

# A PREDICTIVE MODEL FOR IDENTIFYING THE SECURITY HEALTH STATE OF IT ASSETS

Mohamed Yerou

VU Amsterdam  
Master Business Analytics

Amsterdam, 31 January 2014

# Title page

**Author** Mohamed Yerou  
**Student number** myu300  
**E-mail** m.yerou@student.vu.nl  
**Academic institution** Vrije Universiteit Amsterdam

**Supervisor VU**  
Herbert Bos  
Mark Hoogendoorn

**Supervisor Podictive**  
Aziz Ahrouch

**Vrije Universiteit Amsterdam**  
Faculty of Sciences  
Master Business Analytics  
De Boelelaan 1081a 1081 HV Amsterdam

**Podictive B.V.**  
IT Security & Advisory Services  
Notweg 38-a  
1068 LL Amsterdam

[www.vu.nl](http://www.vu.nl)

[www.podictive.com](http://www.podictive.com)



**Internship subject** A predictive model for identifying the security health state of IT assets

**Internship period** May 2013- January 2014

VU Amsterdam  
Master Business Analytics

Amsterdam, 31 January 2014

## Preface

This internship report was written as a part of the Business Analytics Master program at the Vrije Universiteit in Amsterdam. The goal of the internship is to encompass the program components Business, Mathematics and Informatics in practice at a business, industry or research facility that has its common ground in Business Analytics. This internship report contains the results of my research performed at Podictive, an IT company that is actively involved in researching and developing new services whereby the use of business analytics to innovate its IT Security services is one of its research topics.

I would like to thank Aziz Ahrouch for giving me the opportunity to undertake my internship at Podictive and for providing me the support and guidance to accomplish my research. Furthermore, I would like to thank Herbert Bos and Mark Hoogendoorn from the Vrije Universiteit for their supervising role.

Mohamed Yerou

Amsterdam, January 2014

# Contents

	<b>1</b>
<b>Title page</b>	<b>2</b>
<b>Preface</b>	<b>3</b>
<b>1. Introduction</b>	<b>5</b>
1.1 <i>Internship program</i>	5
1.2 <i>Research Company</i>	5
<b>2. Research subject</b>	<b>7</b>
2.1 <i>Understanding the problem</i>	7
2.2 <i>Research question</i>	7
2.3 <i>Research approach</i>	9
2.4 <i>Report outline</i>	9
<b>3. Definition study</b>	<b>11</b>
3.1 <i>IT Security stakeholders</i>	11
3.2 <i>Goals and objectives</i>	11
3.3 <i>Conceptualizing the health state</i>	12
3.4 <i>Defining the security health state of an IT asset</i>	21
<b>4. Security metrics in a process approach</b>	<b>24</b>
4.1 <i>Identification of goals and objectives of the metric</i>	25
4.2 <i>Context analysis</i>	26
4.3 <i>Asking Questions and assign metrics</i>	31
4.4 <i>Determine which security metrics to use</i>	32
4.5 <i>Improvement process of the identified security metrics</i>	33
<b>5. Applying the metrics process approach</b>	<b>34</b>
5.1 <i>Security goals</i>	34
5.2 <i>Stakeholders identification</i>	35
5.3 <i>Patch and Vulnerability Management Context analysis</i>	36
5.4 <i>Requirements analysis</i>	42
5.5 <i>Scope determination</i>	43
5.6 <i>Asking Questions</i>	43
<b>6. Modeling the security health state of IT assets</b>	<b>45</b>
6.1 <i>Modeling and Evaluation</i>	45
6.2 <i>Models description</i>	45
6.3 <i>Data manipulation</i>	48
6.4 <i>Models application</i>	61
<b>7. Conclusion and recommendations</b>	<b>65</b>
7.1 <i>Conclusion</i>	65
7.2 <i>Recommendations</i>	66
<b>Appendix I References</b>	<b>39</b>
<b>AppendixII Questionnaire</b>	<b>70</b>
<b>AppendixIII Weka results</b>	<b>73</b>

# 1. Introduction

## 1.1 Internship program

As part of the Master program Business Analytics (BA, students need to perform a six month internship. Business Analytics is a multidisciplinary program, encompassing mathematics, computer science and business management, aimed to improve business processes from the perspective of utilizing information analytics. Goal of the internship is to apply all these three disciplines on a problem defined in accordance with the company where the internship takes place. The internship provides students the opportunity to apply theoretical knowledge from the Master program in practice and at the same time gain working experience. This thesis is written during an internship at Podictive from June 2013 to December 2013.

## 1.2 Research Company

Podictive is an IT security company delivering Advisory and Security services where corporate entrepreneurship and knowledge are keys in bringing success to customers. Podictive support its clients to gain control and become more secured by implementing security measurements to protect the confidentiality, integrity, and availability of customer's data.

Podictive is a specialized and relatively young organization with the ambition to develop itself to what the market needs and what they want to be. Very aware of the fact that employees of Podictive can make the positive difference in any organization, Podictive has experience (with proven successful implementations) at several leading multinationals. Business Analytics is one of the trending topics that Podictive drives to explore a niche market in IT Security. Customers of Podictive can be described as strategic clientele with large scale, complex and custom infrastructure needs in the finance, energy and telecom sector.

### 1.2.1 Mission

Podictive has set its mission to build trusted customer relationships by delivering forward thinking, high-end infrastructure and information intelligence consulting services with talented people from a variety of academic, professional and ethnic backgrounds.

Podictive understands that people, process and technology need to be in harmony with each other in order to have a secure and compliant IT environment against acceptable costs. To achieve this, Podictive invests in expert training and technologies to deliver the right people on the right place to help customers bring their IT security processes and governance to a desired maturity level and adapt technology that effectively supports process objectives.

## 1.2.2 Research and Development

Podictive believes that a continuous investment in Research and Development will enable its competition position with counterparts in the market place. As part of its strategic business development roadmap, Podictive has initiated the development of a Situational Awareness Reporting Tool (SART) that aims to generate logical and mathematical intelligence on existing security and process data within an organization's IT environment. SART will be able to provide insight for organizations in their security health state of IT assets.

## 1.2.3 Company services

### *IT Advisory Services*

Podictive helps organizations to achieve their desire for an optimal alignment between business and IT while maintaining an unremitting focus on controlling costs. Podictive has the capabilities to advise clients in setting up a roadmap for their IT processes to reach a desired maturity level.

### *Security Management Services*

Security management services ensure that service delivery for IT security services from internal and external vendors are aligned to organizational needs and stakeholder requirements.

### *IT Security Service*

Podictive supports its customers in validating their security posture and help setting up an effective Vulnerability Threat Management process. Podictive offers capabilities for customers to identify vulnerabilities in their network with high-end tools that have proven their effectiveness in large scale and complicated network environments.

## 2. Research subject

### 2.1 Understanding the problem

Nowadays, information technology security is becoming one of the more important concerns that organizations of all forms and sizes are focusing on. Banks, hospitals, universities, governments, national and international organizations are all susceptible to be compromised by hackers and malicious users because of the amount of sensitive and worth assets they have. According to Eric Cole, who is an industry-recognized security expert with over 20 years of hand-on experience in his book *Advanced Persistent Threat* [1], the IT security threat landscape is changed where attackers are using and developing more sophisticated methods and techniques to reach their victims while many organizations are still dealing with the emerging threats in a conventional approach by continually investing on security and purchasing various hardware and software solutions in a hoop to protect their data from high organized attacks.

Podictive as IT Security Company has also observed that organizations increase their security budgets while the proportional increase of sophisticated and targeted attacks on organization's IT infrastructure increasingly result in getting compromised. Verizon Enterprise Solutions confirms this threat statement in its publication of *Data Breach Investigations report 2013* [19]. That why Podictive is now seriously and continuously think of innovations that help organizations to evolve with the changing threat landscape and enhance their security processes with proactive and preventive solutions instead of the traditional reactive approaches.

### 2.2 Research question

In recent years, because of the increasing importance of an organization's most valuable assets which is Data, many IT security companies and many applied researchers have become increasingly interested in improving IT security by looking for new solution approaches to deal with the new emerging IT security challenges. SANS institute (<http://www.sans.org/>), one of the biggest and trusted companies that is specialized in computer security training, certification and research, has published different research studies and practices as attempts to developing new security approaches for the new challenges; one of its successful attempts was published in 2001 as a security architecture model which is bases on different data classification and data security models where heterogeneous combination of policies and leading practices, technology, and a sound education and awareness program ware used. The International Business Machines Corporation IBM (<http://www.ibm.com>), which also engaged is developing IT security solutions has lunched many research studies in this domain; one of its efforts was a research study to develop a highly-scalable, run-time extensible, and dynamic cyber security analytics platform to deliver generic analytics capabilities in order to detect threats across multiple data channels [6]. Another research paper was published by the International Journal of Cyber-Security and Digital Forensics (<http://conference.researchbib.com/>) that proposed a solution approach called Security Measurement Based On GQM To Improve Application Security During Requirements Stage which suggest to use a security metrics model based on the Goal Question

assessing (GQM) approach in order to assess security at the requirement analysis stage of the application development life cycle [35].

Unfortunately, although the use of the new security solution organization are still being compromised, that's because the most efforts to deal with the new trend of threads are initiated to use traditional way by reacting on security threats and breaches on the company's infrastructure, whereas there is a need to develop proactive and preventive approaches. According to Eric Cole the new challenging threats are well-organized and data focused; attackers are emerging from standard exploits that take advantage of known vulnerabilities to advanced zero-day vulnerabilities with automated and more targeted methods. Which means reactive security is no longer effective; there is a need for security solutions that can prevent attacks in proactive manner. Eric Cole confirms in his book that Predictive modeling and Intelligence analysis are the innovative concepts that will help achieving preventive and proactive IT security posture.

However, until now little attention has been devoted to improve IT security using predictive modeling and Intelligence analysis; Computers and Security journal has publish an interesting work to model IT security by detecting fraud via regression analysis (Lindsay C.J. Mercer, head of the Group Audit Department of BASS plc and is a frequent speaker at international conferences on computer security and EDP-auditing.) where the functional relationships between variables, which are necessarily numeric values, can be represented is a straight line that can be used to predict the value of one variable based on the value of another or others. Another effort in this context was an attempt done by Alhazmi OH et al., Measuring, analyzing and predicting security vulnerabilities in software systems, Computers & Security (2006), doi:10.1016/j.cose.2006.10.002, to investigate the possibility of predicting the number of vulnerabilities in a software system where the attention was given to the importance of using quantitative aspects of security and identifying metrics that can be evaluated in practice and have a clearly defined interpretation in any attempt to model IT security. From this we can conclude that any attempt to improve IT security, given the emerging threats, needs involving the use of predictive analysis and predictive modeling based on effective IT security metrics that are numerical quantitative measurement. An attempt to do security this way was by one of the biggest IT security vendors, that is Hewlett-Packard Company or HP (<http://www.hp.com>), an American multinational information technology corporation, HP proposed an approach based on Predictive Modeling by publishing a research study in 2009 (Yolanta Beres) about Using Security Metrics Coupled with Predictive Modeling and Simulation to Assess Security Processes in the areas of vulnerability threat management, and identity and access management [34]. This study suggest thus the use of security metrics coupled with security modeling to deal with the emerging threats, but the question is which security metrics to use? Lance Hayden in his book "*Security Metrics: A Practical Framework for Measuring Security and Protecting Data*" [2] answer our question and confirms the recommendations of different IT experts that the improvement of IT security given the new trends has to be done by implement effective security measurements and metrics that are defined and identified in a process approach. He also makes it very clear that before doing any measurements and before using any security data to generate security metrics we have to define and understand well the context of what we want to accomplish "You cannot measure what you do not understand". The framework that Lance Hayden proposed is



intended to identify effective security metrics and not to make prediction. However, researchers can benefit from his effort by using his approach to develop effective security metrics that can be combined with security predictive modeling in order to achieve better IT security improvements. Since there is little information available about how to identify effective security metrics to be used in security prediction, the raising question is now: *How to identify the security health state of an IT asset based on a predictive model?*

## 2.3 Research approach

The main purpose of the present paper is to perform a research study that may contribute in improving IT security by accomplishing a theoretical approach that starts by identifying effective security metrics based on a process approach and end by using those metrics to make IT security prediction.

Believing that Predictive analytics and Predictive modeling are key components to help organizations predicting potential future compromises of their information systems, Podictive as IT Security Company is currently running a project to develop a tool to help organizations to get insight of the security health state of their IT assets, therefore it is raising the question how to use Predictive modeling techniques to assess the security health state of an IT asset. The tool is still in the development phase; therefore part of the functional description will be a theoretical (conceptual) predictive model that can be used to answer the central question posed:

*"How to identify the security health state of an IT asset based on a predictive model?"*

To answer this question there are three sub questions that have to be answered:

1. What is the definition for a security health state of an IT asset?
2. How are security metrics being identified in a process approach?
3. How to build a decision support system/model that can help in the measurement of a security health state of an IT asset?

To answer these questions, the first step that Podictive wants to initiate is a research study to analyze and understand the IT security context in order to give a definition to the security health state of an IT asset; the second step will be to identify effective security metrics based on a process approach, the third step will be then our approach which is the linear regression model for predicting the defined IT security health state of an IT asset, that is to use the identified effective security metrics as input data for the our predictive model to predict the output which is the value that represents the defined security health state of an IT asset.

## 2.4 Report outline

Chapter 3 describes the definitions of fundamental subject and different IT security terms that are needed to understand the IT security context in order to derive an appropriate definition for the security health state of an IT asset.

Chapter 4 describes the process approach for identifying effective security metrics that will be used as input data for the predictive model. The chapter defines a process approach based on the well-known Goal-Question-Metric approach which will be extended by some extra steps and features to fit the IT security context.

In Chapter 5 I apply the metrics process described in the previous chapter to two different IT security processes, namely the patch and vulnerability process, in order to generate a list of effective security metrics.

In Chapter 6 I apply a predictive decision model, which is the linear regression model, where I use the list of pre-defined security metrics that are developed in the previous chapter to generate an input and output data set for this model. Next I use historical training data set to train and test the model. Then the evaluation of the model will be done by comparing our linear regression model to another known predictive model, which is the decision model based on a training and a test dataset results.

### 3. Definition study

In the IT security industry, terms like IT asset, security event, security metric and security health state are often used without even taking trouble to define or to understand what those terms mean. Discrepancies in the misconception of these terms can easily result in a misunderstanding of the purport.

This chapter aims to give a clear definition for a statement that is key in my research: “A Security health state of an IT asset based on a definition of the value of this IT asset using security events that are attributed to effective security metrics.”

To give such a definition we first need to have clear definitions of fundamental subjects and security terms that contribute to above statement. To derive a clear definition of security health state of an IT asset I will conceptualize the following aspects that are relevant to my research in this chapter:

- Health state;
- IT security;
- IT asset;
- Security event;
- Security metric.

Before we elaborate deeper into the definitions related to the IT Security health state, we need to consider and explain two important aspects that are important for the definition of the IT Security health state; these are:

- IT Security stakeholders;
- Goals and objectives of the IT Security health state.

#### 3.1 IT Security stakeholders

Basically, security stakeholders are the individuals that finally will benefit from any security development effort. In our case of developing a definition of the security health state of an IT asset security stakeholders will be the end persons who benefit from any definition made for IT security health state or any IT security model in order to improve IT security, they are also the individuals who will help us specifying the measurement goals for our project. Taking security stakeholders in account in the course of our definition will enable us to focus on the right things and to remain in the context by using specific terms and providing meaningful results [2].

#### 3.2 Goals and objectives

Setting specific goals for your definitions will involve specific measurement of different terms and aspects. Without specific and clear goals, we will run some risk of achieving wrong measurements as result of misinterpretation and misunderstanding of security terms and concepts definition.

For our definition of the security health state of an IT it's very important to set our security goals and objectives upfront in order to determine which approach to adopt and

therefore which definition of security health state best fits to our approach. Generally, it will be enough for security stakeholders to assess their security health state using a security risk assessment approach that finally tells them of their assets “Risky” or “Not risky” like a matrix-based security approach that Lance Hayden states in his book as an inefficient approach [2].

Because we are intended to establish a specific definition for the security health state of an IT asset, we need also to set specific goals for that by involving specific IT security stakeholders. By asking Podictive questions about the goal behind the definition of “Security health state of an IT asset” the answer was: As the trend changed in the IT security industry the goals and objectives of our organization are also changed, there is now a need to find ways to look at IT security, there is a need to correlate security event in such a way that, at any given moment even in the future, we could predict how healthy our IT asset is. Therefore there is a need to find a definition that enables security practitioners to implement effective approaches that keep track of the total value of the IT asset which may be affected by security event over time. Thus there is a need to asset value-based definition of the security health state of an IT asset. In order to achieve this goal we have to take each term and concept that may be covered by our definition in consideration in more detail, that is exactly what we do next.

### 3.3 Conceptualizing the health state

To have a brief understanding of what a health state means even in other sectors different than IT security, we can refer to World Health Organization (<http://www.who.int/en/>) which gives an International Classification of Functioning (ICF), Disability and Health that provides a standard language and framework for the description of health and health-related states. The ICF gives a definition of the health state of human being based on two important qualifiers which are:

***The Performance qualifier** describes what an individual does in his or her current environment. Since the current environment always includes the overall societal context, performance can also be understood as "involvement in a life situation" or "the lived experience" of people in their actual context.*

***The Capacity qualifier** describes an individual's ability to execute a task or an action. This construct indicates the highest probable level of functioning of a person in a given domain at a given moment.*

*When a person has a capacity problem associated with a health condition; therefore, that incapacity is a part of their state of health. To assess the full ability of the individual, one would need to have a “standardized environment” to neutralize the varying impact of different environments on the ability of the individual. In practice, there are many possible environments that we could use for this purpose.*

Fundamentally this definition strictly refers to a health state of a human being. But we can learn one important thing that any definition of the health state of anything will depend on the environment and the context where it function and will include the

impact of different qualifiers and factors. For the concept of health state in our research which is about IT assets, this definition will not suffice. What we are looking for is a definition that describes the health state of an object that indicates its general condition at a certain moment. For the purpose of the research we need a definition of a health state in terms of numerical values based on quantitative metric measurements in order to do calculation, analysis and comparisons, which will help us finally make prediction based on potential prediction drivers.

Prior to look into an object's health state, we will elaborate further on the health state of a person by using a simple real life example. After that, I will extend this definition to define similarly the health state for an object, in this case an IT asset, in terms of security metrics.

### 3.3.1 A persons health state

To get an idea how the health state of something can be defined based on measurements in terms of numerical values we need to take a look at different things from different sides and see how the state of those things changes over time and how they can be affected by other things and external factors. We take here the health state of a person as an example only to simplify thing and get insight how thing work with something that everyone knows.

***Definition: Health state of a person***

*As a result we can now derive the following definition for the health state of a patient who has headache: his health state will be just the condition of his temperature and blood pressure in comparisons with the normal values of temperature and blood pressure that are known to be for a healthy person. The closer his temperature and his blood pressure to the standard values the healthier he is.*

This definition is further explained in following paragraphs.

#### 3.3.1.1 Measuring the health state of a person

To start let's see how to identify the health state of a patient based on two metric variables; if someone for example has a headache then his temperature and his blood pressure will be measured in the first step to get some initial idea how healthy he is. But before you do that you state first your goal and objective, which is in this case in this case to give the patient the right recipe for his headache. You have thus a clear goal and objective and you have two metrics variables to do that; temperature and his blood pressure. This two metrics are known, form experience and researches, to be appropriate to give some ideas about the health state of the patient, or to give some indications about what to measure next to know more. The doctor may stop this measurement here and give the patient a recipe, or may go further with measuring other variables that can help make the picture clearer. Measuring the two variables maybe not enough to determine exactly the patient health state, in that case, the doctor may use other metrics to do other measurements, or may ask the patient some questions to get more information about his health state.

### 3.3.1.2 How to determine the health state of a person?

We are now at the point to ask who decides when to stop or when to go further with measuring activities? What to measure and how to measure it? In our case, not only the doctor that decides how and what, but the patient self can be involved since he may have some feelings about his health that the doctor cannot measure or cannot even think about! Therefore, the problem is that sometimes it is difficult or even impossible to measure all the variables that have to do with the health state of something or someone. From this example we can see that although it's impossible to measure everything there are two stakeholders of this issue, the doctor and the patient, that can help deciding about how and what to do to get a clearer idea about the health state of the patient. These two stakeholders are thus the key factors in deciding which metrics to use and which decision to make; we come here to the point that asking question is also important in this stage to help making right decisions. Thus after setting the goals there are three important steps to do: doing metric measurements, making observations on it or asking questions to stakeholder and then making decisions about the health state of the patient according to some standards values for the metrics that represents the temperature and blood pressure of a healthy person. Normally, the doctor will use standards values resulting from scientific research studies or experts experiences to make comparisons in order to decide.

### 3.3.1.3 Event-based and value-based definition

Let's assume that the value of a person, his temperature and his blood pressure, which we will assume to be numerical values, have initial values,  $V_0$ ,  $T_0$  and  $B_0$  respectively at a given moment  $t_0$  where we start measuring his health state by using the two metrics: temperature and blood pressure. Over time if something, which we intend to call event, happened to the person at a moment  $t > t_0$ , and if the value of the person was affected at that moment  $t$  positively or negatively by that event to take a new value  $V_t$ , then even if we do not know anything about the nature of that event, the new event can be characterized by the new variables namely  $t$ ,  $T_t$ ,  $B_t$  and  $V_t$ . Therefore the new value of the health state of the person at the moment  $t$ , which is  $V_t$ , can be characterize by the variables  $t$ ,  $T_t$ ,  $B_t$ . Formally, if the health state depends only on the these two variables and if we suppose that  $V_t$  is linearly related to  $t$ ,  $T_t$  and  $B_t$  then we can write the new value  $V_t$  as function of the variables  $t$ ,  $T_t$  and  $B_t$  as follows:

$$V_t = a_t * t + b_t * T_t + c_t * B_t + V_{t-1}$$

where  $a_t$ ,  $b_t$  and  $c_t$  are factors that characterize the direction of the change in the initial value  $V_0$  at the moment  $t$  depending on the variables  $t$ ,  $T_t$  and  $B_t$ . The choice of these factors will depend on the our standard thresholds of the metrics temperature and blood pressure that represent the a healthy person, for example if the measured values of the temperature and blood pressure metrics are far away from the standards then we have to choose those factors in a way that represent the affection of person's value, in other words, in a way that causes diminution in the value  $V_0$ . In this case the resulting value  $V_1$  will be a characteristic level of a non-healthy person. The assumption to consider numerical values here will enable us to assume a linear relationship with the value  $V_t$  and the other parameters which may enable us to use linear model like linear regression

to model the health state of persons more effectively and more efficiently since this way enables us to transform the definition of the health state of a person into a numerical function that can be easily calculated and analyzed.

