



**MASTER OF
INFORMATION TECHNOLOGY**

**THE ANNUAL CONFERENCE
ON MANAGEMENT AND
INFORMATION TECHNOLOGY**

ACMIT 2019

**Opportunities and Challenges of Industry
Revolution 4.0**

Tangerang, 10th August 2019

Volume 6

Preface

It is only with the grace of God that the proceedings of the Annual Conference on Management and Information Technology (ACMIT) 2019 is published. This proceedings is the publication of the papers which were presented in the ACMIT 2019. This ACMIT conference was held by Master of Information Technology (MIT) at Swiss German University (SGU) on Saturday, 10th August 2019, with the theme “Opportunities and Challenges of Industry Revolution 4.0”. This theme is considered to be relevant, especially since MIT SGU has the vision to become the training place of future digital transformation architect, which will be really needed to transform the industry to be compatible with Industry Revolution 4.0 era.

The theme was mainly reflected in the plenary sessions with the following speakers:

1. Rahmat Mulyana, President of ISACA Indonesia, who gave a presentation on “Digital Transformation: IT Governance Challenge”
2. Eric Anwar, Vice President IT Automation, Bank Central Asia, who gave a presentation on Robotics in Banking

It is hoped that the published papers in this proceedings can be beneficial for the readers.

Tangerang, 20th August 2019

Dr. Eka Budiarto, S.T., M.Sc.

Chairman of ACMIT 2019 Committee

Table of Contents

PREFACE _____	I
ACMIT 2019 COMMITTEE _____	IV
EVENT RUNDOWN _____	V
SCHEDULE OF PARALLEL SESSIONS _____	VI
PREDICTING OF BANKING STABILITY USING MACHINE LEARNING TECHNIQUE OF RANDOM FORESTS _____	1
Agus Afiantara, Bagus Mahawan and Eka Budiarto _____	1
EARLY DETECTION OF FAILED BANK THROUGH ANALYSIS OF FINANCIAL RATIOS AND BANK SHAREHOLDERS RATIOS USING DATA MINING FOR RURAL BANKS _____	9
H M Agista, E Budiarto, and B Mahawan _____	9
AN ANALYSIS OF MODLE ACCEPTACE FOR STUDENTS IN SMPK2 PENABUR USING UNIFIED THEORY OF ACCEPTANCE AND USE OF TECHNOLOGY MODEL _____	16
D Fernando _____	16
PREDICTIVE ANALYSIS ON EMPLOYEE PERFORMANCE _____	24
S L Tendean¹, M R Mashudi and E Budiarto _____	24
DETERMINING CRITICAL FACTORS PROJECT DELAY AND EFFECTING COST OVERRUNS IN TELECOMMUNICATION MOBILE NETWORK PROJECTS _____	30
Deni Bakhtiar, Mulya R Mashudi and Maulahikmah Galinium _____	30
IMPROVING PERFORMANCE LOAN FRAUD MODEL PREDICTION USING MEAN DECREASE ACCURACY AND MEAN DECREASE GINI _____	36
Arsandi Akhmad, Lukas and Bagus Mahawan _____	36
MEASUREMENT OF IT RISK MANAGEMENT MATURIY LEVEL IN CEC USING IT DOMAIN RISK GOVERNANCE FRAMEWORK _____	42
Annas Iswahyudi _____	42
MONITORING & EVALUATION MODEL FRAMEWORK FOR IT PROJECT MANAGEMENT AT PT. XYZ _____	49

Fattah Hadi Saputro, Mohammad Amin Soetomo and Nuki Agya Utama	49
THREAT HUNTING EARLY EXPERIMENT THROUGH EVENT CORRELATION AND MEMORY FORENSIC	56
Arif D Purnomo, Charles Lim and Burman Noviansyah	56
CORE BANKING SYSTEM SCALABILITY REVIEW	64
Achmad Fakhrudin, Heru P Ipung, M Amin Soetomo and Charles Lim	64
REVEALING AND SHARING MALWARE PROFILE USING MALWARE THREAT INTELLIGENCE PLATFORM	72
Faiz Iman Djufri and Charles Lim	72
METHODOLOGY OF SECURITY TESTING OF IKID WEBSITE AND SECURITY VULNERABILITIES	83
Mustofa Kamil	83

ACMIT 2019 Committee

Steering Committee

- Rector : Rector, Dr. rer. nat. Filiana Santoso
- Vice Rector of Academic Affairs : Dr. Irvan S. Kartawiria, S.T., M.Sc.
- Dean of Faculty of Engineering and IT : Dr. Maulahikmah Galinium, S.Kom., M.Sc

Organizing Committee

- Chairman : Dr. Eka Budiarto, S.T., M.Sc
Secretary : St. Ayu Diana Lestari, S.Pd.
Treasurer : Ingrid Kaunang, S.E.(Finance)
Lestari Nur Wijayanti, S,St (Finance)
- Communication & P.R : Irzan Fahmi, S.Kom – Visual Design & Photographer
Ayu Angela Agusta, M.Si
M. Hilmy Rachmatullah, S.IKom
- Program Support : Yunus Nugraha (ASC)
Somanudin, M.M (SDC)
- IT Infrastructure : Angga Anugrah, B.Sc (ISS)
Billy Siagian, S.Kom (ISS)
Dipl.-Ing. Iqbal Yasin, S.Kom., M.Kom
- Facility : Rina Rahayu (FM)
- Procurement : Latifah Bachrum
- Registration : St. Ayu Diana Lestari, S.Pd
Ayu Angela Agusta, M.Si
M. Hilmy Rochmatullah, S.IKom
- Certificates : M. Hilmy Rochmatullah, S.IKom
Food & Beverage : St. Ayu Diana Lestari, S.Pd
- ### Technical Program
- Paper Reviewer : Dr. Eka Budiarto, S.T., M.Sc
Dr. Charles Lim, M.Sc.,
Dr. Maulahikmah Galinium, S.Kom., M.Sc.
Dr. Ir. Heru Ipung, M.Eng.
Dr. Ir. Moh A. Amin Soetomo, M.Sc
Dr. Ir. Lukas, MAI, CISA, IPM
James Purnama, M.Sc
Kalpin E. Silaen, M.Kom
Burman Noviansyah, MSISPM
- Moderator : Dr. Ir. Heru P. Ipung, M.Eng
Dr. Ir. Moh. A. Amin Soetomo, M.Sc
- Proceeding : Dr. Charles Lim, B.Sc., M.Sc

Event Rundown

Saturday, 10th August 2019

Time	Topic & Speaker
08:30 – 09:00	: Registration
09:00 – 09:15	: Prayers and singing the National Anthem of Republic of Indonesia
09:15 – 09:30	: Welcoming speech by Head of MIT, Dr. Eka Budiarto, S.T., M.Sc as Chairman of Committee
09:30 – 09:45	: Opening Speech by Dean of Faculty Engineering and Information Technology, Dr. Maulahikmah Galinium, S.Kom., M.Sc
09:45 – 10:00	: Photo Session
Keynote Speech	
10:00 – 10:45	: Speech I: Rachmat Mulyana, CISA, CISM, CGET, CRISC, COBIT5F, CSXF President of ISACA Indonesia Topic: Digital Transformation: IT Governance Challenge.
10:45 – 11:30	: Speech II: Eric Anwar Vice President IT Automation, Bank Central Asia Topic: Robotic in Banking
11:30 – 11:45	: Question and Answers
11:45 – 12:00	: Plaque Awarding
12:00 – 13:00	: Lunch Break
13:00 – 17:00	: Parallel Sessions

Schedule of Parallel Sessions ACMIT 2019 – Saturday, 10th August 2019

Parallel Session 1, Room 2013

No.	Time	Presenter	Moderator
1.	13:30 – 13:50	Agus Afiantara	Dr. Ir. Heru Purnomo Ipung, M.Eng
2.	13:50 – 14:10	Deni Bachtiar	
3.	14:10 – 14:30	Hanna M. Agista	
4.	14:30 – 14:50	Annas Iswahyudi	
5.	15:10 – 15:30	Fattah Hadi Saputro	
6.	15:30 – 15:50	Achmad Fachrudin	

Parallel Session 2, Room 2014

No.	Time	Presenter	Moderator
1.	13:30 – 13:50	Steven Leonardo Tendean	Dr. Ir. Moh. A. Amin Soetomo, M.Sc
2.	13:50 – 14:10	Faiz Iman Djufri	
3.	14:10 – 14:30	Arsandi Akhmad	
4.	14:30 – 14:50	Mustofa Kamil	
5.	14:50 – 15:10	Arif Dwi Purnomo	
6.	15:10 – 15:30	Daniel Fernando	

Predicting of Banking Stability Using Machine Learning Technique of Random Forests

Agus Afiantara¹, Bagus Mahawan¹ and Eka Budiarto^{1*}

¹ Department of Information Technology, Swiss German University, Tangerang 15143, Indonesia

* Corresponding author: eka.budiarto@sgu.ac.id

Abstract. The purpose of this research is to create a predicting model of banking stability in Indonesia. Authors use a small set of explanatory indicators primarily related to the banking industry and some relevant economic variables. Among the indicators candidate to be used in this study are the indicator of banking industry, the money markets, capital markets and creditors, and the macro-economic indicator. The source of data in this research is obtained from CEIC and SEKI (Indonesian Economic and Financial Statistics) that published by Central Bank of Indonesia from 2004 and 2011. Principal Component Analysis is used to avoid the multi-collinearity problem when construct the model. Authors train the model using Random Forest Regression with data over the 2004-2007 period, and made predictions of global financial crisis that happened in 2008/9. Python 2.7.10 and scikit-learn version 0.20.0 module has been exploited for simulations and evaluation of the model. Numerical illustration is provided to demonstrate the efficiency of proposed model. As the result nine most components analysis obtained as input for the machine learning model with explained variance ratio around 97%, accuracy around 89%, and precision 91% and mean absolute error around 11%.

1. Background

The stability of financial system is the most important priority of the government concern and its unpredictability is a critical concern of people in the world and try to avoid the events with creating a model that predict and simulate the events, especially the banking sector. The terrible events called “financial crisis” occurred in the past decade in the worlds, there are some events recorded in our memory especially in Asia and the world. In 1994-1995 there has been crisis in Mexico called “Mexican Crisis”, in 1997-1998 there has been crisis in Asia and called “Asian Crisis” (Illing and Liu, 2003), and in 2007-2009 there has been crisis called “Global Financial Crisis” (Popovska, 2014).

In Indonesia, the banking sectors still dominates Indonesia’s financial system. Based on data of BI shown that the market shares of industry reached 78.8% by the end of 2013, up from 77.9% in the first half of the same year. However, the potential risks of this industry remain to be worry of. Financial disturbance such as banking crisis in a country such as Indonesia can costly and affect to the other sectors in economic in a deeply traumatizing way. Consequently, it is important to anticipate the risks and try to prevent disaster and ensure the banking stability.

2. Literature Review

2.1. Banking Stability

Macroeconomic environment with stable condition is very important for the stability of the banking sector, mainly due to uncertainty about shaky macroeconomic and fundamental policies, such as economic growth and inflation, making it a challenge for banks to accurately assess credit risk (Kristína Kočíšová, 2015). Weak economic growth, due to macroeconomic uncertainty or for other reasons, can disrupt the health of banks because it reduces the service capacity of corporate and household debt (Swamy, 2013). The resilience and stability of the financial system, especially the banking sector, as its foundation has attracted public attention even more because of the difficulties facing the global market during the crisis (Popovska, 2014).

The banking system has the important role played in propagating crisis, therefore underway to devise ways of build a strong and resilient banking system (Ghosh, 2011). The Asian crisis for more than a decade ago and even recently, after the banking crisis in Argentina and Turkey, policy makers have made proactive efforts to restructure their banking system (Hawkins and Turner, no date). The idea

behind the strategy is to maintain the banking system back to the position of profitability and solvency using private sector solutions and resolutions assisted by the public sector, or a combination thereof.

2.2. Principle Component Analysis

Principle Component Analysis (PCA) is statistical concept applied to fields such as face recognition and image compression, and common technique for finding patterns in data of high dimension. It covers mathematical concept of standard deviation, covariance, eigenvectors and eigenvalues (Smith, 2002).

PCA can be derived from a number of starting points and optimization criteria. The most important of these are minimization of the mean-square error in data compression, finding mutually orthogonal directions in the data having maximal variances, and de-correlation of the data using orthogonal transformations (Ilin, 2010).

2.3. Machine Learning and Random Forest

Machine learning is one of the technologies with ability to learn and a research field between statistics, artificial intelligence, and computer science and is also known as predictive analysis or statistical learning.

The usefulness of machine learning to predict and model has been found in the several studies. Machine learning techniques have been widely used in predicting pressure on banks and have become a major topic that is widely discussed during the pre-crisis and post-crisis periods (Ronnqvist and Sarlin, 2015). The Random Forest (RF) is one of machine learning technique used to predict bank bankruptcy in a sample of US-based financial institutions (Petropoulos *et al.*, 2017).

The implementation of RF is widely used in many fields especially in prediction, now casting and forecasting. There are samples of implementation of RF, it has the potential to give early warning of recessions (Nyman and Ormerod, 2016), and a superior out of sample and out of time predictive performance in differing the banks of predictors and possible bank insolvencies (Petropoulos *et al.*, 2017), insurers' insolvency prediction (Kartasheva and Traskin, 2013), predicting the direction of stock market prices (Khaidem, Saha and Dey, 2016), and modelling of economic phenomena (David, 2017).

3. Materials and Methods

The machine learning model that generated during research activity based on figure 1. The first step is determined the crisis event represented by the value of indicator tools that called ISSK (Financial System Stability Index), it developed by Central Bank of Indonesia in 2013 as cited in (Wimanda, Maryaningsih and Nurliana, 2014).



Figure 1. The process of Machine Learning Modelling

The crisis event is classified into 4 level of stability, there are normal, alert, standby, and suspected crisis level. The ISSK indicator described in figure 2 (Wimanda, Maryaningsih and Nurliana, 2014).

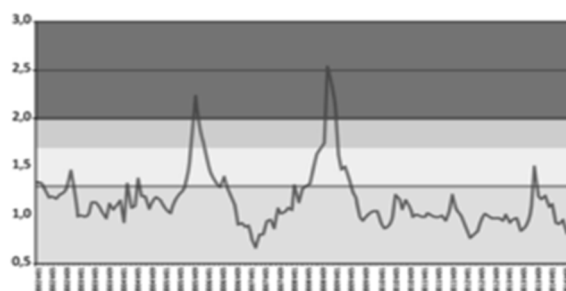


Figure 2. ISSK Indicator Index (Image Source: Macro-prudential Policy Department of Central Bank of Indonesia (2014))

From the ISSK data extraction phase we have collected data series as dependent variable with integer value varied from 0 to 3 as table 1, that represented the stability level of financial system of Indonesia with monthly frequency.

Table 1. ISSK indicator values extraction

Idx	Description	ISSK Values
0	Normal Condition	≤ 1.25
1	Alert Condition	>1.25 and ≤ 1.625
2	Standby Condition	>1.625 and ≤ 1.875
3	Suspected of Crisis	>1.875

The second step, we collected the predictors data or features. The number of features that observed in this study is 33 with period over 2000-2011 and monthly frequency. All features is defined based on our previous research (Stabilitas, Perbankan and Pengantar, 2011). The predictors consist of internal bank, money market and debt, equity market, and macro-economic indicator. The indicators that used in this study is focused area in order to predict the stability of banking system using random forest machine learning algorithm. The model described with simplified linear equation such as equation (1).

$$y(t) = \sum_{n=1}^{n=33} \omega_n X_{n(t-1)} \tag{1}$$

Where $y(t)$ is the dependent variables represented by ISSK value or crisis event at period t , ω_n is the coefficient or weight of each independent variables, and $X_{n(t-1)}$ is the independent variables at the previous of period t . Based on that equation 2.1, the previous period of the predictors is used to predict the next month of condition of banking system purposes.

The third step, is data preparation. In this step performed N/A treatment, transformation and normalization. The firstly is N/A treatment, there are possibilities in dataset in the real-world datasets contain missing values, often encoded as blank, NaN's or other placeholders. A basic strategy to use incomplete datasets is to discard entire rows and/or columns containing missing values. Missing data often hinder the development of robust composite indicators. Due to missing data in several predictors for period in 2000-2003, we used the basic strategy to use is discarded over period 2000-2003. Data in period over 2004-2011 is used as dataset covered in this study. The secondly, Log transformation is performed because of the data size is varies between features in dataset. It could be used to make highly skewed distribution to less skewed and the result of this step would be data pattern more interpretable and the thirdly is Normalization is required prior to any data aggregation as indicators in a dataset often have different measurement units. In this research is used the Standardization (or z-scores) as normalization method. This method is calculated by the equation (2). The result of this method is the variables would be rescaled show they would have properties of a standard normal distribution with $\mu = 0$ and $\delta = 1 - N(0,1)$.

$$z = \frac{x - \mu}{\delta} \tag{2}$$

The fourth step, is data validation. The purposes of data validation is performed in order to get the error rate of the machine learning model, which can be considered close to the actual error rate of the population. If the data volume is large enough to represent the population, validation is not required. Hold-out and K-Fold Cross Validation (KFCV) are the validation techniques covered in this study. In

Hold-out technique data divided into training and test data with portion 80/20 randomly, it means 80% of dataset as training data and 20% as test data (20% of data test have represented enough to the all class of banking stability level). When we used the K-Fold technique as depicted in figure 3, the dataset is divided into 1-fold with sorted period over 2004-2011. The data from period 2004 until 2007 is used as training data, and the data from period 2008 until 2011 as test data, this validation technique made the model learnt the “mini crisis” situation and tried to predict the “global crisis” situation.



Figure 3. K-Fold Cross Validation (K=1)

The next step, is the last step before create machine learning model. PCA is the statistical procedure which used to avoid the multi-collinearity problem and reduce the dimensional of data, and finally the efficiency of resources like time, memory, and communication achieved. Fewer dimension then leads better generalization, elimination of noise to improve quality of data and further processing with machine learning algorithm (Mirkin, 2011).

The Additional step is created the machine learning model. The machine learning model is selected based on business easy of explanation and complexity. Random Forest algorithm is not too complex but not too simple also.

The last step performed in this study is evaluated the model, in this stage we built one or more models that appeared to have high quality from the data analysis perspective. The confusion matrix is a measurement to evaluated the model and known as an error matrix in machine learning, it explained the performance of classification model on a set of test data for which the true values are known.

4. Result and Discussion

In this study we explore the most important parameters of random forest and how they impact our model in term of overfitting and underfitting. The validation curve is used to validate the model described in figure 4, 5, 6 and 7 based on each parameters of random forest.

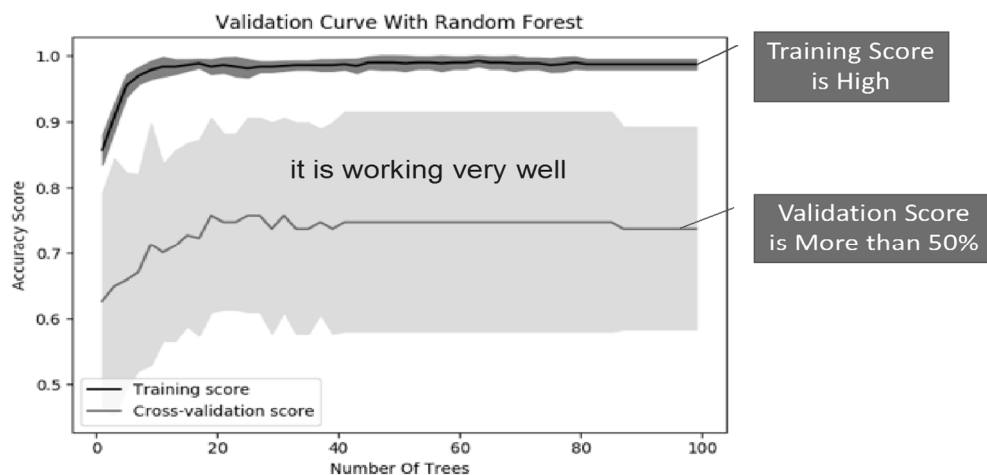


Figure 4. Validation curve based on number of trees

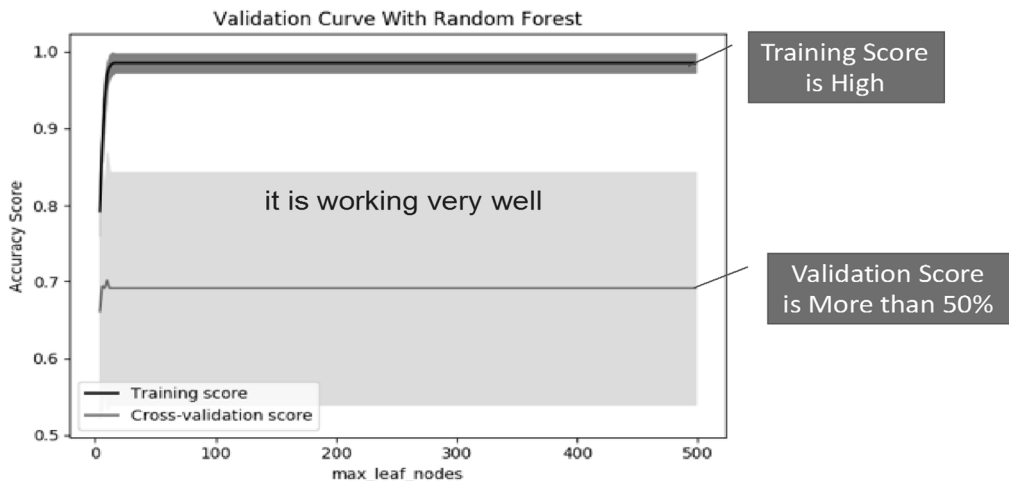


Figure 5. Validation curve based on maximum leaf nodes

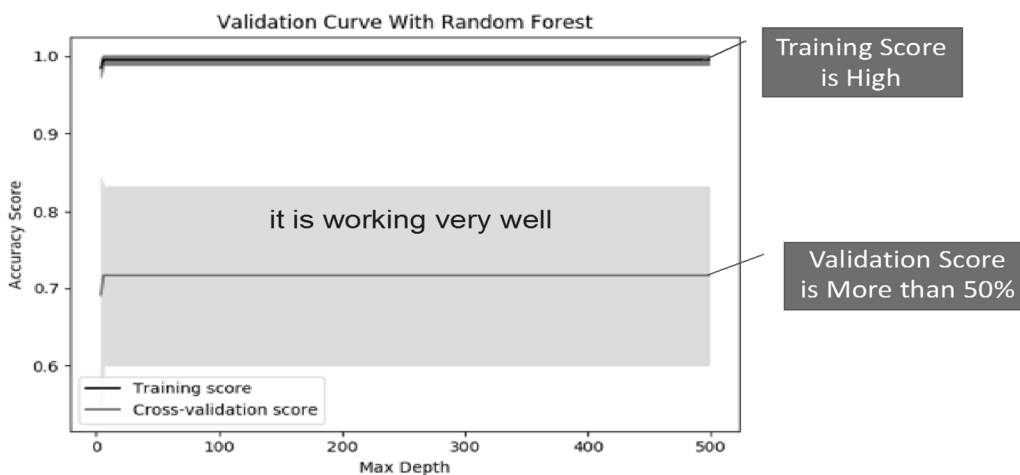


Figure 6. Validation curve based on maximum depth

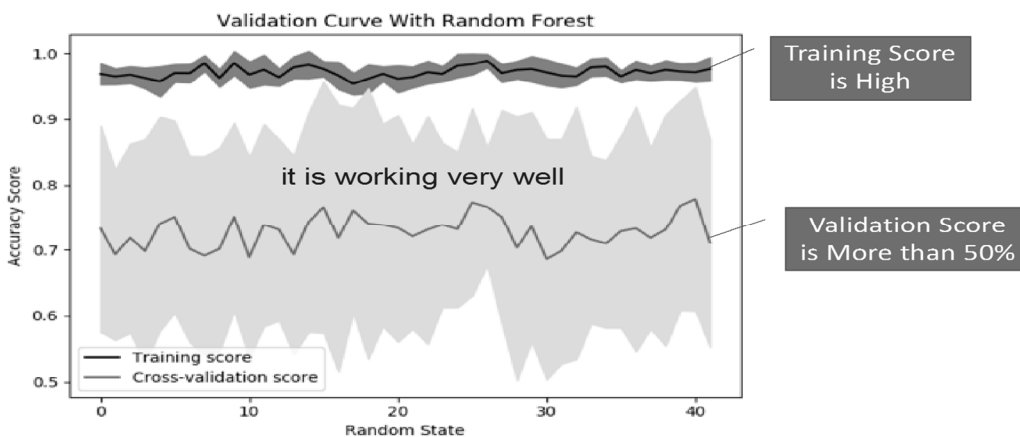


Figure 7. Validation curve based on random state

Based on these figures, the model that we made is fittest based on the rule of thumb. If training score is high and validation score is low then the model is overfitting, else if training score is low and validation score is low then the model is underfitting and else the model is fittest.

The confusion matrix is used to evaluate the model performance could be seen in table 3.1. Based on this table the model number 2 is the best model chosen with criteria number of component analysis is 9, cumulative explained variance around 97%, accuracy around 89%, precision around 91%, recall around 89% and F1-score around 89% with error criteria MAE 0.11, MSE 0.11 and RMSE 0.33.

Table 2. Evaluation of Model Performance

Model No.	RF Config	No. of PCA	Cumm.Exp. Var.	Acc.	Prec.	Recall	F1-Score
1	Max_Depth=2, Random_State=0,7,42	1	60.94%	72%	63%	72%	66%
2	Max_Depth=4, Random_State=7	9	96.91%	89%	91%	89%	88%
3	Max_Depth=8, Random_State=7	7	94.54%	89%	84%	89%	86%
4	Max_Depth=10, Random_State=0	7	94.54%	89%	84%	89%	86%

The confusion matrix of selected model shown in table 2. It means from 18 points of data, only 2 data missed to predict of the class of 1 (bank stability level in alert condition).

Table 3. Confusion Matrix of Selected Model

Predicted		Actual				Total
		0	1	2	3	
0	11	0	0	0	11	
1	2	3	0	0	5	
2	0	0	1	0	1	
3	0	0	0	1	1	
Total	13	3	1	1	18	

The selected model (prediction) is used to predict the ISSK indicator (actual) as the reference of financial system stability index in Indonesia. The prediction model with the random split validation in figure 8. performed the regression of prediction model is earlier than actual in “mini crisis” and “global crisis”. The machine learning model is being a leading indicator for banking stability. The prediction model with k-fold cross validation (k=1) in figure 9. The predictive models have no prior detection ability, it means the amount of data training is more required.



Figure 8. Prediction vs Actual based on hold-out validation

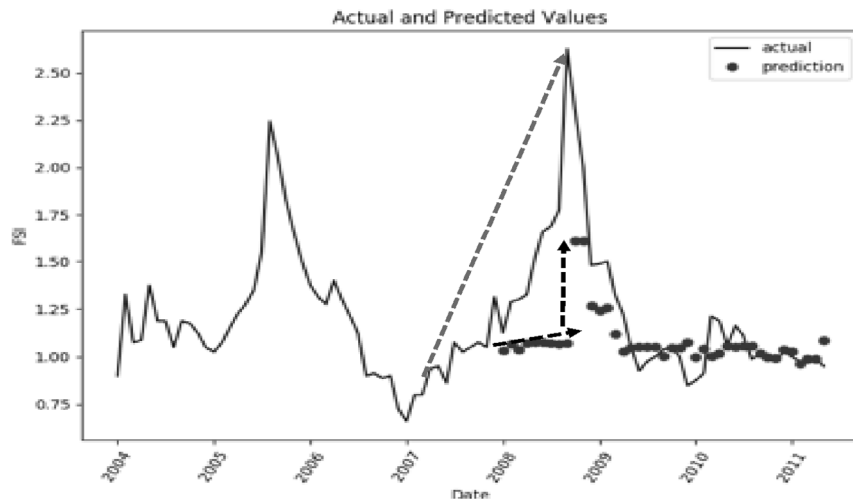


Figure 9. Prediction vs Actual based on 1-fold cross validation

In this last section the selected machine learning model is compared to available indicators that used by regulator of Indonesia, ISSK and BSI. Figure 10 show to us that the selected model is lead earlier detection when stability of banking and financial system getting worst.

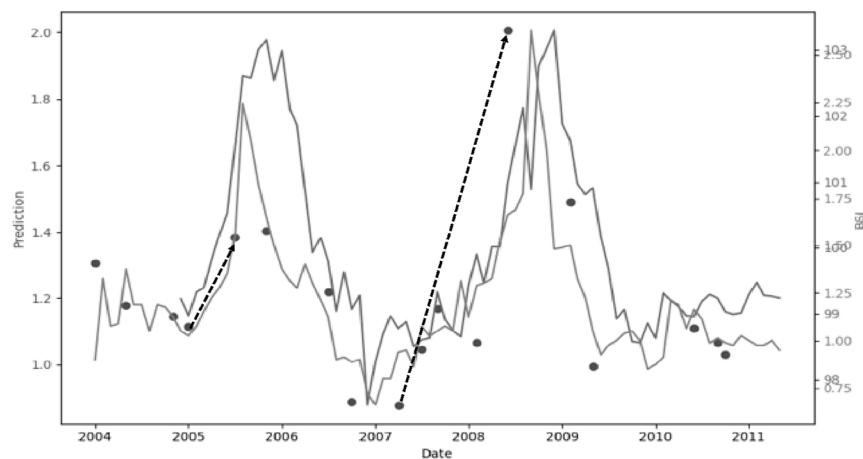


Figure 10. Comparison of BSI, ISSK and selected model

5. Conclusions and Recommendations

The ISSK and BSI are the keys as legal indicator in Indonesia built by regulators, and as initial starting point of this study in order to build the machine learning model and proxied to actual condition of financial and banking stability of Indonesia (Wimanda et.al, 2014). Based on ISSK we classified the level of stability that indicated by integer value of 0 to 3 range of stability level and marked as dependent variable (actual value). RF algorithm could predict the actual condition of banking stability. RF algorithm should be explained and back up with more rigorous computation and data to produce more convincing state of the art. Perhaps using other countries data for test bed.

The data differences could be used to be analysed deeper to measure dynamic processing in order to predict the stability of banking system. Beside that the knowledge of multi system and equations about how variable impact the other variable in the financial system (the transmission channel in financial system) could be nice to explored in the next study.

This methodology in this study could be tried in the next research for outside financial system in the neighbour countries of Indonesia for example, because the contagion effect of crisis should be monitored and to be an input to next indicator in order to make better prediction and study the impact of instability of financial system in other country.

Machine learning algorithm could be used and implemented to detect the stability of overall financial system especially in Indonesia. The amount of data is required and increased in order to our machine learning model more learn with huge data and the prediction should be more accurately and precisely.

Acknowledgement

I wish to thank to my fellows and my colleagues who want to be used as a place to discuss a lot about my thesis, there are two of my friends who give me inspirations in writing they are Hendra Syamsir and Advis Budiman. Thank you for all the information and lessons that have been given. Finally, I would like to all of my student colleagues in SGU, salute to all of you.

References

- David, B. (2017) 'Model economic phenomena with CART and Random Forest algorithms', *EconomiX Working Papers 2017-46*, University of Paris Nanterre, EconomiX.
- Ghosh, S. (2011) 'A simple index of banking fragility: application to Indian data', *Journal of Risk Finance*, 12(2), pp. 112–120. doi: 10.1108/15265941111112839.
- Hawkins, J. and Turner, P. (no date) 'Bank restructuring in practice: an overview'. Available at https://www.researchgate.net/publication/245580482_Bank_Restructuring_in_Practice_An_Overview
- Ilin, A. and Raiko, T. (2010) 'Practical Approaches to Principal Component Analysis in the Presence of Missing Values', *Journal of Machine Learning Research* 11, pp. 1957–2000.
- Illing, M. and Liu, Y. (2003) 'Measuring Financial Stress', *Financial System Review*. Available at <https://www.bankofcanada.ca/wp-content/uploads/2012/02/fsr-1203-illing.pdf>
- Kartasheva, A. V and Traskin, M. (2013) 'Insurers' Insolvency Prediction using Random Forest Classification'. Available at SSRN: <https://ssrn.com/abstract=2364736>
- Khaidem, L., Saha, S. and Dey, S. R. (2016) 'Predicting the direction of stock market prices using random forest', (April). Available at: <http://arxiv.org/abs/1605.00003>.
- Kristína Kočíšová, D. S. (2015) 'Banking Stability Index: New EU countries after Ten Years of Membership', *Working Papers in Interdisciplinary Economics and Business Research*, (December), pp. 1–26.
- Mirkin, B. (2011) 'Principal Component Analysis and SVD', *Core Concepts in Data Analysis: Summarization, Correlation and Visualization*, pp. 173–219. Available at: http://dx.doi.org/10.1007/978-0-85729-287-2_5.
- Nyman, R. and Ormerod, P. (2016) 'Predicting Economic Recessions Using Machine Learning Algorithms'. arXiv:1701.01428. Available at <https://arxiv.org/abs/1701.01428>
- Petropoulos, A. *et al.* (2017) 'Predicting bank insolvencies using machine learning techniques', (April 2017), pp. 1–42.
- Popovska, J. (2014) 'Modelling Financial Stability: The Case of the Banking Sector in Macedonia', *Journal of Applied Economics and Business*, 2(1), pp. 68–91. Available at: <http://www.aebjournal.org/articles/0201/020104.pdf>.
- Ronnqvist, S. and Sarlin, P. (2015) 'Detect & describe: Deep learning of bank stress in the news', *Proceedings - 2015 IEEE Symposium Series on Computational Intelligence, SSCI 2015*, pp. 890–897. doi: 10.1109/SSCI.2015.131.
- Smith, L. I. (2002) 'A tutorial on Principal Components Analysis Introduction', *Statistics, Computer Science Course, University of Otago*, 51, p. 52. doi: 10.1080/03610928808829796.
- Stabilitas, D., Perbankan, S. and Pengantar, K. (2011) 'Laporan Akhir Penyusunan Alat Analisis Risiko Sistemik Dan Stabilitas Sistem Perbankan'.
- Swamy, V. (2013) 'Banking System Resilience and Financial Stability', *Munich Personal RePEc Archive*, (39922), pp. 0–29. Available at: <http://mpra.ub.uni-muenchen.de/47512/>.
- Wimanda, R. E., Maryaningsih, N. and Nurliana, L. (2014) 'EVALUASI TRANSMISI BAURAN KEBIJAKAN BANK INDONESIA'.

