



KCA UNIVERSITY

DATA-CENTRIC SECURITY AND GOVERNANCE FOR DATA

BY

GETRUDE MUKAMI KIMANI

REG: 14/02525

**A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE AWARD OF MASTERS OF SCIENCE IN DATA
COMMUNICATIONS AND NETWORKING IN THE FACULTY OF COMPUTING AND
INFORMATION MANAGEMENT**

NOVEMBER 2016

DECLARATION

I declare that this Research project is my original work and has not been previously published or submitted elsewhere for award of a degree. I also declare that this Research project contains no material written or published by other people except where due reference is made and author duly acknowledged.

Student Name: Getrude Mukami Kimani

Reg No. 14/025252

Sign: _____

Date: _____

I do hereby confirm that I have examined the master's Research project of

Getrude Mukami Kimani

AND have certified that all revisions that the Research project panel and examiners recommended have been adequately addressed.

Sign: _____ **Date:** _____

Dr. Henry Mwangi

Dissertation Supervisor

ABSTRACT

The information security has never been as important as it is today for the business, Health, Banking and education organizations and individuals alike, because Many organizations around the world depend today on reliable information to perform their Daily tasks. So, it is Possible for organizations and individuals to guard themselves by being skilled on the Importance of security and gaining awareness of the possible Security attacks that they may encounter. Basically, if you cannot answer the question where is my sensitive data, you need to first work on a data classification effort for your enterprise.

This thesis represents and describes a new approach used to build a security tool by help of classification tool, that can be used for classifying of data public private confidential and sensitive data depending on the provided attributes define hence reducing intentional data leakage actions through data classification as the first step in ensuring data security by clustering of data to be able to identify the most critical and high risk data to be protected. The main goal of this thesis is to put the basis for a new security application can be used within many organizations to protect and prevent sensitive data from leakage to the wrong hands intentionally or accidentally. DLP solutions detect and prevent unauthorized attempts to copy or send sensitive data, both Intentionally or/and unintentionally, without authorization, by people who are authorized to access the sensitive information.DLP is designed to detect potential data breach incidents in timely manner and this happens by monitoring data.

TABLE OF CONTENTS

DECLARATION	i
ABSTRACT.....	ii
ACKNOWLEDGMENT	vi
TABLE OF FIGURE.....	vii
LIST OF TABLES.....	vii
CHAPTER 1: INTRODUCTION.....	1
1.1Background Information.....	1
1.2Problem Statement.....	4
1.3Objectives.....	6
1.4Justification of the study	6
1.5Significant of the study	7
CHAPTER 2: LITERATURE REVIEW.....	8
2.1Data classification	8
2.1.0Data Classification definition.....	9
2.1.1Definitions of Data	9
2.1.1Classifications for Sensitive Data.....	17
2.1.2Government Classification.....	17
2.1.3Government data classifications	18
2.1.4Commercial Classification	19
2.2.0Granularity of classification.....	20
2.2.1Creating Data Classification Procedures	21
2.2.2Defining Roles and Responsibilities for effective data classification.....	21
2.2.3Review Process	27
2.2.4Summary.....	27
2.2Data Security	28

2.2.0 Access control Models	29
2.2.1 Bell-LaPadula	29
2.2.2 Biba	30
2.2.3 Clark-Wilson	30
2.2.3.1 Tools used in data Security/protection.	31
2.3.2 Encryption	31
2.3.3 Data Security Controls to be consider	31
2.4 Data security awareness training	39
2.5 Technologies used to protect sensitive data	42
2.6 Data monitoring and maintenance	43
2.7 Security policy to be implemented for effective data security.	43
2.7 Data loss prevention	44
2.8 Critique of the literature review	
CHAPTER 3: RESEARCH METHODOLOGY	47
3.1.0 Data Classification Methodology	47
3.2.0 Access control Methodologies	49
3.2.1 Centralized	49
3.2.2 Decentralized	50
3.2.1 Access Control Models	51
3.3.0 Cryptography	52
3.4.0 Security awareness training and education	54
4.0 RESEARCH DESIGN	58
4.1 Input data set	59
4.3 Preprocessing	59
4.4 Classification algorithm	59
4.4.1 J48	60

4.4.2Naive Bayes	60
4.4.3Random Tree (RND)	60
5.0 IMPLEMENTATION	61
5.1Why Weka	61
5.2Identification of the Attributes	62
5.3 Data classification process	63
5.3Data loss prevention and encryption process after data classification	78
6.0 CONCLUSION AND FUTURE WORKS	82
6.1Conclusion	82
6.2Future works	83
REFERENCES	84

ACKNOWLEDGMENT

We would like to thank Almighty for giving us an opportunity to study in KCA University and to work on this Master Thesis of Data Centric security and governance for Data. A very big 'thank you' to our Professors and supervisors for their guidance and support and helping us to complete this required work. To our organization National oil Corporation of Kenya ICT team for mentoring us throughout the thesis with valuable inputs and for their comments and suggestions that have shaped this work. To all our lecturers and friends for very useful comments and suggestions to fine tune thesis. To our families, for their love and support throughout the duration of our course, especially during the Master Thesis course. And to all those who have in their various ways are contributing to our successful studies in KCA University.

TABLE OF FIGURE

Figure 1 Data classification	3
Figure 2 window classification	10
Figure 3 machine learning	11
Figure 4 a) show training data set	13
Figure 6 government classification	18
Figure 7 host-based DLP system	45
Figure 8 Encryption and Decryption Example	53
Figure 9 flow chart classification	56
Figure 10 sensitive data transmission	57
Figure 11 Architecture design of the propose system	58
Figure 12 accuracy Model	58
Figure 13 SSn	78

LIST OF TABLES

Table 1 government classification	18
Table 2 Commercial classification	19
Table 4 classified CIA	48
Table 5 Best Classifier	60
Table 6 classification attributes	63

LIST OF ACRONYMS AND ABBREVIATIONS

DLP-Data loss Prevention

PKI-Public Key Infrastructure

CSV-comma separated value

CIA-confidentiality integrity and authority

DIKW-Data knowledge

SSN-storage service node

SSL-secure socket layer

VPN- virtual private Network

ARFF Attribute-Relation File Format

SBU-sensitive But unclassified

CHAPTER 1: INTRODUCTION

1.1 Background Information

In today's world society are more concern on the security of external factors as compared to internal factors good example are employing security guard at our door gates CCTV surveillance, alarm sensors, deploying of firewalls, door access system among others where by individuals that we are protecting are not train on security measure and the data we are protecting has no security measure given too, hence this research is going to focus on the security of information. The security of information has been a big challenge in the current period of time. The problem of Data security is as a result of active interference by policy, classification, and malicious application, encryption of information, user awareness training and monitoring of information. The traditional methods of how to address information security have all been for ages about protecting the perimeter and the network, protecting where the files are located. While it's good IT practice to have those safeguards in place, organizations are realizing that it's not enough. Data-centric security means that security resides within the data files, wherever they are stored. It doesn't matter if the data are on a server in the organization, your home computer or a storage device, by using data-centric security you're enforcing role-based access control (Cole, 2013).

Data security is built on the confidentiality, integrity and availability of data, and, to a certain extent, on access control and nonrepudiation. Confidentiality means that the data is only available to those who have been authorized to use it by agreed means and in an agreed timescale, and that such information is not revealed or otherwise made known to a third party. Integrity means that the data and data systems are reliable, correct and up-to-date, and have not been changed or damaged as a result of equipment or software defects, natural events or unauthorized human action. Availability means that, from an operational point of view and within an acceptable timescale, the

data and data processing systems are Available and useful to authorized users. Access control means that the data and data system cannot be used without the relevant permission. Non-repudiation means the creation of proof in order to ensure that no party involved in the processing or transfer of data can afterwards deny its part in it.

The fact that sensitive data seems to increasingly follow a pattern of being leaked, lost or stolen, has forced security professionals to rethink how their organizations can keep their most valuable assets safe. Data-centric information security solution that emphasizes on how sensitive information can only be used by those that have express authority view. Even if sensitive data is leaked, it is rendered useless to unauthorized parties that may acquire it. Keeps sensitive data safe and secure regardless of whether it is in transit or stored inside or outside your network perimeter (Krause, 2007).

Data classification, in information security is the classification of data based on its level of sensitivity and the impact to the corporation should that data be disclosed, altered without authorization. The classification of data helps determine what baseline security controls are appropriate for safeguarding that data. Corporation data should be classified into one of three sensitivity levels or classifications: Data should be classified as Restricted when the unauthorized disclosure, alteration or destruction of that data could cause a significant level of high risk to the Corporation or its affiliates .The highest level of security controls should be applied to restricted data. Data can be classified as Private when the unauthorized disclosure, alteration or destruction of that data could result in a moderate level of risk to the University or its affiliates. Data can be.

classified as Public when the unauthorized disclosure, alteration or destruction of a data would results in minimum or no risk to the organization The policy outline essential roles and responsibilities within the corporation for creating and maintaining data from threats to personal, professional and institutional interests and to establish a comprehensive data security program in compliance with applicable law. This policy is designed to ensuring the security and confidentiality of confidential information and to establish protection against unauthorized access or use of this information.

Encryption is a very generic term and there are many ways to encrypt data. Companies need to implement and manage encryption correctly. Using strong encryption and proper key management is a good encryption strategy. Encrypt sensitive data before it is shared over untrusted networks

Training of new employee is crucial to data security. You need to train every member of your corporation about the significance of data security. In that you can't stop cyber criminals at the boundary of your organization, you need to focus specifically on protecting your data. That means protecting your sensitive data at the point of initial and throughout, wherever it is being stored, moved and used. This approach is called data-centric security



Figure 1 Data classification

1.2 Problem Statement

Although information security in organization is a function that is well known as essential, the actual priority allocated to information security is not always commensurate with its advantage. More issues arise on how specific implementation of security practices should occur and how to mitigate the associated impact on work practices

Where lack of a data inventory management and data loss prevention (DLP) tool is purchased. Were company will fail if it does not know where its data resides, who its owners are, whether the data it stores is critical or non-critical and what data security regulatory requirements must met, then procuring and implementing a DLP has no need.

More problem are made when the implementation of a DLP tool is treat as a technology project, not a business program. When a company commits to the implementation of a DLP, it must realize that the effort begins once the tool is in place. Data will have to be discovered, classified and categorized based on a variety of factors on an ongoing basis. It will move from a project to a life term program that must remain maintained for the life of the product.

If this is not done then the tool will eventually fall into disuse as staff is reassigned to other initiatives and executives place other priorities on the information technology department. A DLP tool and its program must be parallel with the business strategy and have a business owner for it to be successful. When not fully tested and prepared to rollout a DLP solution. Where more conflicts between various parts of the enter on DLP. We also frequently see situations where the company implementing DLP hasn't fully classified their data and made decisions about what needs to be protected and to what level. All of these discussions can be time taking and filled with disagreements, so they get avoided and result in an unaccomplished DLP program

Once implemented such data-at-rest discovery and classification solutions have two principal benefits: By precisely locating and tagging sensitive data, post-breach losses can be minimized because the bad guys can't steal what they can't find. And they allow organizations to more efficiently target their security spend: Encryption technologies, for example, can be focused on the locations where sensitive data actually resides, or egress controls can be tuned to monitor certain types of sensitive data that are found to be the prevalent sources of risk for a security organization.

Data is king. We enter, collect, scan, process, analyze, store, print and transmit data all day, every day. It's the heart and soul of most organizations, and they rely on it to achieve their goals and accomplish their missions. But how safe is this most precious asset of the business? How is it being protected? Is enough being done to ensure it is safeguarded? What else can be done? The solution should focus on protecting data, files, documents and folders stored and used by the user community throughout its lifecycle. It should also protect the data when it is in motion and distributed to employees internally, externally and to partner organizations.

A vulnerability is a security hole or weakness in a piece of software, hardware or operating system that provides a potential way to Compromise the, integrity, or confidentiality of the system. An exploit is a highly sophisticated software, piece of data, or a sequence of commands that takes advantage of the system weaknesses to carry out some form of malicious intent. The weakness in the system can be among others design vulnerability, inadequate or wrong configuration of the product, flaws in the implementation etc.

when data classification is done accurately, and in real-time – will reduce both, the residual, non-zero risk of post-breach damages, as well as the security investment required to protect against sensitive data loss

1.3 Objectives.

The objective of data security work in the maintenance organization is to safeguard the uninterrupted use of data systems that are important to the uses of the maintenance organization, to prevent the unauthorized use of data and data systems, the Unintentional or deliberate destruction or distortion of data, and to minimize damage caused. In addition to safeguarding operational data processing at normal times, preparations should be made to deal with threats that might result in the suspension of uses and to recover from such situations. Through administrative, technical and other procedures, the data and data processing systems and services of the maintenance organization can be kept appropriately protected, during both normal times and emergencies. The aim of the maintenance organization is to ensure that data security arrangements are at a good level, both nationally and internationally. The aim is also to ensure that the basic level of data security covers all the data processing of the maintenance organization, taking into account the basic nature of the faculties and their possible need to boost data security.

- i. Identify the attributes that make document either confidential, sensitive, public, private, proprietary.
- ii. Evaluate classification algorithm e.g. Using Weka tool.
- iii. Implement data loss prevention technique on sensitive data
- iv. Implement encryption on sensitive data

1.4 Justification of the study

Data-centric security is the only way to ensure the most important asset of the business the data is protected. The damage caused to the reputation of the organization may never be restored. Best practices that fully minimize risk should revolve around automated data-centric security solution

that features strong data security through data classification process. With support for automated classification and encryption, hence prevent users from disclosing valuable business information to unauthorized recipients. Use data classification to help security solutions recognize and protect your sensitive information. As a front-end to DLP and other security solutions, it identifies sensitive data so that your security team can optimize security policy, and focus on the highest-risk areas.

1.5 Significant of the study

Protect enterprise data by securing files, file names, e-mail messages and attachments regardless of security format or computing platform using strong encryption. Reduce complexity by enabling a seamless user workflow and integration into desktop and office computing applications such as Word, Excel, Outlook, etc. Reduce sensitive data exposure by securing files using PKI encryption. Enforce the use of data protection using a centrally managed security policy in the enterprise. Provide contingency key support to ensure access to all encrypted files by IT security for emergencies, protection against malicious employee behavior and audit purposes. Ensure access to encrypted data on mobile devices.

CHAPTER 2: LITERATURE REVIEW

This chapter provides a theoretical background into information security. It highlights the importance of information to any organization and presents the steps that will improve the effectiveness and efficiency of work. It highlights. The need for organizational information security and the pressure that management and employees put on security components as they attempt to maximize efficiency and minimize Workload. It explains some of the common found Data security issues that have been a big threat to any data.

Best practices that fully minimize risk should revolve around automated data-centric security solution that features classification of data and security controls through policy management. Policy management is an important ingredient that enables the organization to enforce standards and protection on data stored on the devices at the endpoints or the organization. Equally important is the ability to include a contingency key for access to encrypted data by security administrators for auditing purposes or in the event an employee leaves the organization.

2.1Data classification

Understanding your data accurately is the foundation of your sensitive data protection strategy and helps you determine where to apply your data security controls. But once you understand your data, you need to properly classify it. Without security classification, an organization treats all data as if it were the same: You can't know the level of importance of any data because it hasn't been properly classified. Failing to perform classification of data increases the risk of sensitive data being compromised across the data security lifecycle. It also increases the possibility that you could be placing security controls on data that isn't in fact sensitive. Data classification isn't a simple task, even for information security experts. Attempting to change user behavior can tempt anyone to postpone it indefinitely (Cole, 2009).

2.1.0 Data Classification definition

2.1.1 Definitions of Data

By definition, data is a form of information and is represented by the facts, quantities, figures, statements, symbols, and observations that we use for inquiry, reference, or analysis. We produce data every day as we go about our lives. Data has also been conceptualized as part of a hierarchy that includes information, knowledge, and even wisdom. The general idea is that as data is given context through various analytical processes, it transforms through various states or stages. Also contributing to this increased sophistication are the experiences of those dealing with the data and its higher forms, until ultimately wisdom can appear to be an almost intuitive gift for understanding circumstances derived not only from the data at hand but from the insights generated using previous data, information, and knowledge as well.

The rank known as data-information-knowledge-wisdom, frequently used in information science. The DIKW rank is a simple, generalized model for imagining relationships between different ways of understanding the world. It can also be useful to IT security development because it reminds us that data is not the only, or even the most important, aspect of what we are trying to achieve. It represent the core of a larger process of understanding in which we try to constantly learn and improve over time. Moving from metrics data to security wisdom will be one of the goals of the Security Process

2.1.2 Windows Server classification

In window server classification it scan the folder and read the file inside and classify the files within the share folders depending on the set attributes .where most organization they are classified according to the departmental level without considering the sensitivity of the files. The diagram below show how classification is done

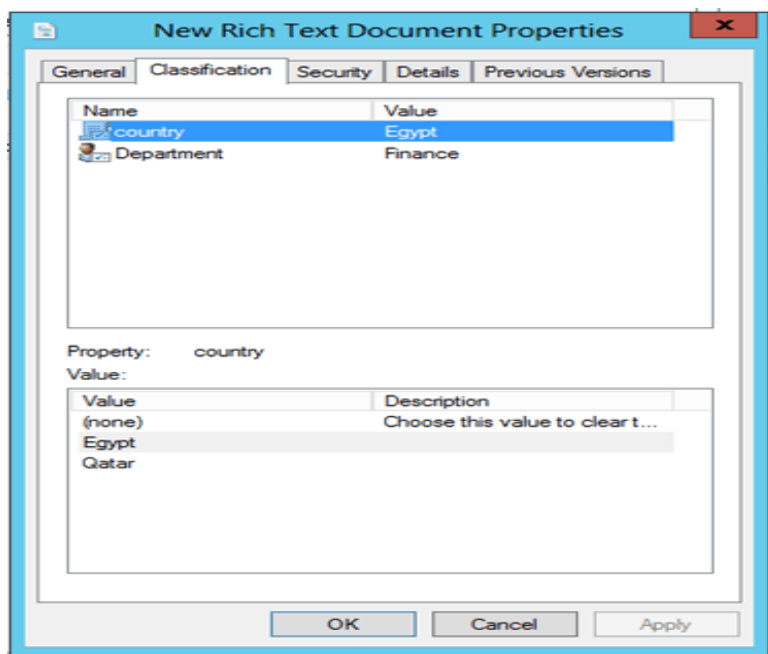


Figure 2 window classification

Why is Data classification very important?