This way of representing the health state of something as linear function of security events characterized by effective security metrics values at the moment of the occurrence of the events  $t$  can be generalized for things that have an initial value  $V_0$ ,  $n$  metrics  $M_i$  where  $i=1, \dots, n$  and value  $V_t$  at the moment of the occurrence of the events  $t$ . Therefore,  $V_t = \sum_{i=1}^n \alpha_i * M_i(t) + \beta$ , where  $\alpha_i$  and  $\beta$  are constant factors that characterize the direction of the affection on the value  $V$  of that thing. Using numerical values for the security metrics  $M_i$  and the assumption that the security metrics and the value  $V_t$  are linearly dependent will ensure the existence of such an expression.

This was a simple example to show how health state of things like persons can be represented as combination of his value and the event that affect this value if we believe that the variable used are linearly dependent. Later we will generalize the idea to derive a definition for the security health state of an IT asset as function of security events attributed with effective security metrics and other IT asset features and characteristics. But now we go further with giving the definitions of terms and concepts that are needed for our definition the health state of an IT asset.

### 3.3.2 IT security

IT security is a term that everyone talks about nowadays, every security expert tries to measure without knowing or understanding what that exactly means [1]. In general, IT security is defined by the preservation of the confidentiality, integrity and availability of information stored or processed in IT assets.

According to PAS 555 the Cyber Security Risk Governance and Management standard published by BSI (The British Standards Institution) in 2013, a general definition to cyber security is given as:

*...the ability to protect or defend the use of cyberspace from cyber-attacks [4].*

NIST defines IT Security as:

*A description of security principles and an overall approach for complying with the principles that drive the system design; i.e., guidelines on the placement and implementation of specific security services within various distributed computing environments.*

In his book “Information Systems” third edition page 162, Steven Alter defines information security as:

*The extent to which information is controlled and protected from inappropriate, unauthorized, or illegal access and use.*

Generally, IT security is the practice to protect the confidentiality, integrity and availability of information that is processed through information systems including the supporting IT infrastructural assets. The Information and IT infrastructural systems that are subject to IT security are indicated in this research as IT Assets.

Although these definitions, the term IT security has not yet a common definition, because each organization gives its own definition that fits and meets his needs and objectives. The definitions that are given to IT security until now are still more business specific, they do not response to the need of the security industry as a standard. As things become big and complex we need a clearer and a standard definition, we need a definition that enables IT security experts to develop more effective approaches to perform IT security more properly and more efficiently. To understand more clearly what IT security exactly means and the deficiency in modeling it we will take a look at some approaches for modeling IT security in following paragraphs.

### 3.3.2.1 Matrix-based security approach

Nowadays, there are a lot of risk assessments approaches for IT security, one common used approach that enterprises use to evaluate their IT security risks is a matrix-based approach which uses a variation of the “Likelihood x Severity” as shown in Figure 1 [2].

		Likelihood of Event		
		High	Medium	Low
Severity of Impact	High	"We're Doomed!"	Bad	Outlier
	Medium	Bad	Not Good	Error
	Low	Annoyance	Typical	"Whatever..."

Figure 1: Generalized risk assessment matrix [2]

In this matrix there are two dimensions that *indicate two* basic parameters; the “Likelihood” of occurrence and the “Severity of impact” which indicates the impact to the enterprise. The important thing to realize here is that the security risk matrix is based on expert judgments which are just opinions about risk as Lance Hayden has confirmed in his book [2]. For example when the likelihood of event is Low and the severity of impact is medium the matrix gives an error, what is an error then? Which decision we can make in this situation? The answer to this question will be pure based on the judgment of the security stakeholder who have this term involved. Therefore assessing risk based on human opinions and judgments is not an efficient way to do things; the fact is that you cannot manage what you cannot measure, thus there is a need to more accurately approaches base on numerical measurements. That is because if cannot measure things numerically then we will not able to analyze it, to compare it or to understand it.



### 3.3.2.2 Risk-based and data-centric approach

Another approach to model IT security is proposed by Jeff Laskowski, a senior IT Specialist with IBM's Software group, in his book “Agile IT Security Implementation Methodology”[3]. The approach that Jeff Laskowski proposes is a Risk-driven and Data-centric security approach. According to him, a data-centric approach will be the easiest one since it is easy to get security data from databases. For this model an assessment of the network topologies, application topology, and business process models are needed to be able to list the data sources with the applications, networks, and business processes and understand their infrastructure in order to use the data more effectively. One more important thing that Jeff Laskowski suggests to is to involve the stake security holders like system owners of the applications in order to better understand the data sensitivity and its architecture and therefore to achieve better results with his approach. Depending on security managers, other approaches can be taken like an application-centric approach, a business process approach, or a network approach. This means that identifying different goals and objectives will lead to different definitions of security health state. Based on his approach, Jeff Laskowski has designed an agile risk model where risk is a combination of future probabilities that cannot exactly be quantified! Moreover, his agile risk model also uses a risk matrix that we already talked about and mentioned its inefficiency and inconsistency, since it is based on human opinions and judgments and not on numerical facts and measurements! What we need thus is a definition of security health state of an IT asset on a consistent, risk-based and data-centric approach.

### 3.3.2.3 Security metrics approach

Andrew Jaquith, who is the program manager for Yankee Group's Enabling Technologies Enterprise group has written his book called “Security Metrics REPLACING FEAR, UNCERTAINTY, AND DOUBT” his purpose from this book was to give an approach to quantify, classify, and measure information security operations in modern enterprise environments based on identifying and modeling security metrics in an effective way [5].

Andrew Jaquith does not give a definition to security health state of an IT asset. Instead, he gives a new approach to improve IT security based especially on metrics data; after defining effective security metrics he suggested to collect and analyze these metrics data and then created a scorecard that aligns and combines everything together. This seems to be an acceptable approach to be adopted in order to find an appropriate definition for the security health state of an IT asset in our case, since this met our purpose to find a definition based on numerical metrics values. That's because Andrew Jaquith has based his approach on data collection and data analysis using different analysis techniques. That is exactly what we also intend to do, we want to find an alternative approach based on effective metrics measurements applied to IT security data that will be analyzed and correlated in some way with the expert opinions and the objectives of security managers.

As we have seen in this example above, there are different approaches devised for the IT security purpose, but they are still business specific attempts to improve the protection of enterprise information assets over time. It would be ideal if we could

combine all the proposed approaches in one to accomplish IT security in an ideal way that met the needs of all enterprises! It may seem difficult to do something new in one attempt, but man has to start trying new approaches for IT security even with small and manageable projects over which he has complete control.

Our purpose from this project is to help Podcitive in here task as an IT security company to provide stakeholders with insight in the security health state of their IT assets. The idea is to use historical and effective security metrics being identified in a process approach in order to generate the input data for our predictive model to predict the security health state of an IT asset.

### 3.3.3 IT asset

Knowing what an IT asset exactly means, especially in the context of our definition of the health state of IT asset, will enable us to identify the right and the effective metrics on it in order to identify, track and control the value of the most critical IT. Defining and identifying IT assets will also enable us to identify easily the standard of measurement for the security metrics we have. That is exactly the same as we did to define the health state of a person; knowing what a person is enable as to talk about his blood pressure as an effective metric for his health, but you cannot use blood pressure as metric for IT asset because of the difference between them. In the following I will give some known definition of the IT asset en then I will derive our own definition that will fit our desired definition of the health state of IT asset.

According to PAS 555 published by BSI (The British Standards Institution 2013)[4] an asset in general is anything that has value to the organization like; information, software such as a computer program, physical such as computers, services, people and their qualifications, skills and experience and intangibles such reputation and image.

ITIL (Information Technology Infrastructure Library) definition of IT asset is:

*In IT Service Continuity Management and in Security Audit and Management, an asset is thought of as an item against which threats and vulnerabilities are identified and calculated in order to carry out a risk assessment. In this sense, it is the asset's importance in underpinning services that matters rather than its cost.*

NIST (National Institute of Standards and Technology) [8] defines an asset in general as:

*A major application, general support system, high impact program, physical plant, mission critical system, personnel, equipment, or a logically related group of systems.*

This are all abstract definition of IT asset, what we need is to define and identify IT assets based on identifiers and features that characterize the IT assets. That because IT Asset identification plays an important role in an organization's ability to quickly correlate different sets of information about assets as it confirmed by NIST (National Institute of Standards and Technology) [8]. Therefore, based on the above abstract definitions for an IT asset we can derive our own definition as follows:

**Definition “IT Asset”**

*An IT asset is thought of as an item that has value to the organization like; data, device, or other component of the environment that supports information-related activities, against which IT security events are identified and correlated in order to carry out a risk assessment on the total value of the IT system.*

### 3.3.4 Security event

Since the value of the IT asset may be influenced by security events over time we need to give a specific definition to a security event that will enable us to implement an approach that will identify the security health state of an IT asset at any instance of time.

As it's generally defined for cyber security event by the BSI (The British Standards Institution 2013) in PAS 555 published in 2013[4]

*A security event is an identified occurrence of a system, service or network state indicating a possible breach of cyber security or failure of safeguards, or a previously unknown situation that may be security relevant.*

The definition of ITIL is as follows:

*A change of state which has significance for the management of a Configuration Item or IT Service.*

The term Event is also used to mean an Alert or notification created by any IT Service, Configuration Item or Monitoring tool. Events typically require IT Operations personnel to take actions, and often lead to Incidents being logged.

The NIST (National Institute of Standards and Technology) [8] definition for event in general is:

*Any observable occurrence in a network or system. NIST gives a definition to security event as a threat event: An event or situation that has the potential for causing undesirable consequences or impact.*

In the light of these definitions, as we are intended to derive a definition for the security health state of an IT asset as function of security events that are attributed to effective security state metrics, we can derive an appropriate definition of that security event as follows:

**Definition “Security event”**

*A security event is identified as occurrence of a system, service or network that causes a change of state which has significance for the management of a Configuration Item or IT Service and which has the potential for causing undesirable consequences or impact and affection of the value of the IT asset now or later.*

We need to define security events to be able to choose the right events that affect the value of the IT asset, to be able to identify the right attributes and the right features that characterize those events in order to correlate and classify them. A vulnerability scan that lead to identifying a zero-day vulnerability on our system, for example, can be considered as security event since it may cause an affection of the value of our IT asset.

This will lead us to consider a number of metrics as characteristics of security events to measure how it will affect the IT asset.

The definition of a security event as occurrence that causes an affection of the value of IT assets will enforce us to identify each security events by features and security metrics that depends on the IT asset self. Doing things this way, will enable us to keep track of how security events are correlated and how they influence IT assets and metrics measurements.

### 3.3.5 IT Security metrics

To give a definition for the security health state of any IT asset we should first be able to measure the IT security of an IT asset in order to identify its health state. But measuring IT security is not an easy task since we cannot manage what we cannot measure, we need therefore to do security measurements with security metrics this will lead us to ask an important question, which is: what are security metrics and how to identify them?

Nowadays, there are several commonly used definitions for security metrics in the IT security industry; for example: NIST (National Institute of Standards and Technology) [8] defines an IT security metrics as:

| *Metrics based on IT security performance goals and objectives [7].*

Whereas ITIL (Information Technology Infrastructure Library) defines it as:

| *Something that is measured and reported to help manage a process, IT service or activity [9].*

The problem is that the IT security industry is not mature enough to identify standard ways of IT security measurement, that why any definition of the IT security metric will still inefficient and ineffective. The most known metrics in the IT security industry are business and goal specific; each company try to find its own metrics to achieve its own objectives. One other problem with identifying IT security metrics is that a lot of companies are measuring IT security but do not make any distinction between different kind measurements. In order to identify effective security metrics we need to understand the deference between security metrics and other benchmarks and standards that are used to measure security.

#### 3.3.5.1 Metrics vs. measurements

To understand metrics well let's how Lance Hayden compares it to Measurement in his book; Lance Hayden said:

| *"I define metric broadly to mean some standard of measurement. I particularly like this definition because it is meaningless unless it combined with an understanding of the word measurement. Recall that metrics are a result and measurement is an activity. Measurement is defined as the act of judging or estimating the qualities of something, including both physical and nonphysical qualities, through comparison to something else. Usually the things being measured are not compared to one another directly, but to some accepted standard of measurement—which circles back around to the original definition*

*of metric. Thus metrics are standards of measurement, and measurement is the comparison of things, usually against standards. Often these standards are expressed in numerical units that provide standard metrics for qualities such as length, weight, or quantity. But metrics don't have to be expressed in this way."*

### 3.3.5.2 Metrics vs. KPI (Key Performance Indicators)

To understand the difference between metrics and KPI (Key Performance Indicators) we can have following definition [33]:

*"A metric is just a number; it can be viewed as a count (number of visitors) or a ratio (conversion rate). All of the data we get from analytics tools are metrics. KPIs are metrics, but not normal metrics. A definition of a KPI is a metric that helps you understand how you are doing against your objectives. In other words, KPIs are a bridge between business objectives and web analytics data."*

Basically, different companies may have a different objective that is why the KPIs tend to be unique and specific to each company. To clarify this I give some examples: on an ecommerce site like [www.24studio.co.uk](http://www.24studio.co.uk), the objective is to sell as much product as possible so the KPIs here could be based on the number of orders, and average size of orders. For the luxury travel site [www.turquoiseolidays.co.uk](http://www.turquoiseolidays.co.uk), the business objective can be sending out so many brochures to encourage a holiday purchase, so one KPI could be the amount of brochures sent out that lead to conversions."

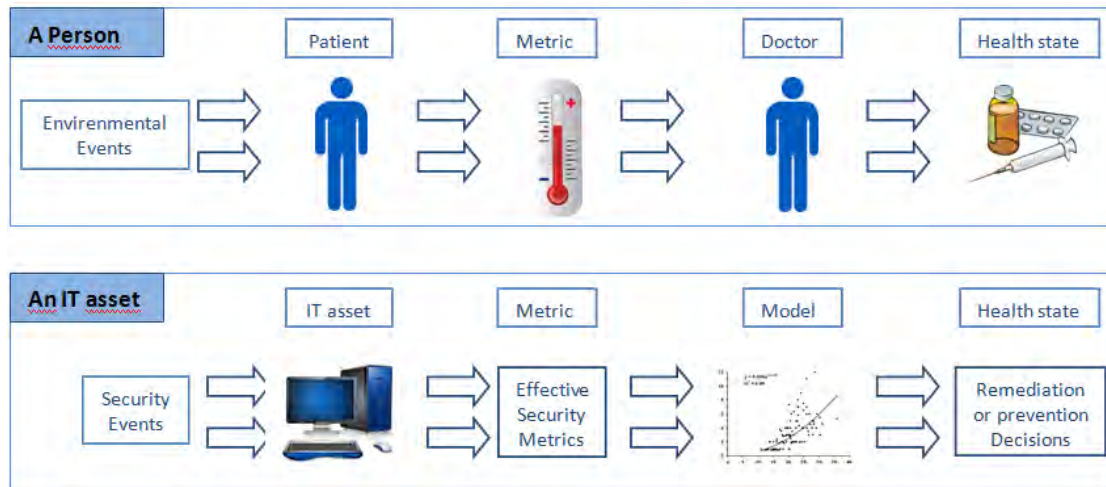
### 3.3.5.3 Process approach for IT Security metrics

The trend is changing in the IT security industry therefore companies have to look forward to find new ways to deal with the new challenges. They have to look for new techniques to identify effective security metrics away from the traditional ones like KPI and other known standards measurements. In order to be able to predict security health state of IT assets effectively we need effective security metrics, for this purpose I have proposed to Prodictive to establish a process approach for identifying effective IT security metrics in order to be able to implement a proactive and preventive IT security.

The proposed process approach for identifying effective IT security metrics will be discussed and described in the following chapter.

## 3.4 Defining the security health state of an IT asset

When we talk about a security health state of an IT asset we can see clear comparisons with the physical health state of a real life person. Having such comparisons helps us using real life examples of a physical health state of a person to explain an abstract definition of what the health state of an IT asset is all about. Measuring the security health state of an IT asset means that security metrics should be available; security events that influence the IT security state of an IT asset should be correlated. The following figure shows the similarity between a person and IT asset concerning the context of the health state.



As it is difficult to identify and measure all characteristics and variables that affect IT assets it is also more difficult if not impossible to determine exactly and perfectly the security health state of any IT asset even if we could define exactly what the security health state of IT asset is. Therefore we have to keep in mind that any attempt to measure the security health state of an IT asset base on any definition and in any context will be just an approximation or estimation of what we desire to achieve. That why we have to make some assumptions in order to come up with an appropriate definition of the health state of an IT asset.

Let's assume that it is possible to identify and measure everything that has to do with the security of IT asset using effective security metrics. The difficult step will be then how to combine all the measurements, how to correlate all security events based on those measurements and how to push then everything together to say something about the IT asset security health state? Basically the problem is almost overall the same, even in the mature industries like finance measurements cannot be done perfectly; Therefore we have now to keep in mind that we cannot measure IT security perfectly and ideally as everyone wishes.

For the ease of use, to give a general definition of the health state of an IT asset based on correlating security events that are in turn based on effective IT security metrics, we will assume ideal IT security industry conditions and mature IT security industry where everything can be measured effectively with effective IT security metrics. In this case, we have the following definition.

***Definition: The security health state of an IT asset:***

*If a security event, as it's defined above, happened at an instant  $t$  then there will be some correlation or affection link between this new security event and of the value of the IT asset and also between all the previous security events. This means that the value of the IT asset will rise above or fall below its some initial value  $V_0$  to get new value  $V_t$  at the instant  $t$  according to the new values of the security metrics that characterize the new event.*

*In the following we will formulate our definition mathematically similarly to what we have done for the health state of a person.*

***Mathematical expression of the definition:***

Assuming that an IT asset has an initial value  $V_0$  at the initial moment  $t_0$ , assuming that we can identify  $n$  numerical effective security metrics  $M_1(t), \dots, M_n(t)$ , for a security event that may affect positively or negatively the value of an IT asset  $V_t$  at a moment  $t > t_0$  and assuming that there is a linear relationship between those metrics  $M_i$  for  $i=1, \dots, n$  and the IT asset value, then the value of the IT asset  $V_t$  can be expressed as linear combination of the metrics  $M_i(t)$  as follows:

$$V_t = \sum_{i=1}^n \alpha_i * M_i(t) + \beta$$

*Where  $\alpha_i$  are constant that characterize the direction of the affection in the value of the IT asset by the metrics  $M_i(t)$  at the moment  $t$  for  $i=1, \dots, n$  and  $\beta$  represent an error or a disturbance factor.*

The determination of constant factors  $\alpha_i$  and  $\beta$  for  $i=1, \dots, n$  will depend on the model used in the predictions and the effective metric  $M_i(t)$ . For example if the measured values of the metrics  $M_i(t)$  at the moment  $t$  are far away from the standards that characterize the IT asset as healthy or secured then the affection of the IT asset by the new security event will cause a diminution in the value of the IT asset. In this case the resulting value  $V_t$  will be a characteristic level of a non-healthy IT asset. The use of effective security metrics, which are numerical value by nature, will enable us to express the health state of an IT asset as linear function of those security metrics, which will enable us in turn to use the linear regression model as a predictive model.

## 4. Security metrics in a process approach

Identifying effective security metrics will help us identify powerful and potential predictive factors. That because attributing events by effective security metrics is more valuable than just saying low medium or high for the value of a metric! Moreover, ineffective metrics could have negative and bad impact on our predictions and therefore our decisions.

It is very important to understand what you are trying to realize very well to be able to derive effective IT security metrics that will drive your measurement efforts. In the context of IT security, it would be great to find a way to build an alignment of what you exactly want to achieve in such a way that you can always be certain you are doing the right things that meet your specific goals and objectives. The best way to do that is to have a business process approach. In the literature we can find different methods and approaches to develop security metrics, one of this approaches is the Goal Question Metric (GQM) approach which was originally developed by V. Basili and D. Weiss [36], and extended by V. Basili, G. Caldiera and D. Rombac in their research study “Goal Question Metric Paradigm” published by the Encyclopedia of Software Engineering, Vol. , pp. 528-532 (1994). The GQM is becoming broadly used to transform business operational goals into useful metrics; for example, as I have stated in the introduction, Lance Hayden has proposed a “Practical Framework for Measuring Security and Protecting Data” based on the Goal Question Metric (GQM) approach, Ala A. Abdulrazeg et al has used this approach in his research “Security Measurement Based On GQM To Improve Application Security During Requirements Stage” to develop security metrics to assess security at the requirement analysis stage of the application development life cycle [35].

GQM is a three-step process for developing software metrics as figure 2 shows, I have adopted this approach and I have extended it with some extra steps and features to fit the IT security context by providing a process design to identify effective security metrics in an IT security environment.

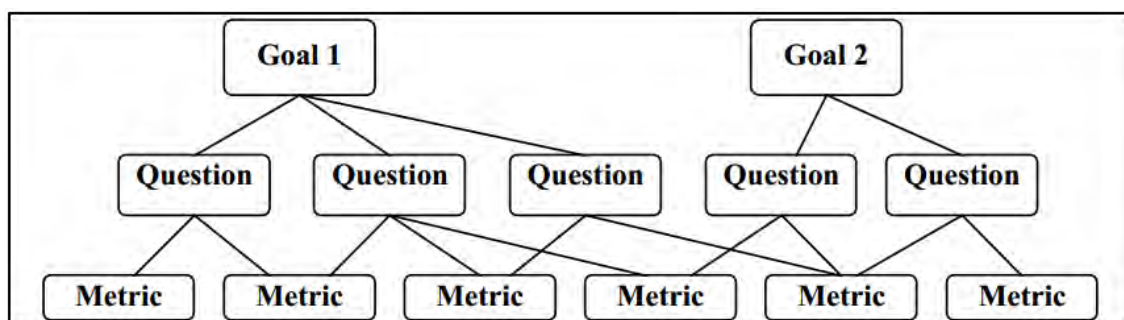


Figure 2: shows the steps of GQM approach [25]

The process starts by setting the goals and objectives that we want to achieve from our measurement. Setting goals will lead us to defining the scope of IT assets that we want to do measurement on. In this context identifying the IT security stakeholders will also be one of the important steps in the process since it will help determining the scope of the IT assets. After setting goals and identifying IT security stakeholders, a context and requirement analysis is needed to be able to determine the scope of the IT asset that we



focus our security measurements on. A context and requirement analysis will also help translating our conceptual goal statements into even more specific questions. This means that asking questions is in turn an important step to do. By answering the questions we will be able to express the components of the goal in terms of clear and executable measurement activities, these measurements activities will be achieved or evaluated in term of a number of security metrics. The final step in the process will be, depending on our need, the selection of a set of these security metrics to be used as effective standards for successful security measurements. I summarize the steps of our process in the following process diagram which may be called the *GCQM (Goal-Context- Question-Metric) approach*.