Early Detection of Failed Bank Through Analysis of Financial Ratios and Bank Shareholders Ratios Using Data Mining For Rural Banks

H M Agista¹, E Budiarto^{1,*}, and B Mahawan¹

¹ Business Informatics, Swiss German University, Tangerang 15143, Indonesia

*Corresponding author: eka.budiarto@sgu.ac.id

Abstract. This study aims to determine the effect of 8 bank financial ratios such as BOPO (operational efficiency ratio), CAR (Capital Adequacy Ratio), NPL (Non Performing Loan), ROA (Return On Assets), CR (Cash Ratio), KAP (quality of productive assets), PPAP (provision for loan losses) and LDR (Loan Deposit Ratio) and another ratio, namely Bank's Shareholder ratio towards bank predictions whether a rural bank will be declared as failed bank or not. Eight financial ratios and another ratio that comparing BOD and BOC to Bank's Shareholders can be obtained from quarterly rural bank's financial reports that have been published on the IFSA website from 2014 until 2018. The data in this research is approximately 1000 rural banks for training dataset. The method to predict rural bank become failed bank is data mining. The training dataset used is an imbalanced dataset. In order to be balanced, the SMOTE method is used. The balance dataset was then analyzed with the data mining process. The data mining methods used are KNN and Naïve Bayes, both are classification method.

1. Introduction

Deposit insurance is being implemented in many countries to protect bank depositors, in full or in part, from losses caused by a bank's inability to pay its debts when due. Deposit insurance systems are one component that promotes financial stability. One of the countries that have deposit insurance corporation is in Indonesia, called IDIC (Indonesia Deposit Insurance Corporation) or Lembaga Penjamin Simpanan in Bahasa Indonesia. One of the IDIC duties is to handle a failed bank (Indonesia Deposit Insurance Corporation, 2018a).

Before becoming a failed bank, that bank would become under special surveillance. IDIC received some notices on under special surveillance bank from IFSA (Indonesia Financial Services Authority) or Otoritas Jasa Keuangan in Bahasa Indonesia. IDIC, together with IFSA, examined those banks and conduct assessment on their financial condition to asses that bank will be a failed bank or not. If become a failed bank, IDIC must handle the failed bank with several methods of resolution bank (Indonesia Deposit Insurance Corporation, 2019). It would be better if IDIC can predict which bank will be assigned as a failed bank so that IDIC can early prepare its resources and become more careful planning to deal with the bank.

To make an early detection or prediction which bank will become a failed bank, this research will use financial ratios and bank shareholder ratio. This study raises the topic of whether there is a relationship between bank financial ratios and bank shareholders to predict which bank will become a failed bank, especially for rural banks. Rural banks are the main focus because there are 91 rural banks out of the total liquidated banks until the end of 2018. (Indonesia Deposit Insurance Corporation, 2018b)

The data source that will be used is the rural bank's publication reports published in the IFSA's website and failed bank status for the label or class (Indonesia Financial Services Authority, 2018)(Indonesia Deposit Insurance Corporation, 2018b). The scope of the study is limited to the rural bank. The data is obtained from rural bank quarterly publication reports with restrictions from the period 2017 to 2018 and rural bank data that are stated as a failed bank. The data collected is used to create data mining models. The model formed will be used to make predictions to determine whether a rural bank will be declared as a failed bank or not. This research is different from previous research that can be explained in Table 1.

Table 1. Compare with Previous Researches

Author and References	Ni Putu Ayu Lisna Purnamandari and I Dewa Nyoman Badera (Purnamandari and Badera, 2015)	Luciana Spica Almilia and Winny Herdinigtyas (Almilia and Herdinigtyas, 2005)	Roberto Christian Widiharto (Widiharto, 2008)	Reny Sri Harjanti (Harjanti, 2011)	This Research
Bank	22 Banks	24 Banks	201 Rural Banks	27 Banks	About 1000 Rural Banks for creating the dataset
Financial Reports Period	2009-2013	2000-2002	2002-2006	2004-2008	2014-2018
Financial Ratios	CAR, BOPO, NPL	CAR, BOPO, NPL, ATTM, APB, PPAP, Fulfillment, ROA, ROE, NIM, LDR	CAR, BOPO, APB, PPAP, Profit Margin, ROA, LDR	CAR, BOPO, NPL, ROA, ROE, NIM, LDR	CAR, BOPO, NPL, KAP, LDR, PPAP, ROA, Cash Ratio
Another Parameter	Bank Size	-	-	-	Compare directors and commissioners with shareholders name
Method	Regression Analysis	Logistic Regression	Logistic Regression	Logistic Regression	Classification using KNN and Naïve Bayes
Prediction	Bankruptcy	Bankruptcy	Bankruptcy	Bankruptcy	Bankruptcy

2. Materials and Methods

2.1. Dataset

Data consisting of 8 bank financial ratios and 1 shareholder ratio compared to directors and commissioners will be used as parameters or attributes. Whereas bank data is declared as a failed bank, will be a label or class. The number of records for this training dataset is 938 which consists of 920 stated as normal banks in 2017 and 18 declared failed banks in 2015 until 2017.

The training dataset which was formed from 9 attributes and 1 class was prepared from financial report data from 2014 to 2017. Each year has 4 reporting periods, namely Quarter 1 (Q1), Quarter 2 (Q2), Quarter 3 (Q3) and Quarter 4 (Q4). Attributes that will be used is last 4 report period x 9 attributes = 36 attributes. The explanation of the reporting period of training dataset can be seen in Table 2.

Table 2 Time Period of Training Dataset

Bank Name	Liquidation M/D/YYYY	Date	Start Report Period	End Report Period
PT. BPR Carano Nagari	7/10/2015		Q4 2014	Q3 2015
PT. BPR Cita Makmur Lestari	12/18/2015		Q4 2014	Q3 2015
PT. BPR Agra Arthaka Mulya	1/14/2016		Q1 2015	Q4 2015

Bank Name	Liquidation M/D/YYYY	Date	Start Period	Report	End Period	Report
PT BPR Mitra Bunda Mandiri	1/22/2016		Q1 2015		Q4 2015	
PT. BPR Dana Niaga Mandiri	4/13/2016		Q2 2015		Q1 2016	
PT. BPR Kudamas Sentosa	4/29/2016		Q2 2015		Q1 2016	
PT. BPR Mustika Utama Kolaka	6/20/2016		Q2 2015		Q1 2016	
PT. BPR Artha Dharma	8/15/2016		Q3 2015		Q2 2016	
PT. BPR Multi Artha Mas Sejahtera	12/20/2016		Q4 2015		Q3 2016	
PT. BPR Nova Trijaya	1/20/2017		Q1 2016		Q4 2016	
PT. BPR Dhasatra Artha Sempurna	2/3/2017		Q1 2016		Q4 2016	
PT. BPR Nusa Galang Makmur	3/7/2017		Q1 2016		Q4 2016	
PT. BPR Indomitra Mega Kapital	6/15/2017		Q2 2016		Q1 2017	
PT. BPR Tri Harta Indah	6/15/2017		Q2 2016		Q1 2017	
PT. BPR Sisibahari Dana	9/5/2017		Q3 2016		Q2 2017	
PT. BPR KS Bali Agung Sedana	11/3/2017		Q4 2016		Q3 2017	
PT. BPR LPN Kampung Manggis	11/29/2017		Q4 2016		Q3 2017	
PT. BPR Sinar Baru Perkasa	12/6/2017		Q2 2016		Q1 2017	
Other Rural Bank (Normal Bank)	<NONE>		Q1 2017		Q4 2017	

The distribution of training dataset categorized as imbalanced dataset because the ratio of normal bank and failed bank is 920 : 18. There are several options to handle imbalance dataset such as collecting more data to get more example of minor class, changing the performance metric (not just using accuracy for measurement), resampling dataset like over-sampling or under-sampling and generate synthetic samples like using the most popular algorithm like SMOTE (Synthetic Minority Over-sampling Technique) (Brownlee, 2015) (Chawla, N.V., Bowyer, K.W., Hall, L.O., 2002).

Facing this imbalance dataset, this research will choose SMOTE using WEKA to balance the training dataset. Using SMOTE, the imbalance training dataset from 938 records become 1838 records. The ratio of normal bank and the failed bank is 920:918. The final training dataset consist of 36 attribute and 1 class these are BOPO_1, CR_1, KAP_1, KPMM_1, LDR_1, NPL_1, PPAP_1, PS_1, ROA_1, BOPO_2, CR_2, KAP_2, KPMM_2, LDR_2, NPL_2, PPAP_2, PS_2, ROA_2, BOPO_3, CR_3, KAP_3, KPMM_3, LDR_3, NPL_3, PPAP_3, PS_3, ROA_3, BOPO_4, CR_4, KAP_4, KPMM_4, LDR_4, NPL_4, PPAP_4, PS_4, ROA_4, and Liquidation.

The balance training dataset then will be processed using data mining classification methods namely KNN and Naive Bayes with validation using 10 cross-validation to get the accuracy results. The model formed from the data mining process will be used to test the dataset by taking a sample of 14 rural banks with 7 banks declared as failed, and 7 banks declared normal. The reporting period used is starting from 2017 until 2018. This distribution of testing dataset can be seen in Table 3.

Table 3 Time Period of Testing Dataset

Bank Name	Liquidation M/D/YYYY	Date	Start Period	Report	End Period	Report
PT. BPR Bina Dian Citra	4/4/2018		Q2 2017		Q1 2018	
PT. BPR Akarumi	4/25/2018		Q2 2017		Q1 2018	
PT. BPR Budisetia	5/25/2018		Q2 2017		Q1 2018	
PT. BPR Mega Karsa Mandiri	6/5/2018		Q2 2017		Q1 2018	
PT. BPR Sambas Arta	7/12/2018		Q3 2017		Q2 2018	
PT. BPR Sinarenam Permai Jatiasih	11/8/2018		Q4 2017		Q3 2018	
PT. BPR Bintang Ekonomi Sejahtera	11/22/2018		Q4 2017		Q3 2018	
Other Rural Bank (Normal Bank)	<NONE>		Q4 2017		Q3 2018	

The accuracy of the prediction results using the data mining method using the testing dataset can be analyzed by comparing whether the reality and predictions are the same or not.

2.2. Data Mining

Financial ratio data and bank management's name can be used as a dataset to predict which banks will become failed bank by using the data mining method. Data mining is the way toward finding insightful, interesting and new patterns, as well as descriptive, understandable and predictive models from expansive scale data. Data mining includes some tasks like anomaly detection, classification, regression, association rule learning, summarization and clustering (Saurkar *et al.*, 2014). The method of data mining to be used in this research is classification. A Classification consists of examining the features of a new object and assigning the new object from a set of classes that have been determined. Classification is portrayed by well-defined classes, and a training set comprising of reclassified examples. The task is to assemble a model that can be connected to data that isn't classified to classify it. Because the data that will be used in this research has a well-defined class, KNN and Naïve Bayes (Irfan *et al.*, 2018) will be chosen for this research as one of the classification methods for predicting failed bank.

3. Results and Discussion

3.1. Evaluation

The data mining process will be evaluated and interpreted based on a certain size that can be seen in Table 4.

Table 4 Evaluation Item

Evaluation	Calculation
TP (True Positive)	The number of true-positive predictions
TN (True Negative)	The number of true-negative predictions
FP (False Positive)	The number of false-positive predictions
FN (False Negative)	The number of false-negative predictions
accuracy (ACC)	$(\text{Correct predictions})/(\text{Number of Examples}) = (TP + TN) / (TP + FP + FN + TN)$
Precision	$(\text{True positive predictions})/(\text{All positive predictions}) = TP / (TP + FP)$
sensitivity (SN) or recall	$(\text{True positive predictions})/(\text{Number of positive Examples}) = TP / (TP + FN)$
F1 score	$2 (\text{precision} * \text{recall}) / (\text{precision} + \text{recall}) = 2TP / (2TP + FP + FN)$

Facing imbalanced dataset, performance metrics that should be used to evaluate beside Accuracy (ACC) is Precision or Positive Predictive Value (PPV), Recall (REC) or Sensitivity (SN) and F1-Score. (Brownlee, 2015)(Paul, 2018)(He and Garcia, 2009)

ACC is calculated as the number of correct predictions divided by the data set's total number. The best accuracy is 1.0, the worst accuracy is 0.0. PPV is calculated as the number of positive predictions divided by the total number of positive class values predicted. PPV can be thought of as a measure of a classifier's exactness. A low PPV can also indicate a large number of False Positives. SN is calculated as the number of correct positive predictions divided by the positive total number. The best sensitivity is 1.0, the worst sensitivity is 0.0. SN can be thought of as a measure of a classifier's completeness. A low SN indicates many False Negatives. F1-Score is the harmonic average of the PPV and SN, calculated as $2 \times ((PPV \times SN) / (PPV + SN))$. A confusion matrix is formed from the four outcomes produced as a result of binary classification.

This research will utilize one of the data mining apparatuses called RapidMiner. RapidMiner can be used as a tool, to do regression, classification and clustering procedures as well as dimension reduction and parameter optimization. These methods can be utilized for different application domains, for example, text, picture, and time series analysis (Land and Fischer, 2012). Table 6 is a confusion matrix

using KNN method and Table 5 is a confusion matrix using Naïve Bayes that was produced by Rapidminer.

Table 5 Confusion Matrix of Naive Bayes using 10 Cross-Validation

	True NO	True YES
Prediction NO	848 (TN)	6 (FN)
Prediction YES	72 (FP)	912 (TP)

Table 6 Confusion Matrix of KNN using 10 Cross-Validation

	True NO	True YES
Prediction NO	900 (TN)	1 (FN)
Prediction YES	20 (FP)	917 (TP)

The performance result of using KNN and Naïve Bayes can be seen in Table 7.

Table 7 Training Dataset Performance Result

Measures	KNN (%)	Naïve Bayes (%)
Accuracy	98.86	95.76
Sensitivity	97.83	99.35
Precision	99.87	92.68
F1-Score	98.87	95.90

3.2. Prediction

From the experiment above, it shows both of the classification methods produce more than 90% accuracy. Both are good to be used to predict which bank will be declared as a failed bank. Using KNN and Naïve Bayes, this research will apply that model using a new dataset or testing dataset. Table 9 is a confusion matrix using KNN method and Table 8 is a confusion matrix using Naïve Bayes that was produced by Rapidminer.

Table 8 Confusion Matrix of Naive Bayes on Testing Dataset

	True NO	True YES
Prediction NO	7 (TN)	1 (FN)
Prediction YES	0 (FP)	6 (TP)

Table 9 Confusion Matrix of KNN on Testing Dataset

	True NO	True YES
Prediction NO	7 (TN)	2 (FN)
Prediction YES	0 (FP)	5 (TP)

The performance result of using KNN and Naïve Bayes can be seen in Table 10.

Table 10 Performance Result on Testing Dataset

Measures	KNN (%)	Naïve Bayes (%)
Accuracy	85.71	92.86
Sensitivity	71.43	85.71
Precision	100	100
F1-Score	83.33	92.31

4. Conclusion

4.1. Conclusion

This research has shown that data mining can be used to get the pattern and predict the bank stated as a failed bank using financial ratios and shareholders ratios data. This research has been done by using KNN and Naïve Bayes on the financial ratios and shareholder ratios data.

When using 10 Cross-Validation on training dataset about 1838 records, KNN is better than Naïve Bayes because it has 3 higher performance value than Naïve Bayes from 4 selected measurement (accuracy, sensitivity, precision, and f1-score). For prediction using testing dataset about 14 records, it turns out that Naive Bayes is better than KNN because it has 4 higher value than KNN from 4 selected measurement.

Moreover, Naïve Bayes is better in determining Sensitivity (SN). The SN score in the testing dataset shows that KNN has a lower score than Naïve Bayes. A low SN indicates it has many False Negatives (FN). FN score means that the actual data is rural bank declared as a failed bank, but the prediction shows that rural bank is a normal bank and this is a big concern for this case.

By using Naïve Bayes and KNN this research shows that both of them can be used to predict which bank will become a failed bank and normal bank. This can be seen in the experiments using 14 records in the testing dataset. Using Naïve Bayes can predict 6 banks will become failed and in fact, they fail. While 7 banks are still being a normal bank and in fact, they normal. Using KNN can predict 5 banks will become failed and in fact, they fail. While 7 banks are still being a normal bank and in fact, they normal.

From the results of the study, it can be concluded that processing financial ratios and shareholders ratios data using data mining methods such as Naïve Bayes and KNN can be used as an early detection to predict that any banks will become failed bank.

4.2. Future Works

This research can be further developed using data mining with other methods and doing research using monthly financial reports data if being published by IFSA in the future so that the report time span is not too far. Also, it can be further developed using several techniques such as under-sampling and over-sampling to handle imbalanced dataset. This research may be applied in any other DIC (Deposit Insurance Corporation) that have similar characteristics, namely those that have many rural banks as IDIC has.

References

- Almilia, S. L. and Herdinigtyas, W. (2005) ‘Analisis Rasio Camel Terhadap Prediksi Kondisi Bermasalah Pada Lembaga Perbankan Periode 2000-2002’, *Jurnal Akuntansi dan Keuangan*, 7(2), pp. 131–147. doi: 10.9744/jak.7.2.pp. 131-147.
- Brownlee, J. (2015) *8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset*. Available at: <https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/> (Accessed: 1 December 2018).
- Chawla, N.V., Bowyer, K.W., Hall, L.O., K. W. P. (2002) ‘SMOTE: Synthetic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research’, *Journal of Artificial Intelligence Research*, 16, pp. 321–357. doi: 10.1613/jair.953.
- Harjanti, R. (2011) ‘Analisis Pengaruh Rasio Keuangan Terhadap Prediksi Kebangkrutan Bank’, Bachelor thesis, Faculty of Economy, Universitas Diponegoro. Available at <http://eprints.undip.ac.id/29313/1/Skripsi015.pdf>
- He, H. and Garcia, E. A. (2009) ‘Learning from Imbalanced Data’, *IEEE Transactions on Knowledge and Data Engineering*, 21(9), pp. 1263–1284.
- Indonesia Deposit Insurance Corporation (2018a) *Functions, Duties & Authorities of IDIC*. Available at: <https://lps.go.id/web/guest/fungsi-tugas-wewenang> (Accessed: 1 December 2018).
- Indonesia Deposit Insurance Corporation (2018b) *Liquidated Bank*. Available at: <https://lps.go.id/web/guest/bank-yang-dilikuidasi> (Accessed: 31 December 2018).
- Indonesia Deposit Insurance Corporation (2019) *Mechanism of Bank Rescue*. Available at: <https://lps.go.id/web/guest/mekanisme-resolusi-bank> (Accessed: 2 February 2019).
- Indonesia Financial Services Authority (2018) *Laporan Publikasi BPR Konvensional*. Available at: <https://cfs.ojk.go.id/CFS> (Accessed: 10 October 2018).
- Irfan, M. *et al.* (2018) ‘Comparison of Naive Bayes and K-Nearest Neighbor methods to predict divorce issues’, *IOP Conference Series: Materials Science and Engineering*, 434(1). doi: 10.1088/1757-899X/434/1/012047.

- Land, S. and Fischer, S. (2012) 'RapidMiner in academic use', *BOOK 54p*, p. V, 1-3, 6.
- Paul, S. (2018) *Diving Deep with Imbalanced Data*. Available at: <https://www.datacamp.com/community/tutorials/diving-deep-imbalanced-data> (Accessed: 1 December 2018).
- Purnamandari, N. P. Y. L. and Badera, I. D. N. (2015) 'Kemampuan Prediksi Rasio Keuangan dan Ukuran Bank Pada Risiko Gagal Bank', 4(6), pp. 1610–1623.
- Saurkar, A. V *et al.* (2014) 'A Review Paper on Various Data Mining Techniques', *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(4), pp. 2277–128. doi: 10.4010/2016.1512.
- Widiharto, R. C. (2008) 'Analisis pengaruh rasio keuangan terhadap prediksi kondisi bermasalah bank perkreditan rakyat (Studi kepada Bank Perkreditan Rakyat di Wilayah Jakarta, Bogor, Depok, Tangerang dan Bekasi)'.

An Analysis of Moodle Acceptance for Students in SMPK2 PENABUR Using Unified Theory of Acceptance and Use of Technology Model

D Fernando^{1*}

¹ Faculty of Engineering and Information Technology, Swiss German University, Tangerang 15143, Indonesia

*Corresponding author: danif98@gmail.com

Abstract. The synergy of combining globalization with new technology has had dramatic learning impacts. Teachers and Students need to work together to successfully implement Moodle as one of learning technologies. Author needs to determine factors that are important to use Moodle. UTAUT was created based on the conceptual and empirical similarities across TAM, TAM2 and TAM3 models. UTAUT model attempts to explain how individual differences influence technology use. Structural Equation Modelling (SEM) is then used to assess the reliability and validity of the model measures. Exploratory Research is conducted in order to determine the nature of the problem and is not intended to provide conclusive evidence. With Uncontrolled Quota Sampling method, respondents filled Likert-scale questions. Analysis of a Moment Structures (AMOS) is used analyse the data using SEM. All hypotheses is verified through different set of experiments. UTAUT is acceptable of analysing respondents' acceptance in using Moodle. Each construct uniquely affects the output, therefore different strategies can be implemented based on respondents' characteristics. Performance Expectancy is important for Students.

1. Introduction

Schools tried to adapt learning technologies which combine text, audio, video, animation and simulation, deliver them through the internet, and students can learn virtually from anywhere and anytime by using their computer and internet connection.

Teachers are responsible for displaying learning resources in more attractive ways, making the learning process more interesting and enjoyable because some students are more receptive to some kind of information than the others (Franzoni, Ana Lidia, and Said Assar, 2009).

Moodle is one of learning technology available, is an open source web-based application. Moodle can be accessed anywhere and anytime, provide features such as storing documents, managing quizzes and information resources. Moodle is also an alternative of providing materials and access to learning activities where face-to-face or classroom activities are limited. Teachers can upload various material digitally to Moodle, then students access those materials depends on their need and learning speed.

Venkatesh (Venkatesh et al, 2003) developed the Unified Theory of Acceptance and Use of Technology (UTAUT) model as a consolidation of previous TAM model. UTAUT serves as a useful tool for decision makers to determine key success factors for new technology implementation and helps them understand factors of acceptance in order to proactively design training materials and marketing, targeted at group of users that may be more difficult to embrace and use new technology. This paper focuses on determining factors that are significant to get more students to use Moodle using Unified Theory of Acceptance and Use of Technology model.

2. Materials and Methods

Currently usage statistics shows that Moodle was accessed by 100 out of 140 students daily. SMPK2 is planning to have Moodle accessed by 140 students next year, therefore we need to examine factors or variables that are significant and important for students to use Moodle, and focuses on Human as root cause of why Moodle is not being used by all students (Figure 1)

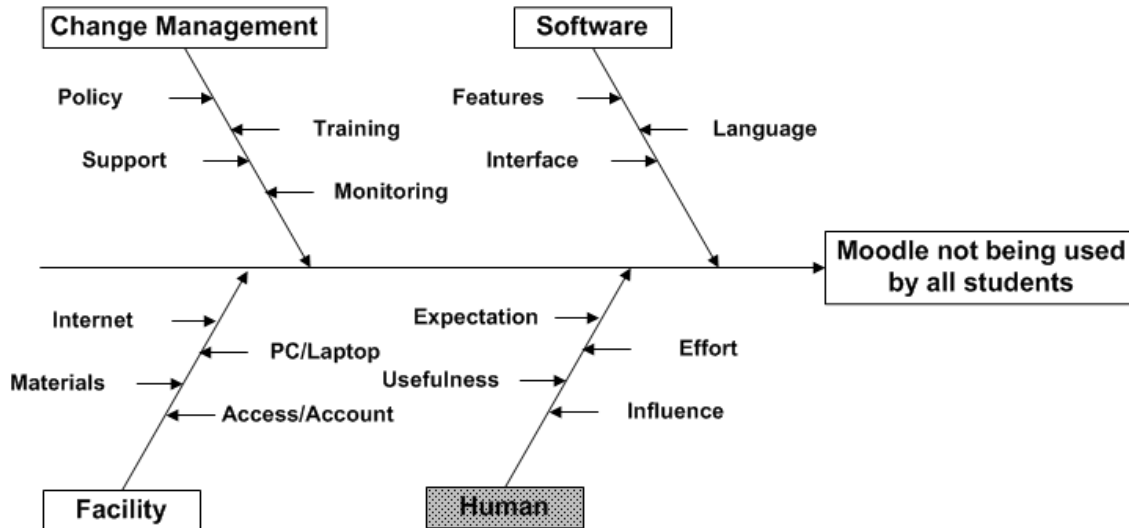


Figure 1 Root Cause Analysis

Hypotheses are visualized in Figure 2, described as follow:

- H1: There is a significant relationship between student's attitudes towards using Moodle and its Performance Expectancy.
- H2: There is a significant relationship between student's attitudes towards using Moodle and its Effort Expectancy
- H3: There is a significant relationship between student's attitudes towards using Moodle and its Social Influence
- H4: There is a significant relationship between student's attitudes towards using Moodle and its Computer Self-Efficacy
- H5: There is a significant relationship between student's attitudes towards using Moodle and its Perceived Enjoyment
- H6: There is a significant relationship between student's attitudes towards using Moodle and its Facilitating Conditions
- H7: There is a significant relationship between student's attitudes towards using Moodle and its Behavioural Intention

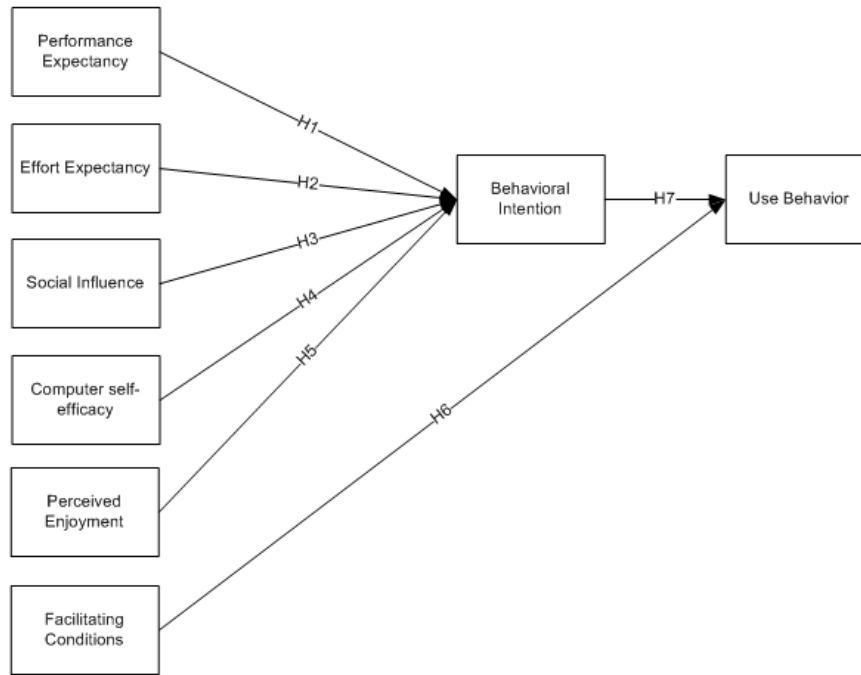


Figure 2 Hypothesis

Unified Theory of Acceptance and Use of Technology (or UTAUT) – suggested in 2003 by Venkatesh, Morris, Davis and Davis, developed based on the theoretical and empiric closeness across existing models. The UTAUT was built with four constructs: Firstly performance expectancy, secondly effort expectancy, thirdly social influence and fourthly facilitating conditions. These constructs are direct determinants or predictors of usage intention and behaviour (Venkatesh et.al. 2003). The predictors are defined as (Venkatesh et al, 2003):

1. Performance expectancy (PE): “is the degree to which an individual believes that using the system will help him or her to attain gains in job performance.”
2. Effort expectancy (EE): “is the degree of ease associated with use of the system.”
3. Social influence (SI): “is the degree to which an individual perceives that important others believe he or she should use the new system.”
4. Facilitating conditions (FC): “is the degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system.”

In the UTAUT model, performance expectancy and effort expectancy were adopted to integrate the constructs of perceived usefulness and ease of use. It is crucial to gain its perceived usefulness from the student angle. The results of this research might help schools make better arrangement and campaign and assist teachers in using technology more effectively (Marchewka, Jack T., and Liu Chang, 2007), by considering these perceptions.

3. Results and Discussion

We include student respondents in this study. We build the model (Figure 3) as described in research model as well as questions prepared to Students, and Table 1 shows the calculated estimates.

Table 1. Calculated Estimates

	PEJ	CE	SI	EE	PE	FC	BI
BI	.163 (H5)	.334 (H4)	.126 (H3)	.041 (H2)	.321 (H1)	.000	.000
Use Behavior	.019	.039	.015	.005	.037	.111 (H6)	.116 (H7)

We summarized those values into Hypothesis Coefficients in table 2 and figure 4.

Table 2. Hypothesis Coefficients

Hypothesis	Effects	Value	Remarks
H1	PE → BI	0.321	High effect
H2	EE → BI	0.041	Low effect
H3	SI → BI	0.126	Medium Effect
H4	CE → BI	0.334	High effect
H5	PEJ → BI	0.163	Medium Effect
H6	FC → Use behavior	0.111	
H7	BI → Use behavior	0.116	

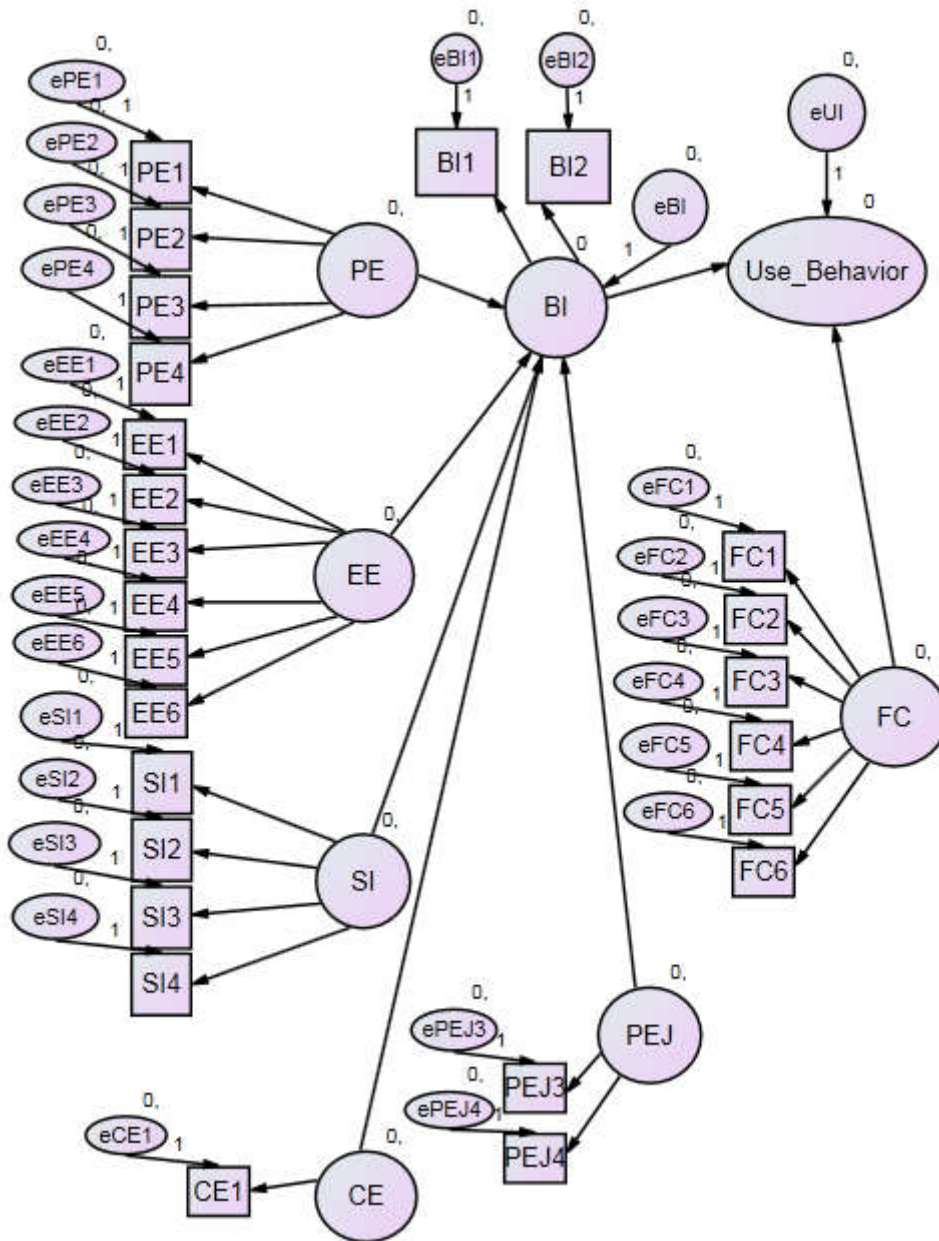


Figure 3 Experiment Model

The greater the hypothesis coefficients, the more necessary the impact it will be. We take Performance Expectancy construct and Computer Self-Efficacy construct as factors significant for respondents. Social Influence and Perceived Enjoyment considered as fair factors, and Effort Expectancy as the least important. We have very slight difference value on Facilitating Conditions and Behavioural Intention.