Different organization have their preferred way of classifying data having various reasons .Some organizations might want to classify data to have easy access or classify due to handling sensitive data that they may be require protection . Classification of data will help in determining baseline security controls for the safeguarding of data (Calder & Watkins, 2012).

2.1.3 Machine Learning Algorithms for Classification

It is important to learn how to make accurate predictions automatically based on the past observation. The diagram shows a classification problem given into set categories. Labeling of training data set goes through machine learning algorithm to be classified. Where test data will also be input to predict the accuracy of classification.

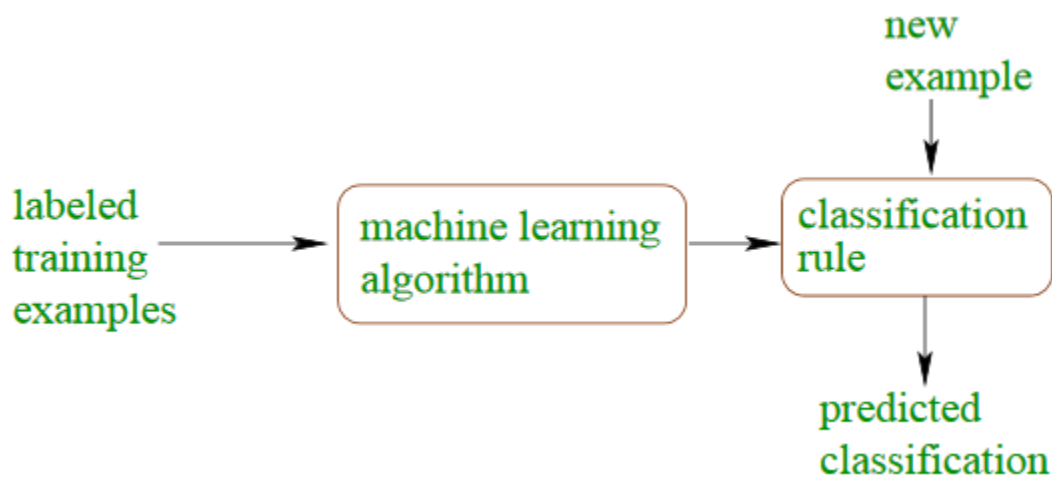


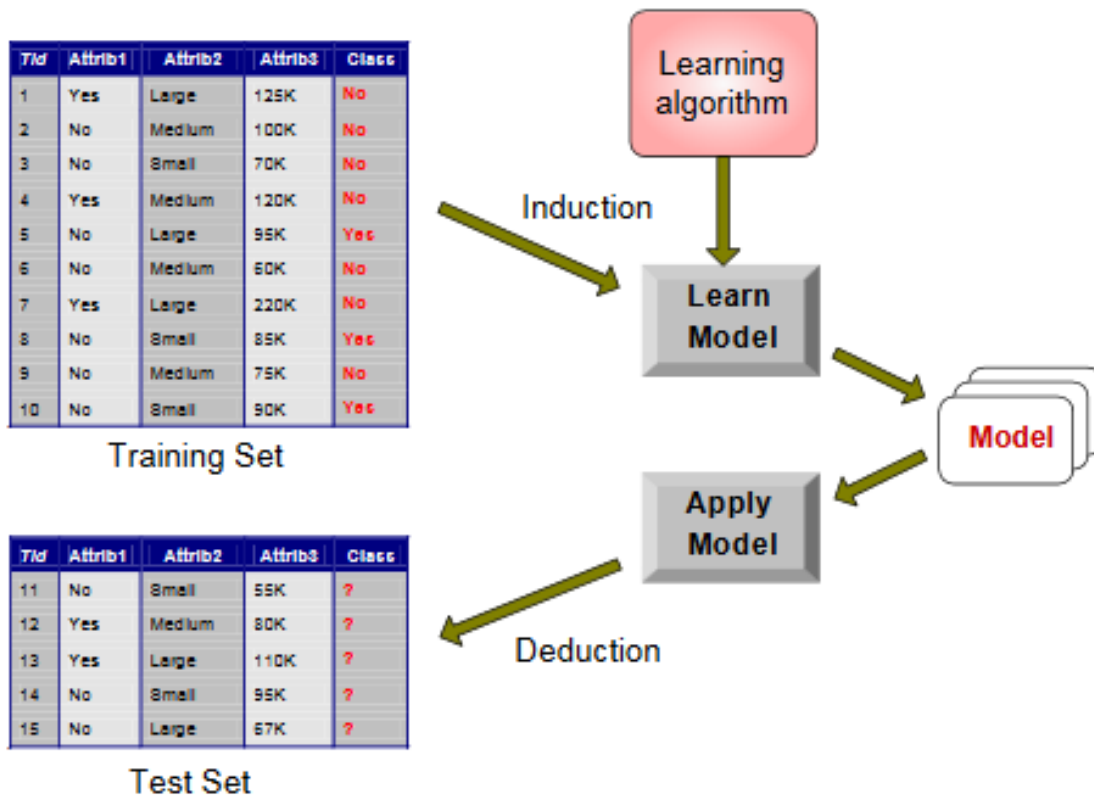
Figure 3 machine learning

Example of classification include:

- ✚ Text categorization either spam mail
- ✚ Market segmentation if customer will buy
- ✚ Fraud detection
- ✚ Weather outlook which commonly used in weka tool

In every record it must contain a set of attributes, where one of the attributes is a class. The data set can be divided into training set which used to build the classification model and test set which is used to determine the accuracy of classification model.

The diagram below show the training data set and test data set. Where by the training data set contain full information ID.The attributes are classified as either No or yes ,while the test set will be load in the tool in order to valid the model accuracy either Yes or No



2.1.4 Classification Techniques

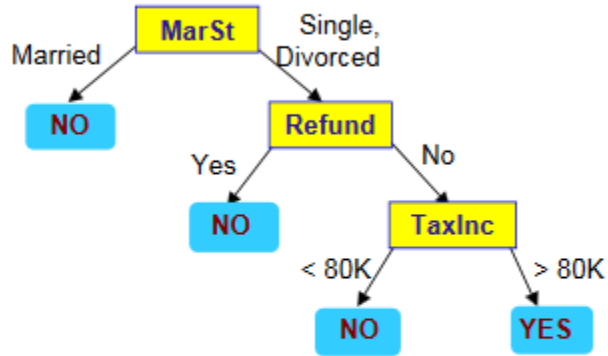
We have different classification technique where from my research I will focus on specific classification as describe in the document

- ✚ Decision Tree based Methods
- ✚ Rule-based Methods
- ✚ Memory based reasoning
- ✚ Neural Networks
- ✚ Naïve Bayes and Bayesian Belief Networks

The below diagram show sample of decision trees on figure 4a) training data set and 4b) on the test data to be validated.

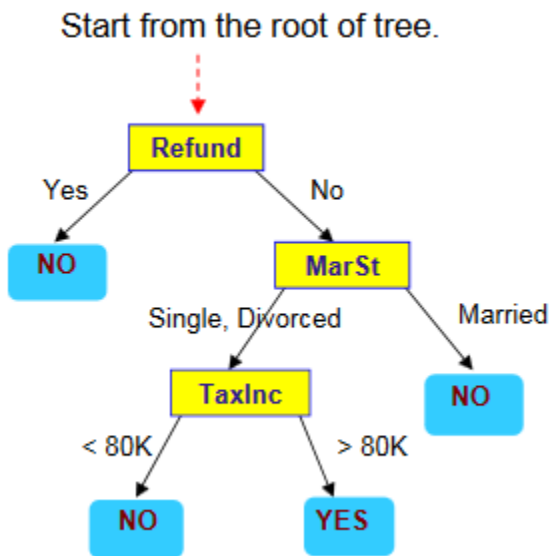
categorical categorical continuous class

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



There could be more than one tree that fits the same data!

Figure 4 a) show training data set



Test Data

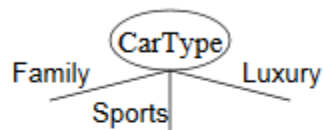
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

Figure 5) show test data set

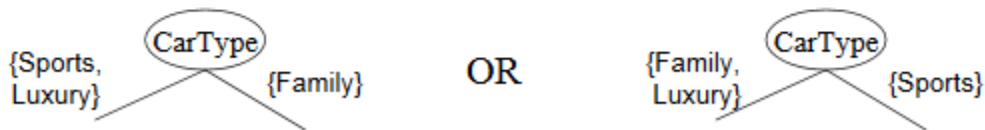
How to specify the test condition

It will depend on the attributes type either continuous, nominal or ordinal. Depending on the number of split binary split or multiway split. The example below show the distinctions between the split

- **Multi-way split:** Use as many partitions as distinct values.



- **Binary split:** Divides values into two subsets. Need to find optimal partitioning.



Measures of Node Impurity

The measure of node impurity can be categories into difference indexing which include Gini index, entropy and misclassification index but our focus will be on how to calculate the Entropy; example is show below.

2.1.5 Entropy and information gain entropy calculation

Definition of Entropy is the sum of the probability of each label multiplies the log probability of the same label as the data become purer and purer, the entropy value becomes smaller and Smaller.

To show that we take an example of classifying a name into female or male group .which are labeled as F or M .it learn the model and predict the gender of unseen first

name	gender
Ashley	f
Brian	m
Caroline	f
David	m

We can try predict the first name Roma

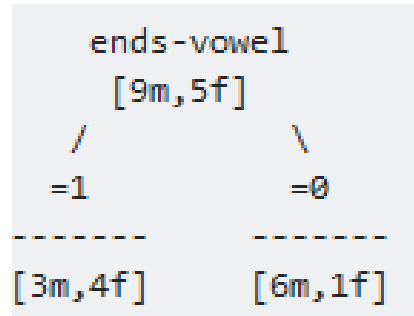
# name	ends-vowel	num-vowels	length	gender
Ashley	1	3	6	f
Brian	0	2	5	m
Caroline	1	4	8	f
David	0	2	5	m

Our main focus is to build a decision tree .example would be as shown below

```
length<7
| num-vowels<3: male
| num-vowels>=3
| | ends-vowel=1: female
| | ends-vowel=0: male
length>=7
| length=5: male
```


Entropy is calculated as show below. With example of decision tree

$$\text{Entropy} = - p(a) \cdot \log(p(a)) - p(b) \cdot \log(p(b))$$



As show above we have 9men and 5 female before the split, after the split we calculate the left, right and after split entropy separately as shown below

$$\text{Entropy_before} = - (5/14) \cdot \log_2(5/14) - (9/14) \cdot \log_2(9/14) = 0.9403$$

$$\text{Entropy_left} = - (3/7) \cdot \log_2(3/7) - (4/7) \cdot \log_2(4/7) = 0.9852$$

$$\text{Entropy_right} = - (6/7) \cdot \log_2(6/7) - (1/7) \cdot \log_2(1/7) = 0.5917$$

$$\text{Entropy_after} = 7/14 \cdot \text{Entropy_left} + 7/14 \cdot \text{Entropy_right} = 0.7885$$

After calculating the split we find out how much information we have gain as show below .we were able to reduce uncertainty by sub dividing the decision tree hence gain small amount of 0.1518

```
Information_Gain = Entropy_before - Entropy_after = 0.1518
```

2.2.1 Classifications for Sensitive Data

The classifications for the sensitivity of data used in government and military applications are top secret, secret, confidential, sensitive but unclassified, and unclassified. The implementation of the classification is based on laws, policies, and executive directives that can be in conflict with each other. Agencies do their best to resolve these conflicts by altering the meaning of the standard classifications. Table explains the types of classifications used by government civilian and military organizations. Data classification is the process of labeling data based on its sensitivity and handling requirements. It help the government ensure that only authorized individuals can access data and that the data is appropriately protected (Seidl, 2015).

2.2.2 Government Classification

Classification of government data is done out of policy for maintaining security and privacy of citizen data. Military and intelligence organizations set their classifications on the ramifications of disclosure of the data. Civilian also look to prevent unauthorized disclosure, but they also have to consider the integrity of the data (Seymour Bosworth, 2014)



Figure 5 government classification

2.2.3 Government data classifications

Classification	Description
Top Secret	Disclosure cause severe damage to national security.
Secret	Disclosure cause serious damage. Less sensitive than data classified as top secret.
Confidential	Exempt from disclosure under.
Sensitive But Unclassified (SBU)	Disclosure would do some harm.
Unclassified	Unclassified is data that has no classification or is not sensitive.

Table 1 government classification

2.2.4 Commercial Classification

Classification of commercial or nongovernment organizations does not have a set standard. The classification used is dependent on the overall sensitivity of the data and the levels of confidentiality desired. The table show typical list of classifications that can be used for commercial classification, from highest to lowest.

Classification	Description
Sensitive	Most limited access and requires a high degree of integrity. Cause most damage if disclosed.
Confidential	Less restrictive might cause damage if disclosed.
Private	might not do the damage but must be keep private for other reasons
Proprietary	reduce the company's competitive advantage,
Public	Least sensitive data cause the least harm if disclosed.

Table 2 Commercial classification

Criteria

Organizations classify data to comply with their requirements of Confidentiality, Integrity and Availability. Data is typically classified according to its type such as medical, financial, personal, to name a few. These will be defined by the organizations or by regulations, policy or law.

	POTENTIAL IMPACT		
Security Objective	LOW	MODERATE	HIGH
Confidentiality Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information.	The unauthorized disclosure of information could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals.	The unauthorized disclosure of information could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals.	The unauthorized disclosure of information could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals.
Integrity Guarding against improper information modification or destruction, and includes ensuring information non-repudiation and authenticity.	The unauthorized modification or destruction of information could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals.	The unauthorized modification or destruction of information could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals.	The unauthorized modification or destruction of information could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals.
Availability Ensuring timely and reliable access to and use of information.	The disruption of access to or use of information or an information system could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals.	The disruption of access to or use of information or an information system could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals.	The disruption of access to or use of information or an information system could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals.

After the classification scheme is identified, the organization must create the criteria for setting the classification. No set guidelines exist for setting the criteria, but some considerations are as follows:

- ✦ Who should access or maintain the data?
- ✦ Which laws or policy should protect data?
- ✦ What will be the effect when data is disclose?
- ✦ Where is the data to be stored?
- ✦ What is the value or important of the data?

2.2.4 Granularity of classification

The sets of information being classified should, in general, be large rather than small. More administrative efforts is required on small unit classification, it involve more decisions and add to complexity, thus decreasing the overall security

2.2.5 Creating Data Classification Procedures

Using this information, your organization can create a procedure for classifying data. Government organizations already have this procedure defined. Commercial organizations have a lot of flexibility in setting the procedures that best describe their preferences. Example of a procedure.

- ✚ Set the criteria for classifying the data.
- ✚ Determine the security controls for classification.
- ✚ Identify the data owner
- ✚ Document any exceptions that might be required for the security of this data.
- ✚ Determine how the custody of the data can be transferred.
- ✚ Create criteria for declassifying information.
- ✚ Add this information to the security awareness and training programs so users can understand their responsibilities in handling data at various classifications.

2.2.6 Defining Roles and Responsibilities for effective data classification

Effective Information Classification program, roles and responsibilities of all participants must be clearly defined. An appropriate training program, developed and implemented, is an essential part of the program. Not all of the roles defined in the sections that follow are applicable for all information classification schemes and many of the roles can be performed by the same individual. (Calder & Watkins, 2012).

- ✚ Information owner: Business executive who is responsible for a company business information asset. Responsibilities include

- ❖ Assign initial information classification and periodically review the classification to ensure it still meets the business needs
- ❖ Ensure security controls are in place commensurate with the classification
- ❖ Review and ensure currency of the access rights associated with information assets they own
- ❖ Determine security requirements, access criteria, and backup requirements for the information assets they own

✚ Information custodian: usually an information systems person, is the delegate of the information owner with primary responsibilities for dealing with backup and recovery of the business information. Responsibilities include the following:

- ❖ Perform backups according to the backup requirements established by the information owner
- ❖ When necessary, restore lost or corrupted information from backup media to return the application to production status
- ❖ Ensure record retention requirements are met based on the information owner's analysis

✚ **Application owner:** Manager of the business unit who is fully accountable for the performance of the business function served by the application. Responsibilities include the following:

- ❖ Establish user access criteria and availability requirements for their applications

- ❖ Ensure the security controls associated with the application are commensurate with support for the highest level of information classification used by the application


✚ **User manager:** the immediate manager or supervisor of an employee. They have ultimate responsibility for all user IDs and information assets owned by company employees. Responsibilities include the following:

- ❖ Inform security administration of the termination of any employee so that the user ID owned by that individual can be revoked, suspended, or made inaccessible in a timely manner.
- ❖ Inform security administration of the transfer of any employee if the transfer involves the change of access rights or privileges.
- ❖ Report any security incident or suspected incident to Information Security.
- ❖ Ensure the currency of user ID information such as the employee identification number and account information of the user ID owner.
- ❖ Receive and distribute initial passwords for newly created user IDs based on the manager's discretionary approval of the user having the user ID.
- ❖ Educate employees with regard to security policies, procedures, and standards to which they are accountable.

