was employed to identify the relationship of the items or Key Areas identified: (1) transparency of processing; (2) Legitimate purpose; (3) Proportionality; (4) Collection; (5) security of personal data; (6) organizational security; (7) physical security; and (8) technical security.A total of 1,002 instances were extracted from the cleaning process. These

instances were used to feed in the data mining tool to identify its relationship or association with the identified "Threat Level" such as Low, Medium, and High. These "Threat Level" was also part of the questionnaire, to find out the level in terms of the items or Key Areas defined.

### 3.3.1Association to "Threat" (Low)

Figure 10 presents the relationship of the item "Technical Security" has an association with "Threat" (Low), with 40 instances out of 891 instances of Technical Security.

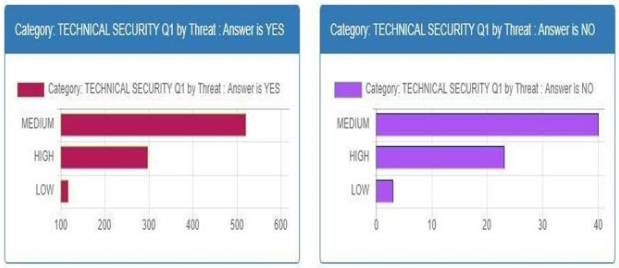


**Figure 10:** "Threat" (Low) of Technical Security

### 3.3.2 Association to "Threat" (Medium)

Based on the results, association with "Threat" Medium reveals organization security, transparency of processing, technical security, and collection. Figure 11 shows a sample.


**Figure 11:"**Threat" (Medium) of Organizational Security

### 3.3.3Association to "Threat" (High)

Figure 12displays the results of the Association Rule Mining that corresponds to Threat "High." This includes organization security, transparency of processing, collection, and technical security. A total of 337 responding with a "High" Threat of items indicated.


**Figure 12:** "Threat" (High) of organizational security

### 3.4 Logistic Regression Algorithm Results

The logistic regression is a robust classification analysis based on literature. It has been used for modeling trends if the target variable is binary depends on multiple regressors. In Figure 13, the different variable was identified with corresponding results based on the level of "Threat" such as "High" and "Medium". The results with the "High" level of threat belong to the variable "*Offices=Provincial Budget Office*". The variable "*Offices=Provincial Human Resource Mgt. Office*", also has a "High" Level of Threat. However, the variable "Offices=Ilocos Sur Community College" has a "Medium" level of threat.

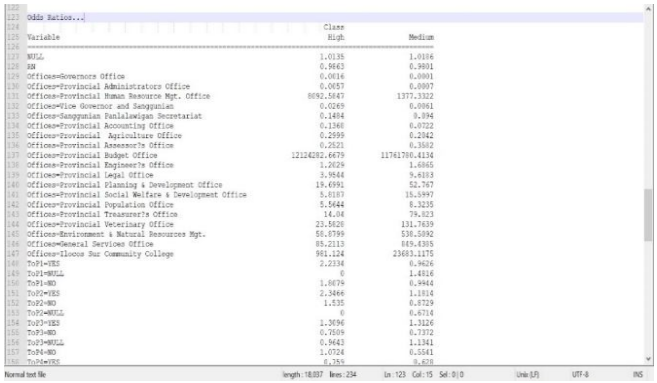


**Figure 13:** Association to "Threat" (High) of Organizational Security, Transparency of Processing, Collection, and Security Technical

### 3.5 Level of efficiency of the proposed Self-Survey System on the compliance of the Data Privacy of 2012 Applying Data Mining Techniques

Figure 14 presents the high scores for both precision and recall, which mean positive results; (high precision) and (high recall). Precision and recall were computed using weka data mining in terms of reliability. Association rule and logistic regression were employed for both precision and recall. In information retrieval, precision refers to the measure of result relevancy, while recall refers to the measure of how many truly relevant results were returned. Concerning recall, a result of 0.856 was established, while precision resulted in 0.998 and 0.85.

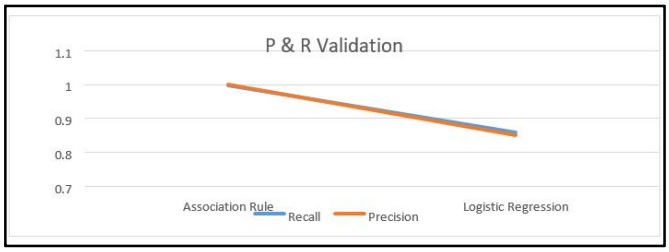


**Figure 14:** Precision and Recall Validation

## 4. CONCLUSION

The developed web-based self-survey system facilitates the conduct of the evaluation determining the level of compliance to data privacy through a visualization where a graphical

representation for each category quickly understands the results of the survey showing patterns and trends. The implementation of data mining techniques, association rule mining, and logistic regression algorithm determine the accuracy levels of each algorithm on datasets. The employment of the association rule identifies the relationship between the eight key areas. Results revealed that 'technical security' has an association with a 'low' threat. In contrast, organization security, transparency of processing, technical security, and collection have an association with 'medium' threat, and organization security, transparency of processing, collection, and technical security have a 'high' threat. The logistic regression algorithm results reveal that 'high' level threat is recorded at the Provincial Budget Office and Provincial Human Resource Management office. At the same time, Ilocos Sur Community College has a 'medium' level of threat. Lastly, in terms of the level of efficiency, precision and recall have high scores, which implies positive results (high precision and high recall). Concerning recall, a result of 0.856 was established, while precision resulted in 0.998 and 0.85.