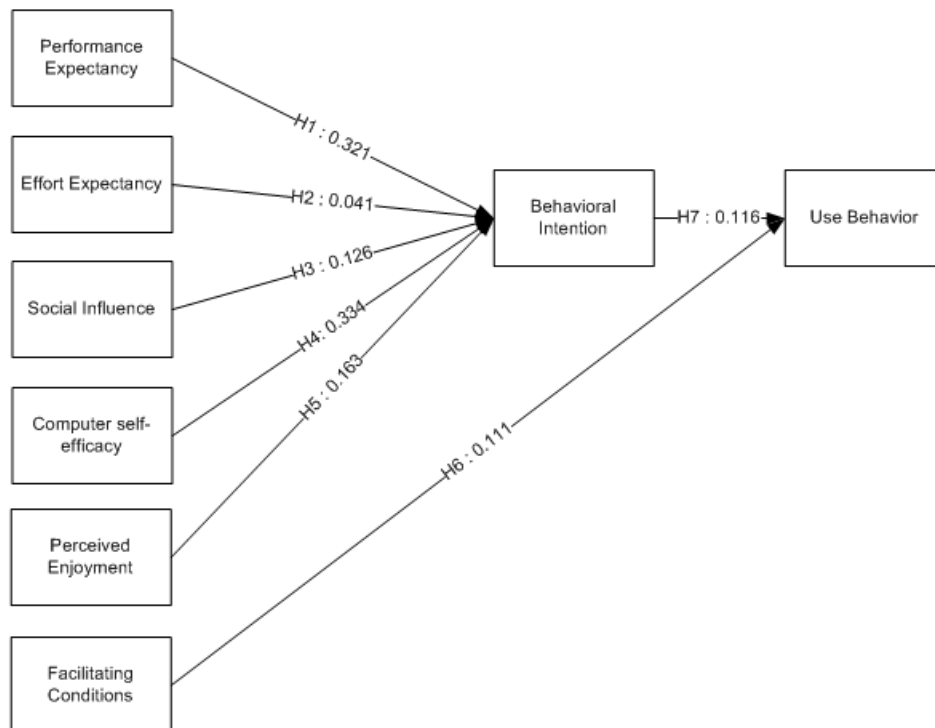


Figure 4 Hypothesis Coefficients

Data from respondents' results are (sorted from the highest) Computer Self-Expectancy, Performance Expectancy, Social Influence and Perceived Enjoyment. Computer Self-Efficacy turns out to be important factor in all the experiment. Therefore the statement in Computer Self-Efficacy question -"I'm an IT-minded Student" became significant. When we observe Performance Expectancy questions, we found that "Moodle useful for understanding topics and achieve higher score" is important as well.

Data from respondents' results "Support from friends, teachers and school" is important. We also discover that "Moodle Easiness" had small effect, meaning Students were convinced in advantage of Moodle even though they might have different level of how to operate Moodle.

"Hardware, Knowledge, Skill, Support Team, Internet and Server Performance" is an important factor that support the Intention. We conclude that Facilities provided by School, will amplify Intention.

The conclusion is all hypotheses verified.

4. Conclusion

It is clear that Unified Theory of Acceptance and Use of Technology is acceptable of analysing Students acceptance in using Moodle. The constructs developed each uniqueness in how it affects the output. These would also can be considered and helpful when creating approaches to ensure more Students interest in using Moodle. Only by implementing different strategies based on student's characteristics, BPK PENABUR Jakarta will succeed in implementing Moodle as a powerful platform for education.

This study is widely opened to improvement. One of the improvements needed is to measure how those factors give us different results on different subjects. This study involved only how respondents accepts Moodle. Parameters at subject level may give us better understanding on how respondents accept specific subject in Moodle environment.

Several recommendations for BPK PENABUR Jakarta in implementing Moodle of other schools as well as SMPK2 new generation as follows:

1. Mind-set comes first. Always state to Students that they are IT-Minded persons and they will be succeeded in using technology when they believe it.
2. Students will accept technology in general, or Moodle in specific, when they know that this technology or Moodle can give them better performances.

Acknowledgement

I wish to thank the members of my committee for their support, patience and broaden view. Their gentle but firm direction has been most appreciated. I would like to thank Dr. Ir. Gembong Baskoro, M.Sc. (Advisor) and Dr. Maulahikmah Galinium, S.Kom, M.Sc. (Co-Advisor) for directing me and guiding me through difficulties and complete this research. From the beginning, they had confidence in my abilities to not only complete a degree but to complete it with excellence. I would like to thank Mr. Supanna Wirija (SMPK2 PENABUR Principal), Mr. Teguh Santoso (Teacher) and Mr. Christophorus Tommy Astanto (Teacher) for providing me access to Moodle usage and respondents, and advised me on research model.

Appendix A: Questionnaire

Constructs	Question	Questions	1	2	3	4	5
PE	PE1	I find Moodle useful in my learning					
	PE2	Moodle helps me to understand topics better					
	PE3	Moodle makes me achieve more					
	PE4	Moodle makes me get higher scores/grades					
EE	EE1	I find that Moodle is easy to use					
	EE2	I can operate Moodle easily					
	EE3	Teachers can organize materials and quizzes/ questions bank easily					
	EE4	Teachers can analyze Students weakness in certain topics or subjects easily					
	EE5	Students can self-prepare themselves before going into classes easily					
	EE6	Students can easily review materials or repeat subjects for those they considered they are weak in					
SI	SI1	Students: My friends think I should use Moodle					
	SI2	My teachers think I should use Moodle					
	SI3	My IT teachers give their support in the use of Moodle					
	SI4	My school supported the use of Moodle					
CE	CE1	I'm an IT-minded Student					
	CE2	I'm an IT-minded Teacher					
PEJ	PEJ1	Teachers enjoy organizing materials and quizzes/ questions bank					
	PEJ2	Teachers enjoy analyzing Students weakness in certain topics or subjects					
	PEJ3	Students enjoy self-preparation before going into classes					
	PEJ4	Students enjoy reviewing materials or repeating subjects for those they considered they are weak in					
FC	FC1	I have hardware required to use Moodle					
	FC2	I have knowledge and skill to operate Moodle					
	FC3	Moodle is compatible with other applications I use (Microsoft Office)					
	FC4	My school provided support team in case of difficulties I experience with Moodle					
	FC5	Moodle enable Students to learn from home					
	FC6	Internet and server hardware performance encourage Students/Teachers to use Moodle					
BI	BI1	I intend to use Moodle in the next (n) days					

Constructs	Question	Questions	1	2	3	4	5
	BI2	I plan to use Moodle in the next (n) days					

Note. PE = performance expectancy, EE = effort expectancy, SI = social influence, CE = Computer Self-Efficacy, PEJ = Perceived Enjoyment, FC = facilitating conditions, BI = behaviour intention.

References

- Franzoni, Ana Lidia, and Said Assar, 2009. Student Learning Styles Adaptation Method Based on Teaching Strategies and Electronic Media. [Online]
Available at: http://www-public.int-evry.fr/~assar/pdf/ETS_Franzoni-Assar.pdf
- Marchewka, Jack T., and Liu Chang, 2007. An Application of the UTAUT Model for Understanding Student Perceptions Using Course Management Software. *Communications of the IIMA*, pp. 93-94.
- Venkatesh et al, 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, pp. 425-478.

Predictive Analysis on Employee Performance

S L Tendean¹, M R Mashudi¹ and E Budiarto^{1*}

¹ Department of Information Technology, Swiss German University, Jl. Jalur Sutera Barat 15, Alam Sutera, Tangerang 15143, Indonesia

*Corresponding author, email: eka.budiarto@sgu.ac.id

Abstract. Human resources management have key roles during running some business. Some good investment in this human resources managements gives some benefit for company to gain more income. Hiring good employee and build some good organization inside company are needed to achieve good running business. The objective in this research is to determine some parameters from some performance record in X company that have key roles for human resources management to give some good hiring affecting business income. The Data that used got from the X company from 2014-2018 with around 14 parameters and 10 human performance record parameters. From this needs, this research describe some models such as multiple regression and random forest as validation focusing on analyzing the human performance record correlated to the revenue data set in PT X Company. As conclusion, this research concluded 3 parameters in human performance record that very important to determine good running business .

1. Introduction

1.1. Background

Around the world, Human Resources have several challenges that they are facing especially in small business. The big company have aware to this condition because they have spent big money to maximize their human resources management works and roles. The main role of human resources management is to build good organization so the company can run more efficient and gain more income. The main challenges that have mostly company facing in their human resources management around the world, which are (Gawke et al., 2018):

- Leadership Development: many small company or small running business doesn't mention and keep eye on their management team. This matter will cause their team lack of responsibility doing their job
- Change of Management Level : Along with the running business slightly become more bigger, the employee inside the company didn't aware about this matter and cannot adapt along the internal main progress and will make the work productivity decreased.
- Recruiting Employees: This problem still maintain in many companies that didn't invest good enough money in their HRM management. Some of company does not have any rules or standard in their human resources management to recruit new people. Some of company just recruit average people and could be become great impact for their running business

In PT X company, the problem occurring is the lack of knowledgement about the human resources management key roles. The manager sometimes didn't know current problem they are facing during running some business example like: In 1-2 years their sales income are fluctuative. The Human Resources Management also doesn't have any standard requirement for hiring new employee based on their performance record data so some of the new employee only works 1-3 months.

1.2. Human Capital and Business

Human capital is very important during running some business. The key of element in human capital were come from the human resource management. They maintain condition of organizational and employee carrer development. Firstly human capital can be described as a key component in moving forward a firm resources and workers in arrange to extend their main income as well as maintain their business. Human capital is also essentially can be described as a way that connected into instruction, preparing, and other proficient activities to build the levels of knowledge, ability, capacities and social resources of an employee

1.3. Regression Analysis

Regression analysis is sort of modelling technique that investigates correlation between a dependent and variable quantity. This technique is employed for prediction, statistic modelling and finding casual result between the variables. This research is using regression analysis for modelling for several reason (Shah and Aman, 2019):

- Regression analysis helps human resources management to ensure skills requirement for the organization and execute plan for recruitment process
- Related to revenue, regression analysis also used to predict number of employee and correlated to human performance record management according to target and current sales condition.

2. Material and Method

2.1. Research Framework

In this research, the experiments is focusing in organization area in PT X Company which is the first step to develop good individual skills. Good individual skills are develop not only from the employee itself but also got some support from human resources management to maintain their satisfaction and their skills during doing some works. The human resources management have key roles to develop good employee skills and good organization to be later can build some good operation and future strategic planning. Applying data mining method to define the selection of parameters is needed from Data Set in PT X Company. The Research framework to define the parameters is provided in Figure 1.

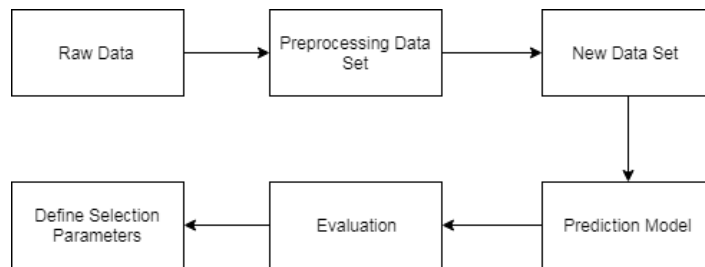


Figure 5 Research framework

2.2. System Overview for Analyze Data Set

In this subchapter, the overview of software system is given based on the research framework. Using data set of human resources management yearly data contains of employee performance and some of their personal data from X company taken around 4-5 years and conducting some propose method of experimental procedure to analyze the data is explained in Figure 2

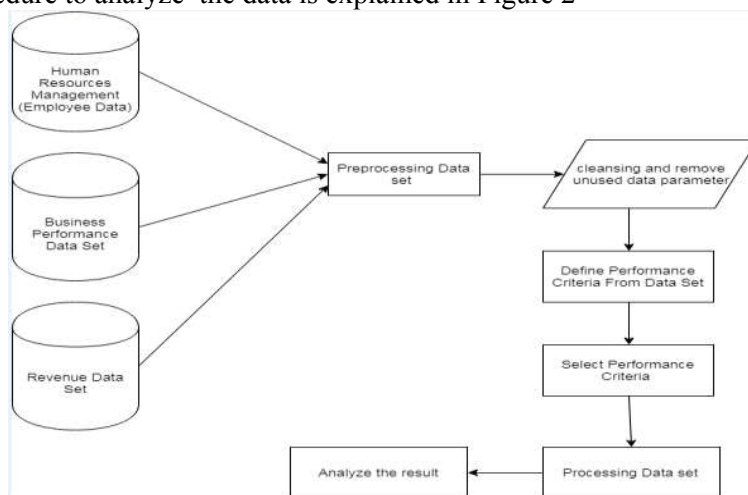


Figure 6 System Overview

2.3. Data Source Collections

In this research, the data are taken from PT. X company as data source. The data content are used by the human resource management for yearly report for employee performance along with some of their personal data. From Performance Data data set and human resources data will be later correlated to Revenue Data taken respectively from 2014-2018 to select the parameters that affecting running business. The Revenue Data set that will be used in this research is shown from Table 1 until Table 3. In this research, multiple and linear regression is best method to analyze the performance data set due to availability of data in numerical in PT X Company for selecting parameters. Revenue Data set is needed in this research to correlate between employee performance especially in Sales Department which are divided into 3 department (Disposable, Electrical and Laboratory) with the revenue in PT X Company

Table 1 Human Resources Data Parameters

No	Features
1	Date Joined
2	Attendance Number
3	Education
4	Place of Birth
5	Date of Birth
6	Age
7	Retirement Age
8	Position
9	Department
10	Address
11	Employee Status
12	Marital Status
13	Phone Number
14	Identity Number

Table 2 Performance Parameters

No	Features
1	Appearance
2	Knowledge
3	Enthusiastic
4	Quality
5	Responsibility
6	Innovation/Initiative
7	Behavior
8	Teamwork
9	Capability
10	Work on Time

Table 3 Revenue Data Set

NO	KETERANGAN	DIVISI DISPOSIBLE	DIVISI LABORATORIUM	DIVISI ELECTRIC	TOTAL
1	PENJUALAN TH. 2014	198,970,761,460	28,089,989,853	7,022,497,463	234,083,248,776
2	PENJUALAN TH. 2015	225,782,369,598	29,882,960,682	9,960,986,894	265,626,317,174
3	PENJUALAN TH.2016	255,218,660,776	29,167,846,946	7,291,961,736	291,678,469,458
4	PENJUALAN TH. 2017	247,192,122,409	37,078,818,361	6,543,320,887	290,814,261,657
5	PENJUALAN TH. 2018	238,813,726,113	35,822,058,917	6,321,539,808	280,957,324,838

3. Result and Discussions

The company starts using performance data set for every supervisor and higher manager in several department to give some scoring according to their team and their colleague regarding their performance during works. The scoring of this performance record later is given to human resources management to give some credit and appraisal. However, in PT X Company, Human Resources Management only uses this performance data to arrange every employee's appraisal and benefit according to their performance. In this chapter, the result of multiple regression is applied to determine selection parameters in performance record correlated to the revenue. The Electrical Department is chosen as testing because the coefficient correlation result between parameters in performance record have highest values

3.1. Regression Result

Table 4 Summary of Regression Correlation Result

Regression Statistic Value	Disposable Department	Laboratory Department	Electrical Department
R-Square Value	0.5073	0.7261	0.4443
Multiple R (correlation Coefficient)	0.7122	0.8521	0.6665

Based on result from Table 4, The performance parameters have been reduced from 10 into 5 by using statistical correlation. From previous study it is indicated that the correlation between performance record and revenue have good relationship with the performance record from electrical Department have high coefficient correlation result and have good R square. According to previous study, it is indicated that the value of R-squared > 0.3898 have good relationship in regression analysis and can be applied as references in financial performance of some company (Shara et al., 2019). The value of R-squared relative small because some of data in company give some similiarity of data distribution and will give some impact for prediction result. From Table 4, the selection of performance parameters cannot be determined because the result of the regression analysis indicated the total value of performance record of 5 selection parameters record. Hence another method multiple regression will be implemented

3.2. Multiple Regression Result

The result of this multiple regression result reduced from 5 parameters which are Pengetahuan Pekerjaan (Knowledge), Tanggung Jawab (Responsibility), Inovasi Pekerjaan (Innovation), Kapabilitas Dalam Berkerja (Capability), and Kerja Tepat Waktu (Work On Time) into 3 parameters which are Tanggung Jawab (Responsibility), Inovasi Pekerjaan (Innovation) and Kapabilitas Dalam Berkerja (Capability) according to P-values and significance values because it is need to be reduced until below the value below 0.05 to give more satisfied result with by reducing some error as shown on Table 5. Significance value below 0,05 means that the regression model can be accepted and each element of data have correlated into the revenue

Table 5 Regression Statistical Electrical Department Result

Regression Statistic	Value Before Parameters Removal (5 Parameters)	Value After Parameters Removal (3 Parameters)
Multiple R Value (Coefficient Correlation)	0.8945	0.890027
R-Squared Value	0.8008	0.792148
Significance	0.1403	0.018042

The statistical of each parameters of multiple regression result after parameters removal is represented in Table 6.

Table 6 Regression Value after Parameters Removal

	Coefficient	P-values
Intercept	17428340686	0.003467
Tanggung Jawab	-3007425308	0.012554
Inovasi Pekerjaan/ Inisiatif	2078877232	0.022566
Kapabilitas Dalam Bekerja	-1560569854	0.062099

Based on result on Table 6, The last 3 parameters have crucial points based on the data in human resource management in PT X Company affecting running business. The multiple regression equation is represented in equation (1) below

$$y = 17428340686 - 3007425308 A + 2078877232 B - 1560569854 C \quad (1)$$

Based on equation above the following parameters indicated:

- Y=predicted Revenue
- A= Performance Record Value of (Tanggung Jawab)
- B= Performance Record Value of (Inovasi Pekerjaan/Inisiatif)
- C= Performance Record Value of (Kapabilitas Dalam Bekerja)

From Equation 1 above, from 3 important parameters, the innovation value have key important for PT X Company to increase their revenue. By providing some training to give opportunity for their employee to develop their skills and know their capability of works.

3.3. Validation Model

In this research random forest will be used to validate our selection parameters and model using regression is good or not to compare the correlation of the coefficient. This method is suitable for PT X Company data set due to small number of data size and this validation method will give more accurate prediction parameters with small number data set and this method have correlated with the multiple regression model based on it's correlation coefficient and current data availability. The result of the comparison model is given on Table 7.

Table 7 Model Validation Result between Multiple Regression and Random Forest

Regression Statistic	Multiple Regression	Random Forest
Multiple R(Correlation Coefficient)	0,89	0,8851
R-Squared Value	0,7921	0,7835

4. Conclusion

Based on purpose of this research is to determine performance record parameters that affecting running business which are revenue in PT X Company. Firstly the selection of performance record from 10 parameters is reduced to 5 parameters based on previous studies(Sutanto, 2004). Using regression analysis model can be determined the correlation between those 5 parameters performance records and revenue however the models shows that the parameters cannot be as individual parameters but as one parameters give better result. Therefore similiar Regression method called Multiple Regression applied to Electrical Department Performance Record Data due to it's coefficient correlation is very good. Further analysis shows that our selection parameters which are 5 attributes from statistical result can be reduced to 3 parameters and give good regression result and can be determined as individual parameters for human resources management in PT X Company.

To validate the regression result, another method which is random forest give nearly similiar result of coefficient correlation result. In Conclusions, our hypothesis to select several parameters can be proved by using model in this research. However based on our hypothesis, knowledge parameters doesnt give correlation to running business in PT X Company. The parameters that should be considered in human resources PT X Company are Innovation, Responsibility and Capability, this parameters also have some correlated with previous study indicated that the innovation should be considered in organization to build employee's skills and capability of their works along with their knowledge (Zarraga-Oberty et al., 2007) (Shah and Aman, 2019).

From the result of this research, it can be concluded that the 3 parameters which are Responsibility, Innovation and Responsibility from human resources performance data set parameters are important to determine good running business in PT X Company. For further research, this 3 parameters can be apply into another small or medium company with bigger data set to determine the accuracy and correlation coefficient to get better result and comparison from this small data set. More over, another method can be also applied to explore more parameters that very important in human resources department to determine good business and revenue.

References

- Shah, M., Aman, Q., 2019. The impact of human resource management practices on leadership styles: The mediating role of employee trust. *City Univ. Res. J.* 9, 58–71.
- Shankar, R.S., Rajanikanth, J., Sivaramaraju, V., Murthy, K.V., 2018. PREDICTION OF EMPLOYEE ATTRITION USING DATAMINING, in: 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA). IEEE, pp. 1–8.
- Shara, Y., Muda, I., Rujiman, R., 2019. Role of Organizational Commitment to the Factor of Performance of Regional Financial Management, in: 1st Aceh Global Conference (AGC 2018). Atlantis Press.
- Sutanto, E.M., 2004. Forecasting: the key to successful human resource management. *J. Manaj. Dan Kewirausahaan* 2, 1–8.
- Thoman, D., Lloyd, R., 2018. A Review of the Literature on Human Resource Development: Leveraging HR as Strategic Partner in the High Performance Organization. *J. Int. Interdiscip. Bus. Res.* 5, 147–160.
- Thomas, R.J., Cheese, P., Benton, J.M., 2003. Human capital development. *Accent. Inst. High Perform. Bus.*
- Tinsley, H.E., Brown, S.D., 2000. *Handbook of applied multivariate statistics and mathematical modeling*. Academic press.
- Van Der Aalst, W., 2016. Data Mining, in: *Process Mining*. Springer, pp. 89–121.
- von Bonsdorff, M.E., Zhou, L., Wang, M., Vanhala, S., von Bonsdorff, M.B., Rantanen, T., 2018. Employee age and company performance: An integrated model of aging and human resource management practices. *J. Manag.* 44, 3124–3150.
- Xia, J., Mandal, R., Sinelnikov, I.V., Broadhurst, D., Wishart, D.S., 2012. MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. *Nucleic Acids Res.* 40, W127–W133.
- XIONG, S., Chen, L., CHANG, L., XIE, A., 2019. Impact Analysis of Financial Early Warning Indicators Based on Random Forest. *DEStech Trans. Comput. Sci. Eng.*
- Zarraga-Oberty, C., Bonache, J., others, 2007. Human factors in the design of revenue management systems in multinational corporations. *Int. J. Revenue Manag.* 1, 141–153.
- Gawke, J.C., Gorgievski, M.J., Bakker, A.B., 2018. Personal costs and benefits of employee intrapreneurship: Disentangling the employee intrapreneurship, well-being, and job performance relationship. *J. Occup. Health Psychol.* 23, 508.
- Cheng, Z., Chen, Y., 2012. Data mining applications in human resources management system. *J. Converg. Inf. Technol.* 7.

Determining Critical Factors Project Delay and Effecting Cost Overruns in Telecommunication Mobile Network Projects

Deni Bakhtiar¹, Mulya R Mashudi¹ and Maulahikmah Galinium^{1*}

¹ Department of Information Technology, Swiss German University, Tangerang 15143, Indonesia

* Corresponding author, email: maulahikmah.galinium@sgu.ac.id

Abstract. Telecommunication Industry has created numerous employment opportunities and playing a role model in economic growth of Indonesia. This industry is facing serious and critical problem of cost overruns in Telecommunication Mobile Network Projects, especially in implementation 4G-LTE Project in Jakarta, Depok, Bogor, Tangerang, Bekasi (Jadebotabek) area. The purpose of this research is to identify and classify critical factors project delay and expound the effect of that critical factors to cost overruns. A structured questionnaire based on 38 factors project delay related to cost overruns (Ullah *et al.*, 2017) was developed and distributed to 100 respondents from 35 selected Contractor Company in Jadebotabek area. The collected data was statistically analyzed with Factor Analysis and Multiple Linear Regression that processed using SPSS. As a result, 4 new Critical Factor Project Delay effect to Cost Overruns in sequence are Lack of cost plan and monitoring, Equipment failure, Extension of time, and Inadequate project preparations were identified.

1. Introduction

Telecommunications industry plays an important role in every economy of a country. Indonesia as a developing country whose telecommunications industry continues to grow and has a stable outlook, recorded double-digit growth for more than a decade and nearly doubled the GDP (Gross Domestic Product) growth rate in 2017. The Central Statistic Agency (BPS) informs that the information and communication sectors (including the telecommunications industry) grew better 9.81% in 2017 compared to the previous year 2016 which reached 8.87%. This growth is far above the national economic growth. Telecommunication industry is one of the main drivers of Indonesia's economy (Telkom, 2017)

Telecommunication Mobile Network Project is a Telecommunication Project that must be supported by large capital, and for developing countries there is limited financial resources for it because it has to do other development. The high complexity in implementation 4G-LTE Technology required proper and accurate planning and budgeting. Inability to complete a project on time and on budget becomes a problem that always appears and worsens. Over the last year only 40% of the project's corresponding schedule, budget and quality, on the other hand only less than one-third of all projects successfully completed on time and within budget (The Standish Group, 2017). There are 65 and 80% projects failing to suit goals, completed with very late or much more costly than planned. In period 2011 until 2018, the project that completed on time are always below 55% as well as the project that completed within original budget are always below 60%. It concludes that time and cost have an important role that causes project delays (PMI, 2018). Project delay not only impede the productivity of the Telecommunications industry, but also perpetrate heavy blows to the national economy.

Due to lack of study on the critical factors project delay and its effect to cost overruns in Telecommunication Mobile Network Project; thus, this research is to identify and classify critical factors project delay and expound the effect of that critical factors to cost overruns in Telecommunication Mobile Network Projects, especially in implementation 4G-LTE Project in Jakarta, Depok, Bogor, Tangerang, Bekasi (Jadebotabek) area.

Project Management

Project is temporary endeavor carried out to produce a unique product, service, or result and Project management is the application of knowledge, skills, tools, and techniques to project activities to meet the project requirements (PMI, 2017).

A very good knowledge of project management is needed to manage projects in the telecommunications industry. It takes good governance and monitoring from a project manager in addition to project planning, problem analysis and new requirements, and always for time and cost efficiency.

Project Delay

Project delay is the project going through delay during the implementation period where the difference gap between actual work progress and work schedule with the ranges of ten to thirty percent (Hamzah *et al.*, 2011).

Cost Overruns

As one of the main considerations throughout the life cycle of project management, costs are also considered a success factor. Cost overrun is the cost of a project during the implementation phase, exceeds (budget) the project budget that is set at the beginning (estimated cost), thus causing significant losses to the contractor.

Over a period, diversified factors have been brought forward that have led to project delay. Previous research can be distinguished into two areas. The first includes research concerned about factors project delay. The second area includes research that identify conditions about factors project delay that effecting cost overruns. Various factors that were raised from the results of previous studies were used as a reference for selecting factors project delay in this research. The framework project delay that effect to cost overruns from table.1 are the points of discussion in this research.

Table 1. 38 Factors Project Delay Related to Cost Overruns (Ullah *et al.*, 2017)

	Code	
Causes by Contractor	1. Inappropriate planning	X1
	2. Improper project management	X2
	3. Lack of contractor experience	X3
	4. Financial difficulties faced by contractors	X4
	5. Relationship between management & labour	X5
	6. Incorrect scheduling by contractors	X6
	7. Errors during implementation	X7
	8. Extension of time	X8
	9. Incompetent subcontractors	X9
	10. Unsuitable implementation equipment	X10
	11. Improper implementation methods	X11
	12. Poor financial control on site	X12
Causes by Consultant	13. Poor contract management	X13
	14. Mistake in design	X14
	15. Underestimate project duration	X15
	16. Slowdown in design preparation	X16
	17. Delay in approval of design changes	X17
	18. Lack of cost plan/ monitoring	X18
	19. Inadequate project preparations	X19
	20. Lack of coordination at design phase	X20
	21. Incomplete design at tendering stage	X21
	22. Lack of consultant teams experience	X22
Causes by Client	23. Financial problems faced by owner	X23
	24. Practice of assigning contract to lowest bidder	X24
	25. Lack of communication with consultant	X25
	26. Change in client requirements	X26
	27. Delay in payment	X27
	28. Re-measurement of provisional sum	X28
Others/ External Causes	29. High cost machineries	X29
	30. Fluctuation in price of raw materials	X30
	31. Unforeseen site conditions	X31
	32. Shortage of site workers	X32
	33. Delay in material procurement	X33
	34. Unpredictable weather conditions	X34
	35. Poor labor productivity	X35
	36. Unskilled labor	X36
	37. Equipment failure	X37
	38. Labor Absenteeism	X38
How 38 Factors Project Delay Effect to Cost Overruns	Y	

Factor Analysis

Factor analysis is used to reduce many variables to fewer. Carried out extraction of the maximum general variant of all variables and placed in general score. It is subsequently used as an index of all variables and this score is used for further analysis.

Classic Assumption Test

The classic assumption tests are Normality test, Multicollinearity test, Autocorrelation test, Heteroscedasticity test and Linearity test. Classic Assumption Test done before proceeding Multiple Linear Regression Analysis (MLRA) and is used to determine validity MLRA.

Multiple Linear Regression Analysis

Multiple Linear Regression Analysis (MLRA) is used to determine a mathematical relationship among a number of random variables, widely used to estimate significance effects of independent variables to dependent variable. In this study, MLRA is used to determine the effect of independent variables (38 original factors project delay (X)) to dependent variable (cost overruns (Y)).

2. Materials and Methods

A quantitative research strategy was carried out in four stages; literature review, data collection through survey, data analysis and conclusions. Research Framework in this research capture in Figure 1.