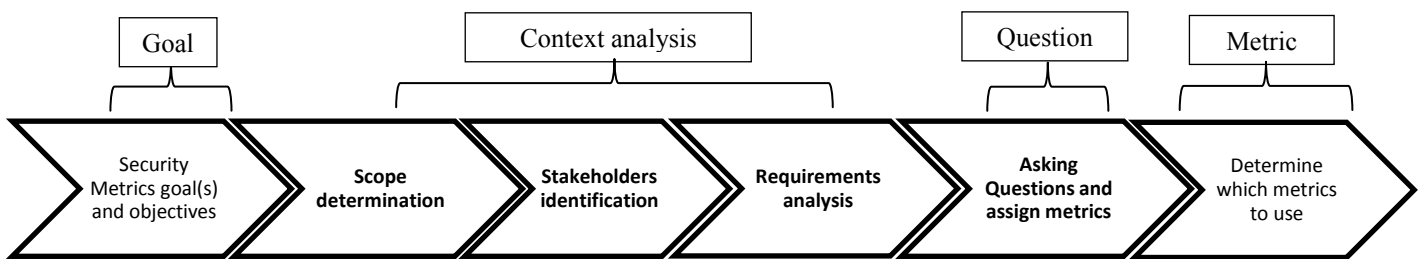


Figure 3: the extended GQM approach by GCQM Goal- Context- Question-Metric approach

#### 4.1 Identification of goals and objectives of the metric

Identifying metrics is not an end in itself, rather any metric process has to be directed by the organization’s goals and objectives, that’s why the first step in the process of identifying security metrics is the determination of the goals and the objectives that the organization sets for its security measurement activities in order to identify the security health state of IT assets and to preserve the confidentiality, integrity and availability of IT assets and information.

It is very important for goal(s) to be well-defined in the beginning, because identifying security metrics and maintaining it may take considerable effort and need resources that may be necessary for other security activities, therefore defining goal(s) up front will save these efforts and resources by enable us to gather and put all required efforts into achieving the objective that has been stated.

Setting appropriate and effective goals is thus one of the most important steps of the metrics process, but it’s not an easy task since there is no hard and fast rule to do dat. Nevertheless, there are several common properties that characterize good goals which can help us make goals more specific; as Lance Hayden in his book [2]; these characteristics can be described briefly as follows:

**Good Goals Are Specific:** Goals have to be clear and specific to be able to measure results easily and precisely, otherwise the value of the accomplishments will be reduced if we keep goals too general or not clear enough even if we succeed.

**Good Goals Are Limited:** Limitation of goals by bounded scope of work according to the context that has also to be well understood up front.

**Good Goals Are Meaningful:** To ensure that goals have meaning you have to construct them in a way that they are both attainable and verifiable.

- **Attainable:** A goal is attainable if it can actually be met; a goal is attainable if it is not open-ended but developed in a particular project that has a starting point and an end point and measurable results.
- **Verifiable:** A goal is Verifiable if it has clear and obvious success and failure criteria that enable us to measure the achievability of the goals.

**Good Goals Have a Context:** Good goals have to be made according to the context in which they are used and measured; they have to be made to meet the desired outcomes stated by stakeholders and according to the unique environment in which they have to be attempted.

**Good Goals Are Documented:** Good goals should have a level of documentation that helps us to capture and organize them according to deferent involved attributes and characteristics. According to Lance Hayden in his book [2] this can be done by broken the goal components down into three elements: Outcome, Elements and Perspective, which can be defined as follows:

- **Outcome:** the outcome can be formulated as the main purpose of the security measurement activity that is what we want to achieve; it can be improvement, assessment of understanding of some business activities.
- **Elements:** These are the component and the objects, such as systems, processes, or characteristics that will be used in our measurement or impacted by the goal. For example: Vulnerabilities, network components, regulatory compliance and system users.
- **Perspective:** This involves perspectives like the point of view that can help to understand the goal. For example: in the case of setting goals for the patch and vulnerability management process this can be: security managers, external attackers or CIO.

After setting the goal components we can combine them to construct a comprehensive goal statement, in the case of the patch and vulnerability management process for example, a goal statement can be given as follows:

*The goal of the patch process is **to assess the remediation priorities for internal servers by identifying the severity of vulnerabilities discovered on internal servers from the perspective of the security manager.***

Making goal statements this way will enforce us to keep our goals limited, specific, and meaningful.

## 4.2 Context analysis

As I said, I have extended the GQM approach by this step which includes in turn three sub steps that's because I believe that a security functional context analysis is necessary to get a clear idea and more insight how things work; a context analysis is needed to understand which measurements are actually used to achieve the preservation of the confidentiality, integrity and availability of IT assets and information. Starting security metric process without a context analysis may lead to spend efforts and resources doing the wrong things. Therefore, to make security metric effective we should determine upfront which scope IT assets needed for measurement activities, we have to analyze and understand which stakeholders have to be involved and which measurements they consider to be important, and finally we have to analyze which requirements are needed for effective security metrics.

The context analysis does not have to be formal or particularly methodical, although in the course of time as your project grows by adding more processes to be measured and as you get more experience, it will be needed to set a formal process of the context analysis.

#### 4.2.1 Scope determination

As I have defined in section 3.3.3. an IT asset is thought of as an item that has value to the organization like; data, device, or other component of the environment that supports information-related activities. If we now start our measurement process considering all kind of asset we have in our organization then at the end of the project measurement we may realize that we was only wasting time and money doing wrong things rather than using it elsewhere to add some value and useful contribution to security.

Starting a security metric process without limiting the scope of the measurement process and without deciding on the scope of the resources and efforts needed, may lead to a lot of frustration and may cause a false starts and ends unsuccessful. That's why it is important to develop a manageable process of limited scope, which can incrementally be developed and improved, than starting by taking too much work and then fail in the middle of at the end of the execution of the process.

It may be that the scope of your process changes considerably as you progress changes in time. Therefore it is important to start with limited scope in a way that you can extend it easily latter if it needed, this means to make it possible for the metric process to add additional assets and resources that are necessary to make the measurement process more valuable and successful at any time of the measurement progress.

#### 4.2.2 Stakeholders identification

I have added this step because I think it is very important to define and identify the stakeholders that have to be involved in the metric measurement since we have limited the scope of our measurements; we have to consider only people who will answer our questions in the course of out metrics identification process and who will provide us with resources and data needed to calculate security metrics. We have identified only the individuals who will benefit from our efforts of security measurement activities. Otherwise we will waste time and efforts by measuring thing for individuals who are not interested in.

In the IT security context, since everyone within an organization is concerned with preservation of the confidentiality, integrity and availability of IT assets and information, everyone can be considered as a security stakeholder, for example CIO, program manager/system owner, security program manager, resource manager, and training/human resources personnel are all security stakeholders, the only difference is that some of them have a greater stake and influence than others. This means that each stakeholder needs a set of metrics depending on the IT security performance needed within their area of responsibility.

Therefore this report proposes to assign each effective security metric to a relevant group of stakeholders as the following table shows [20].

Stakeholder	Metrics
Security/Compliance Officer	<ul style="list-style-type: none"> <li>• Access Accuracy: the number of correctly configured user accounts, against the overall number of accounts created, including badly configured accounts and hanging accounts;</li> <li>• Approval Accuracy: the number of approved provisioning activities, against the overall provisioning activities, including the unauthorized ones.</li> </ul>
Application Owner (Business)	<ul style="list-style-type: none"> <li>• Productivity Cost: these are the costs, in terms of loss of productivity for employees, due to delays during the approval and configuration/deployment phases of the provisioning process.</li> </ul>
IT Operations (IT Budget Holder)	<ul style="list-style-type: none"> <li>• IAM Provisioning Cost: this is the cost of deploying automated IAM provisioning solutions, for a specified timeframe (fixed and variable costs);</li> <li>• Provisioning Effort: this is the actual number of provisioning transactions carried out by the organization, in a specific timeframe, giving an idea of the effort and involved workload.</li> </ul>

*Table 1: Assigning effective security metric to security stakeholders*

More generally, a well-known model and widely used matrix which is the RACI model is used for identifying and assigning roles and responsibilities to stakeholders and individuals who are involved during organizational processes. IT Governance Institute (ITGI™) ([www.itgi.org](http://www.itgi.org)) has designed and created a publication called Control Objectives for Information and related Technology (COBIT®) where different kind of RACI models are presented and directed for chief information officers (CIOs), senior management, IT management and control professionals [22]. In our IT security context of identifying effective security metrics that will be used in a predictive model to assess the security health state of IT asset, we give here three examples of the RACI charts from COBIT concerning the following three security issues: Ensure Systems Security, Identify and Allocate Costs and Manage Service Desk and Incidents.

Activities	Functions										
	CEO	CFD	Business Executive	CIO	Business Process Owner	Head Operations	Chief Architect	Head Development	Head IT Administration	PMO	Compliance, Audit, Risk and Security
Define and maintain an IT security plan.	I	C	C	A	C	C	C	C	I	I	R
Define, establish and operate an identity (account) management process.			I	A	C	R	R	I			C
Monitor potential and actual security incidents.				A	I	R	C	C			R
Periodically review and validate user access rights and privileges.				I	A	C					R
Establish and maintain procedures for maintaining and safeguarding cryptographic keys.				A		R			I		C
Implement and maintain technical and procedural controls to protect information flows across networks.				A	C	C	R	R			C
Conduct regular vulnerability assessments.		I		A	I	C	C	C			R

A RACI chart identifies who is Responsible, Accountable, Consulted and/or Informed.

Figure 4: RACI model for Ensure Systems Security

Activities	Functions										
	CEO	CFD	Business Executive	CIO	Business Process Owner	Head Operations	Chief Architect	Head Development	Head IT Administration	PMO	Compliance, Audit, Risk and Security
Map the IT infrastructure to services provided/business processes supported.		C	C	A	C	C	C	C	R	C	
Identify all IT costs (e.g., people, technology) and map them to IT services on a unit cost basis.		C		A		C	C	C	R	C	
Establish and maintain an IT accounting and cost control process.		C	C	A	C	C	C	C	R	C	
Establish and maintain charging policies and procedures.		C	C	A	C	C	C	C	R	C	

A RACI chart identifies who is Responsible, Accountable, Consulted and/or Informed.

Figure 5: RACI model for Identify and Allocate Costs

Activities	Functions										
	CEO	CFD	Business Executive	CIO	Business Process Owner	Head Operations	Chief Architect	Head Development	Head IT Administration	PMO	Compliance, Audit, Risk and Security, Service Desk, Incident Manager
Create classification (severity and impact) and escalation procedures (functional and hierarchical).				C	C	C	C	C		C	A/R
Detect and record incidents/service requests/information requests.											A/R
Classify, investigate and diagnose queries.				I		C	C	C		I	A/R
Resolve, recover and close incidents.				I	R	R	R			C	A/R
Inform users (e.g., status updates).				I	I						A/R
Produce management reporting.	I			I	I	I		I		I	A/R

A RACI chart identifies who is Responsible, Accountable, Consulted and/or Informed.

Figure 6: RACI model for Manage Service Desk and Incidents

Since we have first set limited, specific, and meaningful goals for the metric process we should not take all information security stakeholders in account otherwise the process will be difficult to manage since you may involve individuals who are not interested in the process goals and are not relevant to the security measurements you care about. Therefore stakeholders have to be identified and prioritized according to the goals and objectives that have been set for the metric process and then a list of the most critical

and interested stakeholders has to be selected as an outcome of this step of metric process, doing things this way will make the process manageable and more successful.

Therefore RACI matrices will be important references which will help us easily identifying to whom we have to ask question about a given goal and metric, since it identify who is Responsible, Accountable, Consulted and/or Informed for a given role or activity in IT organizational processes.

#### 4.2.3 Requirements analysis

In software engineering requirement analysis is an important step that comes at the top of any development project as the following diagram shows.

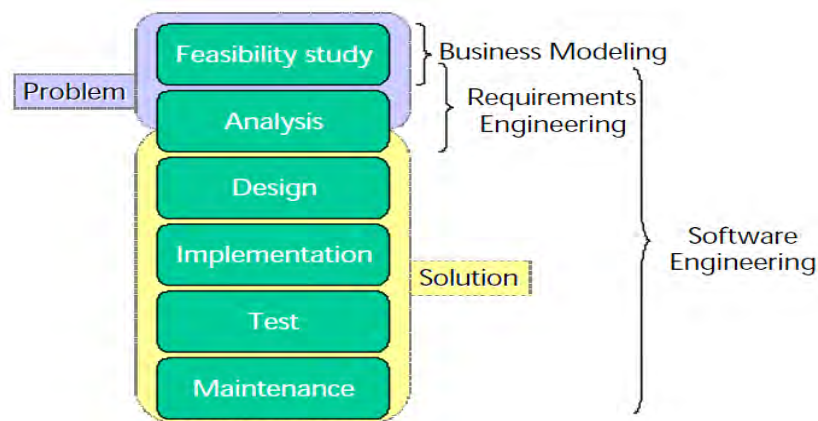


Figure 7: shows the components of the software engineering process [32]

I have added this sub step, which belong to the software engineering tasks, because I believe that this step is very important not only for software engineering but for any project where stakeholders or clients require some tasks or some products to be accomplished.

In our case of identifying effective security metrics a requirements analysis regarding to security context, goals and stakeholders will be an important step in our process. Requirement analysis will enable us to understand how our measurements efforts will be integrated and aligned with the objectives that the stakeholders want to achieve. Requirement analysis will help to save time and efforts by doing the right things to achieve the set goals.

Understanding the motivations and reasons behind our security measurement activities makes it much easier to analyze and improve them, but getting there can be difficult. It is also important to understand the resource requirements of any metrics or data gathering activities.

Starting security metric process without knowing which resources are required may lead to spend efforts and resources and also time doing the wrong things, and may make the process more difficult and unsuccessful. In this context the following ideas can help understanding things well:

- Describe how the measurement of the information security activities and especially using effective IT security metrics will help organizations to be successful.

- Describe how using effective IT security metrics will help and support stakeholders to get insight in the health state of the security of their IT assets.
- Address the relationship between the measurement of the information security activities using effective IT security metrics and its role to identifying unauthorized uses, attackers, and information stealers.
- Define the role of the measurement of the information security activities using effective IT security metrics in helping managing crises and controlling critical business operations.
- Define the role of the measurement of the information security activities using effective IT security metrics in identifying security incidents, in performing IT security investigations, and implementing IT intelligence capabilities in order to defend against unauthorized users, attackers, and information stealers.
- Determine how the measurement of the information security activities using effective IT security metrics will support the organization in making changes and growing.
- Describe how the measurement of the information security activities using effective IT security metrics can help developing the relationship and collaboration with stakeholders external to the organization.
- Describe how the measurement of the information security activities using effective IT security metrics will use technologies and practices in accomplishing the organization goals.
- Using this list of these descriptions can give us more insight wither our goals will be attainable and our process will be successfully accomplishing.

### 4.3 Asking Questions and assign metrics

After defining effective goals, stakeholders, business context and resource requirements that are needed for developing effective security metric, we now came to the second step of the standard GQM approach where the attributes and targets of the goal will be operationally addressed. In this step the conceptual goal statements will be translated into a series of questions that enable the components of the goal to be achieved or evaluated for success. The question here is how would you translate the goal statement into operational questions? Those are questions that will enable us to articulate the goals and objectives in terms of what measurement activities that must be executed and what data must be selected to address the individual components of the conceptual goal in a clear and executable goal statement.

To make things clearer we take here as example the Patch and Vulnerability Management Process. The question is how would you translate the goal statement of Patch and Vulnerability Management Process into operational questions? Several questions are already implied by examining the goal components:

- How vulnerable is an IT asset?
- How severe the vulnerability found in the IT asset?
- How long does it take to identify vulnerabilities from the moment of announcement?
- How much it costs to identify vulnerabilities from the moment that it's discovered?

Developing the operational questions, we are now able to express our goal of security measurement in terms of tangible numerical characteristics of processes involved, these characteristics will be our metrics that have to be evaluated and measured. These metrics will be more effective and more efficient since they are developed in a process approach that enables them to be integrated with the overall organizational security goals and aligned with the original goal of the security stakeholders.

After developing questions in order to define our goals operationally, we are now ready to begin characterizing our goals at data level, which means assigning metrics that will provide answers to the developed questions. To answer the questions for our example of Patch and Vulnerability Management Process we use the security data to determine the number of vulnerabilities per IT asset per severity of vulnerabilities levels like low, medium and high, in this case the metric to answer questions like:

- How vulnerable is an IT asset?
- How severe the vulnerability found in the IT asset?

Can be the (number of vulnerability type) x (Mean severity score per vulnerability type)  
The metrics to answer questions like:

- How long does it take to identify vulnerabilities from the moment of announcement?
- How much it costs to identify vulnerabilities from the moment that it's discovered?

Can be the following:

- The scanning duration for the identified vulnerabilities
- The scanning cost for the identified vulnerabilities

The developed questions and metrics so far will provide us with data that will help increasing security level by increasing security awareness and security efficiency. Using information and data provided by the developed metrics, security managers will be able to analyze data more efficiently and therefore to take the appropriate decisions at the appropriate time to achieve the goals, moreover security managers will be able to make conclusions and get insights in order to improve IT security by doing more repeated measurement activities.

#### 4.4 Determine which security metrics to use

The last step of the GQM approach is the quantitative step (metric) where a set of metrics should be associated to the questions asked in the previous step in order to give answers in a measurable way. Answering these questions will enable the components of the goal to be achieved or evaluated in term of a number of metrics. Depending on our needs a set of these security metrics should be selected to be used.

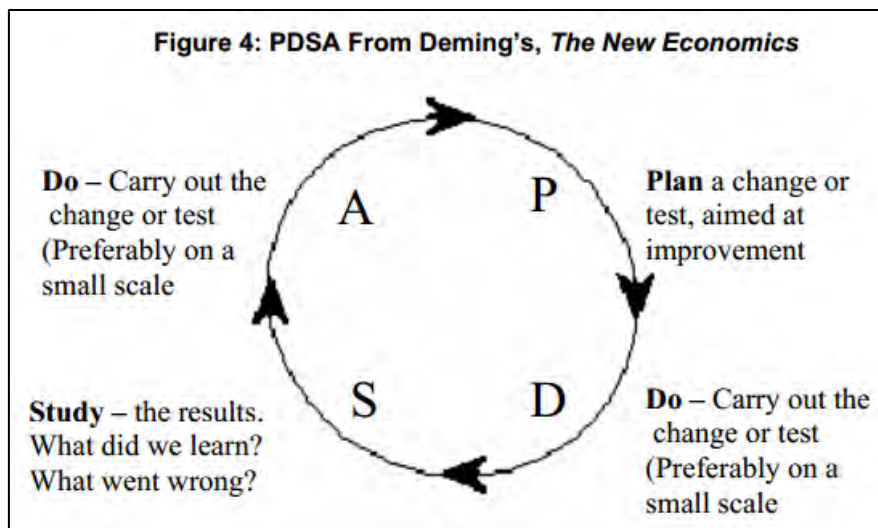
Up to this point, to select IT security metrics, it is more important to know what you are trying to accomplish and to let your goals and objectives drive your measurement efforts then to let the metrics decide this for you. Metrics that have nothing to do with your goals have to be rejected, metrics that add no value to your efforts make things only worse and will be only wasting of resources.



Therefore the selected metrics has to be effective, this means that they have to quantify well your measurements, they have to be more-satisfying, focused and aligned to your strategic security and business goals.

#### 4.5 Improvement process of the identified security metrics

The process approach I have proposed here is more a linear process which we can use in one attempt to identify a security metric for a given security goal. But latter, as security trends may change overtime, the effectiveness of the identified security metric may seem to be no more effective and aligned to the strategic security and business goals. Fortunately, there is a popular tool for doing continuous improvement of processes and products which is Plan-Do-Check-Act Cycle (PDCA); this process is originally developed by Dr. W. Edwards Deming's lecture in Japan in 1950 which is an iterative four-step management method used in business for the control and continuous improvement. The four-step of the PDCA cycle are often represented in a cycle diagram as figure 8 shows, which is for a research study about the Evolution of the PDSA Cycle written by Ron Moen and Cliff Norman as given by W. Edwards Deming's.



*Figure 8: A diagram showing the PDCA Cycle by W. Edwards Deming's*

The PDCA can be used as long as the identified metrics not effective enough en does not meet the expected security objectives.

## 5. Applying the metrics process approach

As I have said before, the challenge that we will face is that difficulty to start trying new approaches of big scope like what we are trying to do here for IT security. The best way to start is to break down the complexity by limiting the scope. Starting with projects of limited scope will enable us to manage them and keep them under control. That's exactly what we will do here in this chapter, we will first start with applying our metrics process approach to one of the most important IT security processes, that is the Patch and Vulnerability Management Process, in order to generate a list of effective security metrics for this process. In the following and last chapter, we will use the obtained list of effective metrics from the Patch and Vulnerability Management Process to develop our desired conceptual model that will help us predict the security health state of an IT asset. Let's now start step by step applying the metrics process.

### 5.1 Security goals

By interviewing the Security Officer of Podictive (see [Appendix II](#) for the interview) I have made a list of goals and objectives for Patch and Vulnerability Management Process. In the following steps we will see how to translate these conceptual goals statements into even more specific questions which will finally be transformed into metrics. We have defined several goals for the vulnerability management and patch management processes.

#### *Goals for Vulnerability Management*

- Goal1:** to get insight how vulnerable an IT asset is.
- Goal2:** to get insight how severe each vulnerability in the IT asset.
- Goal3:** to know how long it takes to identify vulnerabilities from the moment of announcement.
- Goal4:** to know how much it cost to identify vulnerabilities from the moment that it's discovered.

#### *Goals for Patch Management*

- Goal5:** to get insight how efficient the patch management process by measuring the percentage of vulnerabilities that are managed in the patch management process per its severity.
- Goal6:** to measuring the number of patches needed per IT asset.
- Goal7:** to know how long it takes for patch identification from the moment of announcement.
- Goal8:** to know how long it takes for patch execution.

**Goal9:** to know how much it cost to for patch identification from the moment of announcement.

**Goal10:** to know how much it cost for patch execution.

## 5.2 Stakeholders identification

This document provides guidance on creating a security patch and vulnerability management program and testing the effectiveness of that program. The primary audience here is security managers who are responsible for designing and implementing the program. However, this document also contains information useful to system administrators and operations personnel who are responsible for applying patches and deploying solutions (i.e., information related to testing patches and enterprise patching software).

The following RACI charts gives how Podictive IT security functions are assigned to IT security activities in the patch and vulnerability management process.