✚ **Security administrator:** any company employee who owns a user ID that has been assigned attributes or privileges associated with access control systems, such as top Secret,. This user ID allows them to set system-wide security controls or administer user IDs and information resource access rights. These security administrators may report to either a

business division or Information Security within Information Systems. Responsibilities include the following:

- ❖ Understand the different data environments and the impact of granting access to them.
- ❖ Ensure access requests are consistent with the information directions and security guidelines.
- ❖ Administer access rights according to criteria established by the Information Owners.
- ❖ Create and remove user IDs as directed by the user manager.
- ❖ Administer the system within the scope of their job description and functional responsibilities.
- ❖ Distribute and follow up on security violation reports.
- ❖ Send passwords of newly created user IDs to the manager of the user ID owner only.

 **Security analyst:** Person responsible for determining the data security directions (strategies, procedures, guidelines) to ensure information is controlled and secured based on its value, risk of loss or compromise, and ease of recoverability. Duties include the following:

- ❖ Provide data security guidelines to the information management process.
- ❖ Develop basic understanding of the information to ensure proper controls are implemented.

- ❖ Provide data security design input, consulting and review.

- ✚ **Change control analyst:** Person responsible for analyzing requested changes to the IT infrastructure and determining the impact on applications. This function also analyzes the impact to the databases, data-related tools, application code, etc.

- ✚ **Data analyst:** This person analyzes the business requirements to design the data structures and recommends data definition standards and physical platforms, and is responsible for applying certain data management standards. Responsibilities include the following:

- ❖ Design data structures to meet business needs.
- ❖ Design physical data base structure.
- ❖ Create and maintain logical data models based on business requirements.
- ❖ Provide technical assistance to data owner in developing data architectures.
- ❖ Record metadata in the data library.
- ❖ Create, maintain, and use metadata to effectively manage database deployment.

- ✚ **Solution provider:** Person who participates in the solution development and delivery processes in deploying business solutions. Duties include the following:

- ❖ Work with the data analyst to ensure the application and data will work together to meet the business requirements.
- ❖ Give technical requirements to the data analyst to ensure performance and reporting requirements are met.

- ✚ **End user:** Any employees, contractors, or vendors of the company who use information systems resources as part of their job. Responsibilities include:

- ❖ Maintain confidentiality of log-on password(s).
- ❖ Ensure security of information entrusted to their care.
- ❖ Use company business assets and information resources for management approved purposes only.
- ❖ Adhere to all information security policies, procedures, standards, and guidelines.
- ❖ Promptly report security incidents to management.

✚ **Process owner:** This person is responsible for the management, implementation, and continuous improvement of a process that has been defined to meet a business need. This person:

- ❖ Ensures data requirements are defined to support the business process.
- ❖ Understands how the quality and availability affect the overall effectiveness of the process.
- ❖ Works with the data owners to define and champion the data quality program for data within the process.
- ❖ Resolves data-related issues that span applications within the business processes.

✚ **Product line manager:** Person responsible for understanding business requirements and translating them into product requirements, working with the vendor/user area to ensure the product meets requirements, monitoring new releases, and working with the stakeholders when movement to a new release is required. This person:

- ❖ Ensures new releases of software are evaluated and upgrades are planned for and properly implemented.

- ❖ Ensures compliance with software license agreements.
- ❖ Monitors performance of production against business expectations.
- ❖ Analyzes product usage, trends, options, competitive sourcing, etc. to identify actions needed to meet project demands of the product.

2.2.7 Review Process

Each organization may have its own criteria that provide additional levels of classification. It is important to review these classification levels and criteria periodically as rules and regulations change, requiring a change in the way data is classified within the organization.

2.2.8 Summary

Information and software classification is necessary to better manage information. If correctly implemented it will reduce the cost of protecting information because in today's environment, will no longer work within the complexity of most corporation's heterogeneous platforms that make up the IT infrastructure. Information classification enhances the probability that controls will be placed on the data where they are needed the most, and not applied where they are not needed.

Classification security schemes enhance the usability of data by ensuring the confidentiality, integrity, and availability of information. By implementing a corporate-wide information classification program, good business practices are enhanced by providing a secure, cost-effective data platform that supports the company's business objectives. The key to the successful implementation of the information classification process is senior management support. The corporate information security policy should lay the groundwork for the classification process, and be the first step in obtaining management support and buy-in

2.3 Data Security

Data security is crucial for all businesses. Customer and client information, payment information, personal files, bank account details - all of this information is often impossible to replace if lost and dangerous in the hands of criminals. Data lost due to disasters such as a flood or fire is devastating, but losing it to hackers or a malware infection can have far greater consequences. How you handle and protect your data is central to the security of your business and the privacy expectations of customers, employees and partners (Mike Chapple, 2014).

- ✚ Encrypt all confidential info. Keeping sensitive information inaccessible from prying eyes.
- ✚ Use hard-to-guess passwords. Enforcing good password usage is key to stopping hackers crack into your systems.
- ✚ Keep security software up to date. New malware is being released all the time and spreads at alarming rates. Updating your software automatically is key to defending against the latest threats and vulnerabilities.
- ✚ Danger USB! Unauthorized use of USB storage devices could lead to data being lost from your company. Control usage with security software.
- ✚ Knowledge is power. Find out what your local legislative requirements and review your security strategy to ensure you are compliant. They will be able to advise on what type of technologies, processes, and policies are required by law.
- ✚ Prepare for disaster. Create a plan of action to follow if a severe data breach takes place. Swift reaction can make a huge difference to legal ramifications and corporate reputation.
- ✚ Education is key. Find an engaging way to explain to staff the value of data and talk through the technologies, policies and best practice. Have employees be part of the army safeguarding sensitive data rather than keeping them in the dark.

- ✚ Encourage – rather than punish – employees who report potential data loss or breaches.

The information can help you mitigate against costly risks

2.3.0 Access control Models

The formal models of access control are theoretical applications of access control methods. These do not specify methods of controlling access, but rather specific guidelines that should be followed. They work best with static environments and are difficult to implement within dynamic systems that are constantly changing, such as those in most enterprise environments. The documentation on how these models are supposed to be implemented is very limited and does not give any specific examples.

The formal models do provide a good baseline to start from when designing access control systems, however. By ensuring that the guidelines within the formal model most closely related to your needs are followed, you ensure that you have a strong foundation on which to build the rest of the access control system (Harris, 2013).

2.3.1 Bell-LaPadula

David E. Bell and Len J. LaPadula wrote the Bell-LaPadula formal access control model in 1973 for use in government and military applications. This formal model specifies that all access control objects have a minimum security level assigned to it so that access control subjects with a security level lower than the security level of the objects are unable to access the object. Does this sound familiar? The Bell-LaPadula formal model is what the MAC model is based on.

2.3.2Biba

The Biba formal model was written by K.J. Biba in 1977 and is unique as it was the first formal model to address integrity. The Biba model bases its access control on levels of integrity. It consists of three primary rules. The first rule specifies that a subject cannot access objects that have a lower level of integrity than the access control subject has. The second rule states that access control subjects cannot modify objects that have a higher level of integrity than their current integrity levels. The last rule specifies that an access control subject may not request services from subjects that have a higher integrity level.

2.3.3Clark-Wilson

The Clark-Wilson formal model was written in 1987 and updated in 1989 by David D. Clark and David R. Wilson. This model is similar to Biba, as it addresses integrity. The Clark-Wilson model is designed to not only address access to objects, but also to ensure integrity by specifying guidelines for processes which occur using the access control object.

One of the most important guidelines to come out of the Clark-Wilson model is that of *segregation of duties* or *separation of duties*. The principle of segregation of duty states no single person should perform a task from beginning to end, but that the task should be divided among two or more people to prevent fraud by one person acting alone. This ensures the integrity of the access control object by securing the process used to create or modify the object.

2.4.1 Tools used in data Security/protection.

2.4.2 Encryption

The key allows the data to be coded so that to decode it, one would need to know the key that was used to code it. This coding of the given data using a key is known as *encryption*, and decoding of the encrypted data, the reverse step-by-step process, is known as *decryption*. At this stage we point out that the encryption algorithm comes in two flavors: symmetric and asymmetric. Securing data requires a three-pronged approach: detection, prevention, and response. Data normally resides on storage media that are accessible over a network. This network is designed with a perimeter around it, such that a single access point provides a route for inbound and outbound traffic through a router supplemented with a firewall.

Data encryption prevents data from being exposed to unauthorized access and makes it unusable. Detection enables us to monitor the activities of network users and provides a means to differentiate levels of activities and offers a possible clue to network violations. Response is equally important, since a network violation must not be allowed to be repeated. Thus the three-pronged approach is evolutionary, and therefore systems analysis and design principles must be taken into account when we design a secured data network.

2.4.3 Data Security Controls to be consider

The encryption key must be kept secure and known only to those who are authorized to have access to the data. Public/private key algorithms could be considered for maximum security and ease of use.

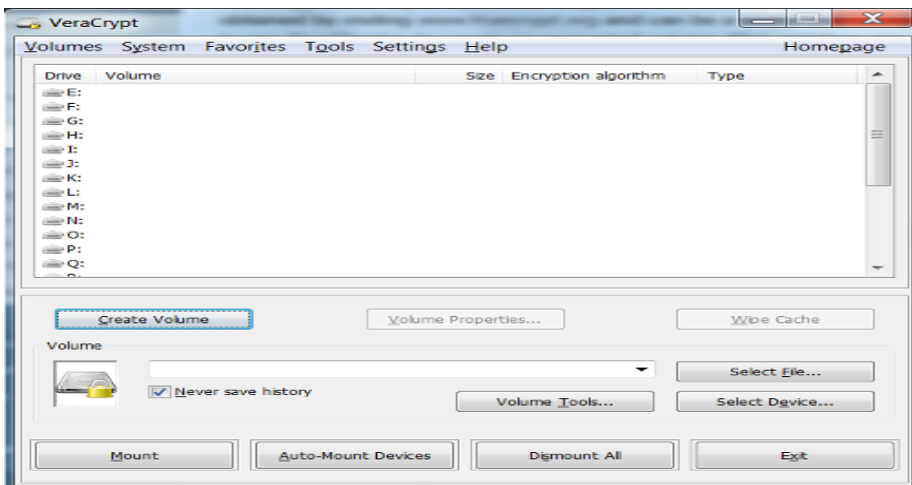
- ✚ Review and approve—this is a procedural control, the intent of which is to ensure that any change data is reviewed by someone technically knowledgeable to perform the task. The review and approval should be done by an authorized individual other than the person who developed the change.
- ✚ Backup and recovery—Depending on the criticality of the data and ease of recovery, plans should be developed and periodically tested to ensure the data is backed up properly, and can be fully recovered.
- ✚ Separation of duties—The intent of this control is to help ensure that no single person has total control over the data entry and validation process, which would enable someone to enter or conceal an error that is intended to defraud the organization or commit other harmful acts. An example would be not allowing the same individual to establish vendors to an Authorized Vendor File, then also be capable of authorizing payments to a vendor.
- ✚ Universal access: none—No one has access to the data unless given specific authority to read, update, etc. This type of control is generally provided by security access control software.
- ✚ Universal access: read—everyone with access to the system can read data with the control applied; however, update authority must be granted to specific individuals, programs, or transactions. This type of control is provided by access control software.
- ✚ Universal access: update—anyone with access to the system can update the data, but specific authority must be granted to delete the data. This control is provided by access control software.
- ✚ Universal access: alter—anyone with access to the system can view, update, or delete the data. This is virtually no security.

- ✦ Security access control software—this software allows the administrator to establish security rules as to who has access rights to protected resources. Resources can include data, programs, transactions, individual computer IDs, and terminal IDs. Access control software can be set up to allow access by classes of users to classes of resources, or at any level of granularity required to any particular resource or group of resources.

Vera Crypt is an open source encryption solution that is easy to use and works on Windows, Mac, and Linux. It can be a useful tool to help protect data. The most common way to use Vera Crypt is to create an encrypted volume (file) and then store files inside the volume that need to be encrypted. Steps have been provided below for the creation and mounting of a Vera Crypt volume.

Creating a Vera Crypt Volume

1. Open up the Vera Crypt application and select "**Create Volume**".



2. Select "Create an encrypted file container" and click "**Next**".



3. Create a name and select a location to save the Vera Crypt volume that you will be creating and select “**Next**.”



4. Vera Crypt provides some options for encryption. Select the desired encryption algorithms and select “**Next**”. Vera Crypt uses the AES algorithm by default and is recommended for selection.



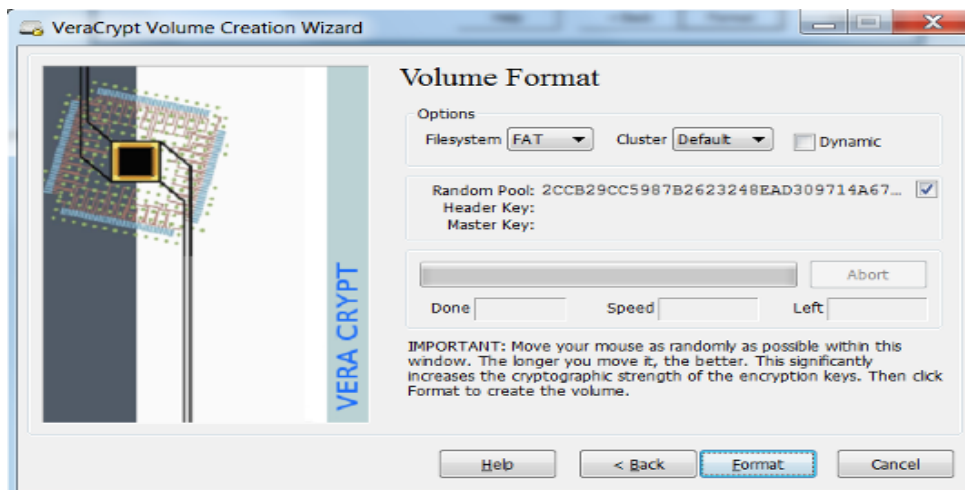
5. Create the size of the volume that you would like to create. Click “Next”.



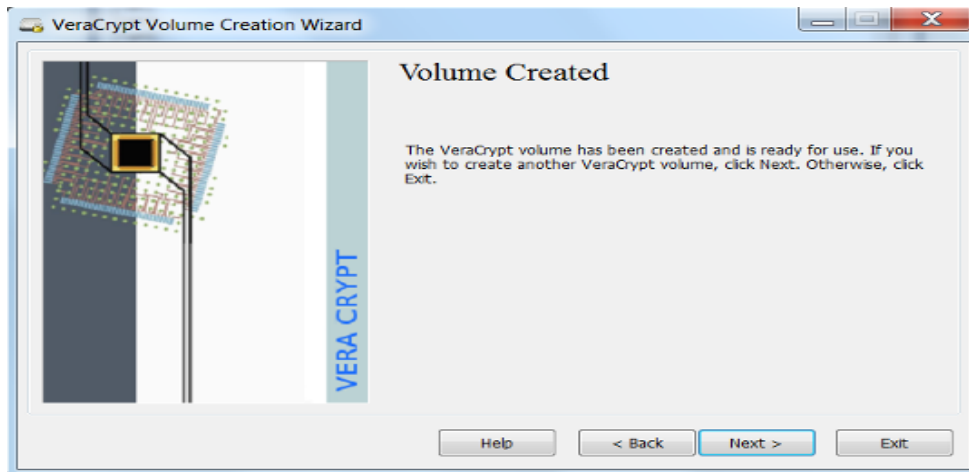
6. Create a password for your encrypted volume. Click “Next”.