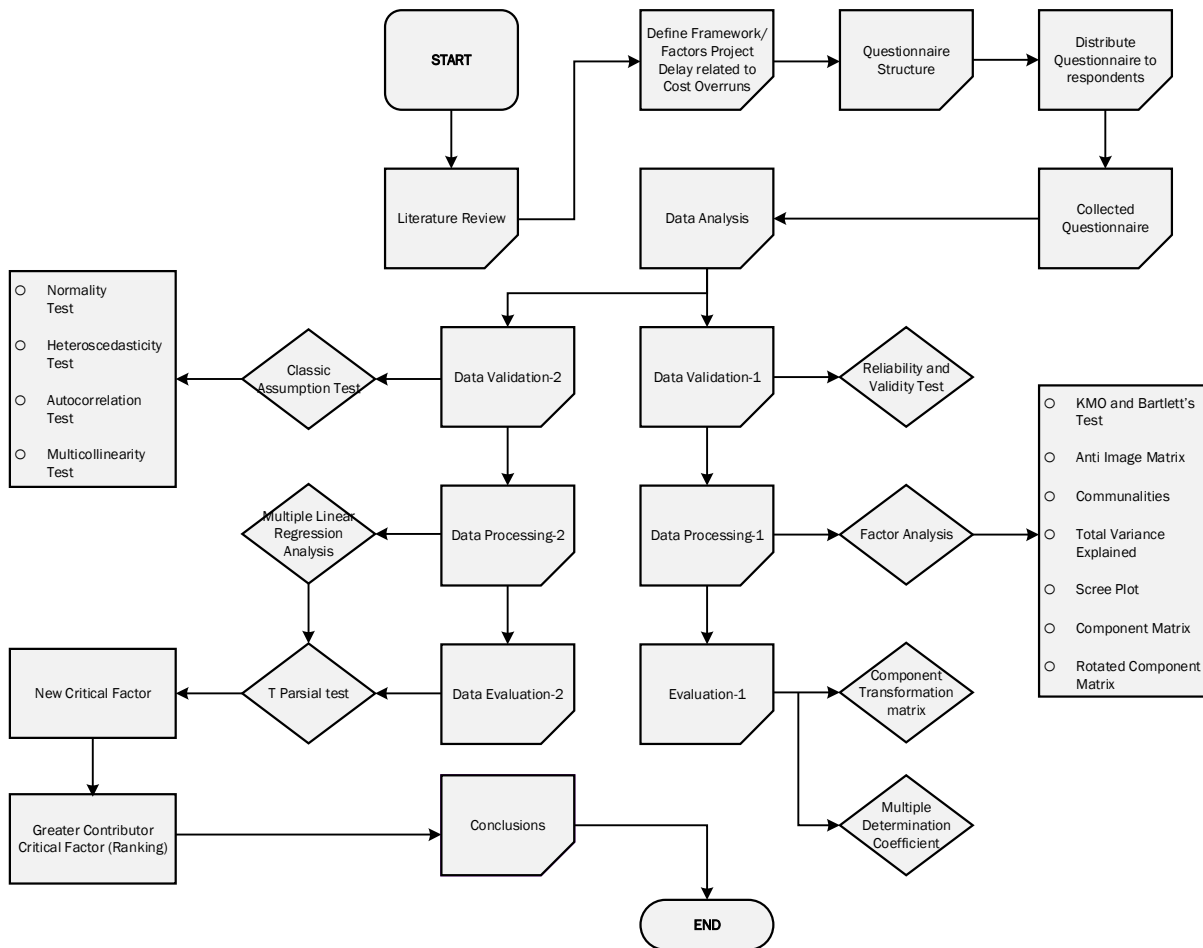


Figure 1. Research Framework

In this research, data was collected from 100 respondents from 35 selected Contractor Company in Jadebotabek area based on a structured questionnaire from 38 factors project delay related to cost overruns (Ullah *et al.*, 2017). Factor Analysis and Multiple Linear Regression Analysis are utilized in data analysis to identify and classify critical factors project delay and expound the effect of that critical factors to cost overruns.

3. Result and Discussion

Reliability and Validity Test used in data validation. N = 38 items question with Cronbach's Alpha value = 0.919, the result test Cronbach's Alpha for each 38 items question > 0.7, then 38 items of questionnaire question are reliable. N = 100 respondent, $\alpha = 5\%$, r table = 0.197, the result validity test for each 38 items question > 0.197, then each question in questionnaire this research is valid or feasible to be used as a research question.

Kaiser-Meyer-Olkin (KMO) and Bartlett's Test is to assure feasibility for each variable, it can be processed in factor analysis or not. The result KMO = 0.770 > 0.5 and Bartlett's Test = 0.000 < 0.5, then the Factor Analysis can be continued.

Anti-Image Matrix is used to determine which variable are feasible to use in Factor Analysis. The result Measure of Sampling Adequacy (MSA) for each 38 variables > 0.5 then all 38 variables can be used in Factor Analysis.

Communalities is used to determine the value for each 38 variables that used in this research able to explain the factors or not. The result value for each 38 variables > 0.5, then all 38 variables can be used to explain factors.

Total Variance Explained is used to extracted 38 variables. There are 10 components or factors as a result with eigen value for each component > 1. Scree Plot used to determine the number of factors formed. From the Scree Plot image in Figure 2, the result are 10 component points that have Eigenvalue > 1, then there are 10 factors that can be formed.

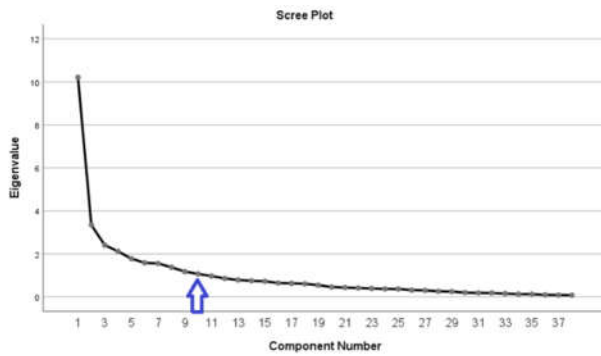


Figure 2. Scree Plot

Table 2. Summary Rotated Component Matrix

New Factor	Original Variable
1	X4, X7, X8, X10, X12, X13, X14, X17, X18, X19, X20, X37
2	X2, X5, X9, X22, X27
3	X25, X36, X38
4	X23, X24, X28, X29, X35
5	X1, X33
6	X31, X32
7	X26, X30, X34
8	X6, X11, X16
9	X3, X21
10	X15

Component Matrix and Rotated Component is used to determine correlation or relationship among each factor formed. The results summarized in Table 2.

Component Transformation Matrix is used to evaluate the result of extracted factor (10 new factors). The result are 10 new factors have correlation value > 0.5 then 10 new factors feasible to summarize 38 original variables that analyzed in this research.

Multiple Determination Coefficient done before proceeding classic assumption test and is used to determine from 10 new factors (Y1, Y2, Y3, ..., Y10), which one is the best factors project delay effected to cost overruns. The results are 12 independent variables from new factor 1 or Y1; X4, X7, X8, X10, X12, X13, X14, X17, X18, X19, X20, X37 are the best to explain, about 92.3% effect to cost overruns.

The purpose of Multiple Linear Regression Analysis in this research is to determine the magnitude of the effect of 12 independent variable X4, X7, X8, X10, X12, X13, X14, X17, X18, X19, X20, X37 to dependent variable; cost overruns (Y).

Table 3. Multiple Linear Regression Analysis Result

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.961 ^a	.923	.912	.29614540

a. Predictors: (Constant), X37, X4, X20, X10, X14, X18, X19, X13, X8, X12, X7, X17

b. Dependent Variable: Y1

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	91.370	12	7.614	86.818	.000 ^b
	Residual	7.630	87	.088		
	Total	99.000	99			

a. Dependent Variable: Y1

b. Predictors: (Constant), X37, X4, X20, X10, X14, X18, X19, X13, X8, X12, X7, X17

Table 4. Summary Multiple Linear Regression Analysis Result

Coefficients ^a						
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			
1	(Constant)	-2.560	.105		-24.313	.000
	X4	.104	.036	.106	2.855	.005
	X7	.102	.039	.118	2.610	.011
	X8	.038	.035	.042	1.069	.288
	X10	.072	.031	.082	2.327	.022
	X12	.219	.038	.259	5.845	.000
	X13	.099	.034	.120	2.923	.004
	X14	.158	.034	.173	4.596	.000
	X17	.235	.038	.282	6.104	.000
	X18	-.066	.034	-.073	-1.937	.056
	X19	-.022	.036	-.024	-.610	.543
	X20	.155	.029	.191	5.368	.000
	X37	.036	.031	.043	1.159	.250

a. Dependent Variable: Y1

Variable	Coefficient Regression	t	Sig.
Constant	-2.560		
X4	0.104	2.855	0.005
X7	0.102	2.610	0.011
X8	0.038	1.069	0.288
X10	0.072	2.327	0.022
X12	0.219	5.845	0.000
X13	0.099	2.923	0.004
X14	0.158	4.596	0.000
X17	0.235	6.104	0.000
X18	-0.066	-1.937	0.056
X19	-0.022	-0.610	0.543
X20	0.155	5.368	0.000
X37	0.036	1.159	0.250

F = 86.818 Sig. 0.000

R² = 0.923

Evaluation Multiple Linear Regression with T Partial test is utilized to determine 12 independent variables that has significant effect to dependent variable cost overruns (Y). The results of the T Partial test are the new critical factors in this research. The result is 4 independent variables; X8, X18, X19, and X37 that has the most significant independent variables and effect to dependent variable cost overruns (Y). Based on the result of Evaluation Multiple Linear Regression with T Partial test has results a regression equation model as described in equation (1).

$$Y = -2.560 + 0.038X8 - 0.066X18 - 0.022X19 + 0.036X37 \quad (1)$$

The Multiple Linear Regression equation (1) above can be described as follows:

1. Constant **-2.560** indicates that if there is cost overruns without no change or due to none critical factors project delay in 4 independent variables (X8, X18, X19, X37) then the cost overruns will be reduced at the value of 2.560 points.
2. Coefficient **0.038X8** indicates that there is positive correlation from Extension of time. Positive correlation corresponds that if there is cost overruns due to Critical Factors Project Delay Extension of time by 1 point then the cost overruns will be at the value of 0.038 points.
3. Coefficient **-0.066X18** indicates that there is positive correlation from Lack of cost plan/ monitoring. Positive correlation corresponds that if there is cost overruns due to Critical Factors Project Delay Lack of cost plan/ monitoring by 1 point then the cost overruns will reduce at the value of 0.066 points.
4. Coefficient **-0.022X19** indicates that there is positive correlation from Inadequate project preparations. Positive correlation corresponds that if there is cost overruns due to Critical Factors Project Delay Inadequate project preparations by 1 point then the cost overruns will reduce at the value of 0.022 points.
5. Coefficient **0.036X37** indicates that there is positive correlation from Equipment failure. Positive correlation corresponds that if there is cost overruns due to Critical Factors Project Delay Equipment failure by 1 point then the cost overruns will be at the value of 0.036 points.

Ranking has been done for 4 new critical factors project delay to find out which new critical factors project delay that have been greater contributor variable as predictor to cost overruns. According to (Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, 2014), if sig. value is smaller and t value is larger, then the predictor has the greater influence on its contribution. The result is the greatest contributor variable as predictor is Lack of cost plan and monitoring (X18). The second greater contributor variable as predictor is Equipment failure (X37), the third greater contributor variable as predictor is Extension of time (X8) and the fourth greater contributor variable is Inadequate project preparations (X19).

4. Conclusion

After the researchers conducted the analysis in the previous chapters then the researchers can give some conclusions as follows:

1. The critical factors project delay in Telecommunication Mobile Network as the result from Factor Analysis techniques are **Y1**; X4, X7, X8, X10, X12, X13, X14, X17, X18, X19, X20, X37, **Y2**; X2, X5, X9, X22, X27, **Y3**; X25, X36, X38, **Y4**; X23, X24, X28, X29, X35, **Y5**; X1, X33, **Y6**; X31, X32, **Y7**; X26, X30, X34, **Y8**; X6, X11, X16, **Y9**; X3, X21, **Y10**; X15.
2. The critical factors project delay in Telecommunication Mobile Network that effect to cost overruns as the result from Multiple Linear Regression analysis results obtained one equation model as described in equation (1).

$$Y = -2.560 + 0.038X8 - 0.066X18 - 0.022X19 + 0.036X37 \quad (1)$$

3. There is significant effect of 38 critical factors project delay in Telecommunication Mobile Network Projects, it can be seen from 10 new factor as the result of factor Analysis.
4. There are major critical factors that impact to cost overruns as the greater predictor contribution from equation (1), in sequence are lack of cost plan/ monitoring (X18) as first Critical Factor Project Delay effect to Cost Overruns, equipment failure (X37) as second Critical Factor Project Delay effect to Cost Overruns, extension of time (X8) as third Critical Factor Project Delay effect to Cost Overruns, and Inadequate project preparations (X19) as fourth Critical Factor Project Delay effect to Cost Overruns.

In the next study it is expected to conduct a similar survey with the wide scope, area, sample, population, limitation, feature tools or technology and heuristic knowledge. This research is reference for any Companies (Operators, Vendors, Contractors, Subcontractors, Suppliers, 3rd parties, etc.) in Telecommunications Industry, particularly that handling implementations 4G-LTE projects in order to avoid project delay and cost overruns, Professionals that works in Telecommunications Industry (Project Manager, Coordinator, etc.) and researchers.

Acknowledgement

The authors would like dedicate this research activity to my God, Allah SWT, and also my family. Acknowledge and appreciate would like to express my special thanks of gratitude to Swiss German University (SGU), Mr. Dr. Mulya R. Mashudi, ST, MEM, Mr. Dr. Maulahikmah Galinium S.Kom., Mr. Dr. Ir. Mohammad Achmad Amin Soetomo, M.Sc. and Mr. Dr. Eka Budiarto, S.T., M.Sc.

References

- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2014) *Multivariate data analysis*. 7th ed. UK: Pearson Educated Limited. Available at: www.pearsoned.co.uk.
- Hamzah, N. *et al.* (2011) 'Cause of construction delay - Theoretical framework', *Procedia Engineering*, 20(December), pp. 490–495. doi: 10.1016/j.proeng.2011.11.192.
- PMI (ed.) (2017) *A guide to the project management body of knowledge - PMBoK*. 6th edn. USA: Project Management Institute; Sixth Edition, Sixth edition edition (September 22, 2017).
- PMI (2018) 'Success in Disruptive Times'. Available at: <https://www.pmi.org/-/media/pmi/documents/public/pdf/learning/thought-leadership/pulse/pulse-of-the-profession-2018.pdf>.
- Telkom (2017) *Accelerating The Indonesian Digital Economy*. Available at: <https://www.telkom.co.id/servlet/Satellite>.
- The Standish Group (2017) *Chaos Report, Standish Group - Chaos Report*. doi: 10.1145/1145287.1145301.
- Ullah, K. *et al.* (2017) 'Theoretical framework of the causes of construction time and cost overruns', *IOP Conference Series: Materials Science and Engineering*, 271(1). doi: 10.1088/1757-899X/271/1/012032.

Improving Performance Loan Fraud Model Prediction Using Mean Decrease Accuracy and Mean Decrease Gini

Arsandi Akhmad¹, Lukas¹ and Bagus Mahawan¹

¹ Department of Information Technology, Swiss German University, Tangerang 15143, Indonesia

Abstract. The purpose of this research is to develop a fraud detection model on loan transactions at failed banks in the context of deposit and deposit guarantees mandated to the Indonesia Deposit Insurance Corporation (IDIC). The data used in this study is the data of a bank in the Jakarta area that had liquidated at the end of 2015. Meanwhile, data on loan transaction ranges ranged from 2010 to 2015. This research also focuses on improving the performance of detection models by using feature selection. With the feature selection, it is expected that the impact of the reduced performance of the model exposed to high variance and high bias due to the many features used can be handled better.

1. Introduction

1.1. Background

Banks are business entities that collect funds from the public in the form of deposits and channel them to the public in the form of loans and/ or other forms in order to improve the lives of many people (Indonesia, 1998). Banking provides various types of alternative loan funds for customers, one of which is through providing loans in the form of loans to customers. Indeed, in granting credit to customers, the bank experiences various problems or risks. One of the problems or risks experienced by banks in granting credit is the behaviour of customers who do not pay instalments on time or delay up to several months in instalment payments which ultimately causes bad credit rating. This is a serious problem that needs to be considered by the bank because the regulator determines the maximum percentage of Non-Performing Loans (NPL) (Kustina, 2018). If the level is above the set limit, the bank is declared as special handling status.

Therefore, some banks in special handling status and experience normal liquidity difficulties commit fraudulent financial records to keep NPL levels low. One method that is commonly used is to do window dressing transactions that are usually if related to credit; there will be transactions that are not supposed to be like the existence of credit instalment payment transactions without accompanied by the actual flow of funds. When the bank's condition gets worse so it must be liquidated, the fraudulent action causes the under-stated loan balance, and credit quality is not a real condition. The impact is that in the process of guaranteeing customers' funds, it will lead to wrong decision making regarding guaranteed customer funds.

To be more careful in determining customer funding guarantees, there is a need for a calculation method with an intelligent computing system to assist the Indonesia Deposit Insurance Corporation (IDIC) in selecting which customers are eligible to pay. The methods used can be Regression Analysis, Neural Networks, Genetic Algorithms, Decision Tree, and Random Forest.

1.2. Random Forest

The Random Forest was first officially published by Leo Breiman in 2001 (Breiman, 2001). The Random Forest was developed to improve Decision Tree methods that are prone to overfitting. In its development, Random Forest has become one of the popular methods in the field of machine learning. This is due to the ease of use of Random Forest, which is capable of achieving high accuracy without the need to do many tuning parameters.

The Random Forest learning process uses the CART (Classification and Regression Tree) method. CART is a non-parametric method that is useful for getting a group of accurate data as a characteristic of a classification. The CART classification method consists of two methods, namely the regression tree method and the classification tree. If the dependent variable has a specific type, then CART produces

classification trees. Whereas if the dependent variable is of a continuous or numerical type, CART produces regression trees (Breiman, 1996).

1.2. Related Work

Detection of fraudulent financial statement can be done in various ways, and many studies have been conducted related to this. The most commonly used method is to use the concept of the Fraud Triangle when analysing documents in the audit process that is done in a worksheet and looking at the pattern of fraud (Huda et al., 2014). The study of transactional fraud especially related to financing is usually carried out by examining entities or variables from credit data such as jobs, homeownership, collateral, number of dependents and others (Xu et al., 2016). Also, a study of credit financing fraud can be done using detailed transaction data such as fraud on credit cards. Choi et al. (Choi et al., 2013) In their study developed an individual level utility concept which would later be used to evaluate the model. The study conducted by Johannsson et al. (Jóhannsson, 2017) uses transaction behaviour profiling and performs time analysis of transaction data and makes variables used in the detection model.

The problem that is often faced in building fraud detection models is the number of features used in the dataset so that the prediction model becomes biased, and the model's performance decrease. Several studies have been done to reduce bias and improve performance in detection models such as Hong Han et al. (Hong Han et al., 2016) makes a study to reduce the error rate by making feature selection using the Mean Decrease Gini and Mean Decreasing Accuracy where iterates the feature reduction with that method. Hasan et al. (Hasan, 2017) also made a study of how to improve accuracy by making feature selection using Forward / Stepwise Regression implemented in prediction models based on SVM. Improving performance model also can be done by calibrating the probability values used in the prediction model. Alejandro Bahnsen (Bahnsen et al., 2014) study about calibrates probability values that will be used in the ROC curve to improve the performance of the model. Louppe et al. (Louppe et al., 2013) conducted a study of the dataset feature by calculating the mean deviation impurity (MDI) score that can be used to select features.

2. Materials and Methods

The data to be used in this study are from the XYZ Rural Bank, located in Tangerang. The number of datasets to be used in this study is 6448 records. Loan data used as a basis for predicting loan fraud is obtained when Bank XYZ is liquidated by IDIC and also records of which credit fraud perpetrators are banned without the flow of funds. Variables used in predicting credit risk in this study consist of 12 input variables and one target variable, namely:

Table 1. List of features in the dataset and its description

Feature	Description
Sb	Loan Interest rate
Plafond	Loan Amount
PLF	Features created with the criteria if the new loan is realised in the range of 30 days before up to 30 days after the previous credit is repaid. If yes value = 1 and if no value = 0
PLF1	The Comparison between the new amount of loan and the previous amount of loan. If new loan higher than the previous loan, the value will be set as 1 and if not, the value will be set as 0
PLF2	the Features created with criteria if another loan was realised less than one year from the previous realisation loan.
PLF3	Feature created as a flag if repaid date greater than 60 days from its schedule
PLF4	closing date in month/ divide total month in the schedule
Col2	sum transaction that days between two transactions is between 90 until 180 days
Col3	sum transaction that days between two transactions is between 180 until 360 days
Col4	sum transaction that days between two transactions are greater 360 days

Feature	Description
Rata-rata.Jarak.Angs	averages of days between two instalments transaction
Tunggakan	is any arrears in instalment transaction
dialling	instalment payments are bailed out (fraud flag)

Framework in this study follow CRISP-DM (Chongwatpol et al., 2012) that contains the life cycle of six-phase data mining activities, but in this study follow until the fifth phase. One of the advantages of this model is that the phase stages of the model are not a rigid stage. Moving forward and backwards between each different phase can always be done. This is by the nature of data mining, where the data mining process is not completed when a result is found because the process of data mining is a continuous learning process continuously. The experimental design model can be explained in several stages:

1. Collection of credit customer datasets where customer credit datasets was collected at XYZ bank, which amounts to 6448 records.
2. Distribution of the dataset will be divided into two parts, namely, data transfer and testing data. Data sharing will be carried out stratified.
3. Applying the Random Forest Algorithm after data sharing, training data will be processed to obtain the prediction model.
4. Evaluate prediction models using data testing by looking for measurement values of models such as Accuracy, Kappa, Specificity and Sensitivity.
5. Using Mean Decrease Accuracy - Mean Decrease Gini as a Featured method Selection and chose the top ten scores then create a new dataset.
6. It is creating a Random Forest model based on data that has been reduced by the variable and validating it using the testing data.
7. Compare the performance of the non-feature selection and feature selection models.

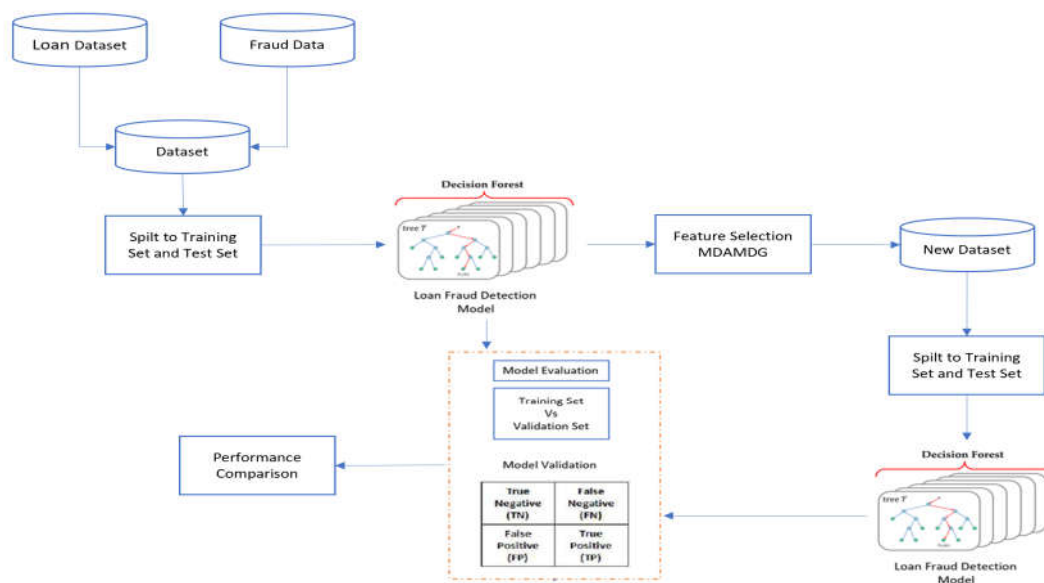


Figure 1. Research Framework

3. Results and Discussion

The test and the experiment were conducted using R 3.6.1 for Windows OS version. Test and experiment where conducted base upon:

- a. Bank Loan dataset without using feature selection
- b. Tuning the model and analyse variable using Mean Decrease Accuracy and Mean Decrease Gini as Feature Selection than create a new dataset.

c. Bank Loan dataset with feature selection.

Each dataset was applied to Random Forest Algorithm to create a classification model. To get the minimum out-of-bag error, tuning classification needed for every model.

3.1. Feature Selection.

To obtain the calculation score of Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG), the Random Forest package can use the importance function and visualise it can use the varImPlot function. The value of Mean Decrease Accuracy and Mean Decrease Gini scores are shown in Table 2 and already ordered from the largest and smallest. From Table 2 can be seen features sb, plafond, PLF2 dan Rata-rata.Jarak.Angs as the most important feature. In Random Forest, Gini impurity affects the impurity of the data so that if the feature with the MDG value omitted, the data will be increasingly difficult to classify. Likewise, with the MDA value, which if the feature with the highest value omitted, the value of accuracy will be decreased. For better performance model, feature with the highest score in the dataset is selected by using Mean Decrease Accuracy and Mean Decrease Gini. While for the MDA and MDG values that are small, the feature is not too influential or even causes a bias in the model.

Table 2. Mean Decrease Accuracy and Mean Decrease Gini Score

Variable	MeanDecreaseAccuracy	MeanDecreaseGini
sb	164.21330	961.47057
plafond	87.13234	366.61910
PLF	11.42071	15.00216
PLF1	18.22853	22.04726
PLF2	84.62336	277.34986
PLF3	16.80891	18.38644
PLF4	59.63332	232.29434
Col2	37.97315	141.87118
Col3	25.24739	39.97915
Col4	28.34939	77.40978
Rata-rata.Jarak.Angs	73.80904	297.08782

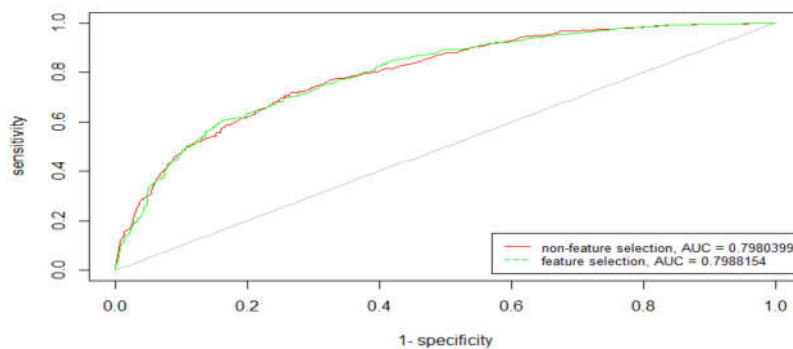
3.2. Performance Metrics

This study uses the statistical measurement for performance model test that conducted. The statistical measure includes accuracy, Kappa, sensitivity, specificity, true-positive rate, false-positive rate and P-value. In Tables 2 and 3, there is a comparison of the value of performance statistics between non-feature selection and feature selection. If we look further, there is an increase in the value of accuracy from 0.7395 to 0.7599. Also there is an increase in the value of Cohen's Kappa from 0.4168 to 0.4348, which indicates an increase in the degree of agreement. The sensitivity (True Positive Rate) value increases from 0.8006 to 0.8658, which indicates that the detection rate for fraud class increases even though the Specific Negative Rate decreases from 0.6178 to 0.5492.

Table 3. Performance measure comparison between non-feature selection and feature selection

Measurement	Before Feature Selection	After Feature Selection
Accuracy	0.7395	0.7599
95% CI	(0.7173, 0.7607)	(0.7383, 0.7806)
P-Value [Acc > NIR]	8.47E-11	< 2.2e-16
Kappa	0.4168	0.4348
McNemar's Test P-Value	0.7327	6.31E-07
Sensitivity	0.8006	0.8658
Specificity	0.6178	0.5492
Pos Pred Value	0.8066	0.7927
Neg Pred Value	0.6088	0.6727
Prevalence	0.6656	0.6656
Detection Rate	0.5329	0.5763
Detection Prevalence	0.6607	0.7270
Balanced Accuracy	0.7092	0.7075

The performance of the prediction model can also be measured using the ROC curve and the AUC value. In Figure 2, it can be seen that the AUC value of the non-feature selection model is 0.7980399, and the model with the selected feature is 0.7988154. There is an increase in value even though it is not too significant so that the ROC curve between the two models is not too different in shape.

**Figure 2. ROC curve non-feature selection and feature selection model**

4. Conclusion

In this paper, a fraud prediction study in credit instalment payment transactions can be done using the Random Forest algorithm and from the results of experiments and tests that predict the model can predict credit fraud is good enough it can be seen from the accuracy value of 0.7395 and some statistical values of model performance measurement. In addition to improving performance, a feature selection is performed by using Mean Decrease Accuracy and Mean Decrease Gini where the results of the prediction model by reducing the less important features have an impact on increasing the accuracy value to 0.7599 and several other performance measurement values. We hope that further research can improve the performance of the detection model even better.

References

Bahnsen, A.C., Stojanovic, A., Aouada, D., Ottersten, B., 2014. Improving Credit Card Fraud Detection with Calibrated Probabilities, in: Proceedings of the 2014 SIAM International Conference on Data Mining. Presented at the Proceedings of the 2014 SIAM International Conference on Data

- Mining, Society for Industrial and Applied Mathematics, pp. 677–685.
<https://doi.org/10.1137/1.9781611973440.78>
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140.
- Choi, K., Kim, G., Suh, Y., 2013. Classification model for detecting and managing credit loan fraud based on individual-level utility concept. *ACM SIGMIS Database* 44, 49.
<https://doi.org/10.1145/2516955.2516959>
- Chongwatpol, J., Sa-ngasoongsong, A., Woratanarat, P., Sa-ngasoongsong, P., 2012. Prognostic Analysis of Hip Fracture in Elderly Women with Data Mining Methods 11.
- Hasan, M., 2017. PREDIKSI TINGKAT KELANCARAN PEMBAYARAN KREDIT BANK MENGGUNAKAN ALGORITMA NAÏVE BAYES BERBASIS FORWARD SELECTION. *Ilk. J. Ilm.* 9, 317. <https://doi.org/10.33096/ilkom.v9i3.163.317-324>
- Hong Han, Xiaoling Guo, Hua Yu, 2016. Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest, in: 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS). Presented at the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), IEEE, Beijing, China, pp. 219–224. <https://doi.org/10.1109/ICSESS.2016.7883053>
- Huda, S., Ahmad, T., Sarno, R., Santoso, H.A., 2014. Identification of process-based fraud patterns in credit application, in: 2014 2nd International Conference on Information and Communication Technology (ICoICT). Presented at the 2014 2nd International Conference on Information and Communication Technology (ICoICT), IEEE, Bandung, Indonesia, pp. 84–89.
<https://doi.org/10.1109/ICoICT.2014.6914045>
- Indonesia, P.R., 1998. Undang-undang Republik Indonesia nomor 10 tahun 1998 tentang perubahan atas Undang-undang nomor 7 tahun 1992 tentang perbankan. BP. Cipta Jaya.
- Jóhannsson, J., 2017. Detecting fraudulent users using behaviour analysis 66.
- Kustina, K.T., 2018. MSMEs Credit Distribution and Non-Performing Loan towards Banking Companies Profit in Indonesia. *Int. J. Soc. Sci. Humanit. IJSSH.*
<https://doi.org/10.29332/ijssh.v2n1.72>
- Loupe, G., Wehenkel, L., Sutera, A., Geurts, P., 2013. Understanding variable importances in forests of randomized trees 9.
- Xu, J., Chen, D., Chau, M., 2016. Identifying features for detecting fraudulent loan requests on P2P platforms, in: 2016 IEEE Conference on Intelligence and Security Informatics (ISI). Presented at the 2016 IEEE Conference on Intelligence and Security Informatics (ISI), IEEE, Tucson, AZ, USA, pp. 79–84. <https://doi.org/10.1109/ISI.2016.7745447>

Measurement of IT Risk Management Maturity Level in CEC Using IT Domain Risk Governance Framework

Annas Iswahyudi^{1*}

¹Graduate Student of Master Information Technology Program, Swiss German University, Tangerang 15143, Indonesia

*Corresponding author: annasipteknet@gmail.com

Abstract. IT Risk Management has long been adopted and implemented in CEC. This is inseparable from the high need for reliable and trusted information technology services at CEC as a government institution that has primary task for eradicating corruption. With a good IT risk management is expected to reduce the impact if the IT risk occurs and impacted to overall business process in CEC. However, up to 15 years after the implementation of IT risk management has never been measured how the level of IT maturity risk management. In this research, Author will use the IT Risk Framework with the risk governance domain approach as a standard IT risk management framework to evaluate the implementation of IT risk management in CEC. The process of evaluating the level of IT maturity is based on the maturity model that has been defined in the IT risk framework.

Keyword: IT Risk Management, IT maturity, risk governance domain

1. Introduction

CEC was established in 2002 to address against corruption in Indonesia. Become super body Institution, CEC has main duty to eradicate corruption as the extraordinary crime in Indonesia CEC was established to revitalize national anti-corruption efforts. CEC has adopted ISO / IEC, ISO / DIS 31000, Risk Management Standard since 2008 in the IT Department, which is reviewed annually by independent auditors (CEC Enterprise Architecture Review, 2014). In carrying out its role information technology risk management has been works very well because it is able to provide added value in order to achieve organizational goals, possible risks to technology information that can cause failure in running the information system function so that it can causing the impact of loss and reputation risk for the organization. IT Risk Framework provides a framework comprehensive to control and manage business based information Technology. Risk IT provides a framework to assist organizations in identifying, determine, and manage information technology risks. Therefore an analysis of risk management in CEC uses the risk domain IT Risk domain framework Governance.