Activities	Functions				
	Security Officer	Compliance Officer	Vulnerability Engineer	Asset Owner	IT System Engineer
Designs the vulnerability management process and ensures its execution and implemented as planned	R	R			
Responsible for the configuration of the vulnerability scanner and for the scheduling of various vulnerability scans	C		R	I	
Responsible for the scanned IT asset for vulnerabilities and decides about the mitigation of the identified vulnerabilities	C		I	R	C
responsible for the execution of the remediation and for the detected vulnerabilities	I			C	R

Figure 9: RACI model for Patch and vulnerability management process

For each goal of the goals stated in previous section above and for both patch and vulnerability management processes we give the stakeholders that will benefit from measuring these goals in the following table.

Goal	Stakeholders
Goal1: to get insight how vulnerable an IT asset is	<ul style="list-style-type: none"> <li>• Security/Compliance Officer</li> <li>• Vulnerability Engineer</li> </ul>
Goal2: to get insight how severe the vulnerability found in the IT asset.	<ul style="list-style-type: none"> <li>• Security/Compliance Officer</li> <li>• Vulnerability Engineer</li> <li>• Asset owner</li> </ul>
Goal3: to know how long it takes to identify vulnerabilities from the moment of announcement.	<ul style="list-style-type: none"> <li>• Security/Compliance Officer</li> <li>• Vulnerability Engineer</li> </ul>
Goal4: to know how much it cost to identify vulnerabilities from the moment that it's discovered.	<ul style="list-style-type: none"> <li>• Security/Compliance Officer</li> <li>• Vulnerability Engineer</li> </ul>

Goal5: to get insight how efficient the patch management process a by measuring the percentage of total assets that are managed in the patch management process.	<ul style="list-style-type: none"> <li>• Asset Owner</li> <li>• IT System Engineer</li> </ul>
Goal6: to measuring the number of patches needed per IT asset.	<ul style="list-style-type: none"> <li>• Asset Owner</li> <li>• IT System Engineer</li> </ul>
Goal7: to know how long it takes for patch identification from the moment of announcement.	<ul style="list-style-type: none"> <li>• Asset Owner</li> <li>• IT System Engineer</li> </ul>
Goal8: to know how long it takes for patch execution.	<ul style="list-style-type: none"> <li>• Asset Owner</li> <li>• IT System Engineer</li> </ul>
Goal9: to know how much it cost to for patch identification from the moment of announcement.	<ul style="list-style-type: none"> <li>• Asset Owner</li> <li>• IT System Engineer</li> </ul>
Goal10: to know how much it cost for patch application.	<ul style="list-style-type: none"> <li>• Asset Owner</li> <li>• IT System Engineer</li> </ul>

Table 2: Assigning stakeholders to each security goal

### 5.3 Patch and Vulnerability Management Context analysis

As de fined by NIST (National Institute of Standards and Technology) [8] “Patches are additional pieces of code developed to address problems (commonly called “bugs”) in software. Patches enable additional functionality or address security flaws within a program. Vulnerabilities are flaws that can be exploited by a malicious entity to gain greater access or privileges than it is authorized to have on a computer system.”

Vulnerabilities are weaknesses in IT systems that can be used by hackers or attackers to get unauthorized access into the system in order to cause problems, security patches are pieces of software designed to fix those problems.

As the growth of cyber-crime increased, cyber security became a concern that enforces organizations to allocate resources and give more attention to information security in order to protect themselves. Therefore, there was a need to patch and vulnerability management to avoid the use of IT vulnerabilities that exist within an organization by malicious user. Creating a patch and vulnerability management process in organizations enables them to gain insight of vulnerabilities in their IT environment and the risks associated with them and therefore to identify and mitigate vulnerabilities in order to prevent attacks from penetrating the organization’s networks and stealing information (Williams and Nicollet, 2005 [12]). In the following a Patch and Vulnerability Management Context analysis will be given where I will try to explain what this process is and how it is used in enterprises according to SANS (<http://www.sans.org/>), NIST (National Institute of Standards and Technology) [8] and Podictive.

#### 5.3.1 SANS’s Patch and Vulnerability Management process [12]

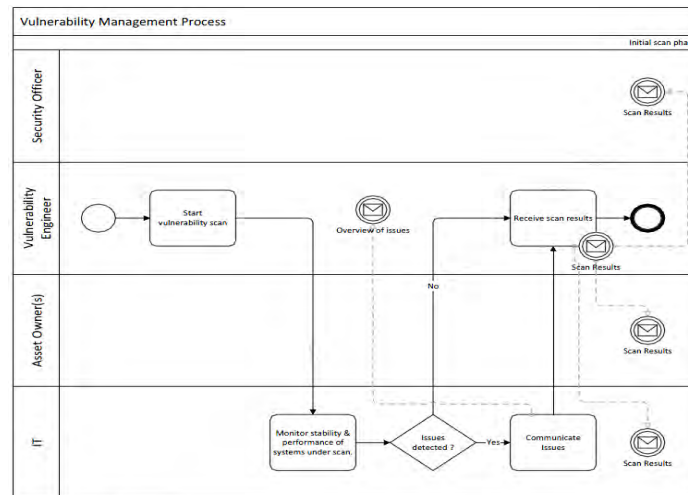
The Patch and Vulnerability Management process of SANS consist of five steps:

##### 1. Preparation

This phase is the first phase in a vulnerability management process, the purpose of this step is to define the scope of the vulnerability management process by determining the number of systems that have to be scanned or the limiting the number of vulnerabilities that have to be scanned.

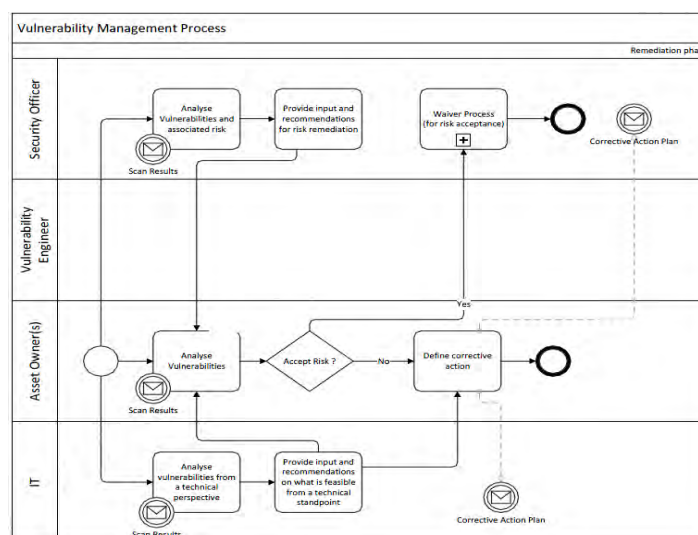
## 2. Vulnerability scan

During this phase vulnerability scans are executed using scanning tools that offer different scan reports. This reports will be used by management and the security officer to understand and analyze the identified vulnerabilities in order to get an overview of how secure and level of risk they are and therefore to make prioritizations and recommendations for mitigating them which will happen in the following phase. the following activity diagram illustrates activities of the initial vulnerability scan that SANS uses [12].



## 3. Define remediating actions

In this phase the remediating actions will be defined and vulnerabilities will be analyzed in order to determine the associated risks for remediation. Vulnerabilities will also be analyzed from a technical perspective to determine the availability of patches and whether they will still supported by the vendor. Furthermore, in this phase the remediation actions should be planned by making a remediation timeframe and setting clear deadlines for the implementation of the remediation actions. The remediation phase as illustrated by SANS [12] is shown in the following activity diagram.

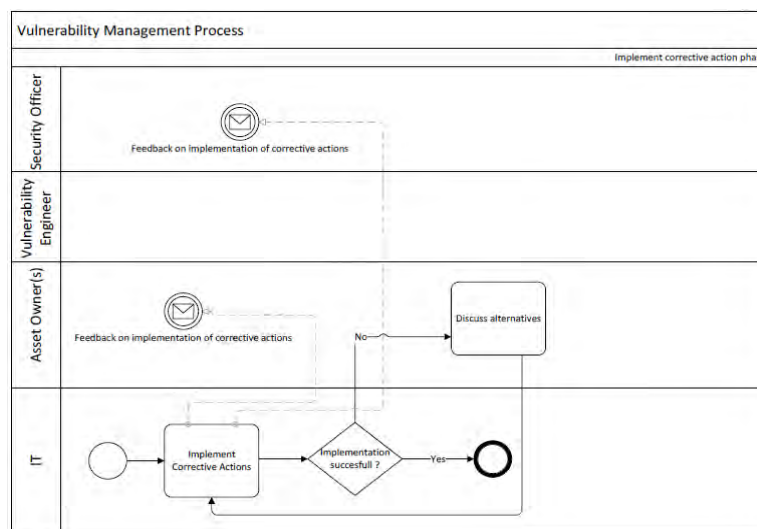


The following table shows an example of corrective remediation actions of detected vulnerabilities.

L	Scan Date	Vulnerability Detected	Risk Rating	Corrective Action	Implementation Date
192.168.4.56	1/02/2013	PHP "safe mode" - Restriction Bypass Vulnerability	4	PHP upgrade. This can only be deployed after code migration is complete.	15/12/2013
192.168.4.56	1/02/2013	Apache prior to 2.2.15 - Multiple vulnerabilities	4	Upgrade apache to newer version	15/02/2013

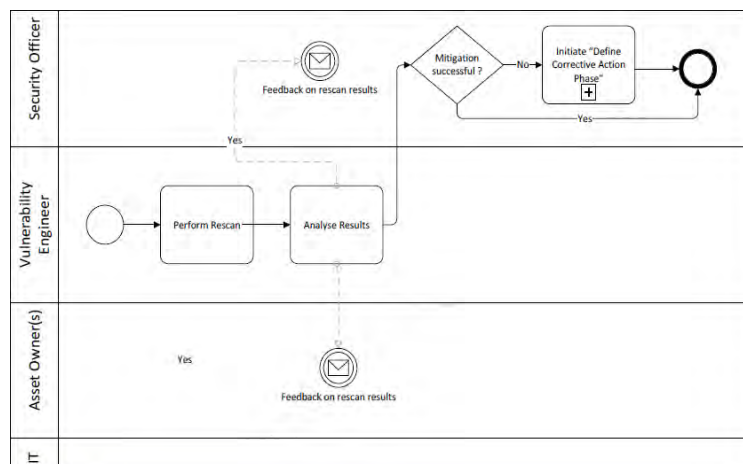
#### 4. Implement remediating actions

This phase is devoted to the execution of the remediation action in line with the planned timeframe. If the execution of the remediation actions failed, other alternative possible action will be defined and implemented. A possible implementation remediation action as SANS illustrates is shown in following activity diagram.



#### 5. Rescan

The rescan phase is intended to verify whether the remediating actions are successfully implemented. This means that new vulnerability scans should be executed using the same vulnerability scanning tools and in the same conditions of the initial scan to be sure that the detected vulnerabilities are mitigated. Following activity diagram illustrates a possible rescan action that SANS uses to verify the implementation of the remediation action.



### 5.3.2 NIST's Patch and Vulnerability Management process

The Patch and Vulnerability Management process of NIST [8] consist of 11 steps, before starting the process NIST recommends organizations that want to implement this process to create a group of individuals, called the patch and vulnerability group (PVG), this group will guide the implementation of the Patch and Vulnerability Management process, communicate and coordinate between different local administrators.

The patch and vulnerability group should be chosen based on the following features:

- Representatives from information security and operations;
- Individuals with knowledge of vulnerability and patch management, as well as system administration, intrusion detection, and firewall management;
- Specialists in the operating systems and applications most used within the organization;
- Personnel who already provide system or network administration functions, perform vulnerability scanning, or operate intrusion detection systems.

The central task of the chosen PVG will be thus to ensure the implementation of the vulnerability remediation efforts during eleven steps which are listed below.

#### 1. System Inventory

The purpose of the creation of system inventory is to be able, at each moment, to take an overview of which hardware equipment, operating systems, and software applications are used within the organization. Here the PVG plays an important role by maintaining manually the system inventory, prioritizing and grouping information technology resources which will facilitate monitoring for vulnerabilities, patches, and threats to response rapidly and effectively. The following table shows an example of an inventory list that PVG can create according to NIST [8].

Software configuration	Hardware configuration	General IT items
Operating system and version number	Central processing unit	Associated system name
Software packages and version numbers	Memory	Property number
Network services	Disk space	Owner of the IT resource (i.e., main user)
Internet Protocol (IP) address (if it is static)	Ethernet addresses (i.e., network cards)	System administrator
	Wireless capability	Physical location
	Firmware versions.	Connected network port

Table 3: An inventory list that PVG creates according to NIST [8]

#### 2. Monitoring Vulnerabilities, Remediations, and Threats

After creating a system inventory there are three types of security issues for which IT security sources should be monitored and prepared by the PVG using the available monitoring resources, these three security issues are:

- **Vulnerabilities** - These are weaknesses in the security of IT system that may be exploited by malicious and unauthorized users;
- **Remediations** - The remediation security activity contains three steps which are: installation of a software patch, adjustment of a configuration setting, and removal of affected software;

- **Threats** - Threats are possible dangers methods that malicious users can use to exploit vulnerabilities in order to breach security and therefore to attack and cause harm to IT systems.

### **3. Prioritization of Vulnerability Remediation. In this step vulnerability**

Remediation and threats should be prioritized according to its potential impact and significance by focusing on the systems that are very crucial and important for the functionality of the organization.

### **4. Creation of Organization-Specific Remediation Database**

In this step it is recommended to create a remediations database that the organization need to applied. The maintenance of this data base can be done manually but NIST strongly recommends automated patching products that contain such databases.

### **5. Remediations testing**

It's the responsibility of the PVG to test patches and non-patch remediations on IT all devices with standardized configurations but they should work closely with local administrators avoid redundant testing by local administrators.

### **6. Deployment of Vulnerability Remediations**

The PVG should deploy vulnerability remediations to all devices that have the vulnerability. The process of vulnerability remediation should be done through there step which are: the installation of a software patch, the adjustment of a configuration setting, and the removal of the affected software.

### **7. Distribution of Vulnerability and Remediation Information**

The distribution of vulnerability and remediation information is necessary to inform local administrators about vulnerabilities that are occurred in the system and the proposed remediation for them, this can be done directly by means of enterprise patch management software or by reporting directly local administrators about it.

### **8. Deployment of Patches**

Like vulnerability remediation, patches should also be automatically deployed to all IT devices using enterprise patch management tools, which will facilitate the work for system administrators to lunch different updates on different systems.

### **9. Configure Automatic Update of Applications.**

It is recommended to have a locally distributed automated update process which makes it possible to use available patches from the organization's network which will provide update for applications from the local network instead of from the Internet.

### **10. Verifying Vulnerability Remediation**

It's also the task of the PVG to ensure that vulnerabilities are remediated and mitigated as planned. This can be done by the executing the following activities:

- **Vulnerability Scanning** - Performing vulnerability scanning using capable vulnerability scanners that are commonly used in organizations to detect and identifies known and associated vulnerabilities.
- **Reviewing Patch Logs** - Using Log files that keep track of the history of a system to review patch logs in order to verify if patches are successfully installed.



- **Checking Patch Levels** - Perform penetration tests to evaluate the system security by simulating attacks as attempts to exploit the vulnerability and breach security using possible threats in the system.

## **11. Vulnerability Remediation Training**

The last step of the patch and vulnerability management process of NIST is the vulnerability remediation training for administrators and other users in the organization in order to be able to apply vulnerability remediations.

### **5.3.3 Patch and Vulnerability Management process**

The vulnerability scanning process consists of four cyclic phases which are described in following paragraphs.

#### **5.3.3.1 Scoping**

In this phase the scope of the IT assets on which the vulnerability scanning needs to be performed will be determined to ensure that the right information assets (servers & databases) is identified to be used in the vulnerability scanning process in question. The scoping of the IT asset is done on quarterly basis and based on criticality and risk level of applications on which vulnerability scanning on the monthly basis needs to be performed. In the course of the time if an IT asset was not scoped but needs to be scanned then an Ad-hoc request can be done for vulnerability scanning. After the scope of IT asset is determined the asset will be registered using a tracking system for future use in order to facilitate the life cycle management of vulnerability scanning findings, finally the scan schedule will be deployed for further processing and executing of the vulnerability scan.

#### **5.3.3.2 Assessment**

This phase starts with the creation of a scan profile of a particular platform (e.g. Windows, Solaris, AIX and MS-SQL) in the vulnerability scan tool (Nessus) for the scoped IT asset. This is important because the vulnerability of IT asset depends on the technology components used on these systems. If this is done then a scheduling of vulnerability scan is needed to be carried out. If vulnerability scans are finished then scan results are imported and reported in XML format that will be used to provide overviews of critical and high findings which in turn will be used in the analysis phase.

#### **5.3.3.3 Analysis**

In this phase the critical and high findings of vulnerability scan from the assessment phase will be analyzed in order to get insight whether those findings are either false positives (to ensure that they do not appear as open findings in further scans) or have an accepted risk that has to be tracked in the future scans. In this phase will be also determine which findings are of accepted risk to ensure that they do not appear as open findings in the further scans.

#### **5.3.3.4 Remediation**

In this phase, remediation and corrective actions will be lunched to remediate and solve the critical and high findings for the vulnerabilities identified during vulnerability scan,

after that a resolution status and action tracking will be provided and will contain the following status: **Solved**, if the finding is solved and successfully is implemented and. **Solvable**, if the finding is not yet solved but can be solved in further scheduled implementation. **Not resolvable due to issues**: if the finding is not solved and cannot be solved due to issues that will be registered for further analysis.

#### 5.3.4 Conclusion

From this analysis of the context of the patch and vulnerability management process of the three security providers, SANS, NIST and PODICTIVE we can conclude that there are four important steps of this process which are: IT asset scoping, vulnerability scanning, vulnerability remediation and results reporting. Since we are intended to identifying effective metrics for the patch and vulnerability management process that will be implemented in a predictive decision model to identify the security health state of an IT asset, and because the metrics that we want to derive will be finally calculated based on vulnerability scanning and remediation reports, a further requirement analysis of the four steps will be very important to understand what we want to accomplish and which metrics data and resources we need to understand how decisions are made, which will help us to get an idea how security events can be correlated in our model. That's what I will try to address in the following paragraph.

#### 5.4 Requirements analysis

In this step of the metrics process I will go further with the analysis of findings and result reports step. Since I will used the case of a client of PODICTIVE as a case study I will concentrate on how PODICTIVE analyze scan results to make decisions. Here, our drivers and motivations behind the identification of effective metrics through a process approach are the key factor of what we want to accomplish; we want effective security metrics as attribute for security events that will correlated to be the engine of our IT asset health state decision model, that's because the more effective the security metrics are the more powerful and accurate the model will be. Therefore,

During the scoping phase IT assets are registered using an integrated action tracking system, this will be very important in building a powerful decision model because using this registered historical data that identifies the IT assets in the system will enable us to easily keep track how the value of IT asset is affected by security events.

Moreover, during the assessment phase of vulnerability management process, a scan profiles is generated based on the existed platform (e.g. Windows, Solaris, AIX and SQL server) in the system and the vulnerability scan tool (Nessus) used to carry out the scans. Therefore security events will be correlated depending on the IT technology used. Thus the final scan reports should include information about the platform on which the security event is occurred. Basically, all findings (or security event as I would like to call) are registered. This registration is more based on the level of criticality of vulnerabilities like High, Medium and Low. Having these registrations in hand will enable our decision model to classify events more easily base on the platforms where they are happened.

In the analysis phase of the process, Podictive does corrective actions for high and critical findings by assessment to identify which findings are either false positives or have an accepted risk; the decision model has to have the ability to take corrective action automatically or manually. Using historical data that can tell us how corrective actions are made in the past will enable us to build a model that has analytical capabilities to decide automatically when findings or events are false positives or have an accepted risk.

Finally, security data about how remediation is done and how resolution status is tracked is also required in order to make correlation between events stronger; correct data that contains registers about when some problem is considered to be solved or not is very required, that will help our model to acquire intelligence capabilities during the training phase.

## 5.5 Scope determination

Since we are intended to drive effective metrics, we need general historical scan reports that contains details registers about the detected vulnerabilities in the past; on which platforms and IT assets they are detected and how frequent they are detected, how they are solved and with which costs they are scanned and solved and how long it took to be solved. Therefore, the scope of our study will be the configuration management database (CMDB) that Podictive uses as a data warehouse to register all the information concerning IT assets in an organization, as well as the relationships between such assets and how they influence each other in order to keep track of the statues of each IT asset.

## 5.6 Asking Questions

The following table shows the questions that can be asked in order to translate the identified security goal statements of Patch and Vulnerability Management Process into operational questions.

Goal	Question
Goal1: to get insight how vulnerable an IT asset is	How vulnerable is an IT asset?
Goal2: to get insight how severe the vulnerabilities found in the IT asset.	How severe the vulnerabilities found in the IT asset?
Goal3: to know how long it takes to identify vulnerabilities from the moment of announcement.	How long does it take to identify vulnerabilities from the moment of announcement?
Goal4: to know how much it cost to identify vulnerabilities from the moment that it's discovered.	How much it cost to identify vulnerabilities from the moment that it's discovered?
Goal5: to get insight how efficient the patch management process by measuring the percentage of vulnerabilities that are managed in the patch management process per its severity.	What's the percentage of the vulnerabilities that are managed in the patch management process per its severity? This represents the residual risks for vulnerabilities.
Goal6: to measuring the number of patches needed per IT asset.	How many patches are needed for the given IT asset?
Goal7: to know how long it takes for patch	How long it take for patch identification for the

identification from the moment of announcement.	given vulnerability from the moment of announcement?
Goal8: to know how long it takes for patch execution.	How long it takes for patch execution for the given vulnerability?
Goal9: to know how much it cost to for patch identification from the moment of announcement.	How much it cost for patch identification for the given vulnerability from the moment of announcement?
Goal10: to know how much it cost for patch application.	How much it cost for patch execution for the given vulnerability?

Table 4: Translating goal statements into operational questions

Several questions are already implied by examining the goal components: Developing the operational questions, we are now able to express our goal of security measurement in terms of tangible characteristics of processes involved.

Goal	Question
Goal1: to get insight how vulnerable an IT asset is. Goal2: to get insight how severe the vulnerability found in the IT asset.	(Number of vulnerabilities per IT asset per severity low, medium and high) x (Mean CVSS[23] severity score) that is: (number of vulnerability type) x (Mean severity score per vulnerability type)
Goal3: to know how long it takes to identify vulnerabilities from the moment that it's discovered.	The vulnerability scan duration for the identified vulnerabilities
Goal4: to know how much it cost to identify vulnerabilities from the moment that it's discovered.	The vulnerability scan cost for the identified vulnerabilities
<ul style="list-style-type: none"> <li>Goal5: to get insight how efficient the patch management process by measuring the percentage of vulnerabilities that are managed in the patch management process per its severity.</li> <li>Goal6: to measuring the number of patches needed per IT asset.</li> </ul>	$(\sum(\text{number of vulnerability type that are still open}) \times (\text{Mean severity score per vulnerability type})) \times \text{number of patches executed}$
Goal7: to know how long it takes for patch identification from the moment of announcement.	The duration of patch identification
Goal8: to know how long it takes for patch execution.	The duration of patch implementation
Goal9: to know how much it cost for patch identification from the moment of announcement.	The cost of patch identification
Goal10: to know how much it cost for patch application.	The cost of patch implementation

Table 5: Translating goal statements into effective metrics

## 6. Modeling the security health state of IT assets

As we are intended to build a security predictive decision model we will select security metrics and collect raw data in order to manipulate it and let it fit our model. In this chapter we will introduce our proposed predictive decision modeling; then we will manipulate the data we get from patch en vulnerability management process to fit it to our predictive decision model, after that we will train and test the model to evaluate its performance and its accuracy using the manipulated data set.