7. Select the file system and click “**Format**”. The default is suitable in most cases.

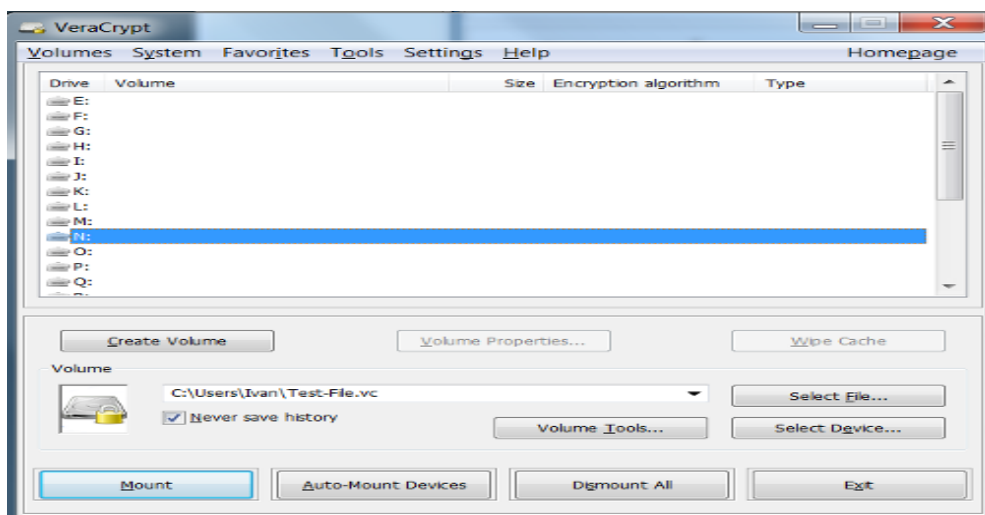


8. Your volume has been created. Click “**Exit**”

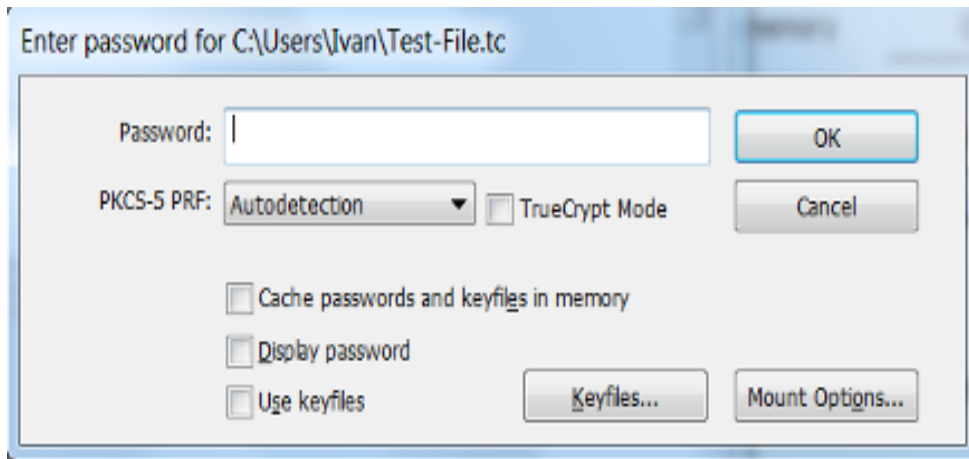


Mounting a VeraCrypt Volume

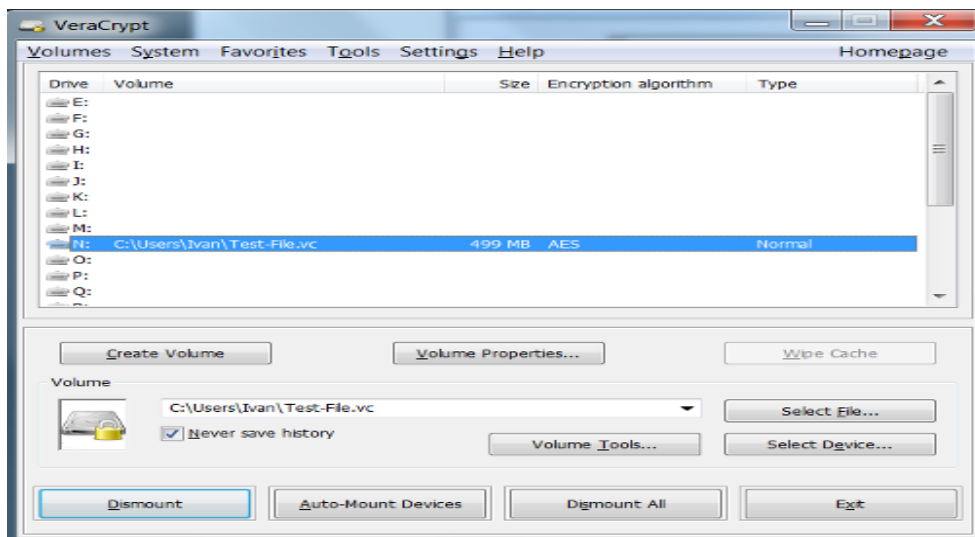
1. Select the VeraCrypt volume that you would like to mount. Select the letter drive location you would like to use. Click **Mount**.



2. Enter your password for the Vera Crypt volume. Click **OK**.



3. The drive is now unencrypted and ready for use at the drive location you have selected. Save files and documents as you would normally do with a mounted drive.



2.5 Data security awareness training

The importance of *security awareness training* and education cannot be overstated. By taking the policy, standards, and procedures and teaching all the stakeholders about their roles in maintaining the information security environment, they will embrace the policy as an integral part of their jobs. This is not easy. One problem is that over the last decade, the commitment to Data security by industry-leading companies has been viewed as lacking. The results are products that have insufficient security measures being installed into environments that further weaken the information security program.

Security awareness training requires clear communication. One thing you might consider for your organization is hiring a technically competent communicator for the security department. This person would do the training, educate the department to the concerns of its users, and act as a liaison between users and the department. Having someone who can communicate helps raise the confidence level users should have for the department.

Mandating that training be required for anyone with access to an organization's information assets is reasonable. Human resources should have complete records, including information on training courses required and taken as well as all signed documents showing acceptance of defined corporate policies.

Understanding the management role of information security means understanding how the information security process interfaces with the rest of the organization. It is not enough to just set policies—security is a process that must be molded into the business process to support its functions. Management must support these processes with commitment and training.

Understanding what is to be protected is an important beginning of the management process. A risk analysis is used to determine the information assets that need to be protected and how they can be best protected. The risk analysis takes into consideration the costs of the assets to determine not only the countermeasures, but also whether the assets are worth protecting.

Using this information, policies, guidelines, standards, and procedures can be created to reach the security goals. Policies can be described as the goals of the information security program. Guidelines are suggestions, and standards are the specific security mechanisms that can be used. Procedures use the guidelines and standards to implement the policies.

Access methods and protection mechanisms are used to manage the access and movement of data. A typical access method paradigm is to set the roles and responsibilities for access to the data. Protection mechanisms are used to compartmentalize access to data and processes. Layers are used to prevent unauthorized access to protected resources and data, whereas abstraction and data hiding are used to protect data.

Knowing who your users are is as important as setting their access rights to information assets. Employment policies enforce background checks during the hiring process to prevent hiring those who might be security risks. They can also set termination procedures to prevent the terminated user from destroying systems and data out of malice.

Change control and configuration management can be used to prevent unauthorized changes to the network. Change control policies can be used to maintain the configuration of all information assets to prevent them from being used to attack your organization.

The only way to really demonstrate management support of the policies and procedures is to require and support security awareness training. Through training, users come to understand their roles and responsibilities in the security environment. Training is the only way for the users to understand their responsibilities.

Security awareness training teaches employees to understand system vulnerabilities and threats to business operations that are present when using a computer on a business network. A strong IT security program must include training IT users on security policy, procedures and techniques, as well as the various management, operational and technical controls necessary and available to keep IT resources secure. In addition, IT infrastructure managers must have the skills necessary to carry out their assigned duties effectively. Failure to give attention to the area of security training puts an enterprise at great risk because security of business resources is as much a human issue as it is a technology issue. Technology users are the largest audience in any organization and are the single most important group of people who can help to reduce unintentional errors and IT vulnerabilities. Users may include employees, contractors, foreign or domestic guest researchers, other personnel, visitors, guests and other collaborators or associates requiring access. Users must:

- ✚ Understand and comply with security policies and procedures.
- ✚ Be appropriately trained in the rules of behavior for the systems and applications to which they have access.
- ✚ Work with management to meet training needs.
- ✚ Keep software and applications updated with security patches.
- ✚ Be aware of actions they can take to better protect company information. These actions include: proper password usage, data backup, proper antivirus protection, reporting any suspected incidents or violations of

2.6 Technologies used to protect sensitive data

Keeping sensitive data safe from inappropriate access and disclosure is of the utmost importance. Virginia Tech has many policies, procedures, and standards in place to protect sensitive data. It is the responsibility of everyone handling sensitive data from Virginia Tech to be familiar with these policies, procedures, and standards. It is important to find out what sensitive data you are handling and what steps are needed to protect it.

Data in motion

While protecting information at rest is important, it is also critical to protect any information that goes over an untrusted network. Utilizing VPNs when data is sent over an untrusted network is critical to make sure the information cannot be intercepted or compromised. While many organizations typically do a pretty good job of making sure their laptops have VPN clients installed when they communicate over the Internet, the area that we see organizations have trouble is with untrusted clients. More and more people are accessing and connecting to corporate resources from the Web using an untrusted client. Two of the most common examples are personal computers from home and computers at international airports. Many executives travel internationally and many airlines have lounges where people can wait between flights. To make it easier for them to check email or work, many airlines often have computers that can be used to make it easier so someone does not have to turn on their computer between flights. The problem in these cases is the communication is usually encrypted with SSL but there is no protection of the data at rest. When users use the Web and SSL to access sensitive information, information it is often saved to the local hard drive of the untrusted system without the user realizing it. Since the system is untrusted someone else could potentially use the computer and access the sensitive information

that the user inadvertently saved to that system. Therefore to provide both data in transit and data at rest protection, SSL VPNs are often used. An SSL VPN is an SSL connection that creates an encrypted RAM drive on the untrusted computer and all of the activity with the session is stored in the encrypted RAM drive. Now any file or information that is saved is in an encrypted area and because it is in memory, it is removed when the system is turned off. Now by using SSL VPNs an organization can ensure that their information is protected when it is going across an untrusted network and when it is stored on an untrusted computer.

2.7 Data monitoring and maintenance

Ongoing Monitoring once the information processes have been implemented and data classified, the ongoing monitoring processes should be implemented. The internal audit department should lead this effort to ensure compliance with policy and established procedures. Information Security, working with selected information owners, Legal, and other interested parties, should periodically review the information classifications themselves to ensure they still meet business requirements.

The information owners should periodically review the data to ensure that it is still appropriately classified. Also, access rights of individuals should be periodically reviewed to ensure these rights are still appropriate for the job requirements. The controls associated with each classification should also be reviewed to ensure they are still appropriate for the classification they define.

2.7 Security policy to be implemented for effective data security.

- ✚ Create use of flash disk.
- ✚ Password policy
- ✚ Authorization form allow access to data

- ✚ Transfer of data over the network
- ✚ User right delete, edit, update, full control
- ✚ Back up of data BCP DRP
- ✚ User induction training on the use of information

2.8 Data loss prevention

The information security industry has responded to the need to control data leaving systems and networks by creating technical solutions called data loss prevention (DLP) systems. These systems include both software installed on computers and devices, and network devices that are placed between networks and the network edge. Integrated hybrid systems combine protection for both networks and individual systems, along with reporting and management tools to provide insight into the status of an organization's data usage and protection, where it can be successfully implemented after data classification (Seidl, 2015).

The below figure shows the installation of a typical data loss prevention system for a network-based DLP system monitoring traffic at a network boundary. When systems attempt to send outbound e-mail, or file transfers, the DLP device reads the traffic and checks it for data that meets its filtering rules. If the rules are matched, the DLP takes the programmed action. In Figure below, the DLP is set to block traffic that carries a confidential /sensitive information.

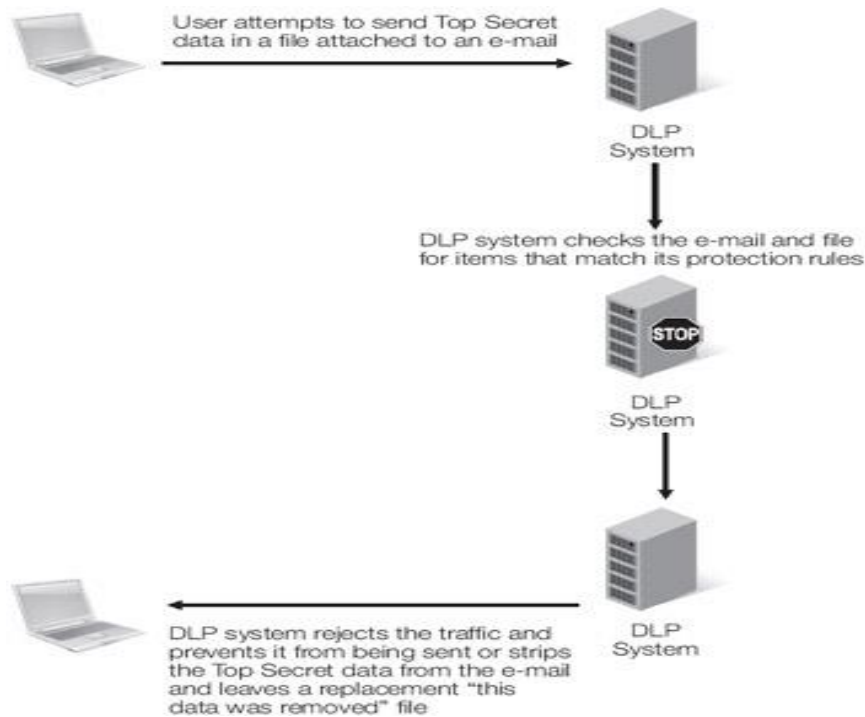


Figure 6 host-based DLP system

The figure below shows a typical host-based DLP system. Host-based DLP often includes both discovery capabilities and data loss prevention technology. In Figure the host-based DLP system detects a user attempting to copy to data to drives a file that matches protection rules. It prevents the transfer, notifies the user, and sends a notification message to a security administrator.

Host-based DLP provides a greater level of control than does network-based DLP alone. However, because of the large number of systems found on a typical network, it also increases the overall workload for security administrators

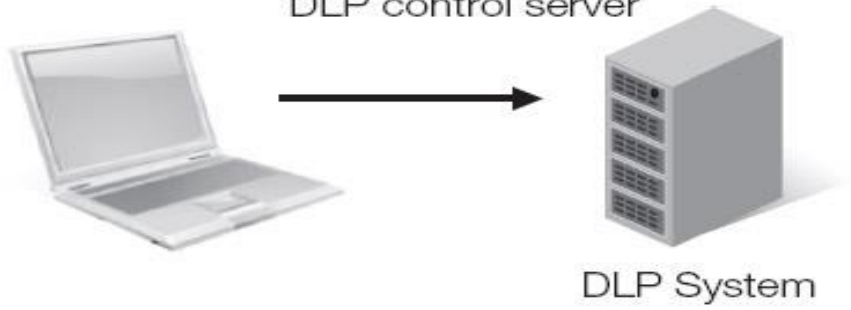
User attempts to copy a Top Secret file to a USB thumb drive



DLP software prevents the transfer



DLP sends notification to central DLP control server



CHAPTER 3: RESEARCH METHODOLOGY

This chapter describes the research method used in carrying out the study. It describes the tool to be used in the research detailing the data collection, processing techniques and tools used for the research. Given time, resources and motivation any attacker can break into almost any system. Security procedures and technologies currently in use cannot guarantee the safety of the data there are many procedures and technologies in use to ensure that our data are on safe hands but that is not enough. Equally, there are many tools that can be used for data classification to Dr. Paul Dorey, The Director of Digital Business Security Information security provides the management processes, technology and assurance to allow businesses' management to ensure business processes can be trusted

The COBIT process framework for IT security comprises 34 generic IT processes grouped in four domains; Plan and Organize, Acquire and Implement (AI), Deliver and Support (DS), Monitor and Evaluate (ME). These processes endeavor to provide information that is effective, Confidential, Efficient, available with integrity, compliant and reliable using IT resources that comprise people, infrastructure, applications and information.

3.1.0 Data Classification Methodology

The methodology presented here is adapted from the Federal Government's Federal Information Security Management Act information security framework and supporting Federal Information Processing Standard and national Institute of Standards and Technology guides and publications.

3.1.1 Data is classified on the Basis of Confidentiality, Integrity and Availability Impact Levels

Table 1: Information and System Security Objectives Security Objectives	FISMA Definition [44 U.S.C., Sec. 3542]	FIPS 199 Definition
Confidentiality	“Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information...”	A loss of <i>confidentiality</i> is the unauthorized disclosure of information.
Integrity	“Guarding against improper information Modification or destruction, and includes ensuring information non-repudiation and authenticity...”	A loss of <i>integrity</i> is the unauthorized modification or destruction of information.
Availability	“Ensuring timely and reliable access to and use of information...”	A loss of <i>availability</i> is the disruption of access to or use of information or an information system.

Table 3 classified CIA

FIPS 199 defines three levels of potential impact on organizations or individuals in the event of a breach of security (i.e., a loss of confidentiality, integrity, or availability). The application of these

definitions must take place within the context of each organization. Table 2 below provides FIPS 199 potential impact definitions

3.2.0 Access control Methodologies

Access control systems work by using two basic methods of operation: *centralized* and *decentralized* access control. Most organizations actually end up using both methods in different situations, as both offer specific benefits to the overall access control system. This section examines each method and how it works.

3.2.1 Centralized

A centralized access control system is based on the concept of all access control queries being directed to a central point of authentication. The central authentication system performs the authentication and forwards the authorization data back to the requesting system. This type of system allows for a single point of administration for the entire access control system. This decreases the administrative effort, but also raises costs as each computer system using the centralized access control system must be able to communicate with the central administration point at all times.

Implementing a centralized access control system is more difficult than implementing a decentralized system, but the benefits are typically worth the extra effort. Some examples of a centralized access control system are Kerberos, RADIUS, and TACACS, which were discussed earlier in this chapter. Using a centralized access control system is usually a requirement for handling the access control needs of large enterprise systems due to the decreased administrative effort required for ongoing maintenance tasks. Making a change within the centralized system

allows for that change to be reflected on all computer systems using the access control system almost immediately.

3.2.2 Decentralized

It is not always possible or desirable to have a single reference point for all access control requests. When an access control system is configured so that multiple authentication systems are responsible for the access control requests for a small group of computer systems, it is considered to be a decentralized access control system. This basically means that the access control system is not centralized to a single computer system or group of systems. Some examples of this are a Windows workgroup where every member of the workgroup handles access control, or a database system that handles its own authentication. These systems do not rely on any other system to perform access control for them.

When working with decentralized access control systems, the individual computer systems performing access control will typically keep a local database of accounts, passwords, and permissions. All access control decisions are made based on this data. This offers the advantage of providing for access control system functionality in cases where connectivity to a centralized access control system may be impossible or intermittent.

It takes a great deal more administrative effort to work with and maintain a decentralized access control system compared to a centralized access control system. If there is a requirement for users to be able to authenticate against multiple computer systems in a decentralized access control system, the user will have to have an account on each computer system. This can easily cause an

administrative nightmare when trying to perform password resets or access control troubleshooting.