.

## REFERENCES


[1] Maria Garnaeva, Jornt van der Wiel, Denis Makrushin, Anton Ivanov, YuryNamestnikov, **Kaspersky Security Bulletin 2015. Overall Statistics for 2015**, available :athttps://securelist.com/kaspersky-security-bulletin-2015-overall-statistics-for-2015/73038/

[2] National Privacy Commission, **About NPC Vision and Mission,** available : athttps://www.privacy.gov.ph/about-us/#visionmission

[3] National Privacy Commission, **Implementing Rules and Regulations of the Data Privacy Act of 2012**,available athttps://www.privacy.gov.ph/implementing-rules-regulations-data-privacy-act-2012/

[4] GieDela, **Privacy and Data Protection Laws in Southeast Asia,**May 15, 2018, available athttp://ateneo.edu/udpo/article/Privacy-and-data-protection-laws-southeast-asia

[5] J. Krumm, "**A survey of computational location privacy**", Personal and Ubiquitous Computing, 13(6):291–399, 2009 https://doi.org/10.1007/s00779-008-0212-5

[6] A. F. Westin, Privacy and Freedom. New York: Atheneum for the Assoc. of the Bar of the City of New York. 1967

[7] Warren Chick, **The Singapore Personal Data Protection Act and an Assessment of Future Trends in Data Privacy**, Singapore Management University, 2013 https://doi.org/10.1016/j.clsr.2013.07.010

[8] Republic Act. No. 10173, Ch. 1, Sec. 3, **An Act Protecting Individual Personal Information in Information and Communications Systems in the Government and the Private Sector**, available athttps://www.officialgazette.gov.ph/2012/08/15/republic-act-no-10173/

[9] K. Sumiran, **An Overview of Data Mining Techniques and Their Application in Industrial Engineering**, Department of Computer Science and Engineering, Indian Institute of Technology, Roorkee. Roorkee- Haridwar Highway, Roorkee, Uttarakhand, 247667, 2018.

[10] S. Bhise and S. Kale, **Efficient Algorithms to find Frequent Itemset Using Data Mining**, Department of Information Technology, RMD Sinhgad School of Engineering,Warje, Pune, MH, India , 2017

[11] J. Asenjo, **Data Masking, Encryption, and their Effect on Classification Performance: Trade-offs Between Data Security and Utility**, Nova Southeastern University, 2017, available at https://nsuworks.nova.edu/gscis_etd/1010

[12] Vijay Kotu and BalaDeshpande, **Data mining for the internet of things: Literature review and challenges**, TIBCO Software Inc. All Rights Reserved, 2019, available at http://www.statsoft.com/textbook/data-miningtechniques

[13] Jiawei Han, **Data Mining Concepts and Techniques**, University of Illinois at Urbana–Champaign, 2016

[14] M. Rogalewicz and R. Sika, **Methodologies of Knowledge Discovery From Data And Data Mining Methods In Mechanical Engineering.**, Poznan University of Technology Chair of Management and Production Engineering Piotrowo 3, 61-138 Poznań, Poland, 2015

[15] TIBCO ®Data Virtualization Reference GuideCopyright © 2004-2019 TIBCO Software Inc.

[16] Paul Nelson, **Quality Analysis in Data Mining Projects Cruising the Data Ocean**, Blog Series - Part 6 of 6, 2017

[17] M. Rodriguez, and M. Piattini. **Software product quality evaluation using ISO/IEC 25000**, ERCIM NEWS (October): 39-40, 2014

[18] W. Koehrsen, W. **Beyond accuracy: Precision and recall**, 2018, available at https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c

[19] D. Olson and D. Dursun,**Advanced Data Mining Techniques,** Springer, 1st edition, page 138,ISBN3-540-76916-1, 2008

[20] Irma T. Plata and John M, Facun, **Development and Implementation of Web-based Paperless Student Evaluation for Teachers (PSET)**, International Journal of Advanced Trends in Computer Science and Engineering, Vol.9 No. 1, 2020 available at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse28912020.pdf https://doi.org/10.30534/ijatcse/2020/28912020

[21] Irma T. Plata, **Development and Implementation of Web-Based Pupils' FORM 137-E Information System to Primary Elementary Schools**. International Journal of Scientific & Technology Research, Vol. 9, Issue 3, Page. 5177, ISSN 2277-8616, 2020. Available at http://www.ijstr.org/final-print/mar2020/Development-And-Implementation-Of-Web-based-Pupils-Form-1

37-e-Information-System-To-Primary-Elementary-Schools.pdf

[22] Mark Otto, "**Bootstrap 4.1.2 released**"."Search · stars:>1". GitHub, 2018.

[23] Mark Otto, "**Bootstrap in A List Apart No. 342**". Mark Otto's blog. Archived from the original on October 28, 2016. Retrieved February 23, 2018

[24] S. Masehwari, **Emerging trends in expert applications and security**, Proceedings of ICETEAS 2018, 2019.

[25] D. Kabakchieva, **Predicting Student Performance by using Data Mining methods for classification**, Cybernetics and Information Technologies, 2013 https://doi.org/10.2478/cait-2013-0006