### 6.1 Modeling and Evaluation

As defined in <http://en.wikipedia.org/wiki/Predictive>, “predictive modeling is the process by which a model is created or chosen to try to best predict the probability of an outcome. In many cases the model is chosen on the basis of detection theory to try to guess the probability of an outcome given a set amount of input data, for example given an email determining how likely that it is spam. Models can use one or more classifiers in trying to determine the probability of a set of data belonging to another set, say spam or 'ham'.”

An appropriate definition that's best related to our case of using security metrics in predictions is given by Caroline Wong [14], led security teams at Zynga and eBay: “The predictive security model describes the role that security metrics play within an information security program and how this role relates to the other functional areas. The predictive security model is one way of looking at the interaction among different components of an information security program”

From these two definitions we can conclude that security predictive modeling is the process of finding an approach or a model that uses historical security metrics in correlating security events in order to predict the probability of outcome of future behaviors. That is exactly what we will do here; we will collect historical security metrics and then based on our definition of the health state of an IT asset we will use predictive analytics techniques like data mining and machine learning techniques to predict the health state of an IT asset. Basically, I will use the linear regression model to make prediction on the target variable which will be the defined value of the health state of the IT asset in question. Additionally, in order to be able to assess the performance of our models I will use percentage split technique as a model validation technique and also another predictive modeling technique, which is the decision tree model, to make the prediction.

### 6.2 Models description

We are intended here to find an approach that will help measuring and predicting the health state of an IT asset based on the definition of the security health state of an IT asset we have made, the approach should be able to correlate security events to keep track how they influence the IT asset value. If we now go back to our definition of the health state of an IT asset which is as follows:

$V_t = \sum_{i=1}^n \alpha_i * M_i(t) + \beta$  where  $\alpha_i$  and  $\beta$  are constant factors that characterize the direction of the affection on the value of the IT asset for the metric  $M_i(t)$  at the moment  $t$  where  $i=1, \dots, n$ . Then we see that at a given moment  $t$  our model should be able to predict the value  $V_t$  of the IT asset as a linear combination of  $n$  effective security metrics  $M_i(t)$   $i=1, \dots, n$ .

As I have mentioned earlier, using numerical values, that characterizes effective security metrics for the security metrics  $M_i$  will enable us to assume a linear relationship between these security metrics and the value  $V_t$ .

Since we are starting small by using only the patch en vulnerability management process, the idea is thus to use historical security data from this process and then to manipulate it in order to fit it to our model. That is, for each available historical data record that represent a security event, we have to calculate the value of each effective metric that we have got from the process approach of identifying security metrics. In addition, we have to be able to estimate the value of the IT asset in question for each security event. Then for a given moment  $t$ , now or in the future, when a new security event takes place we can apply approaches that are used in statistics, data mining and machine learning to make prediction. One of this approaches that we believe to be useful and appropriate for our definition of the security health of an IT asset as a linear combination of effective security metrics is the linear regression model. In order to evaluate the performance of linear regression model, another data mining and machine learning model, which is the decision tree model will be used to be able to make comparisons and evaluations.

### 6.2.1 Linear regression model

Given a data set  $\{Y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^n$  linear regression theory assumes that there is a linear relationship between the dependent variable  $Y_i$  and  $p$ -vector of the variables  $x_i$ . This relationship can be formulated as follows:

$$Y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n$$

The unknown model parameters  $\beta_i$   $i = 1, \dots, p$ , which called regression coefficients, can be then estimated from the given data set using statistical estimation techniques such as Least-squares estimation and related techniques [18]. The following figure, from Wikipedia [18], gives an example of the result of the regression model using one independent variable.

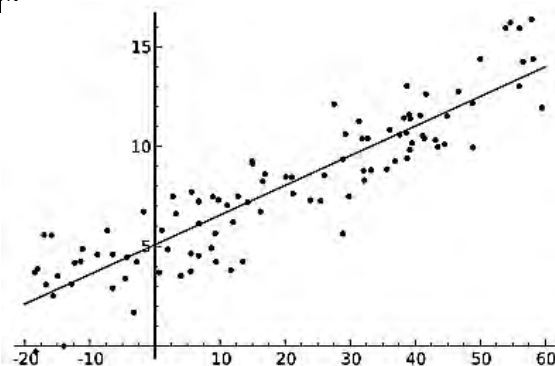


Figure 8: Example of simple linear regression which has one independent variable [18]

After estimating the regression coefficients  $\beta_i$   $i = 1, \dots, p$  we can use them to express and predict the target variable  $Y$  as a linear combination of a given new instance  $(x_{i1}, \dots, x_{ip})$ .

### 6.2.2 Decision tree model

As it's defined in [Wikipedia](#) a decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. The following figure gives an example of a decision tree, as given by [Wikipedia](#) to show the survival of passengers on the Titanic where the numbers under the leaves represent the probability of survival and the percentage of observations in the leaf.

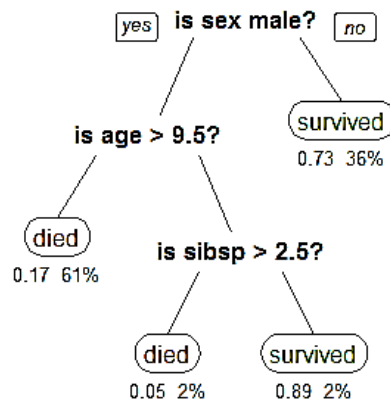


Figure 9: A tree that represents the survival of passengers on the Titanic

To make the decision tree model more clearly we will illustrate here another example as given by Tom M. Mitchell in his book “Machine Learning” [17].

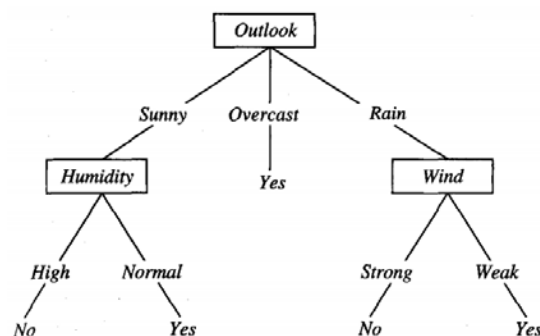


Figure 10: A decision tree for the concept PlayTennis.

In this example the decision tree classifies instances or record by sorting them following a process of top-down induction of decision trees (TDIDT) which is one of the algorithms used to learn a decision trees, given now an instance, each node in the tree will specify a test of some attribute of the given instance to provide a classification of the instance down through the tree which will return the associated predicted value for target variable (in this case of figure 10, Yes or No).

Decision tree is one of the approaches that are used in statistics, data mining and machine learning to make prediction; decision trees are used thus to predict the value of a target variable based on several input variables[16]. That is, given the input source data set as records of the form:

$(x, Y) = (x_1, x_2, x_3, \dots, x_k, Y)$  where  $x_1, x_2, x_3, \dots, x_k$  are input variables that characterizes the target dependent variable Y for each record, the decision tree will be first “learned” by splitting the data set into subsets according to attribute values in each leaf of the tree as figures 2 and 3 shows. To learn decision trees there are several methods and algorithms, in his book “Machine Learning” Tom M. Mitchell discusses the most widely algorithms used in practice such as ID3, ASSISTANT, and C4.5 [17]. After learning the decision tree a given new instance, for which we want to predict the dependent variable Y, will be classified by testing its attributes down through the learned tree which will return the associated predicted value for target variable Y.

### 6.3 Data manipulation

Podictive uses Nessus Vulnerability Scanner, the raw source data from vulnerability scanning looks like the following figure shows.

<b>33850 (1) - Unsupported Unix Operating System</b>
<b>Synopsis</b>
The remote host is running an obsolete operating system.
<b>Description</b>
According to its version, the remote Unix operating system is obsolete and no longer maintained by its vendor or provider. Lack of support implies that no new security patches will be released for it.
<b>Solution</b>
Upgrade to a newer version.
<b>Risk Factor</b>
Critical
<b>CVSS Base Score</b>
10.0 (CVSS2#AV:N/AC:L/Au:N/C:C/I:C/A:C)
<b>Hosts</b>
192.168.1.10 (tcp/0)

Figure 11: An example of record as result of vulnerability scanning

This is just one record from many cases that Nessus provides per vulnerability scanning. The raw source data contains thousands of records as results of monthly vulnerability scans. Each record from the scanning report has deferent characteristics or attributes of the detected vulnerability like the risk factor, CVSS Base Acore [23] and the host ID. Podictive uses Nessus as vulnerability scanner tool which also proposes a remediation solution for the discovered vulnerability as the figure 11 shows for the attribute “Solution”, where an upgrading to newer version of UNIX is proposed as a remediation solution.

I have used Execl to generate a table containing different attribute for all records in the scanning report to be able to perform some analysis on this data. Table 6 shows a snapshot of the resulting vulnerability data which including confidential data given by using letters “x”.



Host ID (IP)	Description	Solution	Risk Factor
x.x.x.4	The PUT method allows an attacker to upload arbitrary web pages on the server.	Disable the PUT and/or DELETE method in the web server configuration.	High
x.x.x.6	The remote host appears to be running a version of Apache 2.x which is older than 2.0.48. Such versions are reportedly affected by multiple vulnerabilities.	Upgrade to Apache web server version 2.0.48 or newer.	Critical
x.x.x.10	The remote host appears to be running a version of the Apache web server which is older than 1.3.29.	Upgrade to Apache web server version 1.3.29 or later.	High
x.x.x.11	The remote host is running Apache web server 2.0.51. It is reported that this version of Apache is vulnerable to an access control bypass attack.	Upgrade to Apache web server 2.0.52 or newer.	High
x.x.x.20	The remote host is running a Compaq Web Management server. The remote version of this software is vulnerable to an unspecified buffer overflow that may allow an attacker to execute arbitrary code on the remote host with the privileges of the web server pr	Upgrade to HP HTTP Server version 5.96 or later or to the System Management Homepage Version 2.0 or later.	Critical
x.x.x.40	According to its version, the remote web server is obsolete and no longer maintained by its vendor or provider. A lack of support implies that no new security patches are being released for it.	Remove the service if it is no longer needed. Otherwise, upgrade to a newer version if possible or switch to another server.	High

Table 4: Snapshot from the resulting vulnerability data

Before using this raw data to make prediction, this data has to be first manipulated and treated to fit the predictive model that will be used, that's because the source data may contain missing values or outliers that may disturb the prediction results. Therefore, a standard step of any data analysis research, which is the treatment of missing values and outliers, will be done before using the raw data for any other purpose. Depending on your analysis goals you may decide, in some cases, not to consider attributes that contains missing values. In some other cases, for example; if there are missing values for several cases on different attributes, then you may decide not to delete those cases (otherwise you will lose a lot of your data.) Generally, there are different other alternative ways of dealing with missing data and outliers in the literatures [26][27]. In the following we will discuss this issue where the missing values and outliers in the source data will be treated to fit the input vector for our predictive model.

### 6.3.1 Missing values and outliers

After analyzing the findings of vulnerability scan, we have concluded that there are deferent cases that may be counted as cause for outliers in the source data which have to be treated and manipulated before using it in making prediction otherwise our predictions will be misleading and not accurate enough. From those cases there are two important issues which are known as "false positives" and "false negatives"; a false positive happens when for example vulnerability does not actually exist but is counted by the scanner in its measurement as an open vulnerability, a false negative is when for example vulnerability does exist but is not counted in a measurement. In order to use the source data correctly false positive records have to be removed from the dataset, but false negatives have to added to the dataset, this kind of data treatment will be done according to the security stakeholders who can identify these cases. From the historically source data and according the security stakeholders (Asset Owner and IT

System Engineer) we have concluded that the most cases like these come from host-based vulnerability scanners and network-based vulnerability. Because we are intended to build a predictive model that needs input data to be generated automatically, a corrective action has to be done automatically during the vulnerability scanning process to prevent those outliers from happening, which can be easily done by fixing and configuring the used scanning tools to keep track of known false positives and false negatives in order to remove or add the necessary data in the future scan results.

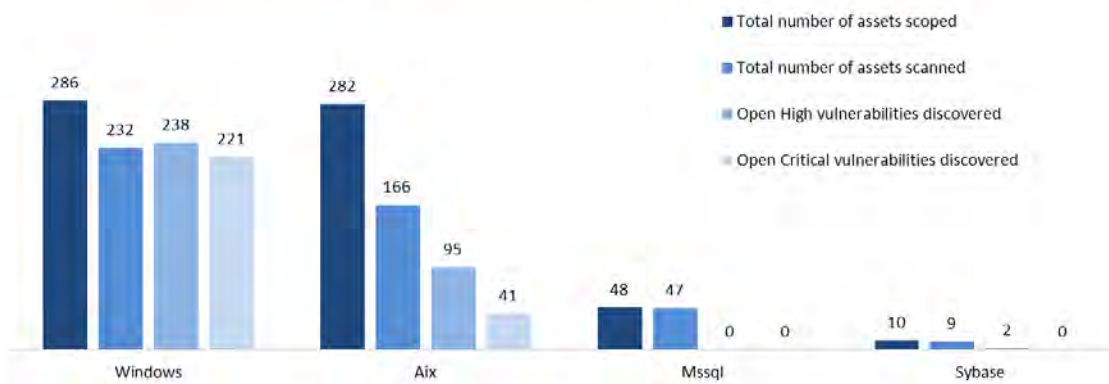
One other case that also may causes outliers in the source data was a case known as “Accepted Risk” where the scanner counts some vulnerabilities with some level of risk whereas from the security manager’s perspective that may be considered as accepted risk, this enforces us to change the risk value for such cases in the source dataset in order to use it correctly in training the predictive model otherwise the result of our predictions will be not accurate and may be not aligned with the consideration of the security managers which may lead to bad conclusions and decisions. Moreover, in order to have good and appropriate input of the model (and also for future use) the scanning tool has to be configured to fix and keep track of the cases of Accepted Risk in order to ensure that this case do not appear as open vulnerabilities in the further scans.

There is some kind of missing values that has to do with the ability of the scanning tool to assign appropriate values to some attributes of the discovered vulnerability. As example; the ability of the scanning tool to propose remediation solution for some vulnerabilities since some vulnerabilities may not have solutions yet at the time of the scanning process like zero-day vulnerabilities. To treat these cases, the missing values will be replaced by the values that have been used by the security managers to identify and solve those problems. For example, to replace the missing values concerning the proposed solution for the discovered vulnerability we asked the security stakeholders (Asset Owner and IT System Engineer) to provide us with this information since they are responsible for the patch management process.

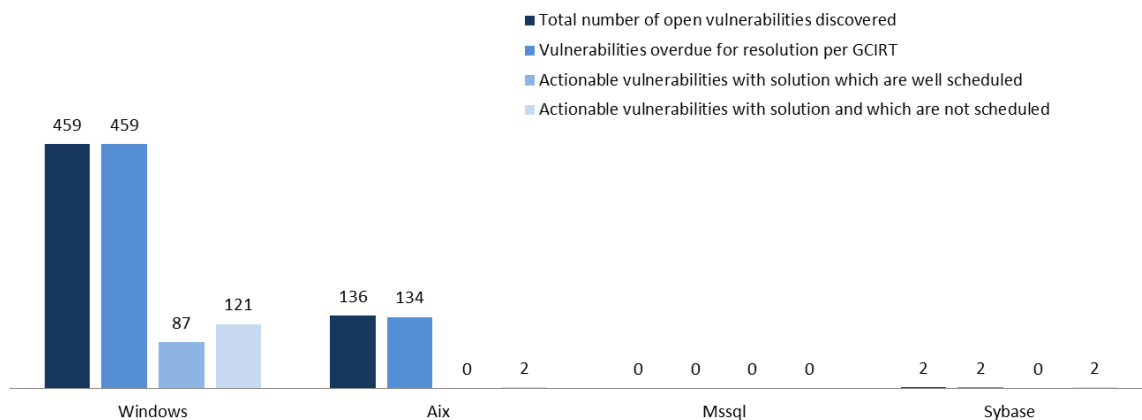
### 6.3.2 Data scoping

Podictive uses the configuration management database (CMDB) as a data warehouse to register all the information concerning IT assets in an organization including the historical vulnerability scanning reports, this data warehouse provides thousands of records concerning different kind of vulnerabilities on different type of IT assets. Because using all the source data from all type of IT assets may lead to unmanageable and uncontrollable project, I have advised the security managers of Podictive to start small by limiting and restricting their attempts using input data about one type IT assets, after that if the model works good with an acceptable accuracy then the model can be extended incrementally to be applied for other type of IT assets. I have discussed this with the security managers, we have decided to apply the predictive model on data from the high critical type of IT assets where the most critical vulnerabilities are discovered. To do that we have done some analysis to decide which high and critical IT asset to consider, the results are given in the following graphs.

## DASHBOARD OPEN VULNERABILITIES FOR HIGH & CRITICAL ASSETS



## Remediation action for open vulnerabilities



From these two graphs we can conclude that the IT assets with Windows platforms are the most critical IT assets with high number of critical discovered vulnerabilities. This leads us to scope the input data for our model only on data from IT assets that use Windows as platforms.

### 6.3.3 The input data set description

Using the available source data from IT assets with Windows platforms, we have calculated for each record in the dataset, which means for each security event, the value of each security metric that I have determined in the metric process approach which are given in the following table.

M1	Number of open "Low" severity vulnerabilities
M2	Number of open "Medium" severity vulnerabilities
M3	Number of open "High" severity vulnerabilities
M4	number of "Low" severity vulnerabilities overdue (still open)

M5	Number of "Medium" severity vulnerabilities overdue (still open)
M6	Number of "High" severity vulnerabilities overdue (still open)
M7	Cost impact for open vulnerabilities before remediation process
M8	Cost impact for open vulnerabilities after remediation process
M9	Cost of remediation and patch management process

Table 7: The definition of security metrics that will be used as input for the model

Now we have security metrics, the question is which value we have to use to evaluate the health state of an IT asset? Basically we can use any numerical variable that the IT security stakeholders may consider to characterize the health state of their IT assets and which we supposed to have a linear relationship with the identified security metrics. For this purpose we have defined a new metrics M10 which will be modeled in terms of the others. This metric M10 will be defined as follows:

$$M10 = -[M4 * I_{overdue1} + M5 * I_{overdue2} + M6 * I_{overdue3} + ((M4+M5+M6)/(total\ open\ vulnerabilities - (M4+M5+M6))) * remediation\ cost]$$

Where  $I_{overdue1}$ ,  $I_{overdue2}$  and  $I_{overdue3}$  are impact of low, medium and high overdue severity vulnerabilities respectively.

The security managers of Podictive have considered this term to be an acceptable characteristic of the health state of the IT asset for the moment, because it is a function of the overdue vulnerabilities after remediation process and the costs of remediation. This metric is considered to be significant as measure for the security health state of the given IT asset because the bigger the number of the saver vulnerabilities with higher remediation cost the riskier the IT asset is. We use here a negative function to indicate the impact of the overdue open vulnerabilities and the remediation cost on the value of the IT asset.

The input dataset generated from the source data using these metrics looks like the following table shows.

Event ID	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
1355	6	5	1	6	0	2	0	1000000	130000	-3378667
1356	12	6	0	3	0	1	0	1000000	150000	-1697500
1357	7	4	2	1	1	0	0	500000	100000	-722727
1358	2	4	0	3	4	3	0	1000000	60000	-5552857
1359	8	3	5	5	5	1	0	1000000	70000	-4602222
1360	5	3	1	1	1	0	0	500000	210000	-723750
1361	6	2	4	1	0	1	0	1000000	130000	-1203571
1362	6	4	5	1	0	2	0	1000000	170000	-2160000
1363	2	2	2	1	1	2	0	1000000	80000	-2652222
1364	2	6	0	3	1	3	0	1000000	30000	-4122500
1365	0	2	2	1	0	1	0	1000000	240000	-1207143
1366	5	6	2	1	0	0	0	250000	30000	-240000
1367	1	1	1	2	4	3	0	1000000	160000	-5292941
1368	5	4	0	2	0	0	0	250000	120000	-480000
1369	3	5	0	2	0	0	0	250000	90000	-482000
1370	3	7	7	2	2	0	0	500000	60000	-1448571
1371	5	2	2	1	4	0	0	500000	180000	-2172857

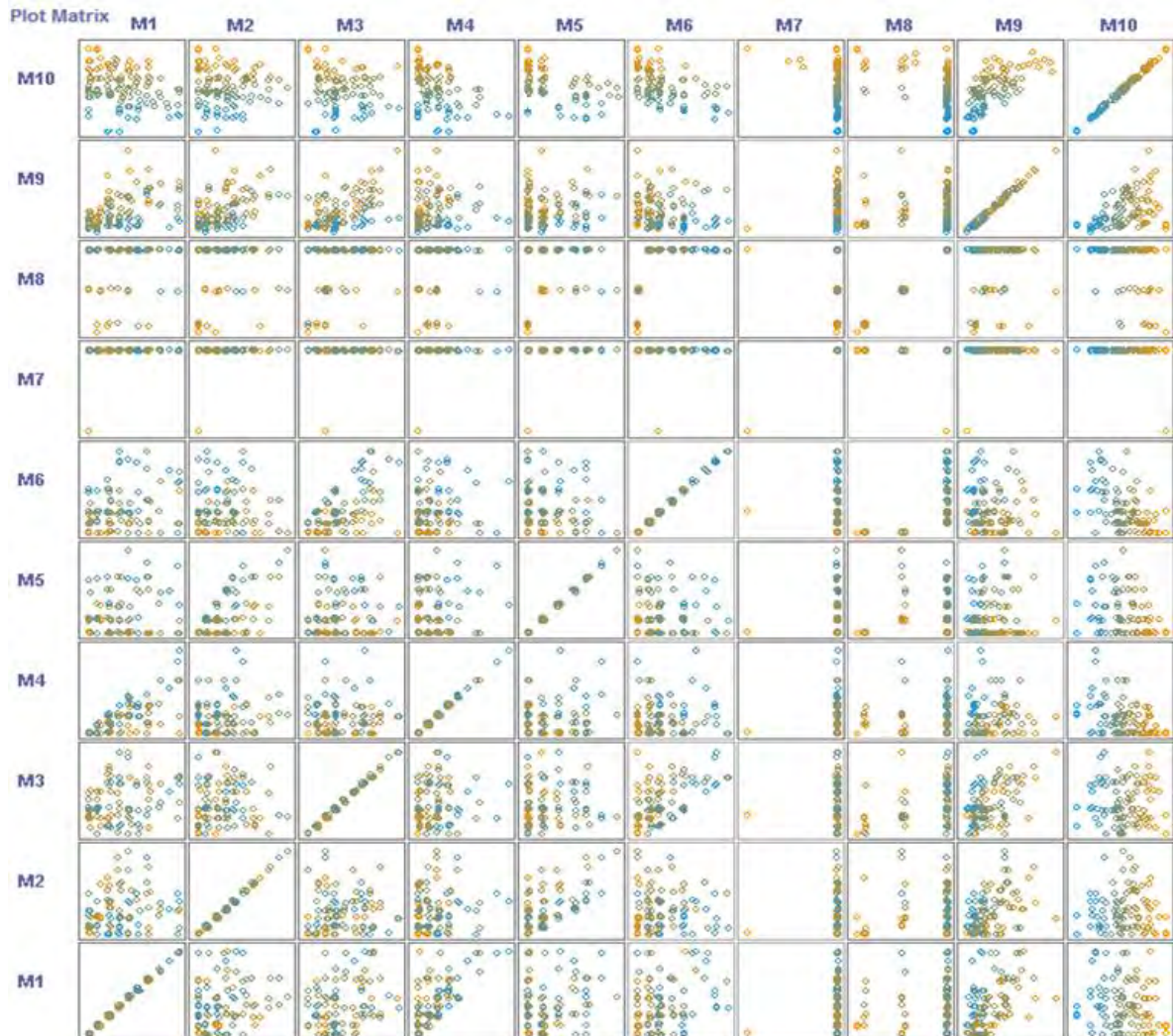
Table 8: A snapshot from the generated dataset using the defined security metrics

In this table each record (security event) contains the attributes M1, M2 and M3 which denote respectively the numbers of "Low", "Medium" and "High" severity vulnerabilities discovered in the IT asset before the remediation process is executed. The record contains also M4, M5 and M6 which are respectively the number of "Low", "Medium" and "High" severity vulnerabilities that are still open after the remediation process. Moreover, cost of impact before remediation, cost of impact after remediation and the total cost of remediation is also given respectively by M7, M8 and M9. At the end, we have calculated for each record/event the value of M10 according to the formula given above. Now, given a new security event characterized by M1, M2, and M3 we want to get some idea what would the security health state of the IT asset in order to make effective vulnerability remediation instead of randomly remediating! Therefore, if we could predict the health state of the IT asset based on the known number of vulnerabilities in the system, then we would be able to perform cost effectively remediation by looking for optimal prioritization of vulnerability remediation.