Access Control Methods should be used by all system administrators. With security in today's world, individual restrictions need to be applied to all roles and models on all forms of information used in businesses. In order to protect assets in a company or corporation, security analysts and programmers can set access control to individual components on a computer. These components can be the operating system, programs or even hardware settings. It is a necessary function that many end users often see as a burden.

Access control can be used with the following items when implementing security:

- ✚ Objects - Files or hardware settings (restriction of network settings, usb ports or individual files and folders)
- ✚ Subject - a process function (such as opening a file or folder or program) a subject can be an end user
- ✚ Operation - is the process of an end user trying to modify or delete an object

3.2.1 Access Control Models

There are four major access control models that should be embedded within applications for access control to prevent malicious users from accessing key functions within an application. These access control models are:

Discretionary Access Control (DAC): A system that uses discretionary access control allows the owner of the resource to specify which subjects can access which resources. Access control is at the discretion of the owner.

Mandatory Access Control (MAC): Access control is based on a security labeling system. Users have security clearances and resources have security labels that contain data classifications. This model is used in environments where information classification and confidentiality is very important (e.g., the military).

Non-Discretionary (Role Based) Access Control Models: Role Based Access Control (RBAC) uses a centrally administered set of controls to determine how subjects and objects interact. Is the best system for an organization that has high turnover? (Dubov, 2011).

3.3.0 Cryptography

Cryptography is simply the science of encrypting and decrypting information. Cryptography has multiple uses in a Data Center: the encryption of transactions from client to server, encryption of communication between a user and a managed device, encryption of the communication channel between two sites, and so on (Vacca, 2014).

Encryption is the process of transforming data into a format that cannot be read by anyone except the intended receiver. Encryption uses algorithms called *ciphers*, which are based on mathematical transformations applied to the original information. In this process, the original information (*plaintext*) is processed along with an encrypting key to produce a *cipher text*, or the resulting scrambled data (Cole, 2013).

Decryption is the reverse process in which the receiver obtains the original plaintext from the cipher text, which is possible with the right set of credentials or decryption keys. A key is a randomly generated string used to encrypt and decrypt information. Figure 5-10 illustrates the processes of encryption and decryption (Cao, 2013).

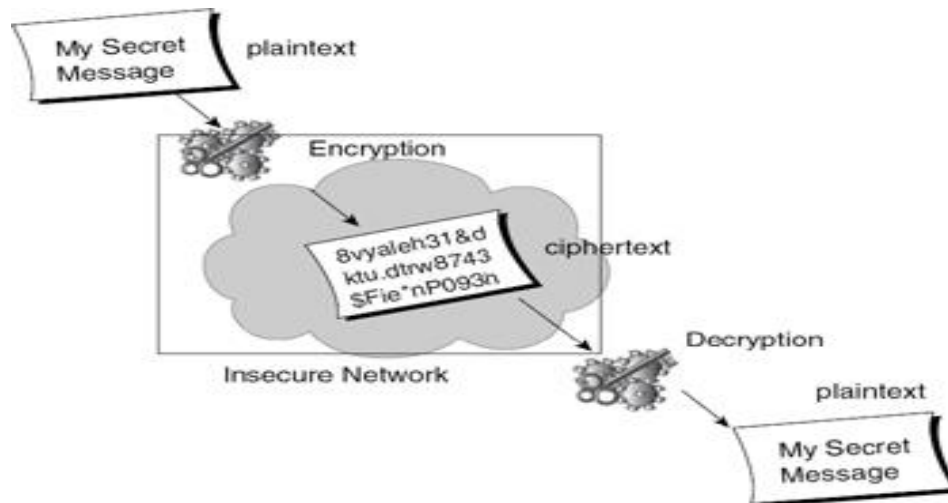


Figure 7 Encryption and Decryption Example

In Figure 8 above the message that needs to be encrypted is considered plaintext. The plaintext is encrypted using the encryption algorithm and the encryption key, producing an encrypted message called cipher text. The cipher text is then sent over the network; it remains confidential because the data cannot be seen without decrypting it and decryption cannot occur without the proper keys. Once the cipher text arrives at its intended destination, it is decrypted and presented in plaintext to the application.

Cryptography is typically associated with confidentiality but also provides for integrity, nonrepudiation, authentication, and ant replay protection (Murphy, 2015):

- + **Confidentiality**—ensures that the information cannot be read by anybody but the intended receiver. The information is encrypted in a format that cannot be understood by anyone but the entity holding the appropriate key to decrypt it.

- ✚ **Integrity**—ensures information is consistent. The encryption software signs each packet with a secret key. When the intended destination receives the packet, the signature is checked to make sure the packet has not been changed in transit and that it was signed with the right key.
- ✚ **Nonrepudiation**—Public-key cryptography provides nonrepudiation when the sender uses a private key that is only known and controlled by the sender, and only the corresponding public key can decrypt what was encrypted with the private key. Because the owner of the key is unique, you know the message was sent by that owner.
- ✚ **Authentication**—allows the sender and receiver to confirm each other's identity to make sure they are communicating with the intended party.

3.4.0 Security awareness training and education

Effective security will always be depended on people, as a result security can only be effective if employees know what is expected of them and what their responsibilities are .They should know why various security measures, such as data classification (sensitive data) and use of logon IDs, are in place and the repercussions of violating security. Promoting security awareness is a preventive control. Through this process, employee becomes aware of their responsibilities for maintaining good physical and logical security .This can be a detective measure, because it encourages people to identify and report possible security violations. Training should start with the new employee orientation or induction process .ongoing awareness can be provided in company newsletters' through visible and consistent security enforcement and short reminders during staff meetings. The security administrator should direct the program to determine the effectiveness of the program, the IS auditor should interview sample of employees to determine their overall awareness.

Managers and employees within an organization often tend to consider information security as a secondary priority if compared with their own efficiency or effectiveness matters because these have a direct and material impact on the outcome of their work. For this reason, strong leadership, direction and commitment by senior management on security training is needed. This commitment should be supported with a comprehensive program of formal security awareness training. This may require special management –level training since security is not necessarily a part of management expertise. The security training for different functions within the organization needs different functions have different levels of risk. For example, infrastructure staff needs technical security training, whereas security management requires training that will show the link between information security management and the organization.

3.5.0 Conceptual model

A conceptual model is a high level representation of how a system is organized and operates. It comprises of system inputs, processes alongside their interrelationships and the outputs. The proposed model in this research work involves the following;

This Flow Chart should be used to help determine which data classification each piece of information should be classified as. Note that it is possible for one piece of information or document to have different classifications throughout its life time. For instance commercially sensitive information may become less sensitive over time. Where one set of information contains a range of data, such as database, the highest classification should be applied to the whole set of information. Flow chart for determining Data Classification

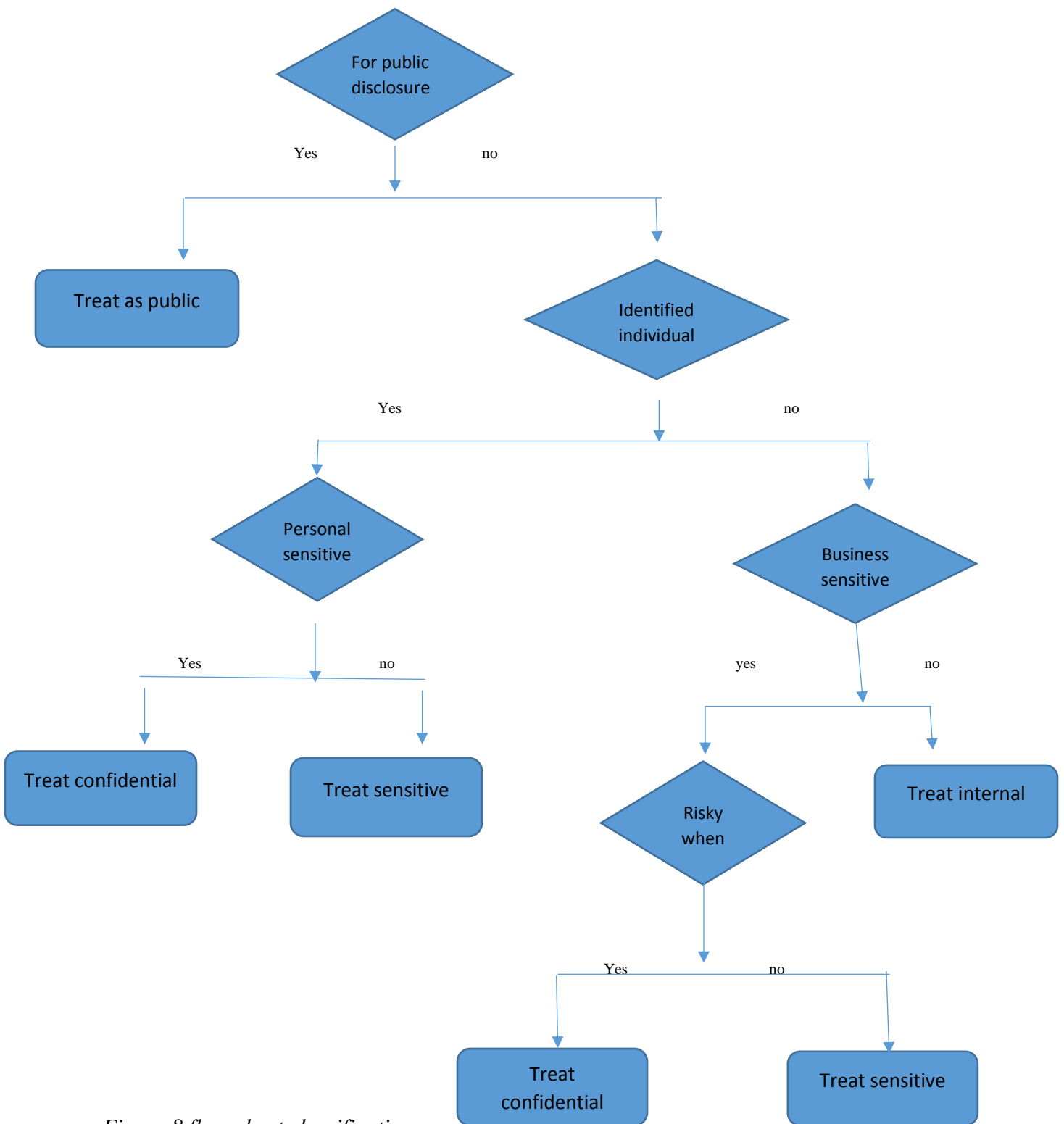


Figure 8 flow chart classification

The figure below show a well self-explained data flow diagram of how sensitive data can be monitored. When attempt is made to transmit sensitive data.

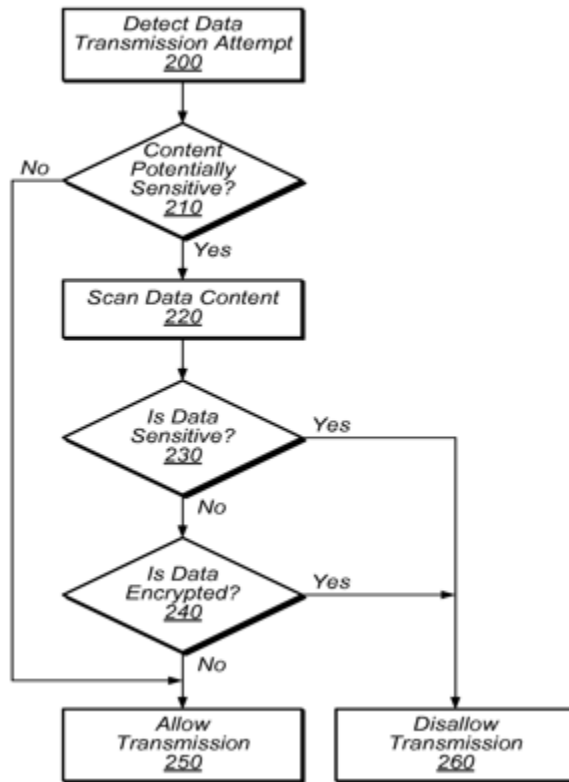


Figure 9 sensitive data transmission

4.0 RESEARCH DESIGN

The overall design of the proposed system is as show on the diagram below .where each steps is described briefly.

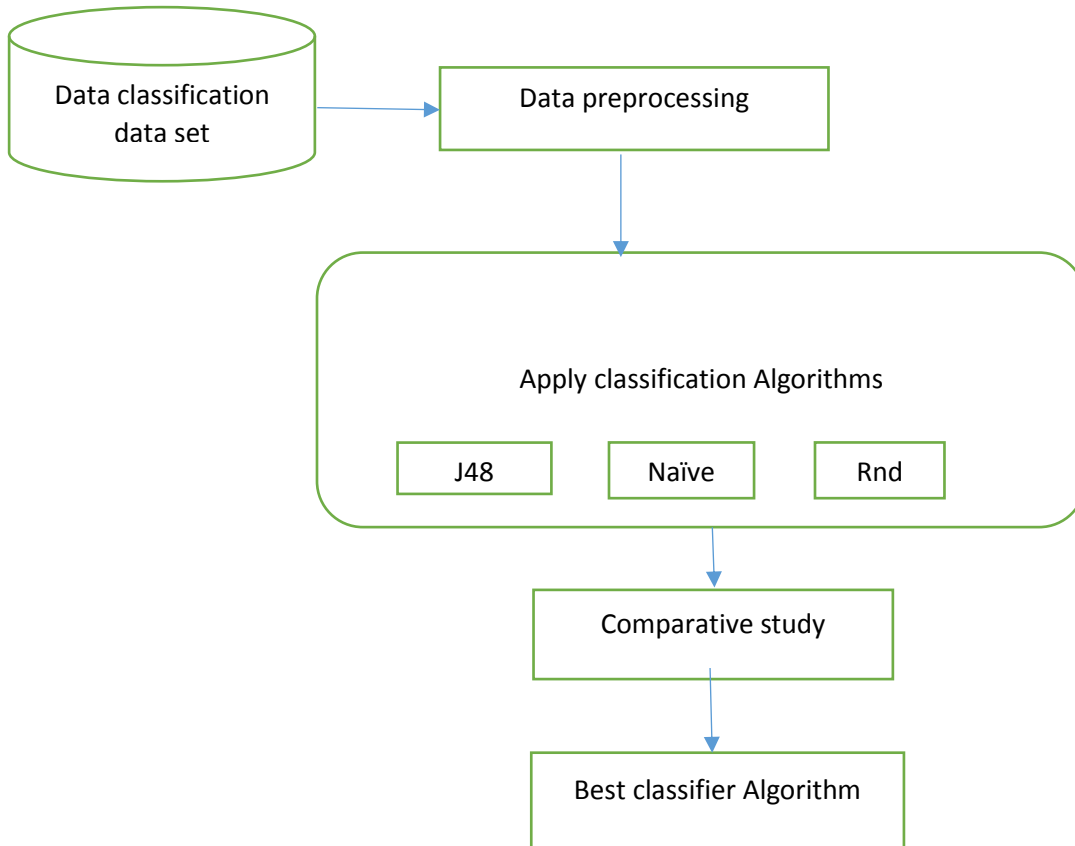


Figure 10Architecture design of the propose system

The diagram below show the processes of training the data set and validating the accuracy of classification using test data.

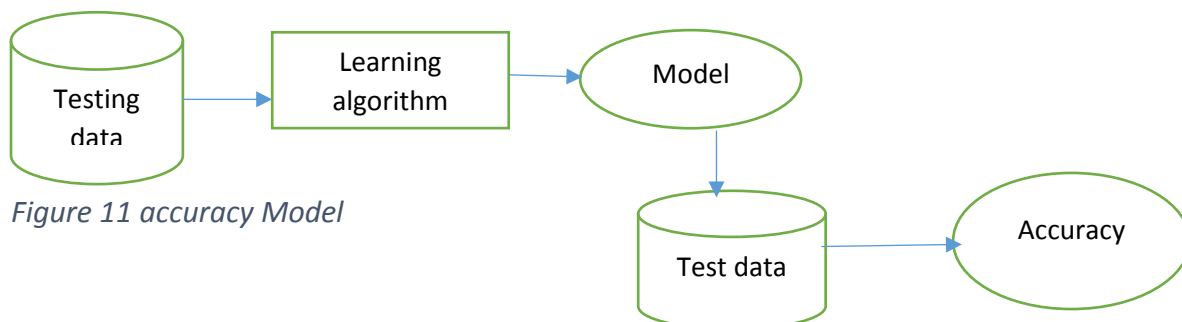


Figure 11 accuracy Model

4.1 Input data set

The data classification data set was taken from UCI machine learning repository, which consist of 6 attributes with 5 continuous input attributes and 1 discrete the attributes are given on the table below

Data Set Characteristics: Multivariate		Attribute Information:
Number of Instances: 403		SRY(define the salary of employee)
Area: Data center		HS(define Health status of employee)
Attribute Characteristics: Real		NK(define Next of Kin)
Number of Attributes: 5		DPT(define the department which the employee is)
Associated Tasks: Classification		TW(Total weight to define the document weight)
		C(define the classification level of the document)
Class Distribution		
Sensitive: 50		
Public:129		
private: 122		
Confidential 130		

4.3 Preprocessing

Data preprocessing is technique used in data mining that involve converting of raw data into a meaningful /understandable format .real data contain incomplete inconsistency or unfamiliar trends with a lot of errors hence data preprocessing is a clear method of improving it

4.4 Classification algorithm

In order to analyze the data set different classification method can be used, example we have decision tree, J48, Naïve Bayes among others.