2. Materials and Methods

2.1 Literature Study

2.1.1 Risk IT Framework

The IT Risk framework is used to help implementing information technology governance, and company which COBIT has adopted as a governance framework information technology used by Risk IT for improve risk management (Kulkarni, n.d.). Processes must be combined between internal interests and external organization. Internal matters include incidents in IT operations, failures in projects, and the replacement of an IT strategy. External things itself can include changes in circumstances that exist in the market, the existence of new technology and cause regulation on IT. IT risk itself can be said is a business risk where business risks cover in users, owners, ways operate, involvement, influence and adoption of IT in the organization (Alex Pasquini, 2013). The process model in the IT Risk framework has three the domain of Risk Evaluation (RE), Risk Governance (RG) and Risk Response (RR) as described in **Figure 1**

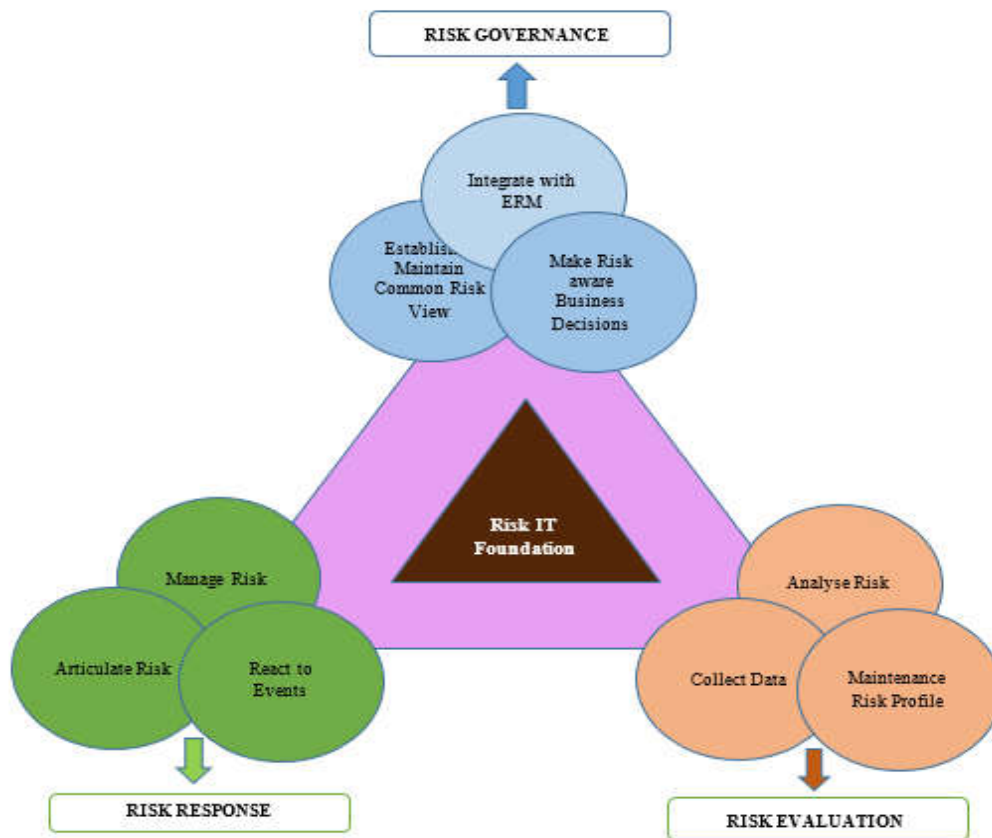


Figure 1. IT Risk Framework (Information Systems Audit and Control Association, 2009)

2.1.2 Risk Governance

At this stage, management practices must be ensured IT risks have been conveyed within the company, for allows for optimal risk adjustment. Risk Governance consists of three processes namely:

1. RG1 Establish and maintain a common risk view
2. RG2 Integrate with ERM
3. RG3 Make risk-aware business decisions

a. RG1 Establish and maintain a common risk view

Ensuring that risk management activities are aligned with the capacity of the company's goals relating to IT losses and leadership has a subjective tolerance for it. Following are the key activities of RG1:

- RG1.1 Perform enterprise IT risk assessment
- RG1.2 Propose IT risk tolerance thresholds
- RG1.3 Approve IT risk tolerance
- RG1.4 Align IT risk policy
- RG1.5 Promote IT risk-aware culture
- RG1.6 Encourage effective communication of IT risk

b. RG2 Integrate with ERM (enterprise risk management)

Integrate IT and operations risk strategies with business strategy risk decisions that have been made. Following are key activities RG2:

- RG2.1 Establish and maintain accountability for IT risk management
- RG2.2 Co-ordinate IT risk strategy and business risk strategy
- RG2.3 Adapt IT risk practices to enterprise risk practice

- RG2.4 Provide adequate resources for IT risk management
- RG2.5 Provide independent assurance over IT risk management

c. RG3 Make risk-aware business decisions

Ensuring that decision making by companies based on opportunities and consequences. Following are the key activities of RG3:

- RG3.1 Gain management buy-in for the IT risk analysis approach
- RG3.2 Approve IT risk analysis
- RG3.3 Embed IT risk considerations in strategic business decision making
- RG3.4 Accept IT risk
- RG3.5 Priorities IT risk response activities

2.1.3 The Risk Maturity Model

COBIT 5 is also designed to be a tool that can solve problems in IT governance in understanding and managing risks and the benefits associated with corporate information resources. In addition, COBIT 5 is also designed to be a tool that can solve problems in IT governance in understanding and managing risks and the benefits associated with corporate information resources. Therefore, a maturity model method is needed to measure the level of process management development, the extent of the management capability. How well development or management capabilities depend on achieving COBIT goals 5 (Arief and Wahab, 2016).

In this way, the maturity models are designed to enable management to focus on key areas needing attention, rather than on trying to get all processes stabilized at one level before moving to the next. The maturity model scales can help management understand where weaknesses exist and set targets for where they need to be (Behara and Palli, 2013). The most suitable maturity level for an enterprise will be influenced by the enterprise’s business objectives, the operating environment and industry practices (Nurpulaela, 2016). Specifically, the level of IT risk management maturity will depend on the enterprise’s dependence on IT, its technological sophistication and, most important, the future role its executives and management foresee for information technology (Pasquini and Galiè, 2013). To create the results easily usable in management meetings—where they should be presented as a means to support the case for future plans to improve risk governance, evaluation and response, a graphical presentation model might need provided as follows in **Figure 2**:

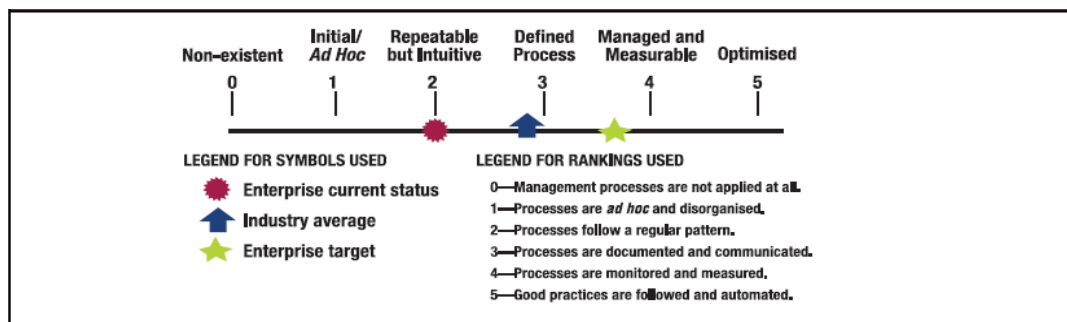


Figure 2. Maturity Model (ISACA, 2012)

There are five levels of application of risk management that can be defined in a series of models according to ISACA as shown in **Figure 2** above, including: 1. Initial: is the starting point for the use of a new or undocumented repetition process; 2. Repeatable: the process is at least adequately documented so that repeated attempts at the same steps can be carried out; 3. Defined: a process defined / confirmed as a standard business process; 4. Managed and Measurable: this process is managed quantitatively according to agreed metrics; 5. Optimized: process management includes intentional optimization / improvement of processes.

2.2 Methods

Analysis of the maturity level of IT risk management governance in CEC by collecting data through:

1. Observation

Observations aims to understand the scope of the implementation of IT Risk management in related business processes with the CEC IT department.

2. Interview

In this study, the objects and materials research are employees with 4 type of employee levels: a. middle management level is IT director or at same level, b. IT governance division head, c. IT risk officer and Staff. With various background level through the composite of different sources to get whole point of view, range of respondent are decided because various range consideration determined accuracy of result

3. Document review

Documents collected are documents that related to IT risk management activities, i.e. IT risk management organizational structure, duties & authorities of the risk management division, IT Risk Register, Risk Profile and Appetite, IT risk policies and controls, guidelines the application of IT risk management. The documents used are limited to the last 3-year version (2014, 2015, 2016) of the document officially designated as a reference document.

4. Develop maturity assessment tool

The rating scale used is 1 to 5. Results from these 6 variables are averaged to obtain the final score. The image above is an example of the tools developed in this research. After observation, interviews and document review, the process of measuring the maturity level of Risk Governance is carried out using a tool to determine the scores for each of the key activities. To assess key activities in the Risk Governance domain, a tool is used to conduct scoring using 6 (six) parameter as follows:

1. Awareness & Communication: is the level of concern of all stakeholders about IT risks and how to communicate in treating these risks
2. Responsibility & Accountability: is the adequacy of the division of tasks, responsibility and audit of each risk charged to each PIC (person in charge) that has been assigned
3. Goal Setting & Measurement: Determination of the final destination and how to measure each risk control that has been set
4. Policies, Standards & Procedures: the adequacy of policies, implementation standards and procedures for IT risks that have been determined
5. Skills & Expertise: quality of human resource management and risk management
6. Tools & Automation: Software, Hardware and other devices used to control IT risks

See detail maturity assessment tool as shown in **Table 1** below:

Table 1. Maturity assessment tool

No.	Key Activities				Maturity Rank Model			Final Score
					Variabel 1	Variable 2	V3...n	
	Process Goals :							AVG(Score 1...n)
1	RG1.1	Perform enterprise IT risk assessment						Score 1
2	RG1.2	Propose IT risk tolerance tresholds					
3	RG1.3...n						
4	RG2.1...n						
5	RG3.1...n							Score n

3. Results and Discussion

3.1 Resume of Result

Detailed scores for all key activities as shown in the following **Table 2**:

Table 2. Results score key activities

RG1	3.21
RG1.1	3.72
RG1.2	3.75
RG1.3	3.83
RG1.4	2.88
RG1.5	2.86
RG1.6	2.25
RG2	2.99
RG2.1	3.27
RG2.2	3.00
RG2.3	3.17
RG2.4	2.50
RG2.5	3.00
RG3	3.00
RG3.1	3.33
RG3.2	2.80
RG3.3	2.80
RG3.4	3.17
RG3.5	2.83
Average Risk Governance Score	3.06

The results in **Table 2** can be explained as follows.

a. RG1 Establish and maintain a common risk view

In this process the CEC already has risk management activities where there has been a workshop on existing IT risk assessment but it has not been followed by all areas in the CEC has also made risk tolerance and policies for IT risk at the CEC have also conducted training for related business units to raise awareness about risk, but for the IT risk discussion activity itself orally is explained to be done if the risk occurs at that time, there is no special planning carried out every periodically to discuss the risk. Existing program activities are still only followed by manager-level positions, not all parties related to the business.

The level of maturity of the RG1 process is Level 3 **Defined Process**, because there is already organizational awareness in discussing and communicating IT risks in the company but the risk tolerance discussed is still only based on technological developments, needs, and skills needed in the company today and there is no regular planning for communication activities that discuss IT risks at CEC.

b. RG2 Integrate with ERM (enterprise risk management)

In this process the CEC has specified responsibility for existing IT risk management in the organization and has considered the effect of IT risk on existing business strategies and has used methods to deal with existing risks using ISO27001, and has a monitoring website that monitors activities on the business, but the related business units do not have risk measurement documents that should be reported to those who handle risk management, namely the risk management division and good corporate governance, because at the time of the incident the related business unit can sometimes solve existing problems. For problems or events that occur also verbally explained is the responsibility of the parties concerned.

The level of maturity of the RG2 process is Level 3 **Defined Process**, because there is already a section that handles IT risk in the organization and the organization's risk management committee that provides risk management guidelines and resources to deal with IT risk but IT risk is still focused on existing risk issues the organization and the IT risk department are not yet fully engaged with the risk management committee in the Organization.

c. RG3 Make risk-aware business decisions

In this process the CEC has conducted trainings on the importance of IT risk analysis but has not been fully attended by all existing leaders and staff, the existing leader assigns tasks to the IT security and quality assurance department at the IT Directorate to consider risk activities. The maturity level of the RG3 process is Level 3 **Defined Process**, because the CEC has considered the effects of IT risks and determined the actions that must be taken in addressing IT risks but for discussion in conducting risk analysis it is still left to the IT department in the organization and consideration of existing risks still based on existing risk issues and only those that occur most frequently in the organization.

3.2 Gap Analysis

Table 3. Standard Deviation Risk Governance Domain

No.	Risk Governance Domain	Standard Deviation
1.	RG1. Establish and maintain a common risk view	0.591
2.	RG2. Integrate with ERM	0.265
3.	RG3. Make risk-aware business decisions	0.221

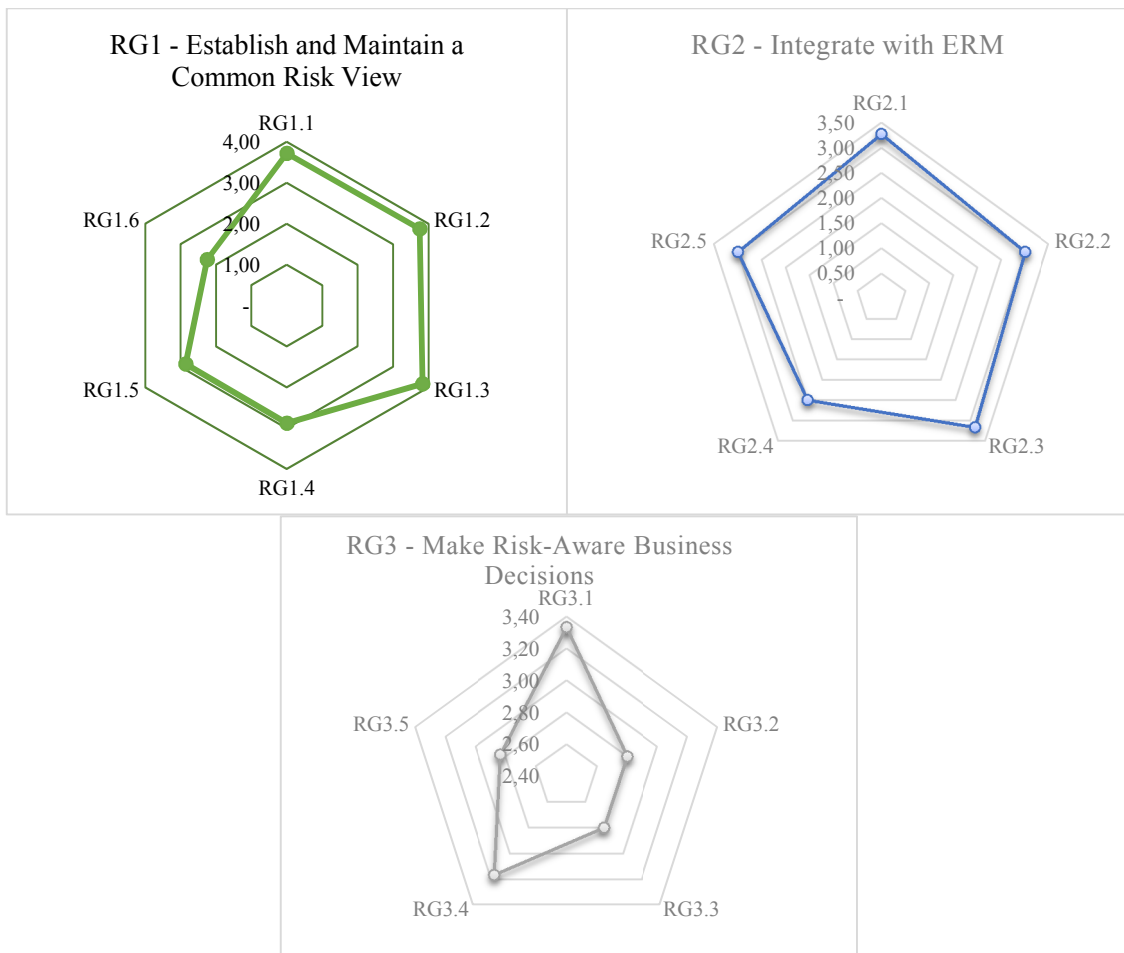


Figure 3. Curve radar comparison RG1, RG2, RG3

Using standard deviation calculations (see **Table 3**) and curve radar comparisons (see **Figure 3**) it can be seen the implementation gap between key activities in the domain of risk governance. From the table xx and xx figures it can be seen that the RG1 domain has the largest standard deviation of 0.591, while the RG3 domain has the smallest standard deviation of 0.221. This indicates that the key activities in RG1 do not yet have an even distribution of maturity values or have a large gap between the high measurement value and the lowest value. Conversely, in RG3 the distribution of key activity

measurement values has a low gap. The lowest value in RG1 is key activities RG1.6 Encourage effective communication of IT risk with a score of 2.25. While the highest score is RG1.3 Approve IT risk tolerance with a score of 23.83.

4. Conclusion

From results and analysis above we conclude that:

1. Based on the results of measurements on IT risk management in CEC using the IT Risk Framework, especially for the domain of Risk Governance, the answers are obtained from the problem formulation that the maturity level of the RG1 Establish and Maintain a Common Risk View, RG2 Integrate With ERM and RG3 Make Risk-aware Business Decisions are level **3 Defined Process**. Level 3 is still in line with the average level in industry or organization best practices that adopt IT risk management.
2. All CEC stakeholders, especially leaders and officials in the IT department need to improve the management process of several key activities that still exist in level 2 and increase even higher levels that already exist in 3 so that in the future it is expected to be at the managed and measurable level (level 4) even if possible achieved at Optimized (level 5). This is necessary considering that CEC is an institution that has a very important task in this country and has a good reputation so IT risk management must be as much as possible.

References

- Alex Pasquini, 2013. COBIT 5 and the Process Capability Model. Improvements Provided for IT Governance Process.
- Arief, A., Wahab, I.H.A., 2016. Information technology audit for management evaluation using COBIT and IT security (Case study on Dishubkominfo of North Maluku Provincial Government, Indonesia), in: Information Technology, Computer, and Electrical Engineering (ICITACEE), 2016 3rd International Conference On. IEEE, pp. 388–392.
- Behara, G.K., Palli, P., 2013. Maturity Assessment for Enterprise Architecture. CEC Enterprise Architecture Review 2014-2016.
- Information Systems Audit and Control Association, 2009. The risk IT framework. ISACA, Rolling Meadows, IL.
- ISACA, 2012. COBIT5-Implementation.
- Kulkarni, G., n.d. Applying the Goals Cascade to the COBIT 5 Principle Meeting Stakeholder Needs 10.
- Nurpulaela, L., 2016. Evaluation of IT governance to support IT operation excellent based on COBIT 4.1 at the PT Timah Tbk, in: Information Technology, Computer, and Electrical Engineering (ICITACEE), 2016 3rd International Conference On. IEEE, pp. 336–339.
- Pasquini, A., Galiè, E., 2013. COBIT 5 and the Process Capability Model. Improvements Provided for IT Governance Process. Proc. FIKUSZ 13, 67–76.

Monitoring & Evaluation Model Framework for IT Project Management at PT. XYZ

Fattah Hadi Saputro¹, Mohammad Amin Soetomo^{1*} and Nuki Agya Utama¹

¹ Department of Information Technology, Swiss German University, Tangerang 15143, Indonesia

* Corresponding author, email: mohammad.soetomo@gmail.com

Abstract. There are projects implementation that have not been run accordance with the objective of desired project. Evaluation Model Framework in project management is very important to get the desire goals such as quality, cost and delivery. The main requirement to start the project implementation is a business case study where investment must be validated with related objectives of the project carried out. Investment in a company is not easy to implement, especially in the planning phase of project management. Nowadays technology is growing rapidly, investment in a company continues growing every year along with the company's business needs to be competitive with appropriate investment. This study aims to propose a new design IT framework on project management, starting alignment business strategy, IT goals, business requirements and business case by validating functional specification based on criteria and the project that have passed business cases meet the expectations of stakeholder needs.

1. Introduction

PT. XYZ has an IT issues in business side that need to be addressed. In previous planning, business users submitted initiatives to the IT Division and directly carried out the execution of the submission for those needs, it causes uncontrolled demand from each business unit and also one of the weaknesses in planning phase before implementing an IT project. For the next 5 years, there are approximately 50 IT Projects/Initiative, with annual budget 2MUSD of investment. For solving the problem, this research will focus to make sure planning phase before IT Project implementation that explain business issues and needs. Annual Activity Planning includes activities and steps of work starting from the preparation business case, allocation annual IT budgets, preparation IT portfolio program. If there is an ad hoc initiative, it will continue to ad hoc activities preparation of annual IT activities.

Every project has a defined scope (performance), schedule (time), and budget (cost). These three project parameters are referred to as the triple constraints of the project. These triple constraints are often illustrated by an equilateral triangle (also known as the Iron Triangle), as shown in Figure 1. The key characteristics of this triangle is that a change in one of the three constraints will affect at least one other constraint. A project manager must balance these constraints as well as the project quality for the success of the project. This balancing act often involves negotiations between the project manager and the project sponsor or owner customer (project owner) (Singh, 2016). Costs might include the costs to design and develop and/or maintain the project or project management improvement initiative, cost of resources, cost of travel and expenses, cost to train, overhead costs, etc (PMI, 2010).

Scope defines the work that must be done to complete the project. Schedule represents the duration of the project. Budget represents the estimated cost of completing the project. Quality stands for how well a product or service meets the pre-defined specifications or requirements and how satisfied the customer is.

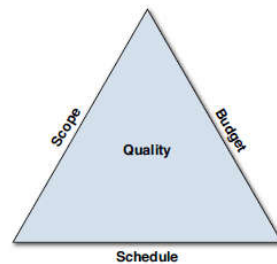


Figure 1. Triple Constraints of a project (Singh, 2016)

2. Materials and Methods

David Norton and Robert S Kaplan using the Balanced Scorecard as a strategic management system, how the company link measures from the four perspective (Kaplan & David, 2007), Balance Scorecard is a strategy performance management tool – a semi-standard structured report, that can be used by managers to keep track of the execution of activities by the staff within their control and to monitor the consequences arising from these actions Aligning Vision, Mission, Destination & Strategy Top to Bottom Approach. Combination of IT strategic alignment with IT governance could be an effective strategic innovation allowing organizations make efficient use of IT (Hosseinbeig, Moghadam, Vahdat, & Moghadam, 2011).

COBIT is a framework that has been developed by IT Governance Institution that part of the Information System Audit and Control Association (Copy & Indrajit, n.d.) Organization are expected to meet quality, compliance and security requirements for their governing and managing enterprise information technology. The major drivers for the development of COBIT include the need to provide more stakeholders a say in determining what they expect from information and related technology (what benefits at what acceptable level of risk and at what costs) and what their priorities are in ensuring that expected value is actually being delivered. Combination of IT strategic alignment with IT governance could be an effective strategic innovation allowing organizations make efficient use of IT (Hosseinbeig, Moghadam, Vahdat, & Moghadam, 2011).

In the previous study, we used some literature from various sources to do alignment within business strategy and IT strategy to get the relationship until the program or initiative used as a project. The success of the project can include additional criteria related to organizational strategy and business delivery. How to align project management practices with the leading IT governance framework, it will be impossible to map directly to the 42 project management processes, 47 program management process, and 14 portfolio management processes into the 37 COBIT processes on the reference model (Rincon, I. (2012)). The right alignment between project management and IT governance provides a strong foundation to support strategic IT initiatives and provide measurable references to enable the business area to make the right decisions. Program and project processes are brought into COBIT via the BAI01 (Manage Programs and Projects) process. Evaluation of IT Project management governance using COBIT 5 framework (Rooswati, 2018). From the mapping done, it produces several processes COBIT 5 is suitable for problems in the company. we use several COBIT processes to analyze the processes related to project management as follows by considering, time, cost and quality.

The approach in this research is based on several process constraints as well as several processes carried out based on the COBIT domain, including BAI01, BAI02, BAI03, APO11 and MEA01 according to the detailed explanation attached. We divide several processes according to needs or project management constraints as described in Table 1, Table 2, and Table 3.

Table 1. Process Alignment IT & Business Strategy

Criteria	COBIT Processes	Remark
Alignment of IT Project and Business Strategy	BAI01	Manage Programmes and Projects
Alignment of IT Project and IT Programmes	BAI01	Manage Programmes and Projects

Table 2. Process Alignment IT Plan

Criteria	COBIT Processes	Remark
Nature and scope of the project are confirmed and developed	BAI01, BAI03	Project Scope & Project Definition
The scope of projects is clearly defined and tied to building or enhancing business capability.	BAI01, BAI02, BAI03	Project Plan (time frame, budget, scope), Feasible Study, Project Charter
Project time, cost, scope is defined	BAI01	Project Plan (time frame, budget, scope)
Prepare and execute a quality management plan	APO11	Quality management plan, Customer requirements for quality management

Table 3. Process Alignment QCD Criteria & Project Management

Criteria	COBIT Processes	Remark
Time	BAI01	Project duration are met with time frame
Cost	BAI01	Project Cost are met with project budget allocation
Quality	APO11	Project scope are met with initial Project Plan Project quality are met with user requirement alongside UAT Document as evidence Each project is monitored, or baby sited after implemented to ensure quality

The last process is to make sure project are maintained and monitored that supported by project progress report as evidence we use MEA01 for controlling the Project Progress Report. the project management metrics of time, cost, scope, and quality have been the most important factors in defining the success of a project. More recently, practitioners and scholars have determined that project success should also be measured with consideration toward achievement of the project objectives.

The evaluation framework can be seen in Figure 2 below.

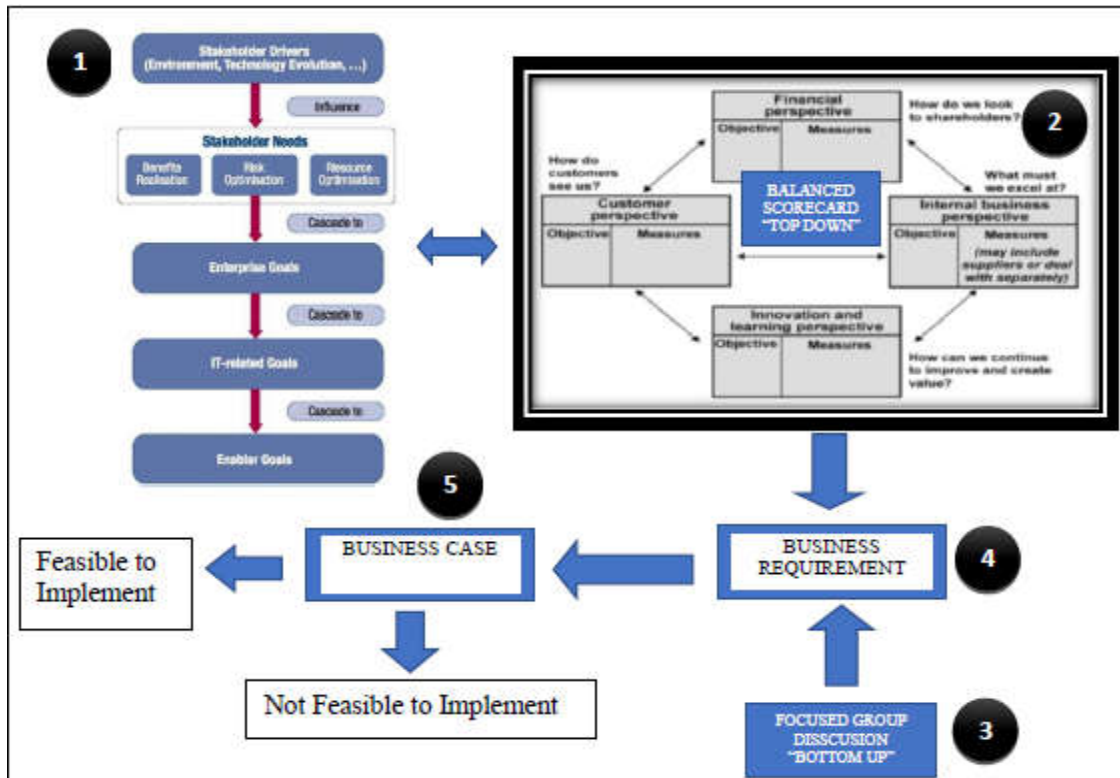


Figure 2. Evaluation Model Framework

As shown in Figure 2, the evaluation procedures consists of the following steps:

1. Step 1: Aligning IT Goals with Enabler Goals
2. Step 2: Access the Balanced Scorecard
3. Step 3: Focused Group Discussion
4. Step 4: Business Requirement
5. Step 5: Business Case Analysis

It can be seen in the IT Alignment Process & Business Strategy shown in Figure 2 that IT Strategy Plan is a combination of top and down approaches. In order get initiative from business user, IT team together with business unit generates another IT or digital initiative that comes from the bottom through Focused Group Discussion (FGD) mechanism. It is followed by creating a business case, which is a structured document that is presented in the form of argument / verbal / presentation that explains the subject matter and business needs, as an analysis material to determine IT approval. From the results of the analysis of the business requirement and business case that has been tested, it can be concluded whether the project is feasible or not.

Table 1 shows where in the process we use best practices in accordance with COBIT 5.0 recommendations on managed programs and projects (ISACA, 2015). In this process it must be ensured that every project that is run supports the company's business strategy that has been set. The evaluation method refers table 1- table 3 must be represented as follows:

- a. The nature and scope of the project are confirmed and developed: Ensuring a general understanding of the scope and objectives of the project among stakeholders has been formalized so that it can become an assessment parameter at the end of the project.
- b. The scope of projects is clearly defined and tied to building or enhancing business capability, that each project carried out has a clear scope of work and ensuring that each project implemented has positive implications for the company.
- c. Project time, cost, scope is defined, ensure that each project has been defined in terms of time frame, scope, cost. This will be a benchmark in ensuring that the project carried out has been well planned and becomes a parameter of assessment when the project has been completed to find out whether the project is carried out in accordance with the plan.

- d. Prepare and execute a quality management plan, ensure that there is a quality management plan in a project.

Evaluation of alignment QCD criteria & Project Management refers to table 3 which is represented as follow

- a. Time

Project duration are met with time frame: In this case, it will be checked between the actual project completion period and the planned timeframe of the project. If the period of completion is on time with planning, it will be assessed "Yes" in the assessment table, but otherwise if the project exceeds the planning period, there will be a "no" in the assessment table.

- b. Cost

Project costs are met with the project budget allocation: In this case, the actual project cost of the project will be checked with the planned cost of the project. If the cost of completion is in accordance with the planning costs, it will be assessed "Yes" in the assessment table, but conversely if the project costs exceed the planning costs, there will be a "no" in the rating table.

- b. Quality

Project scope is met with the initial Project Plan: Ensure that the project scope is carried out in accordance with project planning at the beginning. This is also done to ensure that all the main objectives in the project planning have been carried out entirely.

Project quality is met with user requirements alongside UAT Document as evidence. In this case, it will be checked between the functional quality of the actual project results and the user's request. If the functional quality of the work has fulfilled all user requests, it will be assessed "Yes" in the assessment table, but if the functional quality of the work does not meet the user's request, it will be assessed "no" in the rating table.

Each project is monitored after implemented to ensure quality, ensure that projects that have been implemented are monitored and carried out with close supervision to ensure that the implemented projects are properly implemented and can be used properly by the user. Project Management ensure that there is control and monitoring on each project throughout the life of the project. This can be ensured with a periodic progress report as evidence.

For the validation, Expert Judgment method is used. The used parameters in this study are reviewed by experts who are involved in the related use cases. It is hoped that that the method used in this study can raise the level of accuracy of project implementation as a basis for secondary empirical data evaluation.

4. Results and Discussion

This analysis data uses 70% of Empirical Data. This analysis uses using Evaluation method that supports several criteria as explained in the evaluation method. The results obtained are shown in Figure 3 below.

No.	Project Name	Alignment of IT project and business strategy	Alignment of IT project and IT Programme	Plan				Time	Cost	Project Quality			Project Management
				Nature and scope of the project are confirmed and developed	The scope of projects are clearly defined and tied to building or enhancing business capability.	Project time, cost, scope are defined	Prepare and execute a quality management plan	Project duration are met with time frame	Project Cost are met with project budget allocation	Project scope are met with initial Project Plan	Project quality are met with user requirement alongside UAT Document as evidence	Each project are monitored or baby sitted after implemented to ensure quality	Project are maintained and monitored that supported by project progress report as evidence
1	Automatic Creating Order	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No	No
2	Business Contract Application	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes
3	Project Management System	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4	Manpower Assignment Tracking	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No
5	MDR Creation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	Automatic Condemn Tag	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
7	Customer Perspective Dashboard	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	Material Receiving & Inspection System	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
9	Performance & Goal Management Compensasion	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
10	Recruiting & Onboard	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
11	IFCS Visitor Management System	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12	Covenant Dashboard	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
13	Content Management System - Phase1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Total Yes		13	13	13	13	12	8	12	12	13	13	10	11
Total No		0	0	0	0	1	5	1	1	0	0	3	2
Total Done without evidence		0	0	0	0	0	0	0	0	0	0	0	0
Total		13	13	13	13	13	13	13	13	13	13	13	13
Yes Percentage		100%	100%	100%	100%	92%	62%	92%	92%	100%	100%	77%	85%
No Percentage		0%	0%	0%	0%	8%	38%	8%	8%	0%	0%	23%	15%
Done Without Evidence Percen		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Figure 3. Evaluation Model Framework

Based on the results of the Analysis there are several important points which are of concern as follows:

1. Plan, there is 1 project that has not defined Project time, cost, scope project and there are 5 projects that have not defined the Prepare and execute a quality management plan
2. Time, there is 1 project that is not suitable between the project completion time and the planned time frame
3. Cost, there is 1 project that is not suitable between the cost project and the budget allocation that has been provided
4. Project Quality, there are 3 projects that have not been monitored after implementation

5. Project Quality, there are 2 projects that have not been properly maintained

Based on these results it is known, there are still projects that are not well controlled and monitored, which causes the fulfillment of QCD not to reach 100%.