Our purpose is to use the linear regression model as a predictive model to predict the response dependent variable M10 base on the three variables M1, M2 and M3 the number of "Low", "Medium" and "High" severity vulnerabilities discovered in the IT asset before the vulnerability remediation process starts. These variables are supposed to be independent in order to apply the linear regression model appropriately. In the next section we will get started with analyzing the independency of different variables especially the metrics M1, M2 and M3.

#### 6.3.4 Data correlation

To get insight how the correlation between different metrics is, I have used the data mining software Weka to visualize a scatterplot matrix of the 10 security metrics against each other, the result is the following.



As we can see from this scatterplot matrix, it is obvious that the metrics M1, M2 and M3 are not correlated between each other but they are respectively showing some weak correlation between the metrics M4, M5 and M6 which are also not correlated between each other. The metric M9, which represents the cost of vulnerability remediation, is clearly correlated with the security metrics M1, M2, M3 and M10, but this correlation seems to be not strong enough to conclude interdependency between the variables; the correlation between M9 and M10 maybe normal since we define M10 as function of the remediation cost M9. But the correlation between M9 and M1, M2 is not easy to be confirmed just from this plots that's why we go further with analyzing this correlation in the following section using a more powerful technique which is the correlation matrix.

To get more insight how the ten metrics are exactly correlated I have used the Excel function CORREL(array1;array2), which calculate the correlation coefficient between two variables array1 and array2, to get the following correlation matrix.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
M1	1,000	0,008	0,025	0,629	0,034	0,049	-0,038	0,021	0,317	0,043
M2	0,008	1,000	0,332	-0,004	0,634	0,235	0,173	0,256	0,467	0,045
M3	0,025	0,332	1,000	0,024	0,243	0,698	0,451	0,469	0,463	0,142
M4	0,629	-0,004	0,024	1,000	0,025	0,034	-0,027	0,023	-0,159	0,503
M5	0,034	0,634	0,243	0,025	1,000	0,159	0,123	0,249	0,072	0,522
M6	0,049	0,235	0,698	0,034	0,159	1,000	0,307	0,600	0,080	0,576
M7	-0,038	0,173	0,451	-0,027	0,123	0,307	1,000	0,470	0,210	0,082
M8	0,021	0,256	0,469	0,023	0,249	0,600	0,470	1,000	0,095	0,152
M9	0,317	0,467	0,463	-0,159	0,072	0,080	0,210	0,095	1,000	0,652
M10	0,043	0,045	0,142	0,503	0,522	0,576	0,082	0,152	0,652	1,000

Table 9: The correlation matrix for correlation between different security metrics

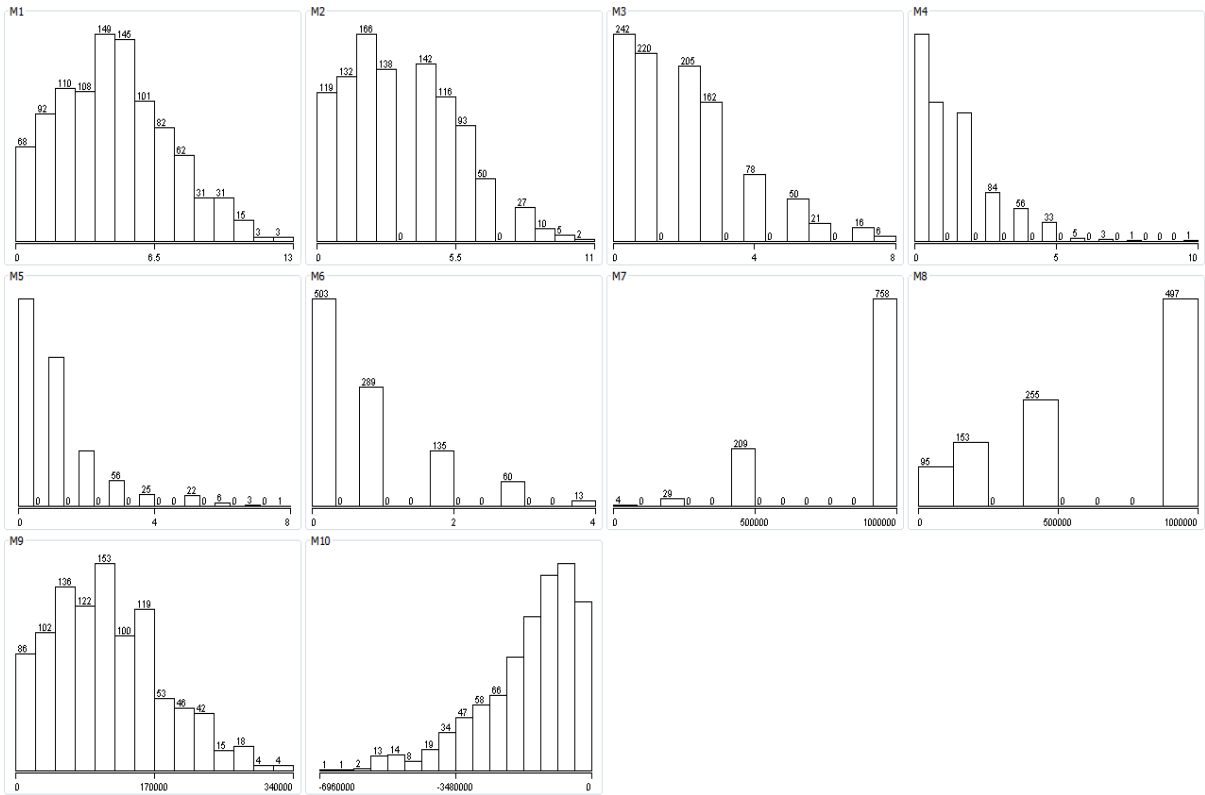
If we look to the correlation matrix we see that correlation coefficients between the security metrics are not very significant and some of them are even negative which means that some of the defined metrics are moving against each other. We can see some significant correlation between some metrics but this correlation may be not strong enough to be able to confirm that these metrics are influencing each other. For example, metrics M1 and M4, M2 are having respectively a correlation coefficients greater than 0,6 with M5 and M3 and M6 colored in red in the table, which means that they are influencing each other in some way; according to our definition of those metrics, the metrics M4, M5 and M6 are respectively the numbers of "Low", "Medium" and "High" severity vulnerabilities overdue or vulnerabilities that are still open after the remediation process, is normal since the overdue vulnerabilities depends on the numbers open vulnerabilities discovered on the system.

Moreover, there is a kind of slightly negative correlation between the metrics M4, M5, M6, M9 with the metric M10, colored in yellow in correlation matrix above, this correlation is good enough that is exactly what we expect since M10 is function of this metrics.

As we are previously claimed, there is some correlation between M9 and M1, M2 and M3 but the correlation coefficients are under 0.5 which make it difficult to confirm a relationship between this metrics. We need to get more insight about the interdependency between all metrics before using it to make prediction therefore we need to do more analysis on this metrics using statistical analysis.

### 6.3.5 Data analysis

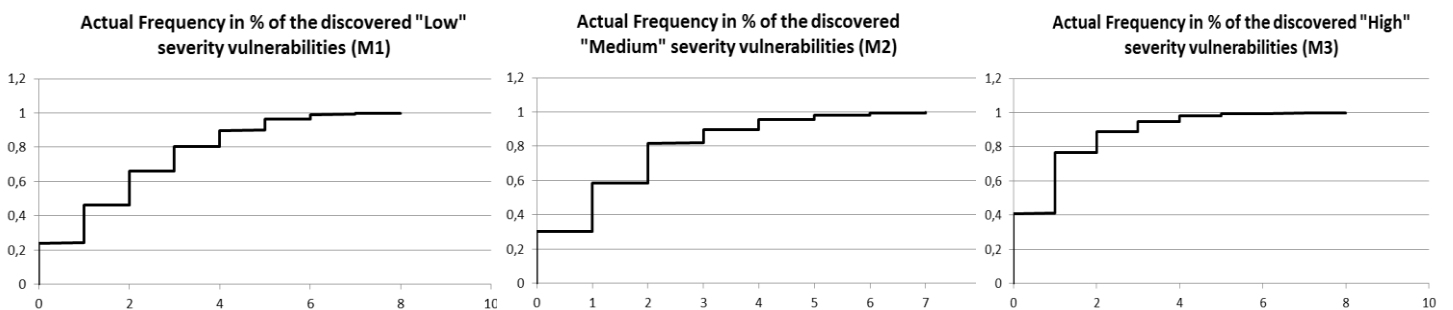
To start the data analysis, I have used the data mining software Weka to visualize the distribution of each metric; the result is the following histograms for each metric.



As we can see in the plots, the distributions of the metrics M1, M2, M3, M4, M5, M6 and M9 are showing almost the same trend; where the most values of those metrics in the data set (more than 60%) are relatively small values. For the first three metrics, which are respectively the number of discovered "Low" severity vulnerabilities, the number of discovered "Medium" severity vulnerabilities and the number of discovered "High" severity vulnerabilities, this is normal because there are no cumulative open vulnerabilities due to the monthly patch and vulnerability management process where Podictive try to remediate all vulnerabilities in the system.

	M1	M2	M3
Mean	4,626	3,498	2,023
Median	5	3	2
Standard Deviation	3	2	2
Minimum	0	0	0
Maximum	14	11	8

Table 10: Statistical results about the discovered vulnerability metrics





From table 10, that shows some statistical results about the vulnerability metrics M1, M2 and M3 and from the three cumulative distribution plots here above we see that almost 80% of each type of the vulnerabilities below the mean. Using similarity in the plots to conclude that these variables are dependent will have no sense, especially because the correlation coefficient between them is not significant at all as the correlation matrix above shows. The only conclusion that we can drive from this observation is that the discovered vulnerabilities are not normal distributed since the normal distribution is characterized by the mean is equal to the media. One more way to ensure the independency of the metrics M1, M2 and M3, which we are intended to use as predictive variables for our proposed linear regression model, was to ask the security managers of Podictive since they have more experience and know how things happen on their systems, they have confirmed the results given by the correlation matrix about the independency of M1, M2 and M3; by ensuring that the dataset we use is generated from different asset with different windows platforms which are in the most cases not connected directly to each other. There are some cases where vulnerabilities of different types may cause each other but these are exceptions that rarely happen and therefore not to be taken in account.

The second three metrics M4, M5 and M6 are having a correlation coefficients greater than 0,6 with the metrics M1, M2 and M3 as given in the correlation matrix in red above, this relationship is also confirmed in the histogram plots above; we see that these variables have respectively almost the same behavior. The metric M9 which represent the remediation cost has the same behavior as these previous metrics which may confirm the results given by the correlation coefficients, that's because the remediation cost depends strongly on the number and the nature of the vulnerabilities discovered on the system.

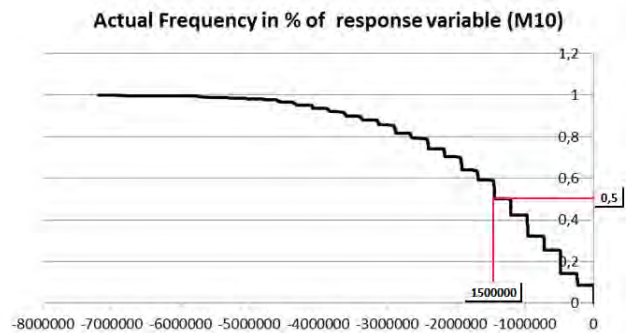
The remarkable one is the metric M10, which we will use as the response variable that has to be predicted using the M1, M2 and M3, this metric seems to behave against the metrics M1, M2, M3, M4, M5, M6 and M9 since it is a negative function of this variables where.

$$M10 = -[M4 * Ioverdue1 + M5 * Ioverdue2 + M6 * Ioverdue3 + ((M4+M5+M6)/(total\ open\ vulnerabilities - (M4+M5+M6))) * remediation\ cost]$$

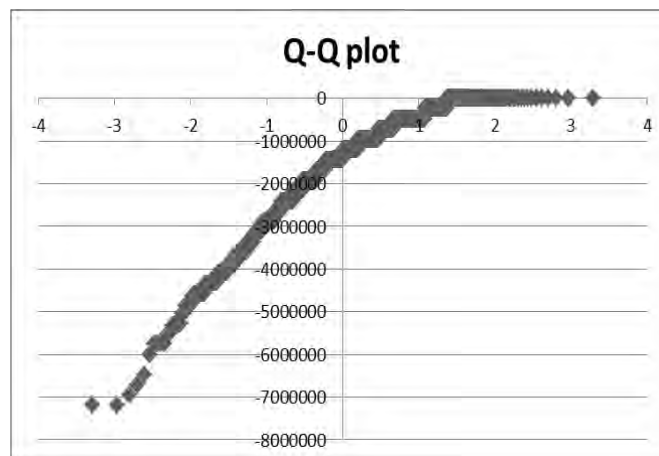
Where Ioverdue1, Ioverdue2 and Ioverdue3 are respectively the impact of low, medium and high overdue severity vulnerabilities.

The definition of M10 contains thus two variable terms which are in turn functions of the metrics M4, M5 and M6, the total number of vulnerabilities and the remediation cost. The purpose of the monthly patch and vulnerability management process is to try to remediate all discovered vulnerabilities at time to minimize the impact on the health state of the IT asset, but since this in the most cases not possible then in the course of the time the value of M10 will tend to be in the middle in between some minimum and the maximum value around.

M10	
Mean	-1614293,269
Median	-1212500
Standard Deviation	1306487,86
Minimum	-7200000
Maximum	0
Confidence Level(95,0%)	81073,69348

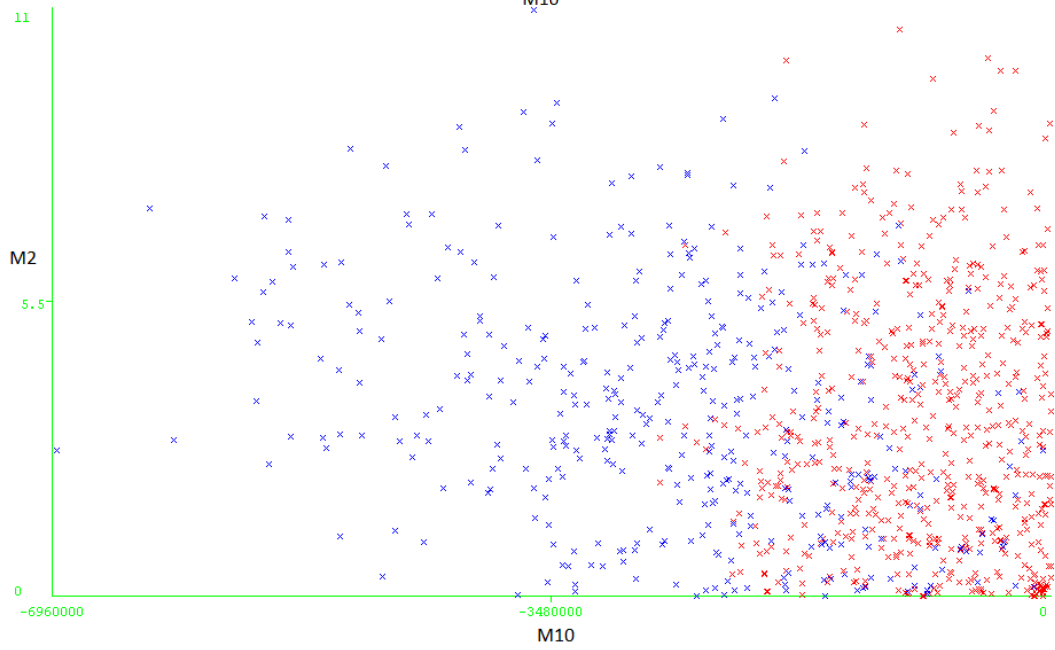
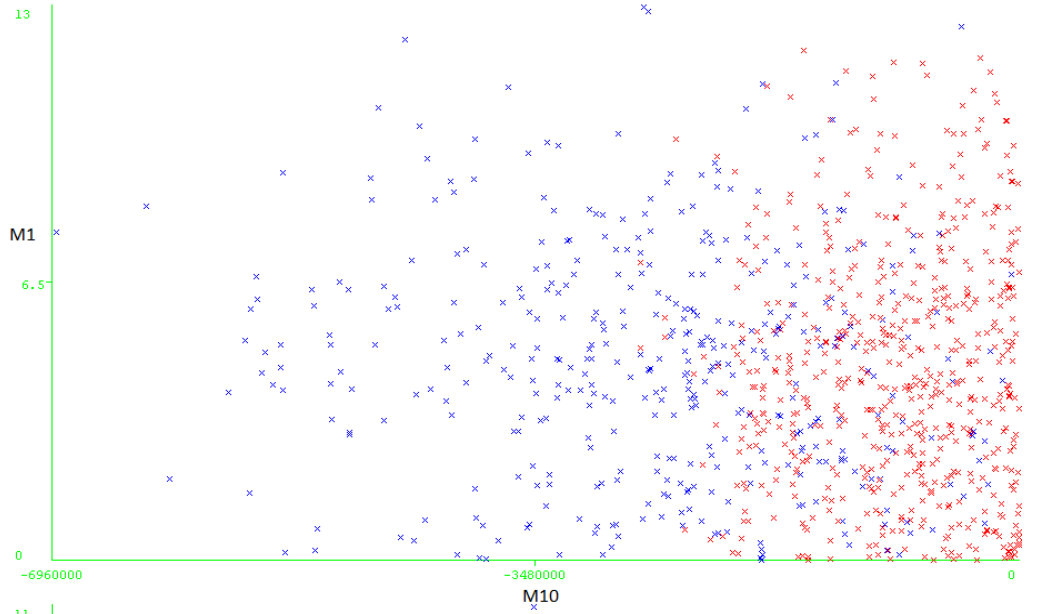


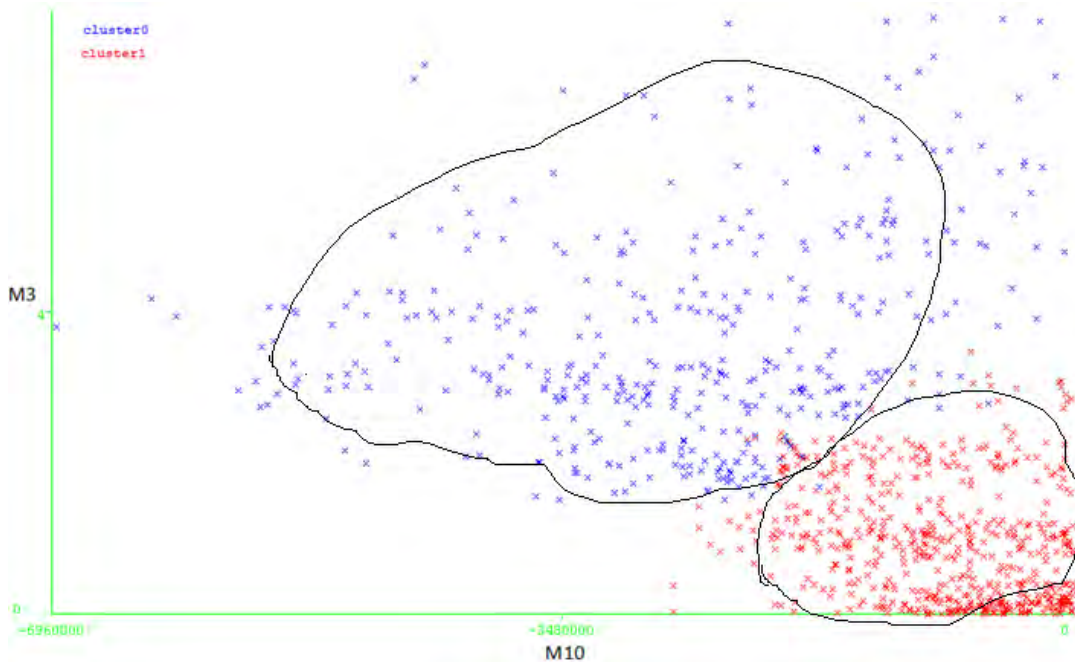
From the statistical results and the cumulative distribution plots of the metric M10 we see that almost 50% of the data is above the mean which is -1614293, this means that the distribution of metrics M10 may be closed to normality even if it is not clear from histogram of M10 above. To have more insight about the normality of M10 tet further examine this issue using another statistical technique which is Q-Q Plot, the following is the Q-Q plot for M10 generated in Excel.