4.4.1J48

It was developed by Ross Quinlan, is java version of C4.5, it is algorithm that can be used to generate decision tree which can be used for classification,

4.4.2Naive Bayes

Naïve Bayes are simple probabilistic classifier based on Bayes theorem with strong independence assumptions within features.

4.4.3Random Tree (RND)

Is learning method of classification regression among other task .which evolve with time

The table below show a comparison between different classifier accuracy precision and recall ,where random tree is the best classifier depending on the data set .where Naïve Bayes is best used when data set are of minimal in number.

Classifier	Accuracy	Precision	Recall
Rnd	100%	100%	100%
J48	98.8%	98.8%	98.8%
REP Tree	95.7%	96.1%	95.7%
Naive Bayes	89.5%	90.1%	89.5%

Table 4Best Classifier

5.0 RESULTS AND DISCUSSION

This chapter gives a detailed explanation of the method for data centric security. Where the class of the attributes are well defined for public, private confidential and sensitive data .classification algorithms are it explains the tools used, a detailed explanation about the implementation procedure and also how the proposed model works.

WEKA

The workflow of WEKA would be as follows:

- ✚ Data →Preprocessing →Data Mining →Knowledge
- ✚ The supported data formats are ARFF, CSV, C4.5 and binary. Alternatively you could also import from URL or an SQL database.
- ✚ After loading the data, preprocessing filters could be used for adding/removing attributes, discretization, Sampling, randomizing etc.

5.1Why Weka

Weka is a machine learning software suite developed in Java. It provides facilities for all the steps involved in solving a machine learning problem- data conversion, preprocessing techniques, classification, categorization and visualization. Weka commands can be carried out via the command line. It also provides a GUI on supporting systems, which makes it very easy in terms of understanding the flow of data, and visualizing the results. However, the GUI component does not work well when dealing with very large data, as was the case in this project. On the other hand, the Weka Simple Command Line Interface (CLI) is lightweight in terms of memory, and provides much more scope for dealing with very large files. Out of all the other machine learning libraries, Weka was the most cited and recommended. It was the most well

documented, and also since it was written in Java, understanding and customization became more convenient

The diagram show database which contain folder which is divided into sub folder and files .hence this folder contain files which are either sensitive ,confidential ,private or public .therefore there is need to classify them. Each folder contain certain weight depending on the level of classification.categories is from zero to one. Zero is less sensitive .Total weight is got from averaging the total of the file contain in the folder

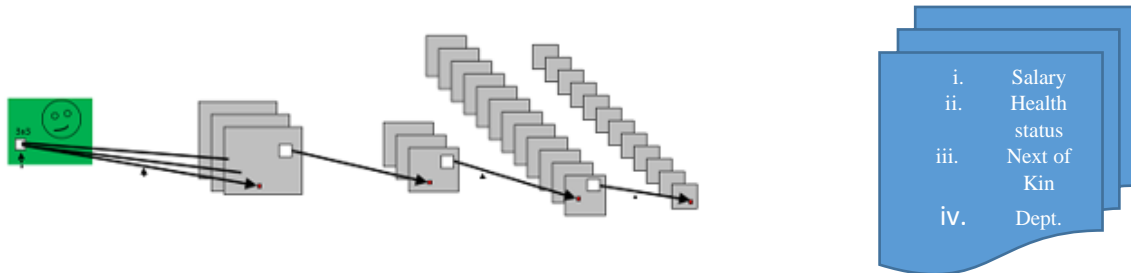


Figure 12 folder identification level

5.2 Identification of the Attributes

SRY	HS	NK	DPT	TW	C
0	0.1	0.5	0.26	0.05	Public
0.05	0.05	0.55	0.6	0.14	private
0.08	0.18	0.63	0.6	0.85	Sensitive
0.2	0.2	0.68	0.67	0.85	Sensitive
0.22	0.22	0.9	0.3	0.9	Sensitive
0.14	0.14	0.7	0.5	0.3	private
0.16	0.16	0.8	0.5	0.5	Confidential
0.12	0.12	0.75	0.68	0.15	private
0.2	0.2	0.88	0.77	0.8	Sensitive
0.16	0.25	0.01	0.1	0.07	Public
0.11	0.29	0.2	0.05	0.66	Confidential
0.18	0.26	0.05	0.4	0.04	Public
0.21	0.32	0.25	0.5	0.8	Sensitive
0.13	0.28	0.18	0.75	0.32	Confidential
0.23	0.29	0.45	0.18	0.88	Sensitive
0.1	0.27	0.35	0.45	0.05	Public

Attribute Information:
SRY (define the salary of employee)
HS (define Health status of employee)
NK (define Next of Kin)
DPT (define the department which the employee is)
TW (Total weight to define the document weight)
C (define the classification level of the document)

Table 5 classification attributes

5.3 Data classification process

Instructions for the CSV to ARFF conversion tool

The online csv2arff tool converts CSV formatted files into ARFF formatted files.

After an experiment data are usually merged into an excel table. From there it can be exported to the CSV (comma separated value) format. In order to use the WEKA data mining software the CSV file needs to be converted to the Weka's ARFF format. Workflow after the input data is properly prepared as a CSV file, the conversion is done in 2 steps. Preparing the CSV file

The CSV file must contain the names of the attributes in the first line. All other lines should contain the data taken from your measurements.

Example: data from a color assignment experiment are stored in Microsoft Excel (or any other spreadsheet software) as a table (click here for the source file): The content of the correct CSV source file looks like this:

id;sensitive;private;public;classification

1;password;internally used; open to all; sensitive

2;encrypted;no encryption; open to all;sensitive

3;high risk; medium risk; low risk; sensitive

4;high risk; medium risk; low risk; private

5;high risk; medium risk; low risk; public

Note: the csv parser allows the use of comma "," or semicolon ";" to delimit the values.

Step 1: uploading of the CSV file

The CSV file is uploaded via the web interface. Before uploading make sure the CSV file meets the format requirements.

Step 2: Define the attributes' types

In the second step choose the attributes you want to include in the ARFF file and their type (numeric or nominal).

Then click the submit button. The server should return you a valid ARFF file that looks like this:

@relation whatever

@attribute public numeric

@attribute private numeric

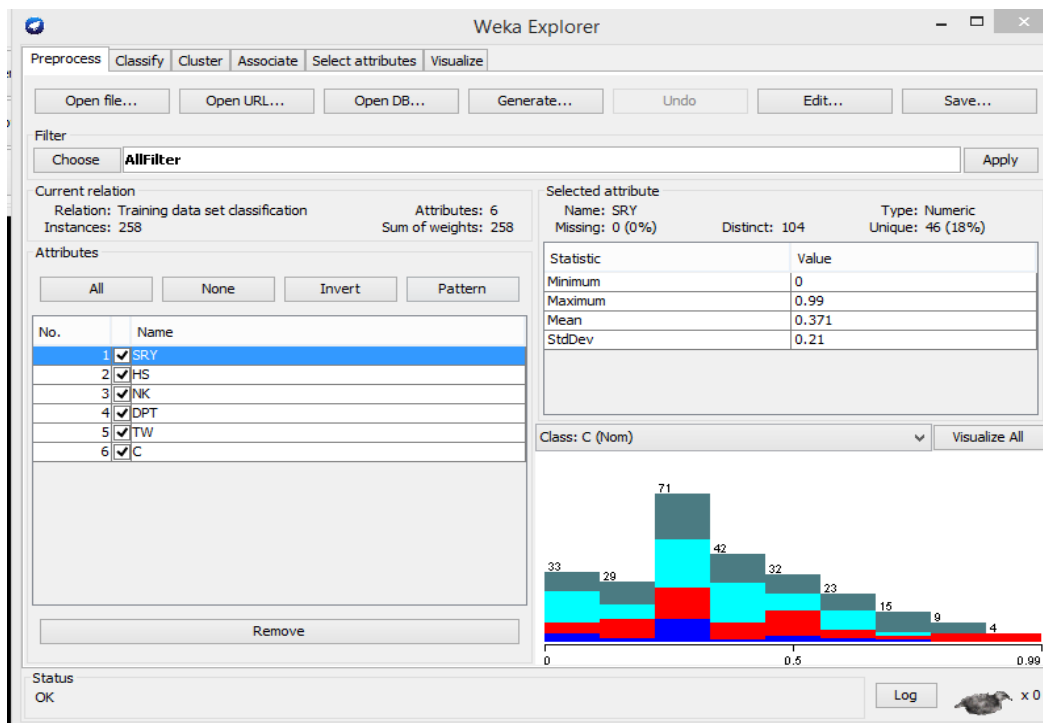
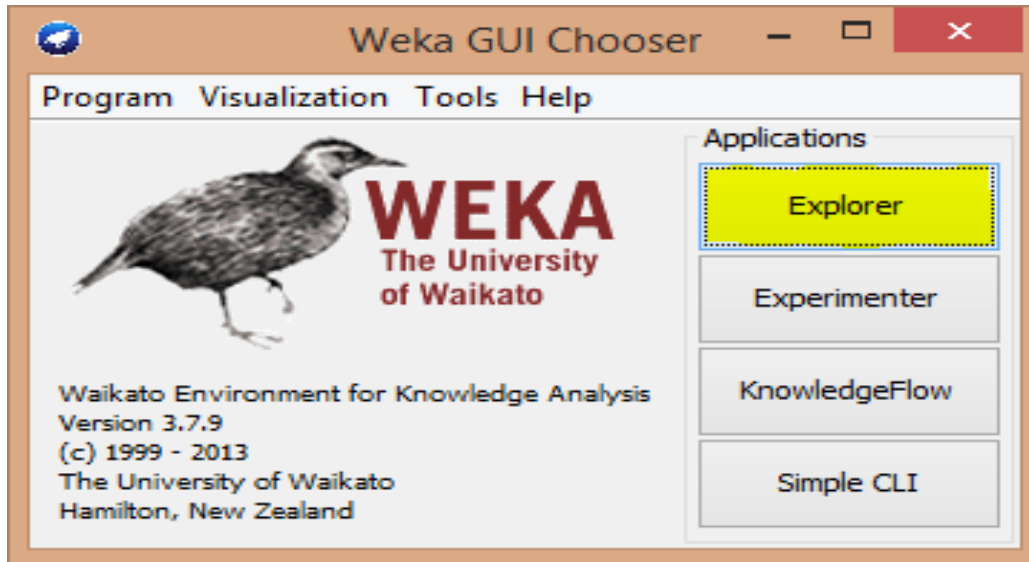
@attribute sensitive numeric

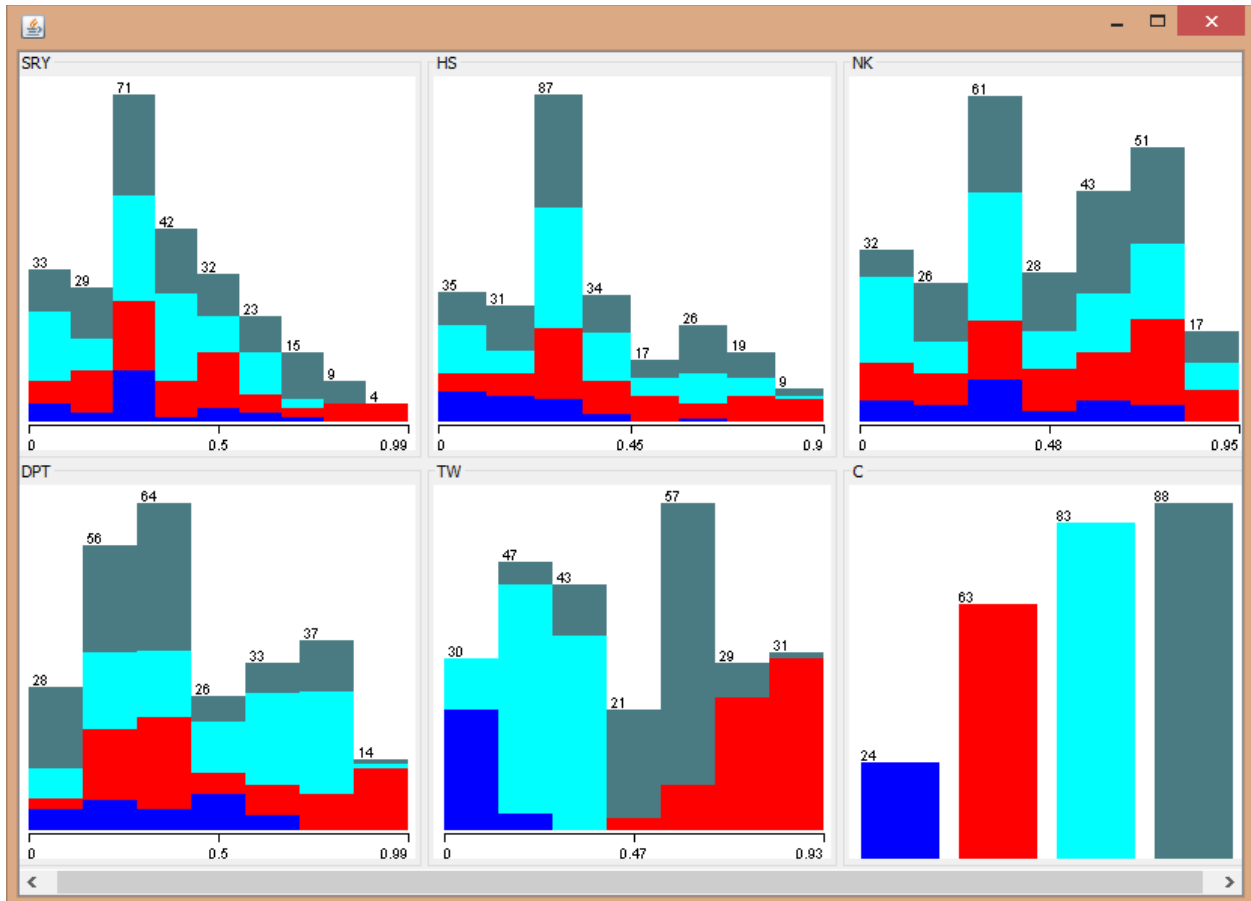
@attribute classification {public, private, sensitive}

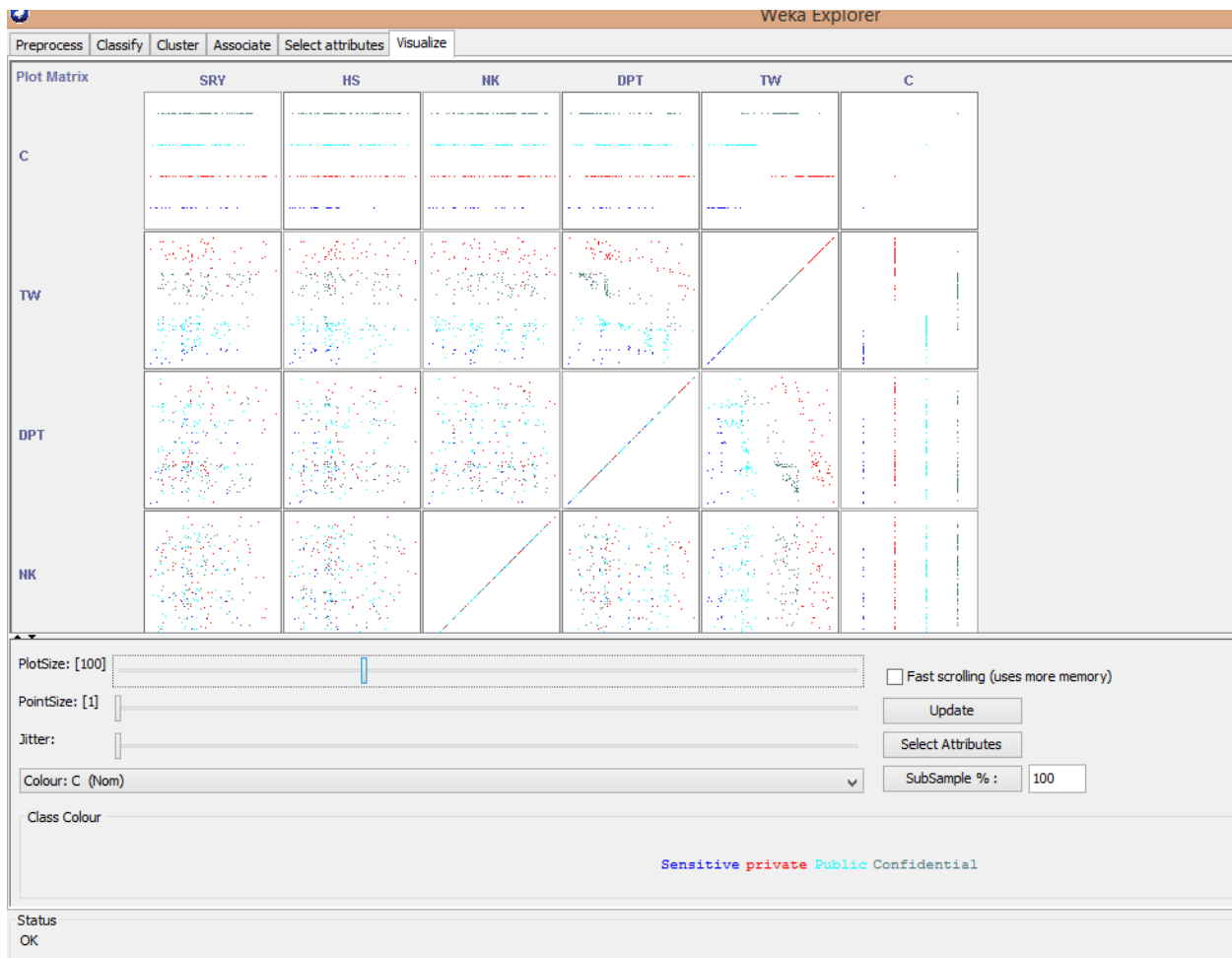
@data

5.4 Results

The result show below are as result of classification of test data set that were run along the training data set (Witten, n.d.).