1. Plan: Prepare and execute a quality management plan (88%)
2. Time: Project duration are met with time frame (92%)
3. Project Quality: Each project is not monitored or baby sitted after implemented to ensure quality with result (92%)
4. Project Management: Project are maintained and monitored that supported by project progress report as evidence (85%)

4. Conclusion

Based on result of the evaluation & modelling framework in project management that has passed this framework has been proven there are 13 projects can be executed in accordance with stakeholder needs which can provide the accuracy of project quality, time, cost based on COBIT criteria. From research conducted with proven through expert judgment, it is known that the method used in this study can raise the level of accuracy of project implementation.

References

- Hosseini, S., Moghadam, D. K., Vahdat, D., & Moghadam, R. A. (2011). Combination of IT Strategic Alignment and IT Governance to Evaluate Strategic Alignment Maturity. Copy, P., & Indrajit, R. E. (n.d.). *Personal Copy of: Prof. Richardus Eko Indrajit*.
- Kaplan, R. S., & David, P. (2007). Using the Balanced Scorecard as a Strategic Management System. *Harvard Business Review • Managing for the Long Term • July–August 2007*.
- Of, A. B., & Practices, C. B. (n.d.). Measures of Project Management Performance and Value. PMI. (2010). The Value of Project Management. *Project Management Institute*, 1–6. <https://doi.org/10.1002/pmj.20105>
- Singh, H. (2016). *Project Management Analytics: A Data-Driven Approach to Making Rational and Effective Project Decisions*.
- Rincon, I. (2012). *COBIT and project management: how to align your project management practices with the leading IT governance framework. Paper presented at PMI® Global Congress 2012—North America, Vancouver, British Columbia, Canada. Newtown Square, PA: Project Management Institute*.
- Rooswati, R. (2018). Evaluation Of IT Project Management Governance Using Cobit 5 Framework In Financing Company, (September), 81–85.
- ISACA. (2015). COBIT 5 Customized Process Reference Guide. *Cobit 5*.

Threat Hunting Early Experiment through Event Correlation and Memory Forensic

Arif D Purnomo¹, Charles Lim^{1*} and Burman Noviansyah¹

¹ Department of Information Technology, Swiss German University, Tangerang 15143, Indonesia

* Corresponding author, email: charles.lim@sgu.ac.id

Abstract. The cyber threat landscapes nowadays are dynamically evolving over time, the cyber security practitioner in corporations need to adapt with more sophisticated way with the latest cyber threat attacks are launched. Cyber Threat Intelligence is one of the tools that can be utilized as a cyber threat detection. Generally, CTI operates by integrating its directory with events collected from Security Information and Event Management (SIEM) to correlates all of the appliances logs within corporation and providing summarized and meaningful information that can be reviewed to identify legitimate malicious cyber threat activity. However, relying only CTI subscription that only contains blacklist domain and ip addresses integrated with SIEM will only provide passive detection for known cyber threats. The needs for proactive cyber threat detection is required to compete with the modern threat landscape. This research work will try to explore the possibility of detecting unknown or undetected cyber threats using network event correlation and memory forensic to validate its existence. Throughout this research time span, we're able to discover malicious network pattern that is proven to be undetected within internal organization endpoint protection. Therefore, this research will provide baseline for threat hunting activity based on network behavioural pattern.

1. Introduction

Many organizations nowadays rely on Indicator of Compromise (IOC), manual analysis, utilizing existing tools in related with detecting internal or external threat. For that reason, they need to get to the point where they could automate those processes on a regular basis and obviously have the skill set and capability to fulfill the maximum potential (Cole, 2016). Another aspect that Organization or Information Security practitioners need to pay more attention is that threat landscape is always changing throughout each year, that's why traditional Security Operations Centre (SOC) that relies only on SIEM cannot cope with the dynamic movement of the threat adversaries. Cyber Threat Intelligence is one of the key points to turn things around against the fast-paced cyber security threats. According to SANS, Cyber Threat Intelligence (CTI) is a "collection, classification, and exploitation of knowledge about adversaries. CTI is analysed information about the intent, opportunity and capability of cyber threats" (Shackleford, 2018).

CTI is just an element that works as a complementary factor for the role of Security Operation Center (SOC) team within the organization, in the end it will take more than just blacklist domain and blacklist ip addresses to conduct proper Cyber Threat Hunting (CTH) activity. Cases where threats from within organization goes undetected by the network and endpoint protection have been around and a new perspective to overcome them should be evaluated. One thing that also needs to be considered is, Cyber Threat Hunting is not all about data feed, CTH should also contain key points that will answer the adversaries' methodologies, tactics, techniques, knowledge gap that will put the defenders the upper hand in combating cyber security threat actors. IT Security Professionals or Security Operations Center (SOC) team must proactively observe and intelligently discover ways to reveal the potentially undetected threats within the organization, so that a descent SOC or incident response team are continuous process in the making.

2. Materials and Methods

2.1 Related Work Materials

Jakhaele (2017) on her journal which titled “*Design of Anomaly Detection Framework by Data Mining Algorithm for Network Flow*” uses data mining techniques with sliding window to capture network traffic, which will then capture the anomaly of the network that is compared with the normal traffic flow. She emphasizes the importance of anomaly-based detection for the unknown threats that will provide better detection rate compare to signature-based detection model.

Al-Mohannadi et al. (2018) on their journal use event analysis log data from cloud service honeypot to uncover attack pattern by the adversaries. They then perform event correlation based on the cyber incident log from AWS platform on Elasticsearch, Logstash, Kibana, which also known as ELK.

John (2017) elaborates all of the techniques required that can be done for detecting APT in his journal “*State of The Art Analysis of Defense Techniques Against Advanced Persistent Threat*”. He explains about the approach of the academic society and commercial solutions for detecting malicious content of APT and how to identify them based on the methodologies of the proposed solution. APT identification techniques such as machine learning classification, attack intelligence, context-based detection framework, honeypots and intrusion kill chain methods are highlighted in this research.

Divakaran et al. (2017) performed a research regarding detection of fundamental anomalous patterns in network traffic. Research conducted by evaluating the regression model from number of gathered malwares and normal traffic.

Mavroeidis and Josang (2018) performed a research back in 2018 regarding threat hunting activity relying on Sysmon log. They created a system that integrate the ontology from sharing threat intelligence platform and compare the sysmon log that is fed into the machine and then compare the activity based on the database of the shared threat intelligence platform.

2.2 Methods

This research work will only focus on the activity of cyber threat hunting by integrating existing event correlation appliance within organization with memory forensic activity to validate the threat hunting activity. The cyber threat hunting activity in this case will be performed by creating use case that is constantly monitoring the common ports used by windows OS service on existing SIEM. The cyber threat hunting activity will also integrate with the cyber threat intelligence on existing SIEM. When interesting traffic occur, the author tries to perform memory forensic to investigate the hypothesis of the undetected threat from the suspicious host captured on the event correlation. This research will be limited within specific organization network environment.

This research will be using existing devices for SIEM appliances, whereas the author also builds a cyber threat intelligence Minemeld from Palo Alto and threat intelligence feeds from OTX Alienvault feeding. The architecture of the research is depicted as Figure 1 below.

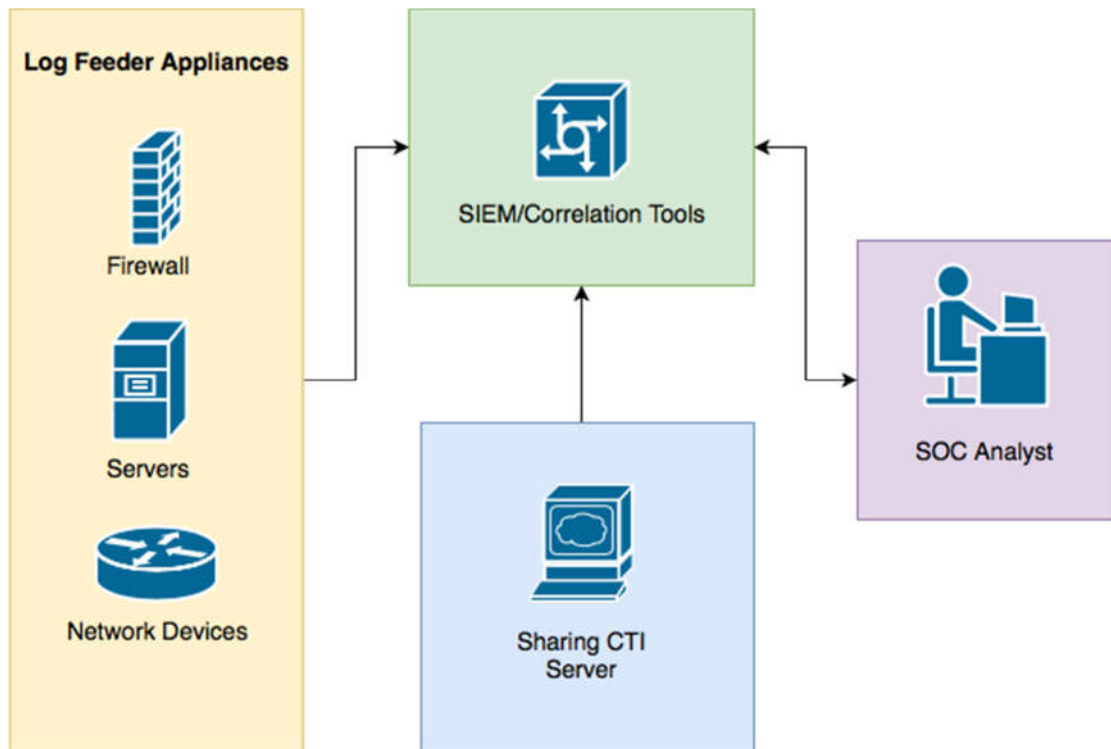


Figure 1. Research Architecture

This research will be using below proposed framework in order to conduct threat hunting program. In which the proposed framework derived from MITRE ATT&CK framework especially from “Lateral Movement” section with a few modifications necessary for conducting this research. The diagram of research methodology framework can be found on Figure 2. This framework represents the action that will be taken during the research, which consists of four phases, Collection, Preliminary Analysis Correlation, Platform Analysis & Evaluation.

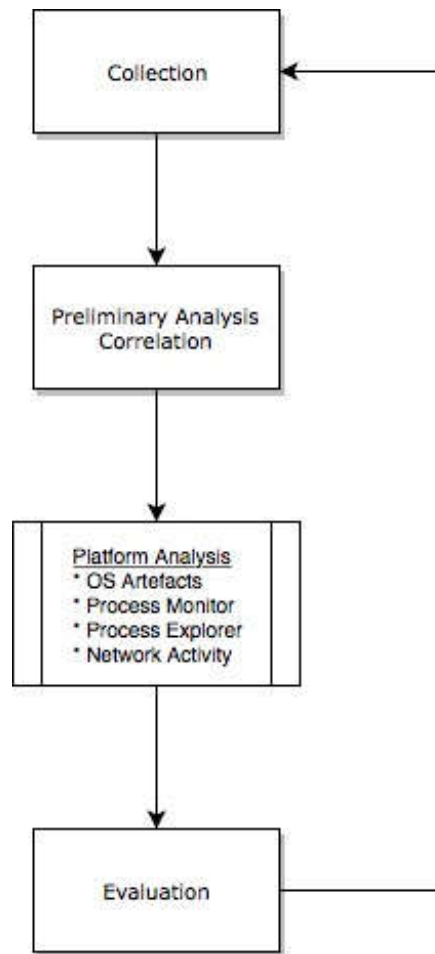


Figure 2. Threat Hunting Research Framework

Collection - In this phase, we gather all log files necessary from disparate log files correlated to our System Information and Event Management (SIEM). Based on this event logs, we should create event correlation that could identify the suspicious events initiated by suspicious host or potentially infected host.

Preliminary Analysis - The process of creating event correlation that could capture the suspicious network traffic will be done in this phase. And based on the captive event, we should be able to identify the platform fingerprint of the suspected host. Define hypothesis if network traffic initiated by specific host is having a suspicious behavior. Each captured event will have different methods to analyze because each captured event will require different method of investigation depending on its Operating System platform. On this phase, we will acquire the running process or any historical event of the suspected host to investigate if there's any malicious activity exist on the suspected host. This phase will analyze if certain conditions are met prior further checking into memory forensic investigation. This research focus on internal hosts within organization and not intended to "BYOD" host in particular.

Platform Analysis - This phase will validate the suspected host is legitimately infected and do remediation and necessarily contain the infected host. On this phase will also confirm if the signature-based detection mechanism running on the infected machine will do the protection feature properly.

Evaluation - This phase will evaluate if the preceding step are properly conducted and meet the objectives stated at previous phases by validating the results of memory forensic analysis to independent & credible 3rd party to identify the finding results of the memory forensic analysis process. The memory forensic analysis results can also be shared to vendor-related party to justify the findings of the memory forensic analysis process. This phase will also review the lesson learned and make room for

improvement of the things need to be done to contain, remediate and mitigate the undetected or even unknown threat from future reoccurrence as part of Security Incident Response Life Cycle Process. And can also be used as a starting point to set up new objectives for the next threat hunting activity.

3. Results and Discussion

3.1 Collection

This phase results in all the related devices and appliances must be sending through its log to correlation tools (SIEM) and should be displayed on SIEM dashboard. This phase will be the fundamental of all remaining phases that come right after this first one.

3.2 Preliminary Analysis

On this Preliminary Analysis phase, the author creates hypothesis of threat hunting. The author filters all the logs collected in hoping that it will display more meaningful information, such as anomaly of network traffic generated by specific host. This phase will mark the event correlation rules created by the analyst, it should represent the intrinsic value of the IOC (Indicator of Compromise) or the TTPs (Tactics, Techniques & Procedure) from Network perspective or Host generated-event regarding the threat that is being hunted. Figure 3 represents the event correlation filter that is created to trigger the undetected or unknown threat.

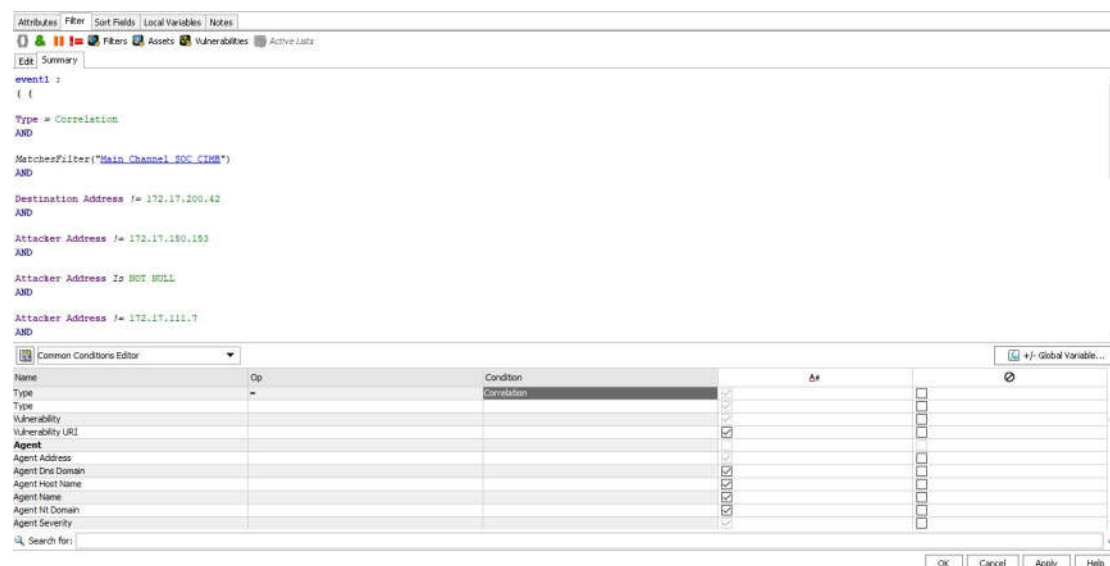


Figure 3. Event Correlation Filter Rule

After the creation of Event Correlation Filter, the analyst should monitor the SIEM dashboard in order to justify the hypothesis of the correlation filter that is previously generated. Figure 4 below depicts the result of event log generation of the previous phase that shows the event of the potentially infected host that remains hidden or undetected if the author would not create the event correlation filter to identify the TTPs of the potentially infected host.

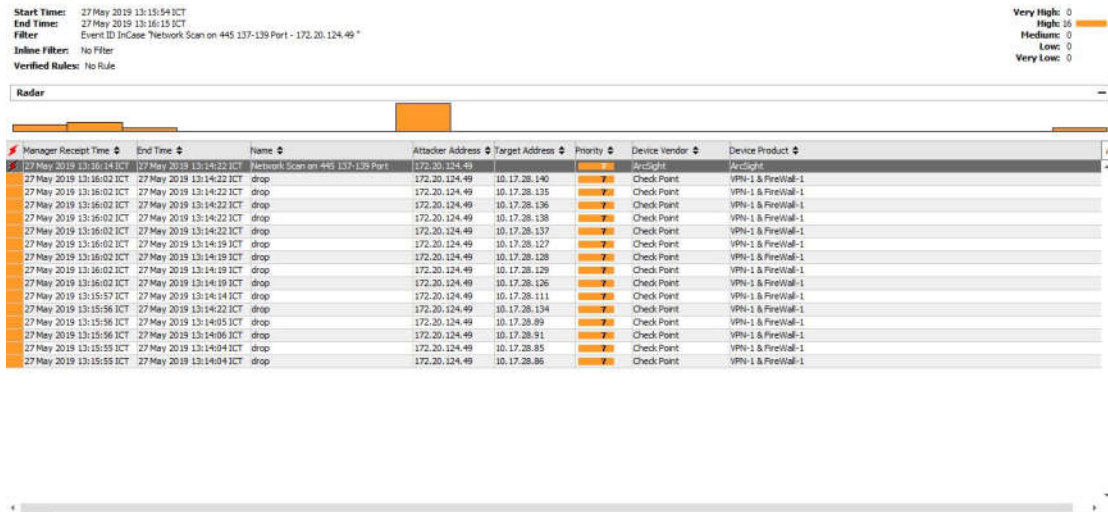


Figure 4. Event Correlation Dashboard

3.3 Platform Analysis

On this phase the author starts the memory forensic activity of the potentially suspected host. Throughout this research, the author managed to gather two captures of live memory acquisition. The one that is generating numerous attempts of network scanning and the other is triggering the rule of threat intelligence. The process of memory acquisition itself conducted by using two different applications, which are *dumpit* and *Belkasoft*. The process of the memory acquisition using the mentioned tools itself is pretty straightforward, the author only run the tools and select the directory of the acquired memory from the host. The memory dump acquired will be further analysed by using different kind of tool. During the memory forensic process, the author uses *Volatility* and its plugins as main tool. Volatility is an open source, advanced tools for memory forensics written on python and runs on Linux, Windows & Mac OS (Ligh et al., 2014). Various processes can be gathered from memory forensic analysis, even potentially malware hidden can be uncover by performing memory forensic analysis.

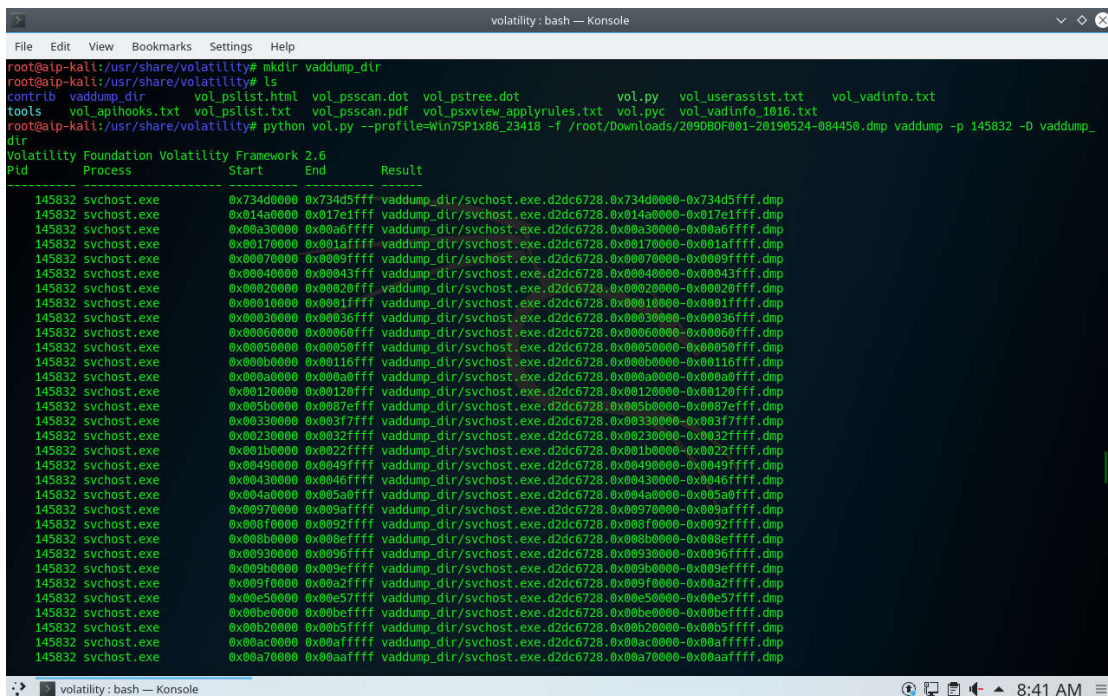


Figure 5. Memory Forensic Analysis using Volatility

3.4 Evaluation

The Evaluation phase of the memory forensic analysis process results in successfully found the undetected or unknown threat for the specific endpoint protection product by the time this writing is published.

DETECTION	DETAILS	COMMUNITY
Acronis	Suspicious	AhnLab-V3 Malware/Win32.Generic.C3177661
CrowdStrike Falcon	Win/malicious_confidence_100% (D)	Endgame Malicious (high Confidence)
FireEye	Generic.mg.c31a28e5c1b21e22	Kaspersky HEUR:Trojan.Win32.Generic
Qihoo-360	HEUR/QVM20.1.B23D.Malware.Gen	SentinelOne (Static ML) DFI - Malicious PE
Sophos ML	Heuristic	ZoneAlarm by Check Point HEUR:Trojan.Win32.Generic
Ad-Aware	Undetected	AeSI-Lab Undetected

Figure 6. VirusTotal results of the suspected host

From Figure 6 can be stated that the threat hunting process generates the expected outcome to identify the previously undetected or even unknown threats within organization.

4. Conclusion

Network Event Correlation that is gathered by mainly network security devices, such as firewalls, IPS, APT and also other appliances, whereas performed throughout this research is able to identify network behavior of potentially infected host. It can also be used as early indication or early detection to identify potential internal threat that cannot be detected by endpoint or network protection within organization. Simple idea such as inspecting common port that is generally used by Windows operating system, can lead to investigation process of potentially undetected malware infected host.

Memory forensic that is performed during this research also complement the network event correlation activity. It acts as a mandatory action to emphasize the legitimacy of early detection that is initially performed by network event correlation. It can become a thrilling activity for IT Security Incident Responder or other Cyber Security Professionals to investigate a cyber security case from different perspective and not solely rely on robust Information Security solutions available on the market. Memory forensic also will broaden the knowledge of the cyber security professionals to better understand the TTP (Techniques, Tactics & Procedures) of the adversaries.

References

- Al-Mohannadi, H., Awan, I., Al Hammar, J., Cullen, A., 2018. Cyber Threat Intelligence from Honeypot Data using Elasticsearch. IEEE.
- Cole, Dr.E., 2016. Threat Hunting: Open Season on The Adversary. SANS.
- Divakaran, D.M., Fok, K.W., Nevat, I., Thing, V.L.L., 2017. Evidence Gathering for Network Security and Forensics. Elsevier.
- Jakhaele, A.R., 2017. Design of Anomaly Detection Framework by Data Mining Algorithm for Network Flow. ICCIDS.
- John, J.T., 2017. State of The Art Analysis of Defense Techniques Against Advanced Persistent Threat. Technical University of Munich.
- Ligh, M.H., Case, A., Levy, J., Walters, A., 2014. The Art of Memory Forensics. Wiley Publishing.
- Mavroeidis, V., Josang, A., 2018. Data-Driven Threat Hunting Using Sysmon.
- Shackleford, D., 2018. CTI in Security Operations: SANS 2018 Cyber Threat Intelligence Survey. SANS Institute.

Core Banking System Scalability Review

Achmad Fakhruddin¹, Heru P Ipung^{1*}, M Amin Soetomo¹ and Charles Lim¹

¹ Department of Information Technology, Swiss German University, Tangerang 15143, Indonesia

* Corresponding author, email: heru.ipung@sgu.ac.id

Abstract. In carrying out the digital transformation, bank has to find their optimal routes to exploit the Open Banking opportunity and minimize or manage any threat. However open Banking models will be making implication on the increasingly hit from the external parties. It will be new and more unpredictable data volumes as third parties request access to the banking system. Bank has to prepare their Banking system on the Enterprise Architecture, and one of the concerns is on the scalability. For the traditional core banking system, the platform of mainframe and other legacy technology could remain the bottleneck to agility and scalability. In this paper, scalability testing and analysis are carried out to verify whether the traditional banking core system running on the power platform can be scalable to support open banking strategy. The results found were that the core banking system could be scalable, but needed a lot of enhancement on the core banking application side, as well as capacity upgrading on the server side.

Keywords: Core Banking System, Scalability, Open Banking, Transaction Volume

1. Introduction

The disruption wave as because of the digital technology are happen in all industry now, including in the banking industry (Accenture, 2018). Some of the existing banks have begun to be affected by this digital change, but most of them already know about the changes that must be made. The digital transformation strategy in each bank can be different, especially related to the condition of the organization and its customers.

The emergence of a number of new players in the financial industry turned out to also change the pattern of services expected by customers while influencing the income they received. Existing banks must make fundamental changes, especially related to the way they operate. Their services must be ensured in very good performance including being able to quickly launch new digital innovation services.

Bank has to implement new digital initiatives including Open Banking. Open Banking is the shifting strategy from closed to open banking models (Deloitte Digital, 2017). Their banking system: data, banking product and other services functions are shared and open to other institutions and third parties with the main objective to generate the additional value.

However there are many implications after implementing Open Banking. The more integrated external parties will create more transaction volume. Bank has to prepare their Banking system on the Enterprise Architecture. It would be suddenly the number of Transaction become huge after any business deal and collaboration with third parties.

Discussing about enterprise architecture system, there are top three concerns: **Scalability, Availability and Performance**. These parameters are really strategic in determining the user experience in transactions and in long term relationship. It also directly impact to business regarding to online revenue, business growth, competitiveness and high business interest (Shailesh Kumar, 2015). For the traditional banks, the platform of mainframe and other legacy technology could remain the bottleneck to agility and scalability (MIT Technology Review Insight, 2018). Banks has to modernize IT for many years to ensure the open banking system readiness.

This paper focused on review traditional core banking system to support open banking system model. The new comprehensive methodology was created based on the previous references. It's used to test the scalability with the targeted business volume number.

2. Materials and Methods

2.1 Open Banking Implementation Challenges

There are challenges in implementation of open banking and other digital initiative. In such country such as Indonesia, it requires an adequate level of internet infrastructure in order to adopt the technology and provide the necessary support for its usage and growth. The internet infrastructure with proper band with in Indonesia is not well distributed and yet to reach many area. Bank has to also strengthen the privacy and security of data transferred over the internet network. Cybercrimes and hacking is a big reality in online banking, and this cause to a decline in trust on bank digital activities (Varda Sardana, 2018). Open banking implementation also has the impact of the additional transaction coming to the banking system. Bank has to prepare their system in the enterprise web application system (Shailesh, 2015).

2.2. Online Banking Performance Parameter

The design of Web site and online banking resulting quick response time to the users should be provided by Banks. Bank customer would be more motivated to access the web system if they have the good performance especially on the response time. (Fida Hussain Chandio *et al.*, 2017). Forrester, sponsored by Akamai Technologies (NASDAQ: AKAM) has conducted study at 2009. They released key findings at e-Commerce web site performance in correlation with behaviours of shopper. The result highlighted that the new threshold response time of web application expected by the user is in two seconds. They will abandon the online web commerce if they have to wait the response more than 3 seconds. That is the new expectation as the similar study conducted 3 years before resulted 4 seconds for the threshold.

The most variables affecting online banking customer satisfaction are: quick responsiveness, reliability of the system, efficient service, security, look and feel site (aesthetic), easy to use and user experience (Mohsen, 2015).

2.3. Computer Performance

Computer performance is majority determined by a combination of 2 parameters: first is throughput and the second is response time. The other aspects related with computer performance are computer power efficiency and systems availability (IBM, 2019). Computer system **response time** is the duration and elapsed time between the request of transactions submitted and the starting of a response to that transaction. At interactive users, the response time is the duration time to get the display result after clicking <enter> button. It's a critical aspect of performance as because of relation to the potential visibility of end users or customers. The computer system **throughput** is a measure of the number of work can be performed by a computer system at over the period of time.

Response time has the closely relation with throughput. As resources of computer (processors, disk, memory and etc) are used by many users and simultaneous transactions, it will result delays to individual transactions because of the similar resources are still uses by other transaction. The computer system resource that has the capability to in shared resource may cause additional response time because of the mechanism.

2.4 Transaction Type

There are 3 transaction types. The first is called **I/O-limited transactions**. It consists of a mix of considerable disk wait times. This transaction type will retrieve and get the most benefit from the good and an efficient of disk I/O subsystem (IBM, 2019). With the technology of disks, setting large disk cache, and also advance speed I/O links will do improve the throughput and response time and of these types of transactions. Application that has the characteristic on this transaction type is suggested to use SSDs as an alternative to previous technology of HDDs.



Figure 1: I/O Limited transactions type

The second type of transaction is **resource limited transaction**. It consists of a considerable mix of processor and wait activity. These transactions type gain benefit most from improving application efficiency. To improve the system, application has to cut, minimize and reduce the impact of critical shared computer resources (between program applications and threads).

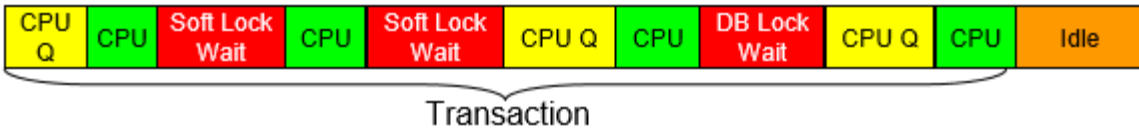


Figure 2: Resource Limited transaction type

The third type is called **processor-bound transactions**. This type requires considerable processor cycles and very little I/O or wait times. The performance of the application with this type will be much better if they use higher processor frequency and also the efficiency of the processor pipeline stages. Utilizing of improved memory and processor will also generate better performance.

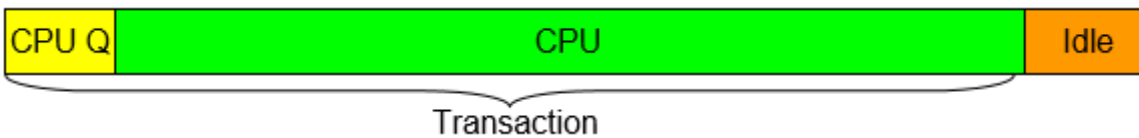


Figure 3: Processor-bound transaction type

2.5 System Scalability

Scalability is related to the capability of system to manage and handle the additional demand and workload without neglecting the overall efficiency (Shailesh Kumar, 2015). This capability is mandatory for the enterprise application system and its related ecosystem. It has different elements related to scalability: Application and its ecosystem, increased workload and efficiency.

Figure 5 explain the scalability layers. It's informed based on *order of contribution to scalability in the chain of request processing*. This figure highlight that the layer of enterprise application give the major impact of the overall enterprise web application system.