As we can see the resulting Q-Q plot is roughly a straight line with a positive slope which means that our observation indeed is true and M10 follows approximately a normal distribution. The normality of the metric M10 may help us seeing something about the health state of our IT asset if the calculated value of M10 based on the realized metrics M4, M5, M and M9 shows anomalies from the descriptive statistics of normality of M10.

Now, before starting applying the predictive model, we are interested in knowing how well each independent variable, M1, M2 and M3 predicts the dependent variable M10. For this purpose I have used Weka to cluster on M10 based on each of the metrics M1, M2 and M3, the results are shown in the following graphs.

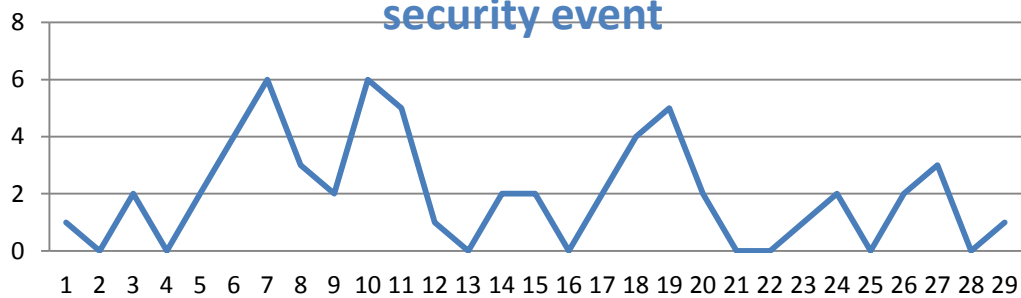




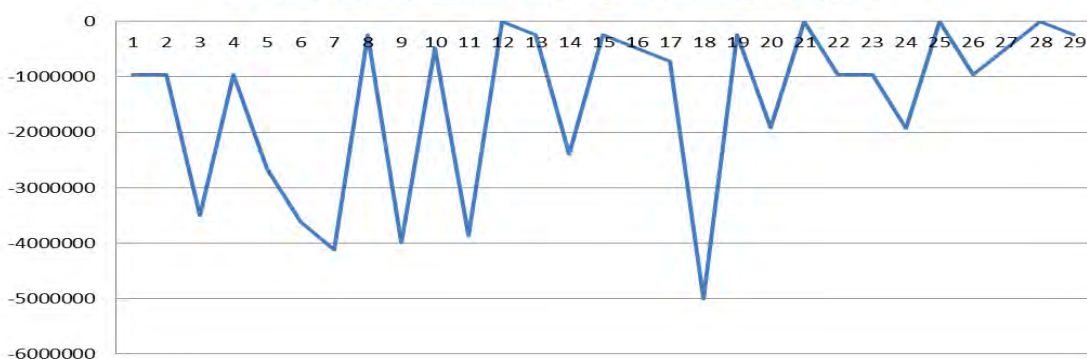
From the two first plots which are clustering M10 according to M1 and M2 we see no remarkable reason to say that these metrics may be significant predictor since there are no clear clusters of data depending on a certain values of M1 or M2. But if we look to the last plot, which is for M10 based on M3, we can see obviously that there are two clusters; one is determined by small values of M3 against small values of M10 and the other gives relatively big values of M3 against relatively big values of M10, which means that M3 representing high critical vulnerabilities may play a roll of significant predictor in our predictions.

To get more insight how the number of high critical vulnerabilities is influencing the value of M10, I have plotted the number of critical vulnerabilities per security event M3 and M10 separately as the following two graphs show.

### Number of critical vulnerabilities per security event



### Asset Value affected by security events



Comparing these two graphs we can see obviously that our asset value M10, that we have defined to represent the health state of the IT asset, moves against the number of critical vulnerabilities discovered in the IT asset, that is: when the number of critical vulnerabilities goes up the It asset value fall down and when the number of critical vulnerabilities fall down the It asset value goes up; this means that the number of critical vulnerabilities on the system is influencing negatively the IT asset value which is logic and normal. We want therefore to predict the IT asset value before remediation actions in order to effectively prioritize remediation. This observation can be also useful and illustrative letter to verify whither our model good or bad, because if the model would not be able to show and hold this logic relation between these two variables then it will be not convenient for making prediction.

## 6.4 Models application

The essential goal is to use the prediction results in order to take preventive actions against bad impact on the security health state of the IT asset; one of possible actions is to search for an optimal prioritizing of the vulnerabilities remediation that gives the minimal impact on the IT asset value and minimal remediation costs. To apply the models described above I have used the data mining software Weka which provide different data mining and machine learning techniques for predictions. For each model used, the same input data set will be used as a training dataset, that has four attributes namely M1, M2, M3 and M10 (the target class), the training dataset will be used in a fitting process in order to optimize the predictive model parameters to make the model fit the training data. Furthermore, we will train the models on 80% of the data set and test on the rest which is 20%.

### 6.4.1 Applying the Linear regression model

To start using the linear regression model which, I believe, will be useful and appropriate for our definition of the security health of an IT asset which will be expressed as a linear combination of effective security metrics:  $V_t = \sum_{i=1}^n \alpha_i * M_i(t) + \beta$  where  $\alpha_i$  and  $\beta$  are constant factors that have to be determined in order make predictions. To determine the model parameters  $\beta$  and  $\alpha_i$  for  $i=1, 2, 3$  I have used the input data set with percentage split technique of 80% of the data set as training set and

```

Linear Regression Model

M10 =

-25526.3834 * M1 +
-77670.8907 * M2 +
-351505.7596 * M3 +
-520966.7133

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===
  
```

20% of the data as test data to be able to evaluate the model accuracy. In our case  $V_t$  is M10 and  $M_i(t)$  are M1, M2 and M3. Using the available input data set and using the linear regression model in the data mining software Weka I have determined  $\alpha_i$  for  $i=1, \dots, 3$  and  $\beta$  to express the target value of M10 as function of M1, M2 and M3 as result here next shows (for more results see [Appendix III](#)):

As we can see from this result the

target class  $M10 = \sum_{i=1}^3 \alpha_i * M_i(t) + \beta$  is expressed as function of the metrics M1, M2 and M3 and  $\alpha_i$  and  $\beta$  for  $i=1, 2, 3$  where

$$\alpha_1 = -25526.3834$$

$$\alpha_2 = -77670.8907$$

$$\alpha_3 = -351505.7596$$

$$\beta = -520966.7133$$

### 6.4.2 Model Evaluation

To evaluate the regression model, it is worth highlighting that in a prediction problem, a model is usually trained on a known data set (training dataset), and tested on an unknown data (or first seen data the testing dataset). The goal of the percentage split technique of Weka that we have used is to define a dataset to "train" the model in the training phase (i.e., the validation dataset) and a "test" data set in order to assess the accuracy of the model based on prediction of the unknown values of M10 in the test data set. Moreover, to get more insight about the accuracy of the regression model I have applied the Decision tree model on the same dataset with the same option of percentage split technique (80% of the data set is used as training set and 20% of the data is used as test data set). Using the decision tree model (REPTree) in Weka, I have obtained the following results (for more results and the visualization of the tree see [Appendix III](#)):

Linear Regression Model		REPTree	
=== Evaluation on test split ===		=== Evaluation on test split ===	
=== Summary ===		=== Summary ===	
Correlation coefficient	0.605	Correlation coefficient	0.6625
Mean absolute error	783492.9724	Mean absolute error	740431.8893
Root mean squared error	1031188.3015	Root mean squared error	964176.4462
Relative absolute error	76.1057 %	Relative absolute error	71.9229 %
Root relative squared error	81.2578 %	Root relative squared error	75.9772 %

### 6.4.3 Discussion

As we can see from the result of our regression model, the correlation coefficient based on prediction done on the test data set of 20% of the whole original data set is 0.605 which means that the regression model can be relatively considered as good prediction model. The result of the correlation coefficient obtained by the decision tree model is quite higher than that of the regression model which means that the decision tree model may be more appropriate for the health state of an IT asset. But as we concerned, the result given by the decision tree just confirm the ability of our regression model to predict the health state of an IT asset. Even if the correlation coefficients obtained using the two models are not very high enough we can conclude that they are reasonably significant enough. We are taking this consideration because of many reasons that having to do with the nature of the problem and different other circumstances; first of all we have only applied these models using source data and effective security metrics from one IT security process while there are other effective security metrics for many other IT security process that have to be involved in order to making the prediction ability of the model more powerful, this because predictive models are based on logical classification and logical reasoning, which means looking only at one side of the

problem implies a lack and shortage of information which in turn lead to insufficiency in classification and therefore in predictions. Second, since the question was to base the model on effective security metrics, I have tried to remain in the context and use only the identified numerical effective security metrics in an appropriate predictive model, while there are many other IT asset security characteristics or metrics, like vulnerability name and vulnerability type, that may help strangely in performing predictions. Finally, this is the first attempt that Podictive does in using predictive modeling in performing IT security processes, that's why this result can be considered as a good enough result which has to be performed by adding more effective metrics from other IT security processes and other IT asset security characteristics.

#### 6.4.4 Decision making

Let's suppose for instance that we have identified enough effective security metrics from different IT security processes for our predictive model and that we could now predict the health state of our IT asset more precisely, then the question would be how to use the prediction results in order to make decision and prevention actions. One way to do that is to look for an optimal prioritization of vulnerability rededication that minimizes the impact on the security health state of your IT asset. The remediation prioritization action may also be based on the same predictive model, which may be used for multifunction rolls depending on the input data and the target class to be predicted! To explain this I will use again our example of input data from the patch and vulnerability management process given in the following table:

Event ID	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
1355	6	5	1	6	0	2	0	1000000	130000	-3378667
1356	12	6	0	3	0	1	0	1000000	150000	-1697500
1357	7	4	2	1	1	0	0	500000	100000	-722727
1358	2	4	0	3	4	3	0	1000000	60000	-5552857
1359	8	3	5	5	5	1	0	1000000	70000	-4602222
1360	5	3	1	1	1	0	0	500000	210000	-723750
1361	6	2	4	1	0	1	0	1000000	130000	-1203571
1362	6	4	5	1	0	2	0	1000000	170000	-2160000
1363	2	2	2	1	1	2	0	1000000	80000	-2652222
1364	2	6	0	3	1	3	0	1000000	30000	-4122500
1365	0	2	2	1	0	1	0	1000000	240000	-1207143
1366	5	6	2	1	0	0	0	250000	30000	-240000
1367	1	1	1	2	4	3	0	1000000	160000	-5292941
1368	5	4	0	2	0	0	0	250000	120000	-480000
1369	3	5	0	2	0	0	0	250000	90000	-482000
1370	3	7	7	2	2	0	0	500000	60000	-1448571
1371	5	2	2	1	4	0	0	500000	180000	-2172857

In this table we see the vulnerability security metrics M1, M2 and M3 that we have used previously to predict the target class metric value M10 which is calculated based on the remediation metrics M4, M5 and M6 and the remediation cost metric M9. These all metrics are calculated based on historical data from different security events that have been happened depending on different security situations and different reactive actions of security managers. If we have now a new vulnerability security event with new metrics value M1, M2 and M3, which characterizes the number of "Low", "Medium" and "High" severity for the discovered vulnerabilities from the vulnerability scan, then

we will run our model to predict the metric M10 which represent the security health state of our IT asset. Depending on the prediction results, security managers may take regular remediation actions, but in the case of a new situation they may not have any idea how to perform cost effectively remediation prioritizations, in this case they should use the predictive model trained on a different input data that have to have attributed on M1, M2, M3, M4, M5, M6 and M9, which is the remediation cost, as the target class. After training the model, we take the new values of M1, M2 and M3 from the new security and then we assign to them all possible remediation metrics M4, M5 and M6. The trained predictive model will then predict, for each situation, the target value M9, the smallest predictive value of M9 will give then the optimal predictive remediation cost which will associated to a given remediation metrics M4, M5 and M6 in the input data, the value of these metrics will help then the security managers to decide how to prioritize the remediation of the new security event.



## 7. Conclusion and recommendations

The purpose of this internship, as we have stated in the problem description, was to find a theoretical (conceptual) decision model that can be used to answer the question "How to identify the security health state of an IT asset based on a decision model?". To be able to answer this question we have break down the central question into three sub-questions:

1. What is the definition for a security health state of an IT asset?
2. How are security metrics being identified in a process approach?
3. How to build a decision support system/model that can help in the measurement of a security health state of an IT asset?

### 7.1 Conclusion

In this section I will present the final conclusions as answers to the questions and finally give answer to the central question.

#### *1. What is the definition for a security health state of an IT asset?*

To answer the first question about the definition of security health state of an IT asset, a description of different definitions of fundamental subject and different IT security terms was needed to be able to derive an appropriate definition for the security health state of an IT asset. Based on the definition of all the terms needed we could derive our final definition based on the value of the IT asset and security events that are attributed with possible effective security metrics. The definition is then formulated in a mathematical expression to fit the chosen predictive model which is the linear regression model as follows: given effective security metrics  $M_1(t), \dots, M_n(t)$  that characterize a security event at the moment  $t$  the security health state of an IT asset can be expressed in value  $V_t$  as function of the given security metrics as follows:

**$V_t = \sum_{i=1}^n \alpha_i * M_i(t) + \beta$  where  $\alpha_i$  and  $\beta$  are constant factors that characterize the direction of the affection in the value of the IT asset at the moment  $t$  where  $i=1, \dots, n$ .**

#### *2. How are security metrics being identified in a process approach?*

The answer of the second question was to create a process approach for identifying effective security metrics. I have answered this question in chapter 4 where I have given a process design and description of a metric process by extending the well-known Goal-Question-Metric approach that fit in the IT security context with some extra steps and features. In the last step of this process I have provided, in section 5.6, a list of effective security metrics for the considered IT security process, which is patch and vulnerability management process. The list of effective security metrics is used as input for the applied predictive model in order to predict the security health state of an IT asset.

3. *How to build a decision support system/model that can help in the measurement of a security health state of an IT asset?*

The last question was answered by proposing the linear regression model, in chapter 6 in section 6.2.1, as an appropriate model to make predictions, since this model works good with numerical values and our identified effective security metrics are all of numerical nature but has to be of numerical nature to be effective. To be able to evaluate our proposed predictive model I have used historical data to calculate the identified effective security metrics in order to train the model. To evaluate the performance of our model I have applied another predictive model, namely the decision tree model. By comparing the results provided by Weka, the data mining software I used to apply the models, we have conclude that our linear regression model performing reasonably well if we take a number of considerations as given in discussion section 6.4.5.

*How to identify the security health state of an IT asset based on a decision model?*

Finally, to answer our central question we can state that finding an appropriate definition of the security health state of an IT asset, that can be expressed mathematically and numerically in order to be able to do calculation will be the key of identifying the security health state of an IT asset. Another key factor in identifying the security health state of an IT asset is the identification of effective security metrics in a process approach and then to express the value of the security health state of the IT asset as function of those effective security metrics, and then to use historical security data to calculate the security metrics in order to generate an appropriate input data set for the predictive model used.

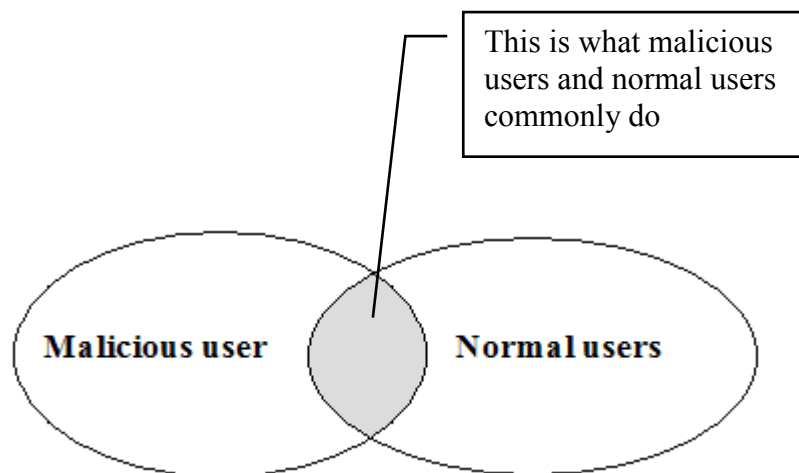
At the end, through this research study I have tried to contribute to improving IT security by providing a complete security approach that starts by analyzing the context of IT security and then identifying effective security metrics in a process approach and then at the end using the identified security metrics as input data for a predictive model in order to make IT security predictions. The advantage of this approach is that other researchers can benefit further from this effort by applying this approach to any other fields where effective security metrics are needed be combined with security predictive modeling in order to achieve better IT security improvements since this approach in not depending on terms defined in advance, but terms definition is part of the approach.

## 7.2 Recommendations

I believe that data and information is the core of any predictive model. Any predictive security model must start beside a process that performs IT security information gathering and manipulation. Not only from vulnerability scans, but also from different other IT security processes like security monitoring, penetration testing or asset value and risk assessments. I recommend management of Podictive to proceed further with this project by executing the process of identifying effective security metrics on other IT security processes as well in order to gather sufficient and effective input data for the predictive model explained in this research. The more effective the security metrics you have the more powerful and accurate the predictive model will be.

Gathering and manipulating effective security metrics data is the core of any model that will transform data to knowledge and then into action. Gathering security data will add no value to an organization if it's not used for actionable decisions and initiation of preventive actions. The purpose of collecting security data for the use of predicting the future is to prevent worse things to happen and let the organization stay out of trouble. Therefore, collecting IT security data without using it does not mean that you are secured but it means instead that you are unsecured and just adding more to your risk and uncertainty. Having important, sensitive and effective IT security data of your IT asset, registered somewhere on your system, may become itself an IT asset that needs to be secured from malicious activities. Therefore I recommend Podictive to focus only on critical data that can lead to concrete results to lower the IT risk level of the organization, because focusing on anything else will lead to distraction and unnecessary investments. Therefore, if you think that collected security data is not to be used in future predictions then IT assets will maybe more secured by removing it.

Looking for specific properties that characterize specific vulnerabilities for example in order to drive predictions will be difficult since the nature of vulnerabilities is continuously changing. However, focusing on identifying effective security metrics that characterize normal activities will help more in predictions since no matter how a malicious user activity will be, it will still differentiating from the activity of a normal user. Classifying security data according to this fact and then using the logic behind the intersection of different data sets, what is not included in B is absolutely included in A as the following figure shows, will be a potential predictive key factor.



## Appendix I References

- [1]. DR. Eric Cole, Advanced Persistent Threat: Understanding the Danger and How to Protect Your Organization, ISBN-13: 978-1597499491
- [2]. Lance Hayden Ph. D. IT Security Metrics: A Practical Framework For Measuring Security and Protecting Data, ISBN: 978-0-07-171341-2.
- [3]. [Jeff Laskowski, Agile IT Security Implementation Methodology, First published: November 2011, ISBN 978-1-84968-570-2]
- [4]. [PAS 555, Cyber security risk-Governance and management- Specification, by BSI (The British Standards Institution 2013), ISBN 978 0 580 78755 3]
- [5]. Andrew Jaquith, Security Metrics REPLACING FEAR, UNCERTAINTY, AND DOUBT ISBN 0-321-34998-9
- [6]. [http://en.wikipedia.org/wiki/Information\\_security](http://en.wikipedia.org/wiki/Information_security)
- [7]. Richard Kissel, Glossary of Key Information Security Terms Editor, Computer Security Division, Information Technology Laboratory, NISTIR 7298 Revision 2
- [8]. Jhn Wunder, , Adam Halbardier, David Waltermire, Specification for Asset Identification, NIST (National Institute of Standards and Technology) Interagency Report 7693
- [9]. ITIL (Information Technology Infrastructure Library) glossary and abbreviations <http://www.itil-officialsite.com/InternationalActivities/TranslatedGlossaries.aspx>
- [10]. Jonathan Berk and Peter, DeMarzo Corporate Finance, ISBN-13: 978-0132745093
- [11]. THE GOAL QUESTION METRIC APPROACH, Victor R. Basili, Gianluigi Caldiera, H. Dieter Rombach, Institute for Advanced Computer Studies, Department of Computer Science, Department of Computer Science
- [12]. Tom Palmaers, Implementing a vulnerability management process, SANS institute InfoSec Reading Room, Accepted: 03/23/2013
- [14]. Caroline Wong, Security Metrics: A Beginner's Guide, ISBN: 0071744002 / 9780071744003, October 20, 2011
- [15]. <http://www.sas.com/offices/europe/netherlands//>
- [16]. Padraic G. Decision Trees for Predictive Modeling, Neville SAS Institute Inc. 4 August 1999

[17]. Tom M. Mitchell, Machine Learning, ISBN: 0070428077, McGraw-Hill Science/Engineering/Math; (March 1, 1997)

[18]. [http://en.wikipedia.org/wiki/Linear\\_regression](http://en.wikipedia.org/wiki/Linear_regression)

[19]. Verizon2013 Data breach investigations report  
<http://www.verizonenterprise.com/DBIR/2013>

[20]. Yolanta Beres, Marco Casassa Mont, Jonathan Griffin, Simon Shiu, Using Security Metrics Coupled with Predictive Modeling and Simulation to Assess Security Processes, HP Laboratories, HPL-2009-142, HP, June 21, 2009]

[21]. Chris I. Cain, Erik, Couture Establishing a Security Metrics Program, SANS Technology Institute

[22]. COBIT® 4.1, IT Governance Institute, [www.itgi.org](http://www.itgi.org), ISBN 1-933284-72-2

[23]. Peter Mell, Karen Scarfone, Sasha Romanosky, A Complete Guide to the Common Vulnerability Scoring System Version 2.0, June, 2007

[24]. Gary Stoneburner, Alice Goguen, and Alexis Feringa, Risk Management Guide for Information Technology Systems, NIST Special Publication 800-30

[25]. [http://en.wikipedia.org/wiki/Asset\\_\(computer\\_security\)](http://en.wikipedia.org/wiki/Asset_(computer_security))

[26]. [http://dss.princeton.edu/online\\_help/analysis/regression\\_intro.htm](http://dss.princeton.edu/online_help/analysis/regression_intro.htm)

[27]. [http://www.uvm.edu/~dhowell/StatPages/More\\_Stuff/Missing\\_Data/Missing.html](http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html)

[28]. Robert M. Kunst, Toward a theory of evaluating predictive accuracy, University of Vienna And Institute for Advanced Studies Vienna, September 3, 2004

[29]. [http://en.wikipedia.org/wiki/Radial\\_basis\\_function\\_network](http://en.wikipedia.org/wiki/Radial_basis_function_network)

[30]. Simon O. Haykin, Neural Networks and Learning Machines (3rd Edition).