=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: Training data set classification

Instances: 258

Attributes: 6

SRY

HS

NK

DPT

TW

C

Test mode: user supplied test set: size unknown (reading incrementally)

==== Classifier model (full training set) ====

J48 pruned tree

TW <= 0.35

| TW <= 0.13

| | DPT <= 0.62: Sensitive (19.0)

| | DPT > 0.62

| | | TW <= 0.09

| | | | SRY <= 0.17: Public (2.0)

| | | | SRY > 0.17: Sensitive (2.0)

| | | TW > 0.09: Public (7.0)

| TW > 0.13

| | DPT <= 0.6

| | | TW <= 0.24

| | | | HS <= 0.255: Sensitive (3.0)

| | | | HS > 0.255: Public (4.0)

| | | TW > 0.24: Public (43.0)

| | DPT > 0.6

| | | TW <= 0.24: Public (21.0)

| | | TW > 0.24

| | | | TW <= 0.27

| | | | | DPT <= 0.66: Confidential (3.0)

| | | | | DPT > 0.66: Public (7.0/1.0)

| | | | TW > 0.27: Confidential (9.0)

TW > 0.35

| TW <= 0.67

| | DPT <= 0.83: Confidential (75.0/1.0)
| | DPT > 0.83: private (10.0)
| TW > 0.67: private (53.0/1.0)

Number of Leaves: 14

Size of the tree: 27

Time taken to build model: 0.05 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.03 seconds

=== Summary ===

Correctly Classified Instances	255	98.8372 %
Incorrectly Classified Instances	3	1.1628 %
Kappa statistic	0.9837	
Mean absolute error	0.0109	
Root mean squared error	0.074	
Relative absolute error	3.0736 %	
Root relative squared error	17.5389 %	
Coverage of cases (0.95 level)	99.2248 %	
Mean rel. region size (0.95 level)	25.6783 %	
Total Number of Instances	258	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Sensitive
	0.984	0.005	0.984	0.984	0.984	0.979	0.995	0.978	private
	1.000	0.006	0.988	1.000	0.994	0.991	1.000	0.999	Public
	0.977	0.006	0.989	0.977	0.983	0.974	0.995	0.985	Confidential
Weighted Avg.	0.988	0.005	0.988	0.988	0.988	0.983	0.997	0.989	

=== Confusion Matrix ===

```

a b c d <-- classified as
24 0 0 0 | a = Sensitive
0 62 0 1 | b = private
0 0 83 0 | c = Public
0 1 1 86 | d = Confidential

```

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: Training data set classification

Instances: 258

Attributes: 6

SRY

HS

NK

DPT

TW

C

Test mode: user supplied test set: size unknown (reading incrementally)

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class			
	Sensitive (0.1)	private (0.24)	Public (0.32)	Confidential (0.34)

SRY

mean	0.3048	0.4223	0.321	0.4002
std. dev.	0.183	0.238	0.1731	0.2129
weight sum	24	63	83	88
precision	0.0096	0.0096	0.0096	0.0096

HS

mean	0.199	0.4227	0.3368	0.3677
std. dev.	0.1327	0.2316	0.1912	0.2063
weight sum	24	63	83	88
precision	0.0102	0.0102	0.0102	0.0102

NK

mean	0.3649	0.5015	0.431	0.5066
std. dev.	0.2122	0.2575	0.2499	0.2271
weight sum	24	63	83	88
precision	0.0116	0.0116	0.0116	0.0116

DPT

mean	0.3592	0.5013	0.4969	0.3432
std. dev.	0.1933	0.2746	0.2234	0.2265
weight sum	24	63	83	88
precision	0.0125	0.0125	0.0125	0.0125

TW

mean	0.0893	0.7721	0.2366	0.5419
std. dev.	0.0544	0.1068	0.0727	0.1262
weight sum	24	63	83	88
precision	0.0118	0.0118	0.0118	0.0118

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.04 seconds

=== Summary ===

Correctly Classified Instances	231	89.5349 %
Incorrectly Classified Instances	27	10.4651 %
Kappa statistic	0.853	
Mean absolute error	0.1032	
Root mean squared error	0.2225	
Relative absolute error	28.9751 %	
Root relative squared error	52.7442 %	
Coverage of cases (0.95 level)	96.8992 %	
Mean rel. region size (0.95 level)	42.7326 %	
Total Number of Instances	258	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.875	0.017	0.840	0.875	0.857	0.842	0.993	0.951	Sensitive
	0.905	0.005	0.983	0.905	0.942	0.926	0.993	0.969	private
	0.952	0.091	0.832	0.952	0.888	0.833	0.948	0.838	Public
	0.841	0.035	0.925	0.841	0.881	0.826	0.918	0.915	Confidential
Weighted Avg.	0.895	0.044	0.901	0.895	0.896	0.854	0.953	0.907	

==== Confusion Matrix ====

```

a b c d <-- classified as
21 0 3 0 | a = Sensitive
0 57 0 6 | b = private
4 0 79 0 | c = Public
0 1 13 74 | d = Confidential

```

The below data show how test data was 89.5% correctly classified from the training data set and 10.5% incorrectly classified with 237 instances and 27 instances irrespectively .

Accuracy

The Accuracy of a classifier was defined as the percentage of the dataset correctly classified by the method.

$$Accuracy = \frac{\text{No. of correctly classified samples}}{\text{Total no. of samples in the class}}$$

Recall

Recall of the classifier was defined as the percentage of errors correctly predicted out of all the errors that actually occurred.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Precision

Precision of the classifier was defined as the percentage of the actual errors among all the encounters that were classified as errors.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

```
=== Evaluation on test set ===
Time taken to test model on supplied test set: 0.02 seconds

=== Summary ===
Correctly Classified Instances      231          89.5349 %
Incorrectly Classified Instances    27           10.4651 %
Kappa statistic                     0.853
Mean absolute error                 0.1032
Root mean squared error            0.2225
Relative absolute error             28.9751 %
Root relative squared error        52.7442 %
Coverage of cases (0.95 level)     96.8992 %
Mean rel. region size (0.95 level) 42.7326 %
Total Number of Instances          258

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.875	0.017	0.840	0.875	0.857	0.842	0.993	0.951	Sensitive
	0.905	0.005	0.983	0.905	0.942	0.926	0.993	0.969	private
	0.952	0.091	0.832	0.952	0.888	0.833	0.948	0.838	Public
	0.841	0.035	0.925	0.841	0.881	0.826	0.918	0.915	Confidential
Weighted Avg.	0.895	0.044	0.901	0.895	0.896	0.854	0.953	0.907	

The confusion matrix below show how data was classified,

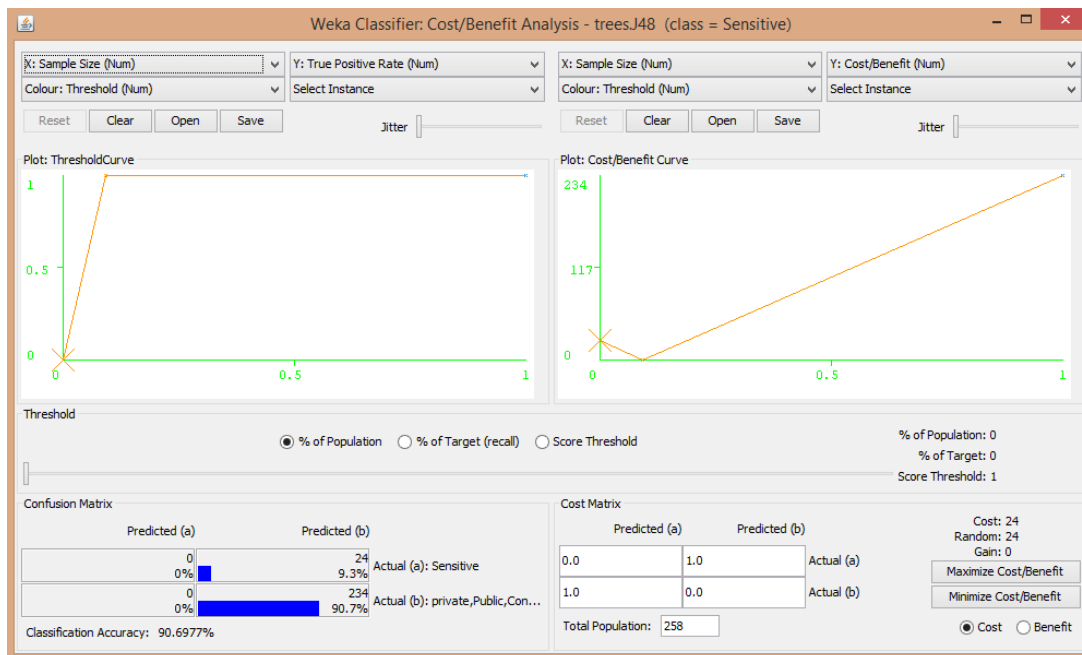
- i. From below data 21 sensitive data was correctly classified 3 incorrectly classified
- ii. On private data 57 was correctly classified and 6 incorrectly classified
- iii. Confidential data was more confused classification where it was classified private 1, public 13 and confidential 74

```

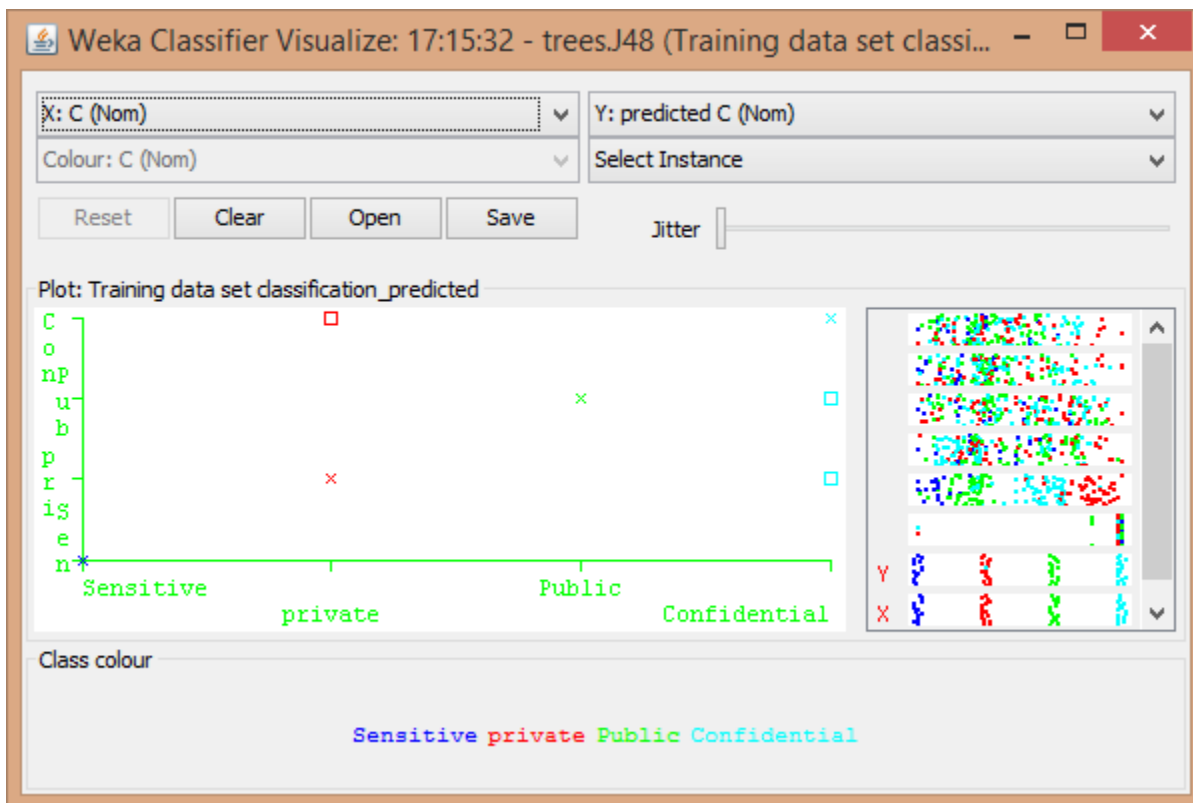
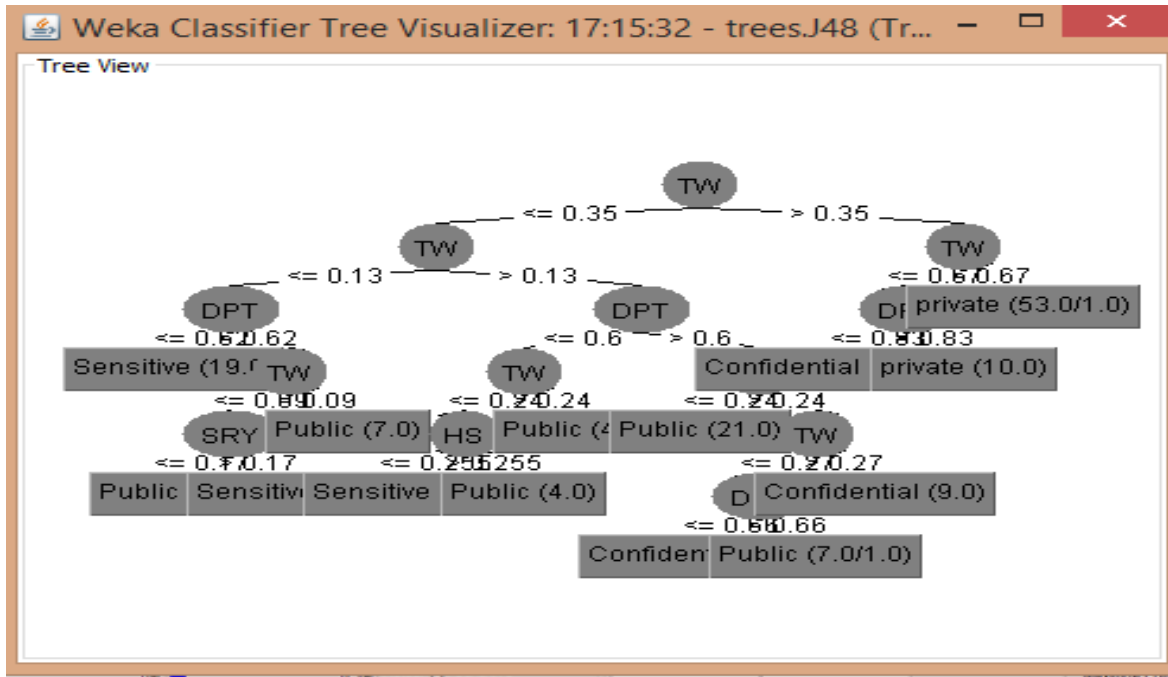
=== Confusion Matrix ===

  a  b  c  d  <-- classified as
21  0  3  0 | a = Sensitive
 0 57  0  6 | b = private
 4  0 79  0 | c = Public
 0  1 13 74 | d = Confidential
    