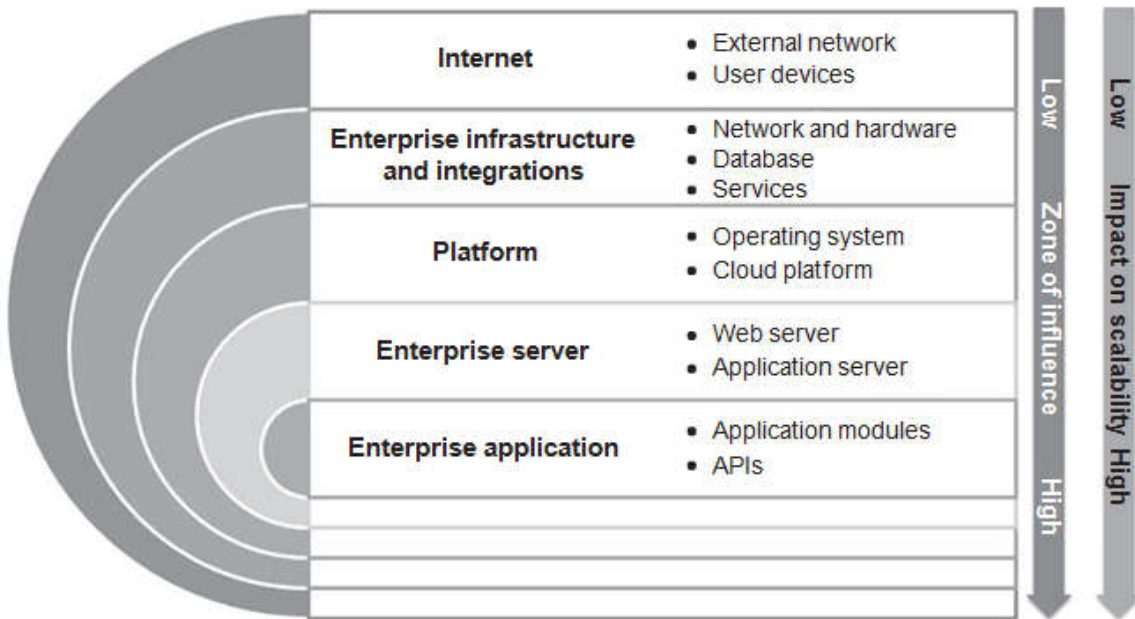


Figure 4: Layers of Scalability and the Impact (Shailesh, 2015)

In the online banking system, the layer of enterprise application is represented by core banking system. When talking about application system performance, one of the major parameter is application speed to response the request. Both server and enterprise applications are the main actors in generating response time (Techempower, 2016).

2.6 Scalability – Vertical or Horizontal

Open Banking models also need the elasticity. This is the degree to which a system can quickly adapt to workload changes by utilizing and upgrading resources in an on-demand manner. During the system application is running, if suddenly any additional workload system must be easy to scale up and if necessary to add the required resource.

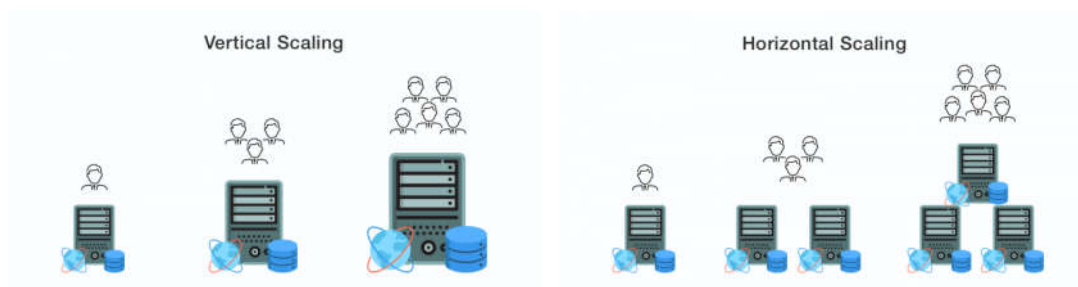


Figure 5: Vertical and horizontal Scalability

When there is an increase in transaction volume that cannot be handled by the existing specs server, it is necessary to increase the existing server capacity. Basically there are two types of scalability: vertical and horizontal. Vertical means with the existing box of server, we add the capacity by upgrading the resources of memory, processor, storage or others depend on the specific required to upgrade. And horizontal scaling is adding the capacity by adding the number of box server. The decision to choose either vertical or horizontal is depend on the related application capability to manage the scalability. Especially for vertical, it also depends on the server technology capabilities used, can it be upgraded up to the desired capacity.

The horizontal scaling now is more popular in the open banking architecture because of several things such as below:

- It's **elastic**.
- It is **dynamic**. The existing resources can be kept still as it is and online during the additional the new server. It doesn't have the negative (risk) implication to the existing or previous server.

2.7 Scalability Testing Methodology

Figure-7 is the new testing methodology for this core banking system scalability review testing. It was modified and adjusted from the previous references of the PSR (Performance, Scalability and Reliability) testing methodology.

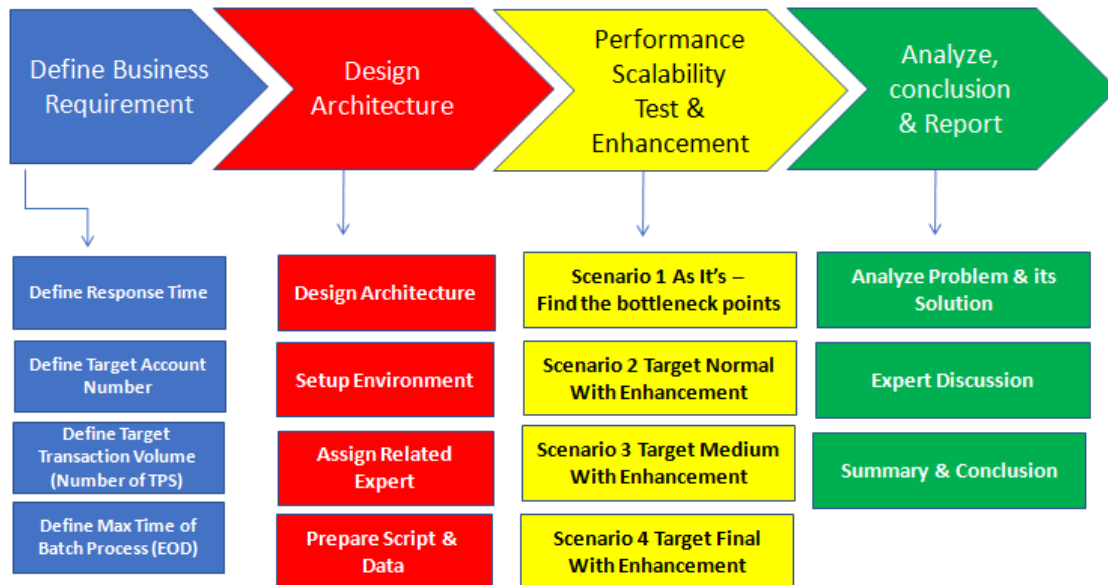


Figure 6: Scalability Performance Testing Methodology

2.7.1 Define Business Requirement

Bank could be lower their competitiveness if they slowly adopt the Digital Banking Technology, but it could worsen their financial performance if they adopt it aggressively (Rosnita, 2018). Thus the business team must decide to what number the system must be scalable in accordance with the business value and targeted growth. These numbers will be a reference for the IT team in designing and preparing the required system scalability.

To ensure the scalability of traditional core banking system, there are parameters should be provided by business team: Expected response time (seconds), Target account number, Target online transaction Volume, Maximum time duration of batch process (EOD and SOD)

2.7.2 Design Architecture

The business number targeted by bank will be main reference to design the IT System Architecture. In this stage, all related components must be prepared and estimated able to cover the business number and its growth (Shailesh, 2015). As this works focus on traditional core banking system scalability, the preparation components is as follows: Design Architecture, Setup Environment, Assign Related expert, Prepare the Data and Test Script.

2.7.3 Performance Scalability Test and Enhancement

It's identified that for the end to end enterprise system architecture, application layer is the main component of which we have the control, and also giving a maximum impact on scalability (Shailesh, 2015). At the application layers, the scalability is contributed by: software, process, and hardware. The challenges on software include potential issues and bottleneck in application code. And the hardware

challenge includes the server and infrastructure components such as memory, CPU, Disk capacity and network.

2.7.4 Analyze, Conclusion and Report

All data and results in this scalability testing are collected and documented. The experts then discussed together including with other related parties and business or user team. Conclusions and suggestions must be made in accordance with the direction of strategy scalability targeted by the company (bank).

3. Results and Discussion

The research carried out was by conducting scalability testing of core banking system based on the methodology that has been made (as per explained in the previous chapter). The first step is defining the target business volume. From the current volume, it's targeted the aggressive growth at year-5 and year-10 (as per stated at table-1). The second stage was testing environment preparation before doing the detail activities of scalability testing.

At the first stage of scalability testing, the focus was finding scalability bottleneck of the current traditional core banking system. All expertises were involved in this activity. Every bottleneck found during test should be solved (application customization or hardware adjustment) and then did the repeated test. This is done back and forth and repeatedly until getting the optimal results. Figure-8 is sample profile application in one step of testing. It explains which program should be enhanced as it consumed high CPU and I/O.

MI complex instruction	Library name	Program name	Module name	Number of threads	Inline CPU percent of total	Inline elapsed time percent of total	Times called	Calls made	MI complex instruction count	Inline CPU usecs	Cumulative CPU usecs	Inline elapsed usecs
	QSYS	QCMD	QCMD	417	.0000	.0001	0	416	10	149,2890	475,575,686.3770	55,10
	VLKMYI	VPRCSTRES	VPRCSTRES	360	.8436	.1983	0	258,945	2	4,034,569.0390	342,137,951.2990	114,6
	VLKMYI	VFMTASTRES	VFMTASTRES	342	1.4667	.4182	20,027	439,017	79,694	7,014,487.5450	302,714,003.1510	241,8
	VLKMYI	VMTBALO	VMTBALO	342	1.2563	.3865	19,941	477,652	59,770	6,008,178.4080	206,841,942.3460	223,4
	QSYS	QDMCOPEN	QDMCOPEN	Q> 347	2.9351	3.5532	458,215	936,265	4,661,912	14,037,501.8830	171,791,179.0330	2,054
	VLKMYI	VDRV	VDRV	V> 20	6.7642	.4002	0	394,127	0	32,350,107.0170	106,573,172.0620	231,4
	VLKMYI	CLDRV	CLDRV	C> 20	0	.0000	0	20	0	106,573,172.0620	941.3	
	QSYS	QDBOPEN	QDBOPEN	Q> 347	1.9617	2.9047	458,186	2,928,863	2,072,362	9,382,046.9560	73,652,876.3120	1,679
*RSLVSP				392	13.4244	3.0069	3,869,197	0	0	64,203,062.8910	64,203,062.8910	1,738
	VLKMYI	VCONVB	VCONVB	V> 19	12.2280	.5767	111,436	0	0	58,481,187.0610	58,481,187.0610	333,4
	VLKMYI	VVLBAL	VVLBAL	V> 342	.6980	.1629	19,885	178,932	59,646	3,338,437.5800	50,616,683.3290	94,17
	QSYS	QDMCLOSE	QDMCLOSE	Q> 347	1.3038	2.3150	417,357	417,354	3,736,263	6,264,392.7440	41,302,797.5440	1,338
*DESCR				347	6.9097	.7321	457,035	0	0	33,046,112.2350	33,046,112.2350	423,3
	VLKMYI	VVLCRD	VVLCRD	342	.3323	.2051	19,907	278,527	59,700	1,589,394.1210	31,857,167.5020	118,5
	QSYS	QDBGETKY	QDBGETKY	355	.9373	.9224	498,689	121,615	935,865	4,482,645.8500	27,714,537.5340	533,4
*CRTDOBJ				346	5.2080	.6639	458,199	0	0	24,907,432.1690	24,907,432.1690	383,9
	VLKMYI	VVLLIB	VVLLIB	342	.1847	.2583	19,939	418,731	21	883,192.7690	22,178,922.5470	149,3
	VLKMYI	VVLACT	VVLACT	342	.1604	.1191	19,887	139,218	59,690	767,303.1290	21,999,903.9530	68,86
	QSYS	QDBXREF	QDBXNOSYNA	346	1.1946	.7650	458,159	916,316	0	5,713,434.4830	20,577,383.6880	442,3
	QSYS	QLRMAIN	QLRMAIN	342	.5914	.6128	218,951	258,823	577,497	2,828,183.7120	19,419,087.8520	354,3
	VISIO>	DGAVBL	DGAVBL	342	.1650	.0989	19,881	119,271	39,759	789,286.5770	19,157,662.9270	57,16
	VLKMYI	VLKFM20	VLKFM20	342	2.1386	.2553	39,642	118,922	39,639	10,228,170.6330	17,669,568.2540	147,6
*SETCR				358	3.4885	.5459	498,728	0	0	16,684,012.3930	16,684,012.3930	315,6
	QSYS	QDBGNSP	QDBGGOBJLN	354	.3056	.7215	458,171	0	916,341	1,461,405.2670	14,531,928.4560	417,2
	QSYS	QWCCRCRC	QWCCRCRC	342	.4977	.4449	99,381	198,765	496,906	2,380,507.8250	13,425,696.5260	257,2
	QSYS	QCADRV	QCADRV	346	.5533	.5618	159,585	478,793	319,175	2,646,019.8970	13,142,304.3050	324,8
	VLKMYI	VLOG	VLOG	8	.3272	.1617	0	258,472	0	1,564,832.1420	12,940,596.6030	93,50
	QSYS	QDBOPEN	QDBOPEN	346	.3577	.6733	458,303	0	0	1,333,481.8300	13,017,336.6550	380,3

Figure-7: I/O and CPU consumption of Programs at Core Banking system

There were many major customization done at core banking application, and also several adjustment and server setting. It's including utilize SSD disk technology that's resulted 3 times speed performance of the system.

Table-1 explained the target business volume and the detail core banking system scalability testing result.

Table-1: Core Banking System Scalability Testing – Target Business and Result

Scenario	Target Business		Achievement		Server Spec
	# Account	TPS	Online Transaction Response time	Batch duration for End Of Day Process	Power 870
Volume-1 (Year-2)	10 M	5,600	< 200 milliseconds	56 minutes	4-cores, Memory 64GB, SSD 24x387 GB
Volume-2 (Year-5)	17 M	8,600	< 200 milliseconds	1 hour 25 minutes	6-cores, Memory 64GB, SSD 24x387 GB
Volume-3 (Year-10)	50 M	13,000	< 200 milliseconds	1 hour 31 minutes	12-cores, Memory 128GB, SSD 24x387 GB

4. Conclusion

Regarding to implementation of open banking models, bank has to review their core banking system scalability. It's reminded by MIT (MIT Technology Review Insight 2018), that bank with traditional core banking system platform might be has more challenges in this scalability. This paper has concluded that:

1. Traditional core banking system running on power platform can be scalable to support open banking strategy. But it is vertical scalability and need enhancement in the core banking application.
2. The detail enhancement of core banking application may vary and differ from one bank to another. It can be identified after scalability performance testing is conducted.
3. The response time of core banking to support online banking can be targeted not more than 250 milliseconds.
4. The scaling of power server depends on the target business volume. The upgrading server specification especially is on number of CPU, memory and disk technology and capacity.

References

- Accenture. Exploiting Inbound and Outbound Trade Opportunities, the Brave new World of Open Banking, 2018
- Akamai. Akamai Reveals 2 Seconds as the New Threshold of Acceptability for eCommerce Web Page Response Times. Available at <https://www.akamai.com/us/en/about/news/press/2009-press/akamai-reveals-2-seconds-as-the-new-threshold-of-acceptability-for-ecommerce-web-page-response-times.jsp>, 2009
- Akshay Kumar, Magapu and Nilhil Yralagadda: Performance, Scalability, and Reliability (PSR) challenges, metrics and tools for web testing, Faculty of Computing Blekinge Institute of Technology SE-371 79, Karlskrona, Sweden, 2016
- Deloitte Digital: Open Banking, What Does The Future Hold? – April 2017
- Fida Hussain Chandio, Zahir Irani, Akram M. Zeki, Asadullah Shah, and Sayed Chattan Shah (2017). Online Banking Information Systems Acceptance: An Empirical Examination of System Characteristics and Web Security. *Informations System Management* 2017, VOL. 34, NO. 1, 50–64
- Hu, Y., Liao, P. Finding critical criteria of evaluating electronic service quality of Internet banking using fuzzy multiple-criteria decision making. *Applied Soft Computing* 11, 3764–3770, 2011
- IBM Corporation: IBM i on Power - Performance FAQ, IBM Power System Performance Book, 2019
- MIT Technology Review Insight: Open Banking - The Race to deliver banking as service, 2018
- Mohsen Mazaheri Asad, Najma Sadat Mohajerani, Mohammad Noursersh: Prioritizing Factors Affecting Customer Satisfaction in the Internet Banking System Based on Cause and Effect Relationships. . 1st International Conference on Applied Economics and Business, ICAEB 2015

- Picarlo Gera: Managing digital disruption in Banking, <https://www.accenture.com/hu-en/insight-perspectives-banking-digital-strategy-drives-new-era> – 2019
- Ping Li, Dong Shi, and Jianping Li. Performance test and bottle analysis based on scientific research management platform. 2013 10th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pages 218–221, 2013
- Rosnita Wirdiyanti: Digital Banking Technology Adoption and Bank Efficiency: The Indonesian Case, Dec 2018
- Shailesh Kumar Shivakumar: Architecting High Performing, Scalable and Available Enterprise Web Applications, Elsevier Book – 2015
- Techempower: think about performance before building a web application, <https://www.techempower.com/blog/2016/02/10/think-about-performance-before-building-a-web-application/>, 2016
- Tiantian Gao, Yujia Ge, Gongxin Wu and Jinlong Ni, “A Reactivity Based Framework of Automated Performance Testing For Web Applications” 9thInternational Symposium on Distributed Computing and Application to Business, Engineering and Science.
- Varda Sardana and Shubham Singhania: Digital technology in the realm of banking: A review of literature, International Journal of Research in Finance and Management 2018; 1(2): 28-32
- Yogita M. Rasal, Sangeeta Nagpure: Web Application Performance Testing Using Reactive Based Framework, International Journal of Research in Computer and Communication Technology, Vol 4, Issue 2 ,February -2015

Revealing and Sharing Malware Profile Using Malware Threat Intelligence Platform

Faiz Iman Djufri¹ and Charles Lim^{1*}

¹ Department of Information Technology, Swiss German University, Tangerang 15143, Indonesia

*Corresponding author: charles.lim@sgu.ac.id

Abstract. Cyber Security is an interchange between attackers and defenders, a non-static balancing force. The increasing trend of novel security threats and security incidents, which does not seem to be stopping, prompts the need to add another line of security defences. This is because the risk management and risk detection has become virtually impossible due to the limited access towards user data and the variations of modern threat taxonomies. The traditional strategy of self-discovery and signature detection which has a static nature is now obsolete in facing threats of the new generation with a dynamic nature; threats which are resilient, complex, and evasive. Therefore, this thesis discusses the use of MISP and The Triad Investigation approach to share the Indicator of Compromise on Cyber Intelligence Sharing Platform to be able to address the newt threats.

1. Introduction

Cyber Security is an interchange between attackers and defenders, a non-static balancing force (Schneier, 2012). As the military strategists Carl von Clausewitz, say in a wartime context, "The Defenders are in the position of the interior." You must defend yourselves against any feasible assault, even against unfulfilled attacks, while assailants must discover only one fault in penetrating the scheme (Schneier, 1998). European Union Agency For Network and Information Security reports Malware is the most common cyber threat recorded last year, involving somehow 30 percent of all occurrences of infringement in Figure 1 (ENISA, 2018).

In 2018, Identity Theft Resource Center reported Records compromised by data breach in 2018 as shown as Figure 1.2. A research by FireEye's M-Trends shows that an organisation has a average moment to find out 78 days of attack (FireEye, 2019).

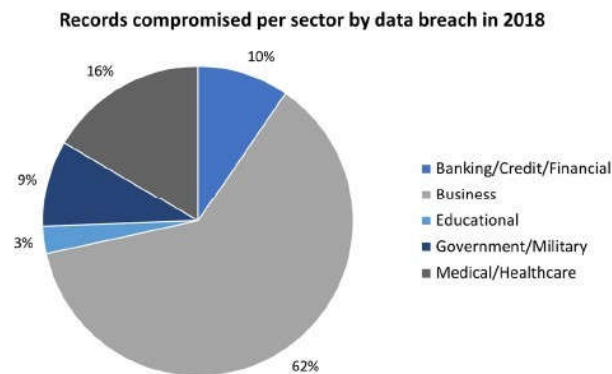


Figure 1 Records compromised by data breach in 2018, per sector (ENISA, 2018)

The increasing trend of novel security threats and security incidents, which does not seem to be stopping, prompts the need to add another line of security defenders. This is because the risk management and risk detection has become virtually impossible due to the limited access towards user data and the variations of modern threat taxonomies. The traditional strategy of self-discovery and signature detection which has a static nature is now obsolete in facing threats of the new generation with

a dynamic nature; threats which are resilient, complex, and evasive. As a result of this new line of security defenders must include cyber attacks today. The traditional heuristic and signed static approach of safety does not match the dynamic nature of the fresh wave of risks, which are identified as evasive, durable, and complicated. Organisation, in fact, must collect and communicate cyber threat data in real time to prevent assaults or at least to implement prompt recovery from disasters. This research's contribution is to fulfill part of Detection and Analysis Process in Incident and Response procedures by analyzing the threats and sharing its information as the effort Security Analyst do, to fight with increasing of new threats and incident indicator in financial sector.

2. Materials and Methods

2.1 Security Operation Center

A Security Operation Center (SOC) is defined as a team comprised of security analysts with the objective to detect, analyze, prevent, respond to, and report on security incidents in the cyber realm (Zimmerman, 2014). The functioning of the SOC assumes that the constituency is compromised at some stage. In addition, those involved include persons who have legally access to the IT resources of the constituency. In accordance with this logic, the SOC shall be able to operate on the integrity, confidentiality and availability of constituency assets and networks without full confidence. Although the SOC must be strongly integrated with constituency IT systems, it must be isolated from compromise. There are some concern point that will become goal of operating SOC (Zimmerman, 2014):

- Achieve close to zero packet loss at specified presence surveillance points.
- Prevent the threat actor from identifying (and avoiding) the existence of (IDS and IPS) surveillance functions.
- Ensuring 100 percent of security events are delivered through SOC monitoring systems from end devices and when necessary, protect them against prying eyes. item Enhance the survivability and discourage unlawful access to SOC resources, even where constituency components are compromised.
- Protect sensitive data and records kept by SOC from disclosure

2.2 The Incident Response Life Cycle

As a team is composed of security analysts with the objective to detect, analyze, prevent, respond to, and report on security incidents, SOC utilize incident handling process to fulfill those objective. Well known organization like National Institute of Standards and Technology (NIST) "has responsible for developing information security standards and guidelines, including minimum requirements for Federal information system" (Cichonski et al., 2012) publish number , provide a publication about Incident Response Process with number document publication NIST.SP.800-61r2

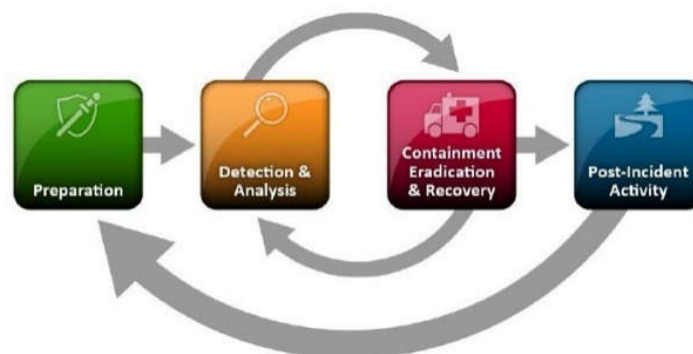


Figure 2 The Incident Response Life Cycle (Cichonski et al., 2012)

NIST Incident Handling Process, which acts as a manual for an event operator, comprises four main stages. The preparation stage is where the organization gathers the required individuals, strategies, information, and instruments to remedy an event rapidly and fully. The second stage of the method includes detecting and analyzing security events from the organization's network perimeter, target

perimeter, system-level activity, application-level activity, or users. It also includes reporting an incident. The organization will start the stage of confinement, eradication, and restoration upon statement of an incident. The threat may proceed its path of intervention during this stage. Incident handlers often switch away and forth between the second and third phases during an incident response as new information about the threat becomes accessible. Finally, the incident response team develops the post-incident document in the post-incident operation stage and develops alternatives to enhance the method for the next incident (Cichonski et al., 2012)

2.3 Cyber Security Threat

In action of Incident Handling, the SOC will deal with the Cybersecurity Threat. SO (2019) define definition of Cybersecurity Threat is "an action on or through an information system that may result in an un-authorized effort to adversely impact the security, availability, confidentiality, or integrity of an information system or information that is stored on, processed by, or transiting an information system. The term does not include any action that solely involves a violation of a consumer term of service or a consumer licensing agreement". There are some common attacks on Cybersecurity Threat (Wang, 2009):

- Eavesdropping.
- Cryptanalysis.
- Password Pilfering.
- Identity Spoofing Attack.
- Buffer-Overflow Exploitations.
- Intrusion such as IP scans and port Scans.
- Denial of Service Attacks.
- Malicious Software (Malware).

2.4 Malware

Malware is "A program that is covertly inserted into another program or system with the intent to destroy data, run destructive or intrusive programs, or otherwise compromise the confidentiality, integrity, or availability of the victim's data, applications, or operating system" (SO, 2019). Commonly, Malware forms include viruses, worms, Trojans, bombs, backdoors and spyware.

2.5 Malware Analysis

Basically, malware analysis is classified into 3 kinds:

1. Static Analysis is a malware analysis technique that looks for malicious string or signature in the body for an executable file to determine if the executable is true. Static analytical methodology basically analyzes the binary without the executable folder and is usually used for ranking by famous antivirus detection engines.
2. In general dynamic analysis is examining suspected binary by executing it. According to Sikorski and Honig (2012), any examination performed after executing malware is called dynamic analysis. Typically dynamic analysis is performed after conducts basic static analysis and it has reached a dead end. Reason for failed static analysis such as obfuscation, packing, or the analyst have exhausted the available static analysis techniques.
3. CCode analysis is also generally classified into static analysis. Code analysis is basically a method of malware assessment that utilizes a reverse engineering technique to determine whether a specific executable file is considered malicious.

Threat Intelligence (TI), is a knowledge of threats which inform decisions for attacking prevention or reduce the gap between compromise and detection, based on evidence (MCMillan, 2013). TI commonly known as cyber threat intelligence is might kind of data that serves to illuminate the risk landscape rather than supporting particular choices (Chismon and Ruks, 2015). Such as in Dalziel (2014); Steele (2014), there occur other terms. A more strict one (Dalziel, 2014) says that TI should be relevant, actionable and valuable. IT can be gathered from a number of technical sources (e.g. local sensor traffic) or from human sources (e.g. discussions in subterranean forums and peer

communication). Thus, threat intelligence encompasses technical indicators, contexts, mechanisms, consequences and actionable advice concerning a threat existing or emerging.

Threat Intelligence is subdivided into some subdivision:

- 1 Strategic threat intelligence
- 2 Operational threat intelligence
- 3 Tactical threat intelligence
- 4 Technical threat intelligence (TTI)

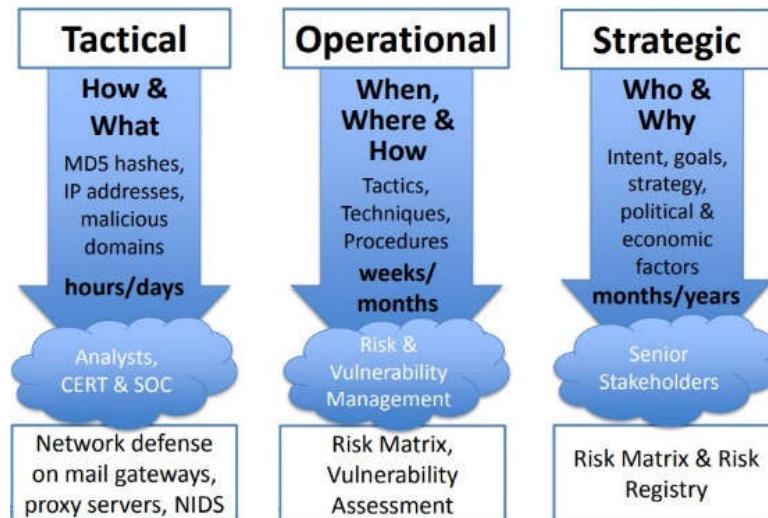


Figure 3 Type of Threat Intelligence (Irving, 2016)

2.6 Pyramid of Pain

The pain pyramid is an significant and elegant idea that can also be used in threat intelligence and threat hunting. The pyramid examines how hard certain attributes of their assault can alter for attackers. It also demonstrates how hard these features are for organisations to discover these attributes. Finding a file with some hash value is simple, but it is completely difficult to uncover illegitimate use in an organisation where PowerShell is frequently used. It is also trivial for assailants to build a fresh file with an other hash, but it is much difficult to relocate or alter the attacker detection method (Bianco, 2014) as shown on Figure 4.

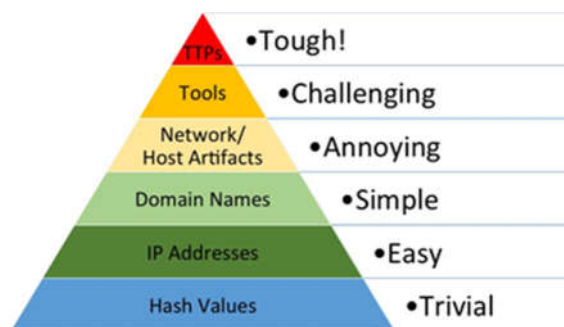


Figure 4 Pyramid of Pain

2.7 Hunting Maturity

The hunting maturity model (HMM) released by Bianco (2015) is one of the most quoted resources in the hunting section. This maturity model focuses on information collection, methods, the development of hypotheses, hypothesis testing instruments and methods, model detection and TTP automation. de Volksbank and Rabobank (2018) combined the Pyramid of Pain an Hunting Maturity Model then called it TaHiTI methodology. TaHiTI methodology provides an overview of where each stage of hunting maturity operates in the pyramid and how TaHiTI should be placed in the HMM. The main focus TaHiTI

methodology is the top 3 layers that also known as High Level IoCs. Then the other layers also known as Low Level IoCs.

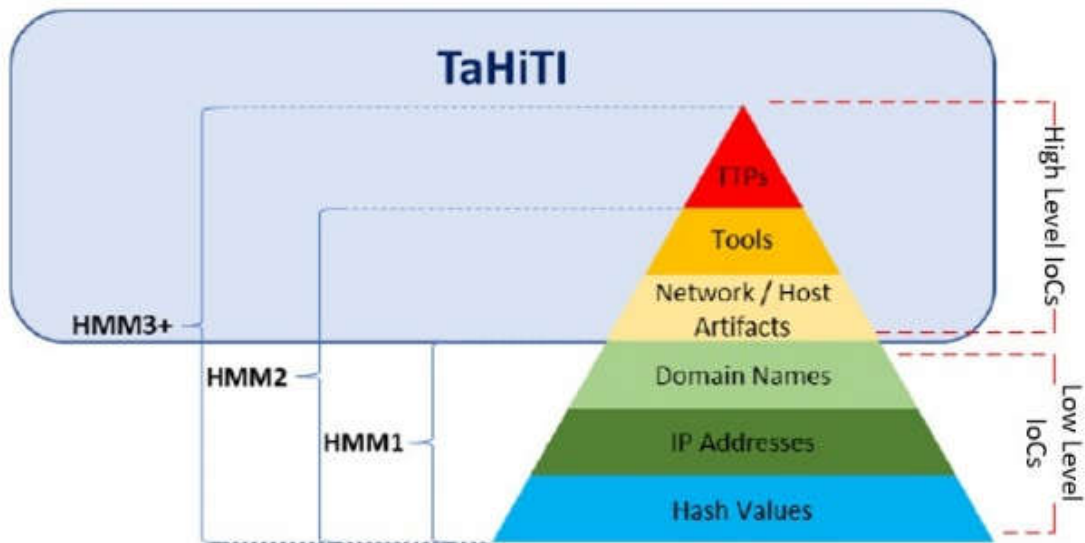


Figure 5 TaHiTI methodology

2.8 Indicator of Compromise

To understand better about Indicator of Compromise (IOC), one should get to know on how current Intrusion Prevention System (IPS) and Intrusion Detection System (IDS) can justify a cyber activity to be an attempt of attack launched by the adversaries. IDS and IPS share the same methods on how they can detect an activity, their detection mechanism is by parsing through previously collected data so that in the end they can generate alert based on currently flowing data compared to the collected data (Sanders To understand better about Indicator of Compromise (IOC), one should get to know on how current Intrusion Prevention System (IPS) and Intrusion Detection System (IDS) can justify a cyber activity to be an attempt of attack launched by the adversaries. IDS and IPS share the same methods on how they can detect an activity, their detection mechanism is by parsing through previously collected data so that in the end they can generate alert based on currently flowing data compared to the collected data (Sanders and Smith, 2014). Indicator of Compromise as also stated by (Pirc et al., 2016) tagged by the remnants of cyber activity that should be recognized on a network or host of an intrusion event. Indicators of Compromise leaving behind the footprints that can lead to exfiltration point of the adversaries, such as ip address, URL and file hash value.