[31]. [http://en.wikipedia.org/wiki/Cross-validation\\_\(statistics\)](http://en.wikipedia.org/wiki/Cross-validation_(statistics))

[32]. Business Modeling & Requirements Engineering, dr. J.F.M. (Hans) Burg, Version: 2.0, 04-12-2001

[33]. <http://www.wmps.com/blog/website-analysis/web-analytics/goals-vs-kpis-analytics-tips/>

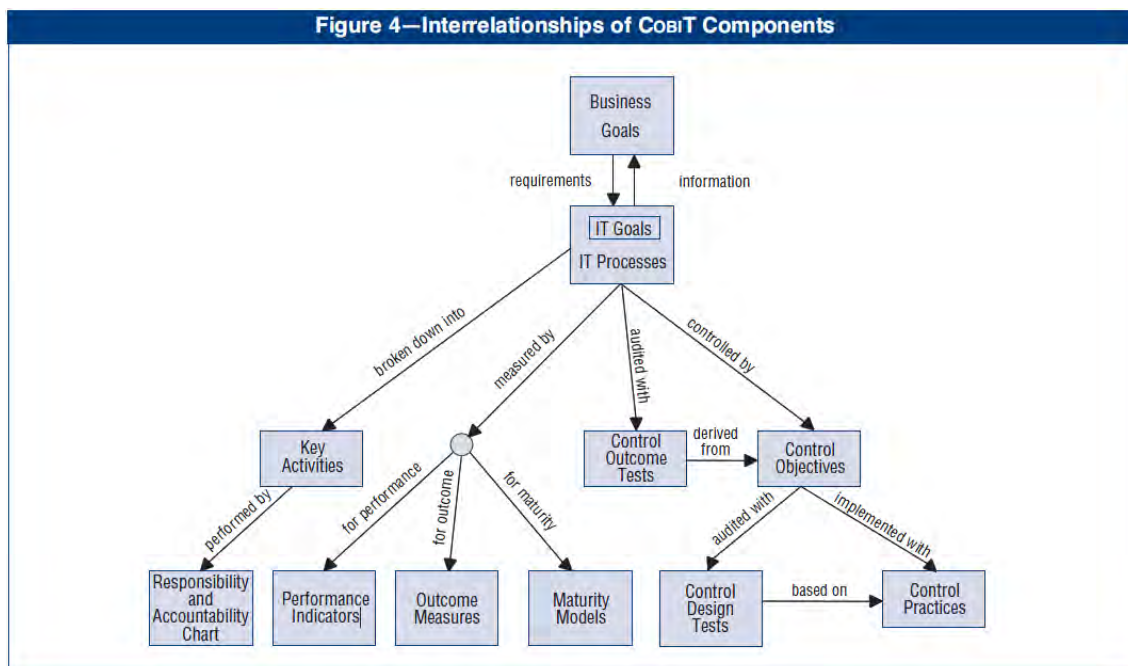
## AppendixII Questionnaire

List of question and answers form the interview with security stakeholders of Podictive about the security goals of the patch and vulnerability management process.

### Q1: What is your mean security goal as IT Security Company?

*Answer:*

*As an IT security company we support our clients to gain control and become more secured by implementing security measurements in order to get more insight on IT security status on their IT assets. Therefore, our IT security goals from executing the patch and vulnerability management process have to be aligned with the IT and strategic goals of our clients business. A good example of how our IT security goals are related to the client's goals can be found in COBIT[22].*



### Q2: I understand from this you are more in the control part of any IT process of your clients. Since we are intended to set effective security metrics for the patch and vulnerability process what is your control purpose from applying this process?

*Answer: The mean purpose of any the patch and vulnerability process is to identify possible vulnerabilities in IT assets and remediate them at time by using appropriate patches and therefore protecting them and preventing its exploitation by malicious user.*

### Q3: Well, what then the mean goal of vulnerability scan?

*Answer: We do monthly vulnerability scans for clients IT assets to get insight how vulnerable each IT asset is, for this we count the number of vulnerably detected in each asset to get an idea how vulnerable it is.*

**Q4: Do you think that just counting the number of vulnerabilities per IT asset will give a good insight about the security status of the IT asset?**

*Answer: No, we don't believe it is easy to do security like this. Basically, to secure IT asset more efficiently, we scan for vulnerabilities; we then classify and prioritize them according to its severity and IT assets criticality. In other words, we do the classification and prioritization according to the following risk matrix:*

RISK MATRIX

SERVER CRITICALITY		RISK FACTOR			
		LOW	MEDIUM	HIGH	CRITICAL
		1	2	3	4
Critical	4	4	8	12	16
High	3	3	6	9	12
Medium	2	2	4	6	8
Low	1	1	2	3	4

$$\text{IMPACT} = \text{SERVER CRITICALITY} \times \text{RISK FACTOR}$$

**Q5: Could the number of vulnerabilities combined with this risk matrix be used as a security metric?**

*Answer: Yes, we use the number of vulnerability per IT asset as security metric the more critical the asset is and the more severe the vulnerability on it the more the impact is, but using it this way seem to be not enough since the risk matrix has been showed its inefficiency in IT security. That is why we are looking forward to find a way to identify effective security metrics in a process approach that will enable us to improve IT security any time that new changes and challenges happen.*

**Q6: I understand that it is a good way to take the severity of the identified vulnerabilities in It asset asses it security status but not in the way the risk matrix do since there only three levels of vulnerability's severity; low, medium, high and critical. Well, do you think that correlating security events (in this case vulnerability scans) well give a more evaluation of the severity of vulnerabilities?**

*Answer: For well-known vulnerabilities the Common Vulnerability Scoring System (CVSS[23]) is used to assess the severity of vulnerabilities. But this is still based on judgments and opinions of people about risk rather than exact and continuous numerical values. It's even worse for zero-day vulnerabilities where no one can have an exact idea about its severity! Yes, if we could correlate security events well then we would be able somehow to evaluate vulnerability's severity more effectively.*

**Q7: As we are talking about the correlation of security events, like discovering new vulnerabilities, do you thing that the discovering duration of a vulnerability from the moment of announcement to the moment of remediating it can used as good correlation factor in the sense that the longer it take until remediating a vulnerability the more saver it is?**

*Answer: Yes and no! Sometimes the duration of the discovering and remediation may characterize the complication and therefore the severity of the issue. But it is not always the case, because sometimes the long duration of treatments is due to technical issues and not to security issues like shortage of staff. Anyway the remediation duration can be used as a good correlation factor between security events, especially if we could take the contribution of other factors like the shortage of staff in it in the sense that two vulnerabilities of the same characteristics have to have the same handling duration. One important factor that affect the remediation duration is that if the remediation of a given vulnerability takes longer than normal this vulnerability may become more severe since it can be used to open back doors or to find other vulnerabilities in the system, which make correlation of security more difficult if not impossible.*

**Q8: The same question but instead of the handling duration of vulnerabilities, this time the cost of handling vulnerabilities in the sense that the severer the vulnerability is the more expansive its remediation is?**

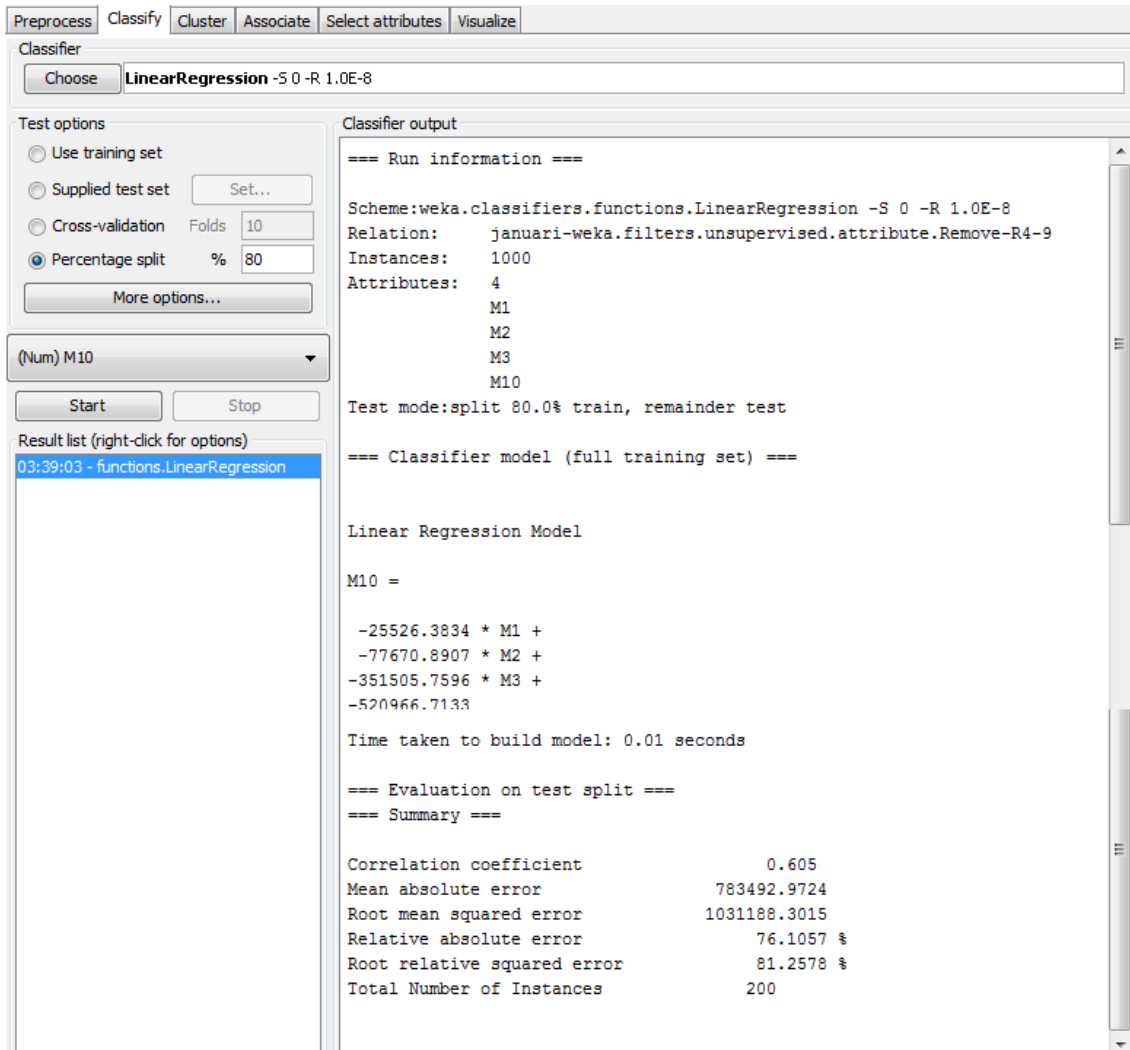
*Answer: since the cost of handling vulnerabilities depend more on the duration of the handling the answer will be almost the same as for the previous question. Therefore costs of remediation can be used in some way as a correlation factor between security events.*



## Appendix III Weka results

Results from training and testing the models used in Weka.

Training the Linear regression model:



The screenshot shows the Weka Classifier window with the 'LinearRegression' model selected. The 'Test options' section is set to 'Percentage split' at 80%. The 'Classifier output' pane displays the following information:

```

=== Run information ===

Scheme:weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8
Relation:  januari-weka.filters.unsupervised.attribute.Remove-R4-9
Instances:  1000
Attributes:  4
             M1
             M2
             M3
             M10

Test mode:split 80.0% train, remainder test

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

Linear Regression Model

M10 =

-25526.3834 * M1 +
-77670.8907 * M2 +
-351505.7596 * M3 +
-520966.7133

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===
=== Summary ===

Correlation coefficient          0.605
Mean absolute error             783492.9724
Root mean squared error         1031188.3015
Relative absolute error         76.1057 %
Root relative squared error     81.2578 %
Total Number of Instances      200
  
```

The 'Result list' on the left shows a single entry: '03:39:03 - functions.LinearRegression'.

Training the Decision tree model:

```

--- Run information ---
Scheme: weka.classifiers.trees.REPTree -M 2 -Y 0.001 -M 3 -S 1 -X -1
Relation: 500ari-weka.filters.unsupervised.attribute.Remove-R4-9
Instances: 1000
Attributes: 4
  M1
  M2
  M3
  M10
Test mode: split 80.0% train, remainder test

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

REPTree
-----
M3 < 1.5
|
| M3 < 0.5
| |
| | M1 < 0.5
| | |
| | | M2 < 4 : 0 (4/0) [3/0]
| | |
| | | M2 >= 4
| | | |
| | | | M2 < 6.5 : -360000 (3/5120000000) [1/2860000000]
| | | | M2 >= 6.5 : -160000 (3/5120000000) [0/0]
| | |
| | | M1 >= 0.5
| | | |
| | | | M1 < 4.5
| | | | |
| | | | | M1 < 3.5
| | | | | |
| | | | | | M2 < 1.5
| | | | | | |
| | | | | | | M1 < 2.5 : -245833.33 (7/17081632653.06) [5/75029796916.37]
| | | | | | | M1 >= 2.5 : -572222.22 (6/13480888888.56) [3/3393611111.11]
| | | | | |
| | | | | | M2 >= 1.5
| | | | | | |
| | | | | | | M2 < 2.5 : -685000 (8/176223437500) [8/270704687500]
| | | | | | | M2 >= 2.5
| | | | | | |
| | | | | | | M1 < 1.5 : -367492.31 (9/5488888888.89) [4/10311111111.11]
| | | | | | | M1 >= 1.5
| | | | | | | |
| | | | | | | | M2 < 6
| | | | | | | | |
| | | | | | | | | M2 < 3.5
| | | | | | | | | |
| | | | | | | | | | M1 < 2.5 : -840000 (5/20736000000) [1/51840000000]
| | | | | | | | | | M1 >= 2.5 : -320000 (3/8120000000) [0/0]
| | | | | | | | |
| | | | | | | | | M2 >= 3.5 : -680000 (7/11820000000) [6/11820000000]
| | | | | | | | |
| | | | | | | | | M2 >= 6 : -600000 (5/9216000000) [1/18662400000]
| | | | | | | |
| | | | | | | | M1 >= 3.5
| | | | | | | | |
| | | | | | | | | M2 < 1.5
| | | | | | | | | |
| | | | | | | | | | M1 < 5.5
| | | | | | | | | | |
| | | | | | | | | | | M1 < 4.5 : -641250 (11/158360330576.51) [5/384650578512.4]
| | | | | | | | | | | M1 >= 4.5 : -611111.11 (6/135013080808.09) [3/10355833333.33]
| | | | | | | | | | | M1 >= 5.5 : -308571.43 (5/105984000000) [2/4096000000]
| | | | | | | | | |
| | | | | | | | | | M2 >= 1.5
| | | | | | | | | | |
| | | | | | | | | | | M2 < 3.5
| | | | | | | | | | | |
| | | | | | | | | | | | M1 < 5.5
| | | | | | | | | | | | |
| | | | | | | | | | | | | M2 < 2.5
| | | | | | | | | | | | | |
| | | | | | | | | | | | | | M1 < 4.5 : -1100000 (3/52195555555.56) [1/177777777.78]
| | | | | | | | | | | | | | M1 >= 4.5 : -97142.56 (13/12486000000) [5/4832000000]
| | | | | | | | | | | | | | M2 >= 2.5 : -1403000 (6/2736000000) [2/13164800000]
| | | | | | | | | | | | | | M1 < 5.5 : -800000 (7/94040816326.53) [5/3601763530.61]
| | | | | | | | | | | | | | M2 >= 3.5 : -675000 (20/9849600000) [12/16828400000]
| | | | | | | | | |
| | | | | | | | | | M1 >= 6.5
| | | | | | | | | | |
| | | | | | | | | | | M2 < 6.5
| | | | | | | | | | | |
| | | | | | | | | | | | M2 < 0.5 : -200000 (6/2720000000) [0/0]
| | | | | | | | | | | | M2 >= 0.5
| | | | | | | | | | | | |
| | | | | | | | | | | | | M2 < 2.5
| | | | | | | | | | | | | |
| | | | | | | | | | | | | | M1 < 7.5 : -600000 (4/1584000000) [2/1440000000]
| | | | | | | | | | | | | | M2 >= 7.5 : -356000 (4/1940000000) [1/2400000000]
| | | | | | | | | | | | | | M2 >= 2.5 : -557142.56 (14/15986930775.1) [14/22146930775.1]
| | | | | | | | | | | | | | M2 >= 6.5 : 0 (2/0) [0/0]
| | | | | | | | | |
| | | | | | | | | | M3 >= 0.5 : -1042653.77 (137/459147231690.32) (83/444929278684.13)
M3 >= 1.5
|
| M2 < 2.5
| |
| | M1 < 3.5
| | |
| | | M3 < 4.5
| | | |
| | | | M1 < 0.5
| | | | |
| | | | | M3 < 2.5
| | | | | |
| | | | | | M2 < 1.5 : -1312777.83 (5/560781735422.24) [1/29127083355.56]
| | | | | | M2 >= 1.5 : -644464.67 (3/38323358585.56) [0/0]
| | | | | | M3 >= 2.5 : -2100000 (4/17822500000) [2/24512500000]
| | | | | |
| | | | | | M2 >= 0.5
| | | | | | |
| | | | | | | M1 < 2.5 : -1705939.11 (16/656829921526.34) [11/1204902392176.89]
| | | | | | | M1 >= 2.5
| | | | | | | |
| | | | | | | | M3 < 2.5
| | | | | | | | |
| | | | | | | | | M2 < 1.5
| | | | | | | | | |
| | | | | | | | | | M2 < 0.5 : -904097 (3/38350308696.09) [1/313082027393.78]
| | | | | | | | | | M2 >= 0.5 : -1208626.4 (5/116699394609.24) [0/0]
| | | | | | | | | | M2 >= 1.5 : -2082424.33 (2/22492218432.28) [1/3872449046132.28]
| | | | | | | | | | M3 >= 2.5 : -2021395.92 (6/933205354270.94) [5/1676681990286.36]
| | | | | | | | |
| | | | | | | | | M3 >= 4.5 : -992467.57 (13/298576699320.08) [7/161424173195.03]
| | | | | | | | |
| | | | | | | | | M1 >= 3.5
| | | | | | | | | |
| | | | | | | | | | M3 < 2.5 : -1642928.82 (35/734850898698.54) [15/1067906155046.03]
| | | | | | | | | | M3 >= 2.5
| | | | | | | | | | |
| | | | | | | | | | | M1 < 6.5
| | | | | | | | | | | |
| | | | | | | | | | | | M3 < 4.5
| | | | | | | | | | | | |
| | | | | | | | | | | | | M1 < 4.5
| | | | | | | | | | | | | |
| | | | | | | | | | | | | | M3 < 3.5 : -2480805.6 (4/829478289793.69) [6/931390472387.48]
| | | | | | | | | | | | | | M3 >= 3.5 : -3670230.75 (4/954520158796.69) [0/0]
| | | | | | | | | | | | | | M1 < 4.5 : -2392056.88 (17/93865191298.29) [8/14833714479.39]
| | | | | | | | | | | | | | M3 >= 4.5 : -1472051.33 (4/26466977115.89) [6/1090012789581.64]
| | | | | | | | | | | | | | M1 >= 6.5 : -1891554.09 (27/174885555874.22) [5/1704856284473.4]
| | | | | | | | | |
| | | | | | | | | | M2 >= 2.5
| | | | | | | | | | |
| | | | | | | | | | | M3 < 2.5
| | | | | | | | | | | |
| | | | | | | | | | | | M2 < 4.5 : -2345719.76 (39/1440740622610.3) [16/1016047701093.67]
| | | | | | | | | | | | M2 >= 4.5 : -1673109.12 (42/56558201807.27) [24/120327157162.71]
| | | | | | | | | | |
| | | | | | | | | | | M3 >= 2.5
| | | | | | | | | | | |
| | | | | | | | | | | | M3 < 5.5
| | | | | | | | | | | | |
| | | | | | | | | | | | | M2 < 4.5
| | | | | | | | | | | | | |
| | | | | | | | | | | | | | M2 < 3.5
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | M3 < 3.5 : -2817463.64 (16/1223670513303.65) [6/2116010554072.22]
| | | | | | | | | | | | | | | M3 >= 3.5 : -4920603.29 (10/271720771486.01) [4/219242369890.49]
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | M2 >= 2.5
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | M1 < 9.5
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | M2 < 4.5 : -2748795.77 (28/1051985136821.47) [9/1898260314666.42]
| | | | | | | | | | | | | | | | | M2 >= 4.5
| | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | M1 < 7.5
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | M1 < 4.5
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | M1 < 4.5
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | M1 < 2.5 : -2950717.85 (5/133155764303.19) [5/3215230217255.36]
| | | | | | | | | | | | | | | | | | | | | M1 >= 2.5 : -4200415.33 (6/2628484187142.67) [3/86118647796.67]
| | | | | | | | | | | | | | | | | | | | | M2 >= 4.5 : -304027.77 (12/3088437256126.64) [1/134478727063.36]
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | M1 >= 6.5 : -4056838.6 (3/46372170150.89) [2/2457591792822.78]
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | M1 >= 7.5
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | M2 < 5.5 : -800492.67 (2/50512096276) [1/3658903306276]
| | | | | | | | | | | | | | | | | | | | | | M2 >= 5.5 : -843525.75 (3/271028120387.58) [1/232397098512.11]
| | | | | | | | | | | | | | | | | | | | | | M3 >= 9.5 : -2412827 (2/12602800000) [3/3117798718076]
| | | | | | | | | | | | | | | | | | | | | | M3 >= 4.5 : -2357440 (14/1372815306344.94) [9/75468588094.53]
| | | | | | | | | | | | | | | | | | | | | | M2 >= 6.5 : -3577535.19 (16/1800944049902.81) [10/1162829764942.15]
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | M3 >= 5.5
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | M2 < 6.5
| | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | M1 < 5.5 : -1308007.69 (11/1414474301649.14) [2/131503075399.36]
| | | | | | | | | | | | | | | | | | M1 >= 5.5 : -2216952.33 (4/858611499313.28) [2/24064934118.25]
| | | | | | | | | | | | | | | | | | M2 >= 6.5 : -3549272.6 (2/130115671940.25) [3/1460010261917.58]

Size of the tree : 113

Time taken to build model: 0.03 seconds

=== Evaluation on test split ===
=== Summary ===
Correlation coefficient: 0.6625
Mean absolute error: 740491.8893
Root mean squared error: 964136.4462
Relative absolute error: 71.8229 %
Root relative squared error: 75.9772 %

```

The visualization of the obtained the decision tree by applying the decision tree model (REPTree) in Weka.