```

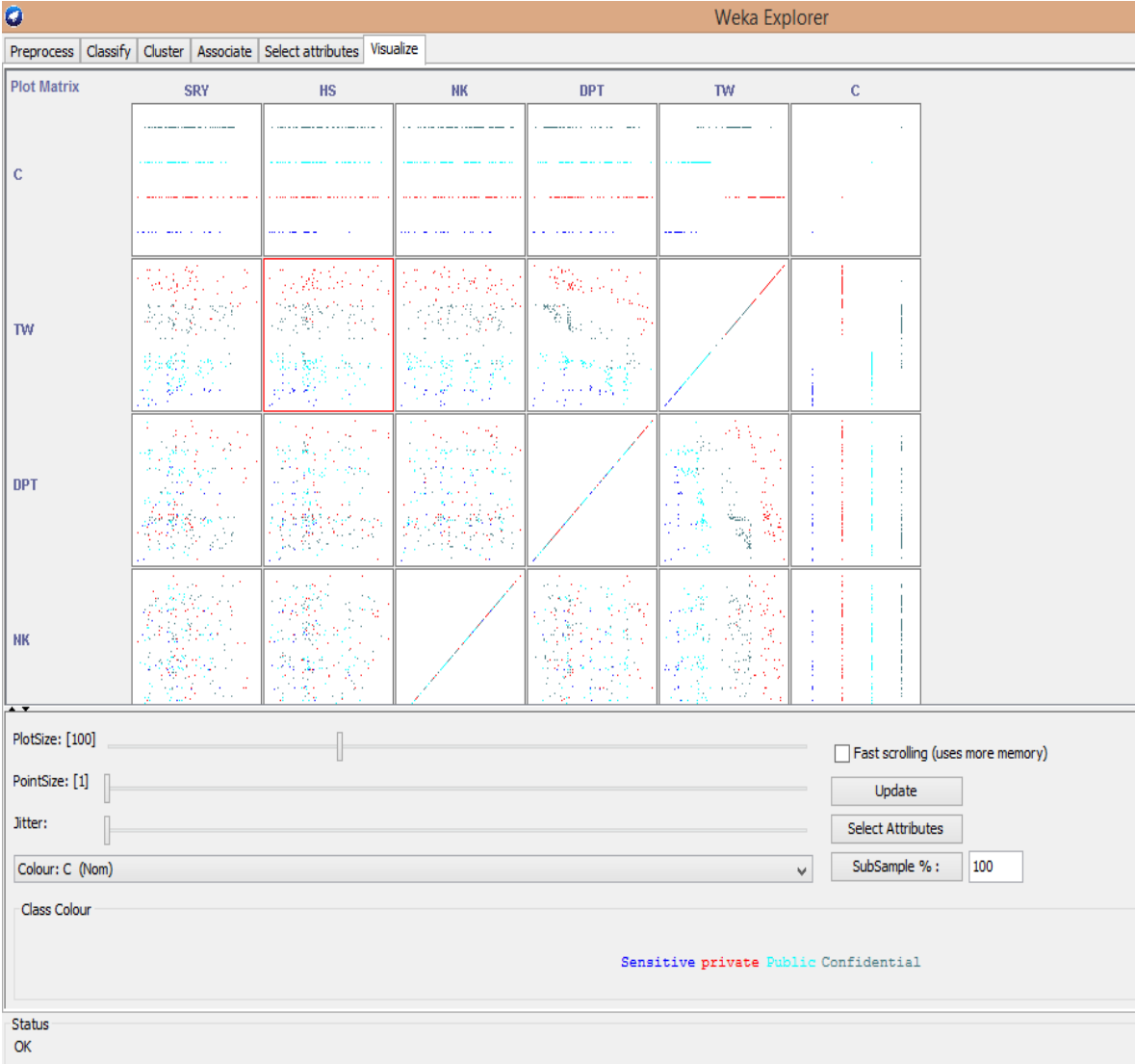
Average Cost/Benefit parameters are observed. The Cost is rather low 24 the Gain is rather low 0 it is sufficient to select only 0% of compounds in order to retrieve 0 % of active



The graphical representation of the tree how weka classify the data .it show in visualization



The cross validation of the training data set and test data set is show below



5.5 Conclusion of the results

The data classification level prediction, help to identify learning curve of the classification level of our store data in order to provide security on our more valuable data assets also provide knowledge on our store data.

5.3 Data loss prevention and encryption process after data classification

It is very important for the organization to identify the sensitive and confidential data in order to provide security on the classified data, after data classification it is important to provide data loss prevention and encryption of data .the diagram below show diagram on where to implement DLP and encryption process (Randy Devlin, 2015).

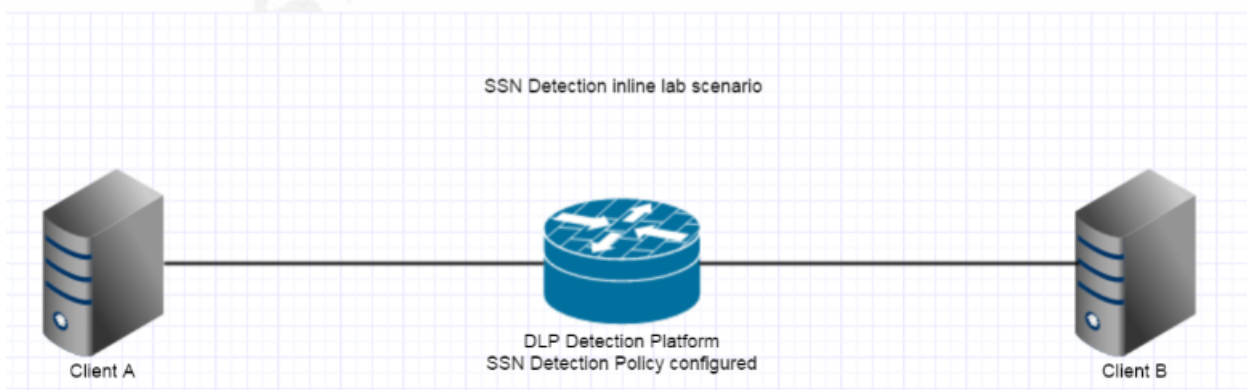
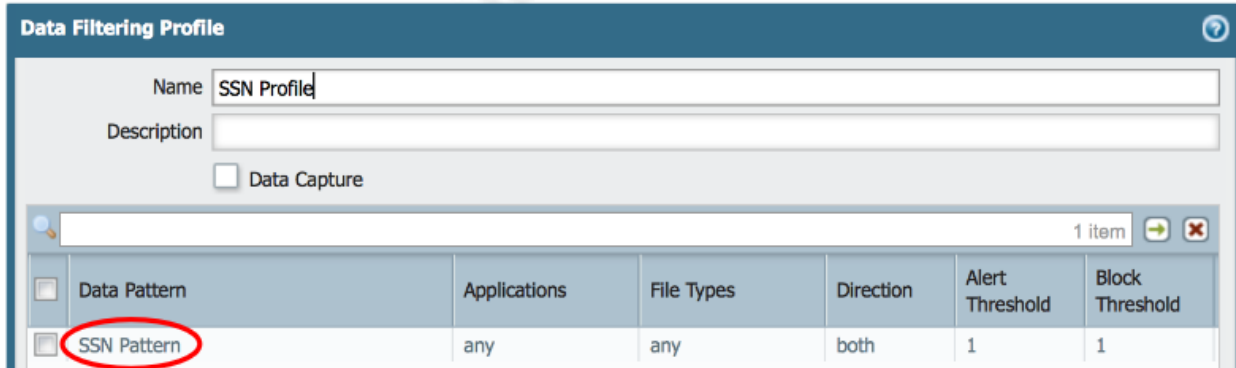


Figure 13 SSN

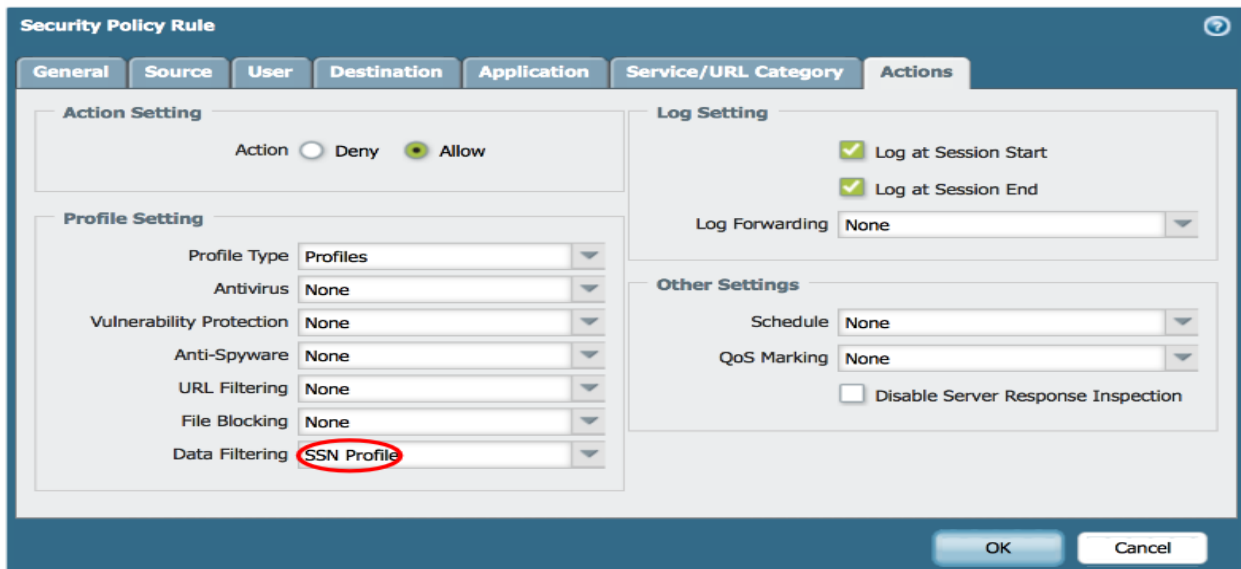
Data pattern definition ,threshold is set to respond when target number exceeded.

The screenshot shows a "Data Patterns" configuration window. It includes a "Name" field with the value "SSN Pattern", an empty "Description" field, and a "Weight (0 - 255)" section. Under the weight section, there are three input fields: "CC#" with the value "0", "SSN#" with the value "1" (circled in red), and "SSN# (without dash)" with the value "0".

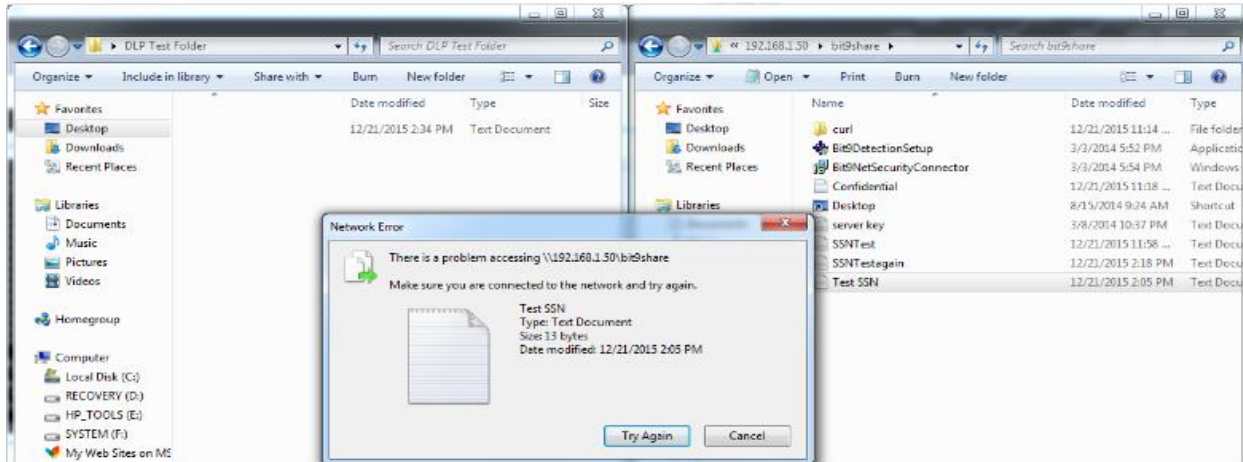
Specific data pattern is inspected by data filtering profile.



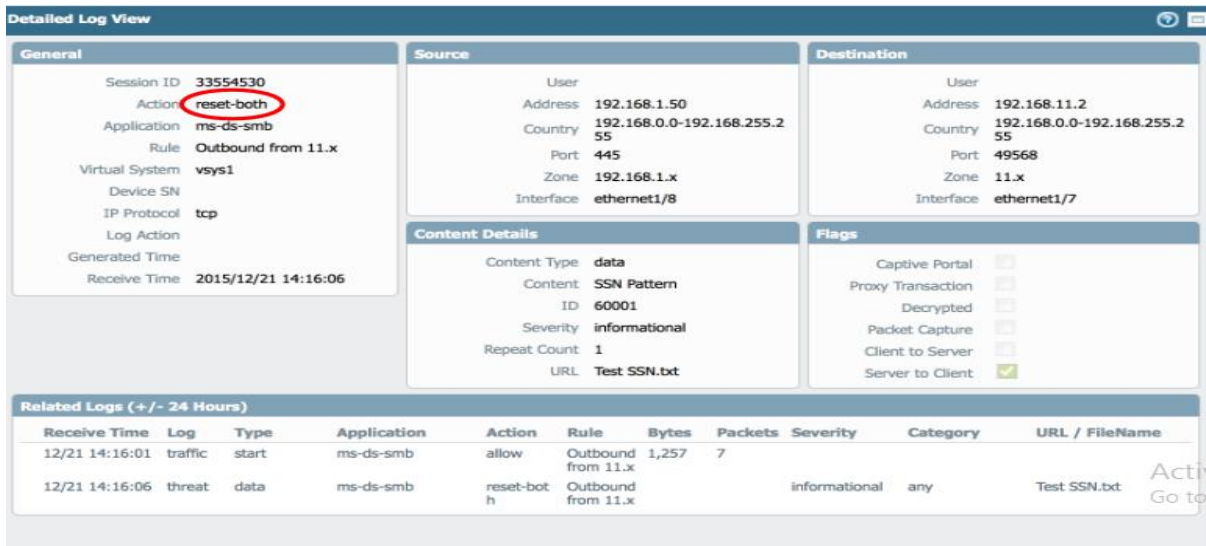
The data filtering profile is then assigned to a security policy rule. Which will allow or deny content specified



The sample below show data that has been set not to be copied by user hence denying user even the rights to copy.



The below diagram give log view snapshot on source and destination IP of the user who attempted to copy sensitive data and was denied permission.



When DLP is properly selected configured and activated it can offer both data loss prevention and encryption of sensitive data (Scarfone, 2013). By offering an integrated Data Loss Prevention solution and data Encryption service, the overall solution, lower the cost, and eliminate the complexity of secure data delivery.

Add users and encrypt


Add Users

Email:

Rights:

User list

Email Address	Rights
*gmukami@nokenya.co.ke	Owner
getrudemukami@gmail.com	Viewer



6.0 CONCLUSION AND FUTURE WORKS

6.1 Conclusion

Information classification is necessary to better manage information. If implemented correctly, classification can reduce the cost of protecting information because in today's environment, the "one size fits all" idea will no longer work within the complexity of most corporation's heterogeneous platforms that make up the I/T infrastructure. Information

Classification enhances the probability that controls will be placed on the data where they are needed the most, and not applied where they are not needed. Classification security schemes enhance the usability of data by ensuring the confidentiality, integrity, and availability of information. By implementing a corporate wide information classification program, good business practices are enhanced by providing a secure, cost-effective data platform that supports the company's business objectives. The key to the successful implementation of the information classification process is senior management support. The corporate information security policy should lay the groundwork for the classification process and be the first step in obtaining management support and buy-in.

All employees of an organization and where relevant third party users should receive appropriate training and regular updates to foster security awareness and compliance with written security policies and procedures. From new employees, this training should occur before access to information or services is granted. A number of different mechanisms available for raising security awareness include;

- ✚ Regular updates to written security policies and procedures.
- ✚ Formal information security training

- ✚ Statements signed by employees and contractors agreeing to follow the written security policy and procedures, including non-disclosure obligations
- ✚ Visible enforcement of security rules and periodic audit

6.2Future works

Once data classification processes have been implemented, the ongoing monitoring processes should be implemented. The internal audit department should ensure compliance established of procedures and work instruction. Information Security, working with selected information owners, Legal, and other interested parties, should periodically review the information classifications themselves to ensure they still meet business requirements. Access rights of individuals should be periodically reviewed to ensure these rights are still appropriate. The controls associated with each classification should also be reviewed to ensure they are still appropriate for the classification they define

The practical real life example on how data classification can work is applied to fraud detection is to create a classification model that can label a record, person or company as being fraudulent or not. The model will be created by analysing a dataset with records classified as fraudulent and non-fraudulent. This process goes on to locate sets of functions that may describe and distinguish those Records that seem to be inappropriate in an efficient manner. After creating the model, one may use it to classify future and unknown data as fraudulent or non-fraudulent, permitting the user to acknowledge which persons or entities are perpetuating fraud hence reducing the fraud.

REFERENCES

1. Calder, A. & Watkins, S., 2012. *IT Governance: An International Guide to Data Security and ISO27001/ ISO27002*. Fifth Edition ed. s.l.:KoganPage.
2. Cao, Z., 2013. *New Directions of Modern Cryptography*. s.l.:s.n.
3. Cole, E., 2009. *Network Security Bible*. 2 Edition ed. s.l.:s.n.
4. Cole, E., 2013. *Advanced Persistent Threat: Understanding the Danger and How to Protect Your Organization*. s.l.:s.n.
5. Cole, E., 2013. *Advanced Persistent Threat: Understanding the Danger and How to Protect Your Organization*. s.l.:s.n.
6. Dubov, A. B. a. L., 2011. *Master Data Management and Data Governance*. 2 Edition ed. s.l.:s.n.
7. Harris, S., 2013. *CISSP All-in-One Exam Guide*. s.l.:s.n.
8. Krause, H. F. T. a. M., 2007. *Information Security Management Handbook*. Sixth Edition ed. s.l.:s.n.
9. Mike Chapple, B. B. T. B. a. E. K. B., 2014. *Access Control, Authentication, and Public Key Infrastructure*. 2 Edition ed. s.l.:s.n.
10. Murphy, S. P., 2015. *Healthcare Information Security and Privacy*. s.l.:s.n.
11. Randy Devlin, 2015. *Data Loss Prevention*. [Online]
Available at: <https://www.sans.org/reading-room/whitepapers/dlp/data-loss-prevention-37152>
[Accessed 15th october 2016].
12. Scarfone, K., 2013. *statetechmagazine*. [Online]
Available at: <http://www.statetechmagazine.com/article>
[Accessed 15th october 2016].
13. Seidl, M. C. a. D., 2015. *Cyberwarfare: Information Operations in a Connected World*. s.l.:s.n.
14. Seidl, M. C. a. D., 2015. *Cyberwarfare: Information Operations in a Connected World*. s.l.:s.n.
15. Seymour Bosworth, M. E. K. a. E. W. (., 2014. *Computer Security Handbook*,. 6 Edition ed. s.l.:s.n.
16. Vacca, J. R., 2014. *Cyber Security and IT Infrastructure Protection*. s.l.:s.n.
17. Watkins, A. C. a. S., 2012. *IT Governance: An International Guide to Data Security and ISO27001/ ISO27002*. 5 Edition ed. s.l.:s.n.

18 <https://digitalguardian.com/resources/data-security-knowledge-base/data-classification>

19 http://www.ijbhtnet.com/journals/Vol_4_No_1_January_2014/6.pdf

20 <http://www.pearsonitcertification.com/articles/article.aspx?p=30287&seqNum=9>

21 <https://www.dhs.gov/sites/default/files/publications/FY%202016%20CIO%20FISMA%20Metrics%20v1.0.pdf>

22 <http://www.maravis.com/library/data-classification-guidelines/>

23 https://security.vt.edu/resources_and_information/veracrypt.html

24 <https://nakedsecurity.sophos.com/2010/07/19/ten-tips-for-protecting-sensitive-data-in-your-organisation/>

25 https://security.vt.edu/resources_and_information/veracrypt.html