2.9 Malware Information Sharing Platform (MISP)

The MISP threat sharing platform is a free and open source software that helps to share data about threat intelligence including cyber security indicators. A Threat Intelligence Platform for collecting, exchanging, distributing and correlating Targeted Attack Compromise Indicators, Threat Intelligence, Financial Fraud Information, Vulnerability Information or even Counter-Terrorism Information

2.10 Research Framework

This research propose Research Framework, divided into three phases of experiments:

- Data Collection, collect malware samples that related to Financial Sector.
- Analysis with Malware Analysis and use the result as input for Cyber Threat Intelligence Sharing Platform
- Creating model based information of Indicator of Compromised.
- Evaluation (Comparison) with an Cyber Threat Intelligence Sharing Platform or research in the past
- Consume and Share the result (Conclusion and Findings).

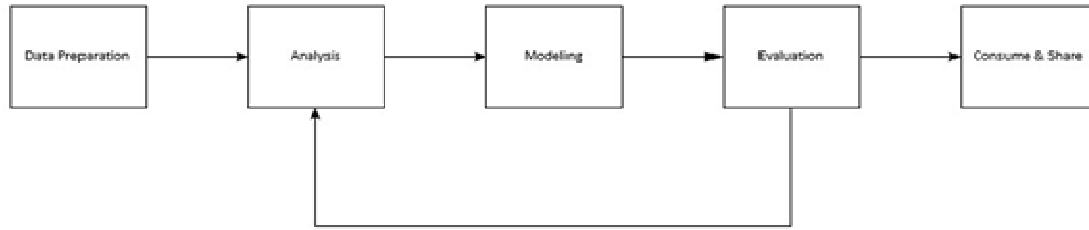


Figure 6 Research Framework

On executing their attack, Malware need some vulnerable part as target and some of them also do some evasion technique to avoid detection by Antivirus, Firewall, Sand- box etc. Based on Miller et al. (2017)’s research, there are some hardened configuration should be configured on Cuckoo Guest to make Cuckoo running well in analyzing malware activity.

(a) Adding Documents in some directory

- In "My Documents" add 5 JPGs, 1 txt, 5 PDFs, and 3 data files.
- In "My Music" add 3 MP3s.
- In "My Pictures" add 6 JPGs and 1 GIF.
- In "My Videos" add 4 MP4s.

(b) Installing New Programs

- Firefox 38.0.5.
- Notepad++ v7.
- VLC 2.2.4.
- 7-Zip 16.02.

(c) Get Recent Documents/Programs by open some documents and programs

- Open all the added documents in multiple times.
- Run each program in multiple times.
- Running some Programs.
- Open Windows explorer.
- Open Notepad.
- Disable all update services for new software.

2.11 System Overview

Proposed the system will use cuckoo as Malware Analysis and feed the result to Malware Information Sharing Platform (MISP)

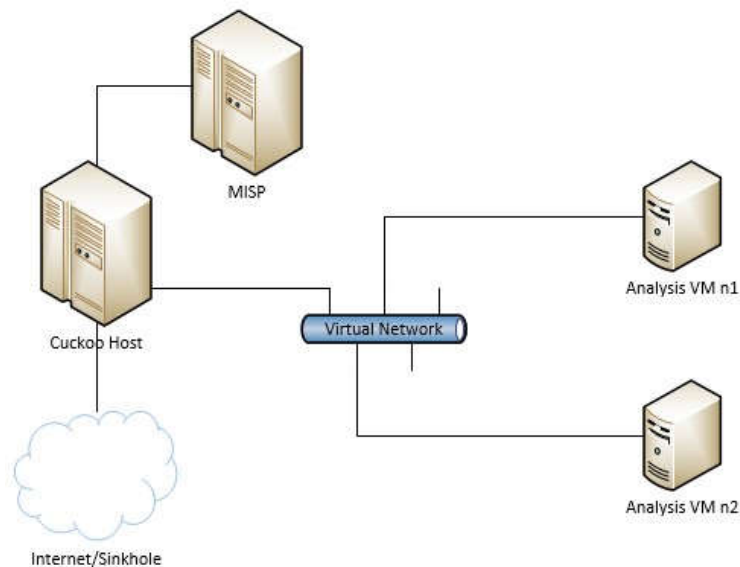


Figure 7 System Overview**2.12 Research Methodology Framework****2.12.1 Data Preparation**

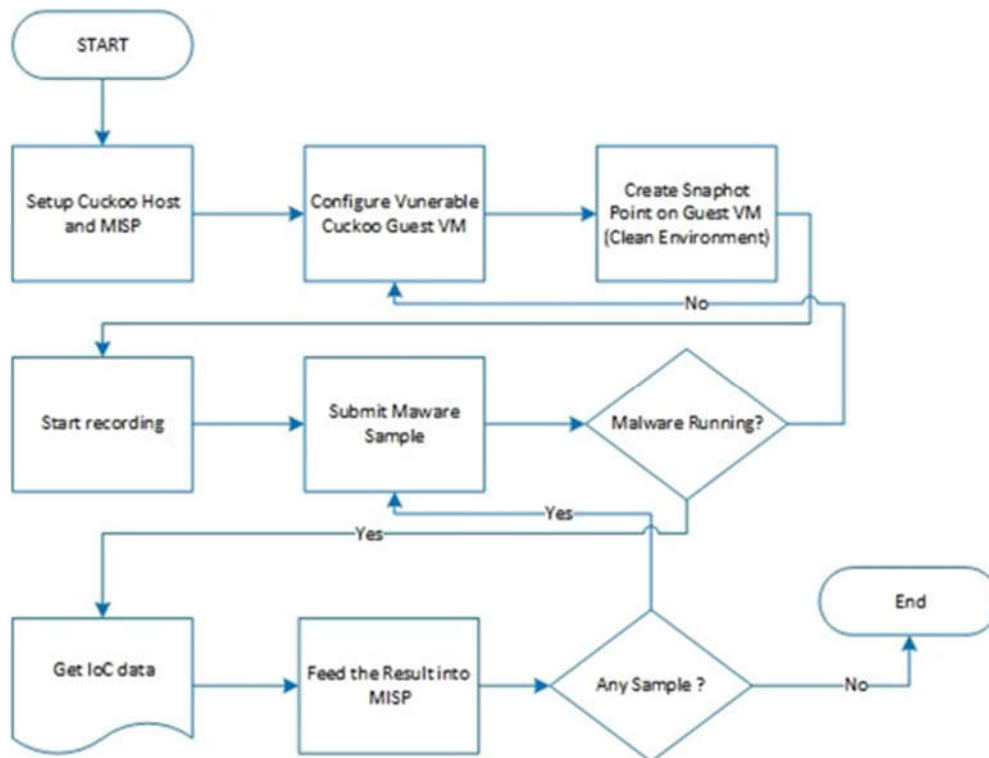
Dataset in this experiment utilize the last five years malware samples which attack Financial Sector during 2014 - 2019. The malware samples are collected from malware repositories such as VirusTotal, VirusShare etc. consist of 2 banking malware families such as Emotet, and Trickbot with detail:

- 30 Emotet samples,
- 29 Trickbot samples

Those Malware samples will be submitted into Cuckoo Sandbox to analyze malicious activity and get the report. By using automation API, Cuckoo feed the IoC data based on previous report into MISP. Each report of malware analysis on Cuckoo will be fed into MISP as one new event. For identifying the event, MISP give ID number on each event with others colums with details:

- Published: publishing status
- Org : organization who create event
- ID : ID number of event
- Atth : Attributes (IoC data)
- Email : username
- Date : date of inputted data
- Distribution : sharing level

As Operational Threat Intelligence, using concept the Triad Investigation is an approach to classify the data into Low Level IoC and High Level. Low Level IoC is data on intrusion response such as IP Address, Hash, Domain, URL etc. On investigation part, this work will focus on tactis, techniques, procedures of Mal- ware while executing attack (main focus of TaHITI Methodolgy). Gathering malware samples is the next step after get the list and put them into dataset which the researcher analyze use Cuckoo Sandbox as Dynamic Malware Analysis tool.

2.12.2 Analysis**Figure 8** Phase 2- Analysis

The analysis is done following the flowchart given in Figure 8:

- Before the malware analysis can be started, Cuckoo and MISP are setup by configuring vulnerable cuckoo using dataset collected in previous phase.
- After that snapshot on Cuckoo Guest VM is created as clean environment which will be reused to simulate the condition before malwares execute the attack.
- While malwares are submitted to Cuckoo Sandbox, the recording process of malware activity starts. When the malware activity run completely, cuckoo will generate the report and feed it to MISP.
- In the next step, with another sample the process starts again on the start recording step until all data collected in Data Preparation phase is tested.

2.12.3 Modeling

In this phase, scalable data model that is provided by MISP is used. It is a collection all types and features for an IoC. MISP also support import and export others data model such as STIX, OpenIoC etc for presenting IoCs informartion.

2.12.4 Evaluation



Figure 9 The Investigation Triad (Alazab et al., 2012)

This research will utilize the investigation triad for for classifying IoC data in MISP into Low Level IoC and High Level IoC.

The triad is composed of the following parts:

a. Intrusion Response

In Intrusion Response part, this research use Cuckoo Sandbox by analyzing malware samples to get data about IP Address, Port, Domain, Hash and futher called low level threat IoC. After get data, low level threat IoC information will be fed into MISP by utilizing automation API MISP for Cuckoo Sandbox. MISP will provide automatic correlating data of IoC for futhering analysis

b. Investigations

In Investigations part, this research work will do deep analysis about tactics, tech- niques and procedures (TTPs) futher called high level threat IoC. Based on result of malware analysis by using Cuckoo Sandbox, the researcher will enrich the existing data (low level threat IoCs) with high level threat IoCs information by mapping them into att&ck mitre manually that MISP provide.

c. Vulnerability

In Vulnerability part, The vulnerabilities information will be added in CVE id for- mat on MISP. After all the data are collected consume and share it to financial sector and move another analysis with new target analysis

2.12.5 Validation

In this research work, the results of Malware Analysis is the most important result because it is the input data for MISP. So, the result of malware analysis has to be valid. Comparing the result on our malware analysis tool with other existing malware analysis such Cape, Hybrid-analysis, malwareanalysis etc will be used as validation method.

3. Results and Discussion

We begin our method by collection respected dataset. dataset will be used as our training dataset before our framework test with malware. After collecting our training dataset, we then run the dataset in our emulation machine. Using Cuckoo Sandbox to analyze malware samples for getting data about IP Address, Port, Domain, Hash and further called low level threat IoC. After getting the data, low level threat IoC information will be fed into MISP by utilizing automation API MISP for Cuckoo Sandbox. Hash Value of Malware. As a tool for sharing threat information, MISP can provide automatic correlation data of IoC for further analysis. The correlation event of IoCs is shown in Table 1.

Table 1: Low Level IoC:URL

No	URL	Correlation Found
1	A	111
2	B	47
3	C	47
4	D	47
5	E	4
6	F	47
7	G	17
8	H	137

Further deep analysis about tactics, techniques and procedures (TTPs) is called high level threat IoC. Based on result of malware analysis by using Cuckoo Sandbox, the researcher will enrich the existing data (low level threat IoCs) with high level threat IoCs information by mapping them manually into att&ck mitre, that is provided by MISP.

Table 3: Low Level IoC:IP Address

No	Destination IP Address	Correlation Found
1	118.98.93.11	20
2	172.217.24.110	138
3	74.125.68.94	138
4	74.125.24.95	138
5	74.125.24.94	138
6	74.125.24.138	138
7	74.125.130.95	138
8	74.125.130.94	138
9	65.52.172.55	63
10	216.58.221.78	138
11	20.189.72.203	88
12	172.217.27.14	138
13	172.217.192.94	138
14	119.110.118.207	139
15	172.217.194.155	138
16	172.217.194.132	138
17	172.217.160.42	18
18	172.217.160.36	138
19	172.217.160.13	138
20	1157.240.13.6	138

No	Destination IP Address	Correlation Found
21	157.240.13.36	138
22	157.240.13.35	139
23	157.240.13.19	139
24	13.107.4.50	76

4. Conclusion

Based on the experimental results, several conclusions can be drawn:

- 1) Using Sharing Level community only on MISP is a fit Cyber Threat Intelligence Sharing model for the financial sector because they want information restricted only to financial sector companies
- 2) Our proposed system successfully is suitable for analyzing Emotet and Trickbot.

References

- Alazab, M., Venkatraman, S., Watters, P., Alazab, M., and Alazab, A., "Cybercrime: the case of obfuscated malware," in "Global Security, Safety and Sustainability & e- Democracy," pp. 204–211, Springer, 2012.
- Bianco, D., "The Pyramid of Pain," 2014, URL <http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html>.
- Bianco, D., "A Simple Hunting Maturity Model," 2015, URL <http://detect-respond.blogspot.com/2015/10/a-simple-hunting-maturity-model.html>.
- Blunden, B., *The Rootkit arsenal: Escape and evasion in the dark corners of the system*, Jones & Bartlett Publishers, 2012.
- Cichonski, P., Millar, T., Grance, T., and Scarfone, K., "Computer Security Incident Handling Guide : Recommendations of the National Institute of Standards and Technology," Technical Report NIST SP 800-61r2, National Institute of Standards and Technology, 2012, URL <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf>.
- de Volksbank, R. v. O. and Rabobank, M. B., "DEF-TaHiTI- Threat-Hunting-Methodology.pdf," FI-ISAC NL Publication, 2018, URL <https://www.betaalvereniging.nl/wp-content/uploads/DEF-TaHiTI-Threat-Hunting-Methodology.pdf>
- ENISA, "ENISA Threat Landscape Report 2018," , 2018, URL <https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2018>.
- Fernandez, G., Nieto, A., and Lopez, J., "Modeling Malware-driven Honeypots," in Lopez, J., Fischer-Hübner, S., and Lambrinouidakis, C. (editors), "Trust, Privacy and Security in Digital Business," volume 10442, pp. 130–144, Cham: Springer International Publishing, 2017, URL http://link.springer.com/10.1007/978-3-319-64483-7_9.
- FireEye, "M-Trends 2019," , 2019, URL content.fireeye.com.
- Irving, R., "irving-threat-intelligence-sharing.pdf," 2016, URL <https://www.first.org/resources/papers/munich2016/irving-threat-intelligence-sharing.pdf>.
- Kime, B., "Cyber Threat Intelligence Support to Incident Handling," p. 31, 2017. Meier, R., Scherrer, C., Gugelmann, D., Lenders, V., and Vanbever, L., "FeedRank: A tamper-resistant method for the ranking of cyber threat intelligence feeds," in "2018 10th International Conference on Cyber Conflict (CyCon)," pp. 321–344, Tallinn: IEEE, 2018, URL <https://ieeexplore.ieee.org/document/8405024/>.
- Miller, C., Glendowne, D., Cook, H., Thomas, D., Lanclos, C., and Pape, P., "Insights gained from constructing a large scale dynamic analysis platform," *Digital Investigation*, volume 22 pp. S48–S56, 2017, URL <https://linkinghub.elsevier.com/retrieve/pii/S1742287617301949>.
- Moser, A., Kruegel, C., and Kirda, E., "Limits of static analysis for malware detection," in "Computer security applications conference, 2007. ACSAC 2007. Twenty-third annual," pp. 421–430, IEEE, 2007.

- Noor, U., Anwar, Z., Amjad, T., and Choo, K.-K. R., "A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise," *Future Generation Computer Systems*, volume 96 pp. 227–242, 2019, URL <https://linkinghub.elsevier.com/retrieve/pii/S0167739X18326141>.
- Roberto and Rodriguez, J., "How to implement and use the MITRE ATT&CK framework – TOP CYBER NEWS," , 2019, URL <https://www.topcybernews.com/how-to-implement-and-use-the-mitre-attck-framework>
- Sikorski, M. and Honig, A., *Practical malware analysis: the hands-on guide to dissecting malicious software*, no starch press, 2012.
- SO, I., "Automated Cyber Threat Intelligence Sharing," 2019, URL <https://www.isao.org/storage/2019/04/ISAO-300-2-Automated-Cyber-Threat-intelligence-Sharing.pdf>.
- Verma, A., Rao, M., Gupta, A., Jeberson, W., and Singh, V., "A literature review on malware and its analysis," *International Journal of Current Research and Review*, volume 5(16) pp. 71–82, 2013.
- Wagner, C., Dulaunoy, A., Wagener, G., and Iklody, A., "MISP: The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform," in "Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security - WISCS'16," pp. 49–56, Vienna, Austria: ACM Press, 2016, URL <http://dl.acm.org/citation.cfm?doid=2994539.2994542>.
- Wang, J., *Computer network security: theory and practice*, Beijing: Higher Education Press, 2009, oCLC: 845386845.
- Ye, Y., Li, T., Adjeroh, D., and Iyengar, S. S., "A survey on malware detection using data mining techniques," *ACM Computing Surveys (CSUR)*, volume 50(3) p. 41, 2017.
- Zimmerman, C., "Ten Strategies of a World-Class Cybersecurity Operations Center", MITRE Corporation, 2014

Methodology of Security testing of IKID website and Security Vulnerabilities

Mustofa Kamil^{1*}

¹Master of Information and Technology, Swiss German University, Tangerang 15143, Indonesia

*Corresponding author: mustofa1981@gmail.com

Abstract. Due to the large amount of data stored in web applications and the increasing number of transactions on the web, the right Web Application Security Testing is very important day by day and web application is an important in business life. By increasing complexity of web systems, Security testing has become a very necessary and important activity of the life cycle of developing web applications, web security testing consists of searching for information about the network, application and looking for holes and weakness.

Keyword: Web Security, Parameters of web security, and Security attack, security vulnerabilities

1. Introduction

Due to customer data going online by adapting to online, and GDPR must be compliant for online business, security web testing must be performed before going live. A web security test to search security databases for known vulnerabilities against the system component as summaries below

- Perform network scans to identify any unexpected and/or unauthorized services and ports are in use.
- Perform targeted scan on in scope Network devices and system to identify vulnerabilities
- Conduct a manual investigation to validate the vulnerability
- Vulnerability ranking is based on the level of threat, potential loss and possible exploitation and the availability of exploitation code.

2. Materials and Methods

There are several testing methods available:

- a. Black box Testing – The test is carried out without prior knowledge about the internal structure or design or implementation of the object being tested.
- b. Grey Box testing – Testing must be done with partial knowledge of the internal structure or design or implementation of the object being tested.
- c. White Box Testing - Testing must be done with knowledge of the internal structure or design or implementation of the object being tested.

Testing activities are divided into three categories: Non-invasive, less invasive, and more invasive. These categories are organized into four stages each of which is based on NIST SP800-115 defined Penetration Testing Methodology.

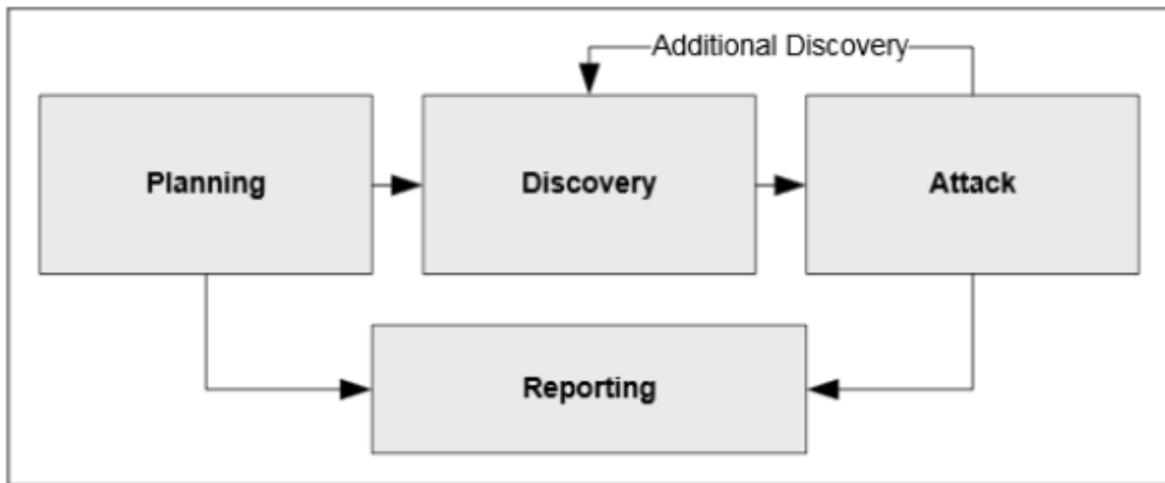


Figure 1 Four-Stage Penetration Testing Methodology

As can be seen in Figure 1, there are several stages in penetration testing methodology.

a. Stage 1 Planning

This stage establishes the basis for successful penetration tests, such as scheduling and resources, where actual testing takes place

b. Stage 2.1 Discovery (Identify Targets and Network mapping)

This penetration test phase involves counting systems in scope and mapping networks and systems from an external perspective to find systems in the target environment, which basically maps the surface of the attack for the remaining phases. Even in a Grey box testing (in which target system personnel provide the testers with a diagram and inventory of target machines), still applies this phase to verify the information provided at the outset of the test. To discover potential target machines. It applies numerous methods and usually do not direct the target environment to abnormal traffic.

- Identify potential Vulnerabilities via Web Research and/or Network Discovery tools.
Information that may be collected from the internet could allow testers to build a profile of how the target environment shows to attackers.
- Network Mapping: This step allows identifying what services are being listened to at the target IP address, previous networks and even hosts behind a firewall. A network diagram shall be built during this step and the asset register. All services identified must be targeted at the next steps.

c. Stage 2.2 Discovery (Vulnerability scanning)

This stage can be divided into two parts, Application level and Network level scanning, and web application level scanning. Application level and Network level scans could be performed to recognize network services that might be vulnerable to attack. All vulnerabilities identified in this stage must be validated by thorough research. Vulnerability scanning suffers from false positives and several vulnerabilities may not apply to the target system. Vulnerabilities must include but are not limited to buffer overflows, network services that are not configured correctly, trust relationships that are not configured correctly, authentication mechanisms that are not safe and outdated network services that are aware of vulnerabilities & etc. This activity does not usually lead to a regularly scanned environment. There are no scanning activities specifically designed to cause disruption to the environment, such as scanning for denial of service vulnerabilities.

d. Stage 3 Attack (Exploitation and password guessing)

This stage forms an interesting part of the penetration test. This is where the testes stand the chance of gaining a footprint on the environment and beating the protection mechanisms. Exploitation can mostly be carried out with certain exploit tools. A few websites can be used to find exploit code.

Manually testing will be conducted via searching vulnerable databases. Password can be conducted with a list of common passwords and potential usernames. These activities while not likely to cause disruption to environment are a more invasive that previous stages.

e. Stage 4 Reporting.

Every item identified, as a problem that may occur, must be followed by exploits that can be proven or not. Problems recognized in one stage must be registered at the end of the stage and carried out for further investigation at the next stage.

Assessment Tools

Following tools will be using for the web security assessment whereas applicable, and they are namely

- Nessus Security Scanner
- Nmap
- Kali Linux
- Burp Suite
- Acunetix Web Vulnerability scanner.

Testing Scenario

This research will use Web Penetration test scenario. In this scenario, pen tester-initiated attacks are done from the external network of the being tested in-scoped system components

Constraints and Assumptions

Penetration testing activities will consist of target identification, network mapping, vulnerability scanning, exploitation of vulnerabilities and password guessing. While most of these activities do not usually have an impact on such environmental operations, stages that involve exploitation of a vulnerability can cause system disruption. There are no tests specifically designed to interfere with computer system operations to be performed. All activities will be carried out in a maintenance window agreed with the operating staff and supporters responsible for the environment as listed above in the Schedule section. All relevant and appropriate penetration testing activities will be informed as they are undertaken and when they have finished. This penetration test is not intended to steal sensitive information and data from the environment. Where access to sensitive information is believed to have been obtained, testing will be terminated, and the vulnerability confirmed with the system owner. Any sensitive information that is accidentally taken by a penetration tester will be safely removed or disguised for reporting purposes.

3. Results and Discussion

Detailed results of open network ports / services identified for the target scope and origin perspective will be listed in Table 1 below.

Vulnerability of cross-site request (CSRF) is a vulnerability might arise when an application relies only on attaching link email or text to identify the user who has issued a specific request. Because the browser automatically adds a cookie to the request regardless of origin, attackers can create malicious websites that fake cross domain requests to vulnerable applications. In order to be vulnerable from CSRF, the next conditions must apply.

Requests can be issued across domains, such as using HTML forms. If the demand contains a header or nonstandard contents, then the demand can only be deleted from pages that are from the identical domain.

This application only relies on http cookies or Basic Authentication to recognize the user who issued the request. If the application displays a section related token using MFA (multifactor authentication) elsewhere in the request, then it might not be sensitive.

- The demand performs several specific actions in the application, which change the application status based on the user's identity issued.
- An attacker can specify all parameters needed to build requests that perform actions. If the demand includes a value that can't be determined or predicted by the attacker, then it is not vulnerable.

Table 1 Vulnerabilities scan result

Findings	Vulnerability name	Risk Level	CVSS Score	Vulnerability Description/Risk
TLSv1 service port is enabled	TLSv1.0 Protocol Detection	High	7	TLSv1 is not longer used anymore, affected software and version of remote code execution.
TRACE /Nessus1046189052.html HTTP/1.1 Connection: Keep-Alive Pragma: no-cache User factor: Mozilla version 4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0) Receive: image/gif, image/x-xbitmap, image/jpeg, image/pjpeg, image/png, */* Receive-Language: En Receive-Charset:iso-8859-1,*,utf-8	HTTP TRACK Methods is active or allowed	Medium	5	Remote web server supports TRACK methods. TRACE and TRACK are HTTP methods used to debug web server connections.
Client-to-server Cipher Block Chaining algorithm. They Are supported: 3des-cbc aes128-cbc aes192-cbc aes256-cbc	SSH Server CBC Mode Ciphers Enabled	low	2	The SSH server is configured to support Cipher Block Chaining (CBC) encryption. This allows the attacker to recover the plaintext message from the ciphertext.

Findings	Vulnerability name	Risk Level	CVSS Score	Vulnerability Description/Risk
blowfish-cbc cast128-cbc Server-to-client Chaining Block Chaining Algorithm They are supported: 3des-cbc aes128-cbc aes192-cbc aes256-cbc blowfish-cbc cast128-cbc on port 22				

CSRF is a one form of penetration techniques in website security gaps. and This technique uses a method for entering data user of a site. with this technique is certainly very detrimental website application users because they can steal personal data, sensitive data exist in an application system data accessed by an attacker. To secure from against Cross-Site Request Forgery vulnerabilities is to include in the request relevant additional tokens that are not sent in the cookie: such as parameters in hidden form fields. where tokens are used to generate random cryptographic keys for sending data securely, so it's not possible for an attacker to assign or set any token value issued to other users. The MFA token shall be opened with a user session, and the application must ratify that the valid token must be received before taking any action that results from the request. Use random validation tokens because Different token values are required for each data delivery form. An optional suggestion, it will be easier to implement, to validate that the Host and referrer Header in the relevant request is present and uses similar host/domain name. Nevertheless, using this is somewhat less powerful: historically, the habits in browsers and add-ons that often require attackers to request cross domain counterfeiting that make use of this header to add such defences. And for vulnerabilities Disabling the use of the TLSv1.0 protocol for cryptographically powerful protocols such as TLSv1.2.

4. Conclusion

There are total ZERO (1) High (H) risk, ONE (2) Medium (M) risk and ZERO (1) Low (L) risk level findings have been identified.

Table 2 Result of Web Application Penetration Test

Web Application penetration test		
Risk Level	Exploitable (suspected) (Yes/No)	No. findings of penetration Test
High	0	0
Medium	Yes	1
Low	0	0
AOI	0	0

A Design Level vulnerability or configuration flaw that makes the network or host network vulnerable to malicious attacks from internal/local or remote users. Vulnerabilities may stand in some areas of the network, such as on firewalls, file transfer protocol servers/file share server or Web servers, and operating systems, depending on the level of security risk. Successful exploitation of the vulnerability may vary from the disclosure of information about host to compromise the host.

Table 3 Severity Level






Severity	Level	Description
 1	Minimal	The intruder can collect information about the host (open ports, services, and so on) and can use this information to find other vulnerabilities.
 2	Medium	The intruder might be able to collect sensitive data or personal information from the host, such as the right software version. With this information, intruders can easily exploit certain software versions of known vulnerabilities.
 3	Serious	Intruders may gain access to storage of information specific to the host and include security settings. Vulnerabilities at this level can include disclosure of a portion of the contents of the file, access to certain files on the host. Directory searches, disclosure of filtering rules or security mechanisms, and denial of service attacks or unauthorized use of services, such as sending emails.
 4	Critical	The intruder might be able to control the host. For example, vulnerabilities at this level can include full read access to files, and potential backdoors in the system or network. And a list of all users on the system and at the host.
 5	Urgent	The intruder can easily gain control of the system or host network, which can cause compromise of the entire network security. Such vulnerabilities at this level can include full access read and write access.

Table 4 Category in Severity Level

Severity	Level	Description
■ LOW	Low	Vulnerability with a CVSS predicate score of 0.0 TO 3.9
■ MED	Medium	Vulnerability with a CVSS predicate score of 4.0 TO 6.9
■ HIGH	High	Vulnerability with a CVSS predicate score of 7.0 TO 10.0

Table 5 Vulnerabilities website score

Vulnerabilities component	Severity Level	CVSS Score
Cross-Site Request Forgery	■ MED	5

Table 6 Vulnerabilities Total

Row Labels	Sum of Avg. CVSS Score	Sum of Total Finding
HIGH	7	1
LOW	2	1
MEDIUM	5	2
Grand Total	14	4

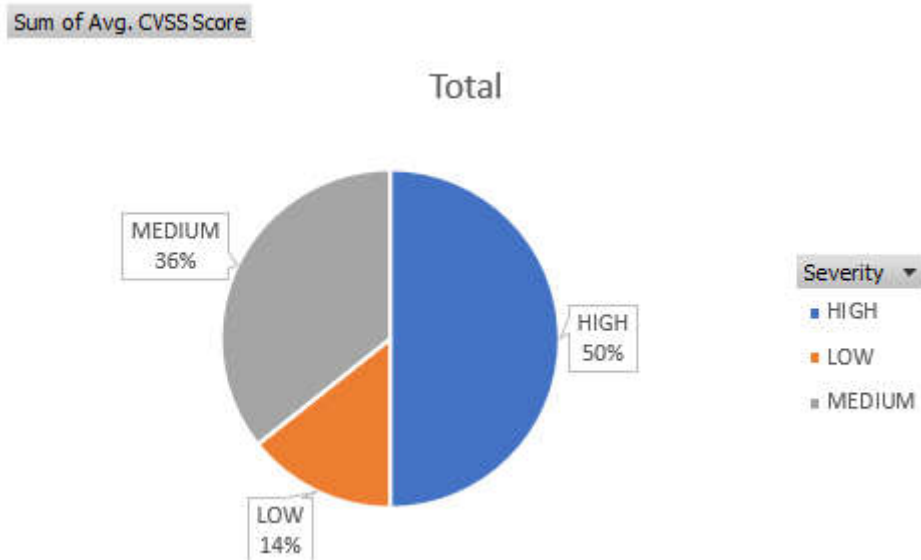


Figure 2 Percentage of CVSS Score

All loopholes are dangerous, whether it's a gap on the website that arises from coding errors. Awareness of security / website security is indeed needed by everyone. Both beginners and experts must always be aware and up to date on the latest malware.

References

- CVSS score. Available at: <https://nvd.nist.gov/vuln-metrics/cvss/v3-calculator>
- Jaiswal, A., Raj, G., and Singh, D. (2014). Security Testing of Web Applications: Issues and Challenges. *International Journal of Computer Applications*. 88. doi: 10.5120/15334-3667.
- Pangalita, R., Noertjahyana, A., Andjarwirawan, J. Program Studi Teknik Informatika, Fakultas Teknologi Industri, Universitas Kristen Petra Available at: <https://media.neliti.com/media/publications/105546-ID-penetration-testing-server-sistem-inform.pdf>.
- Syarifudin, I. Pen testing and analisis keamanan web PAUD DIKMAS. Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta, Depok, Indonesia. at: https://www.researchgate.net/publication/324182514_Jurnal_Irwan_Syarifudin_Pentesting_dan_Analisis_Keamanan_Web_Paud_Dikmas.
- NIST SP800-115 Available at: <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-115.pdf>



SWISS GERMAN UNIVERSITY

The Prominence Tower
Jl. Jalur Sutera Barat No. 15,
Alam Sutera, Tangerang
15143 Indonesia
E-mail: marketing@sgu.ac.id