ANOMALY DETECTION FOR CONTROLLING DATA ACCURACY IN SERVICE INDUSTRY

NURUL ASYIKIN SAMSUDDIN

A project report submitted in partial fulfillment of the
Requirement for the award of the degree of
Master of Industrial Engineering

Faculty of Mechanical Engineering
Universiti Teknologi Malaysia

June 2013

ABSTRAK

Projek in dilakukan bertujuan untuk mengkaji penggunaan pengesan anomali, terutamanya carta kawalan untuk sampel individu, untuk mengawal kualiti data yang dijana oleh sistem pengurusan risiko di dalam industri kewangan. Empat carta kawalan dikaji iaitu carta kawalan individu, carta kawalan jarak bergerak (MR), carta kawalan purata bergerak (MA) dan carta kawalan purata begerak dengan wajar eksponen (EWMA). Kuantitatif dan kualitatif prestasi keempat-empat carta kawalan ini dikaji untuk dua keadaan: aliran lansung dan kajian data. Keputusan kajian dibandingkan dengan anomali yang telah dikenalpasti oleh pakar sistem. Melalui kajian yang dilakukan, carta kawalan individu merupakan pengesan terbaik untuk keadaan aliran lansung dan carta kawalan MR merupakan pengesan terbaik untuk keadaan kajian data. Secara kualitatif, carta kawalan merupakan pengesan yang ringkas, mesra pengguna dan mudah diautomasi sepenuhnya dan dibawa ke produksi berbanding pengesan anomaly yang terdapat di dalam kajian-kajian lain. Di sampling itu, satu kualiti data jaminan dan kawalan dicadangkan berdasarkan keputusan kajian ini.

ABSTRACT

The purpose of this project is to investigate the application of anomaly detection, particularly control charts for individual sample, to control data quality of a risk management system in a financial industry. Four control charts are investigated, namely individual control chart, moving range (MR) control chart, moving average (MA) control chart and exponentially weighted moving average (EWMA) control chart. The quantitative and qualitative detection performance of these control charts is analyzed on two scenarios: live stream and data profiling. Results are compared with expected anomalies determined by system experts. It is discovered that individual control chart performed best for live stream scenario, while MR control chart performed best for data profiling scenario. Qualitatively control charts are simple, user-friendly and easy to fully automate and implement when compared with other detection methods available in literature. In addition, a suitable data quality assurance and control program using the two control charts is suggested.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

ACWP	Actual Cost of Work Performed
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BCBS	Basel Committee on Banking Supervision
BCWP	Budgeted Cost of Work Performed
BCWS	Budgeted Cost of Work Scheduled
BIRCH	Balanced Iterative Reducing and Clustering using Hierarchies
BNM	Bank Negara Malaysia
CBB	Contract Cost Base
CLARANS	Clustering Algorithm based on Randomized Search
CUSUM	Cumulative Sum
DB	Delta Bank
DBSCAN	Density Based Spatial Clustering of Application with Noise
DQA/C	Data Quality Assurance and Control
EAD	Exposure At Default
EL	Expected Loss
ERP	Enterprise Resource Planning
EWMA	Exponentially Weighted Moving Average
FindCBLOF	Find Clustering Local Based Outliers
FN	False Negative
FP	False Positive
IRB	Internal Rating Based

LCL	Lower Control Limit
LGD	Loss Given Default
MA	Moving Average
MR	Moving Range
NCC	Negotiated Contract Cost
PD	Probability of Default
RCWR	Risk Weighted Capital Ratio
RMS	Risk Management System
RWA	Risk Weighted Asset
SOM	Self-Organizing Maps
SVM	Super Vector Machines
TN	True Negative
TP	True Positive
UCL	Upper Control Limit
UL	Unexpected Loss

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter begins with section 1.2, a background of the project, where readers are introduced with the concept of Basel. Thereafter, section 1.3, section 1.4, section 1.5, section 1.6 define the problem statement, research question, scope and significance of the project, respectively. Subsequently, the organization of this report is presented in section 1.7 and conclusion in section 1.8.

1.2 Background

The background consists of explanatory introduction on Basel accord and its relationship with the financial institution, in section 1.1.1, Basel accord implementation and its relationship with data, in section 1.1.2, and the fundamentals to understand the concept of calculating the regulatory requirement of the Basel accord.

1.2.1 Basel Accord and Financial Institution

The Basel Accord was initially drafted in 1988 to improve the safety and the soundness of the global banking system by enhancing the risk measurement and capital adequacy within banks. The Basel committee on Banking Supervision (BCBS) released the first accord in 1988 to increase the focus on risk mitigation by stipulating a required risk weighted capital ratio (RWCR) of at least 8% of Risk Weighted Assets (RWA) for capital provision [1]. Prior to the Basel Accord the progression of capital level was decreasing from more than 10%-20% in the early 1900s to 4.5%-6% in the 1980s [2][3][7].

In 2004, Basel II was issued to provide more flexibility in calculating the regulatory capital mentioned in Basel I of 1988 [4]. The new version of the accord also increases risk sensitivity by providing a more comprehensive approach to measure and managed risk. It incorporates supervisory review and market discipline, which was not present in Basel I [4][5]. The capital (quantitative) requirement, supervisor review and market discipline becomes the 3 pillars of the Basel II framework, seen in Figure 1.1, covers various risks faced by financial institutions: insurance, market, credit, liquidity and operational and other risks such as legal and reputations risks [4].

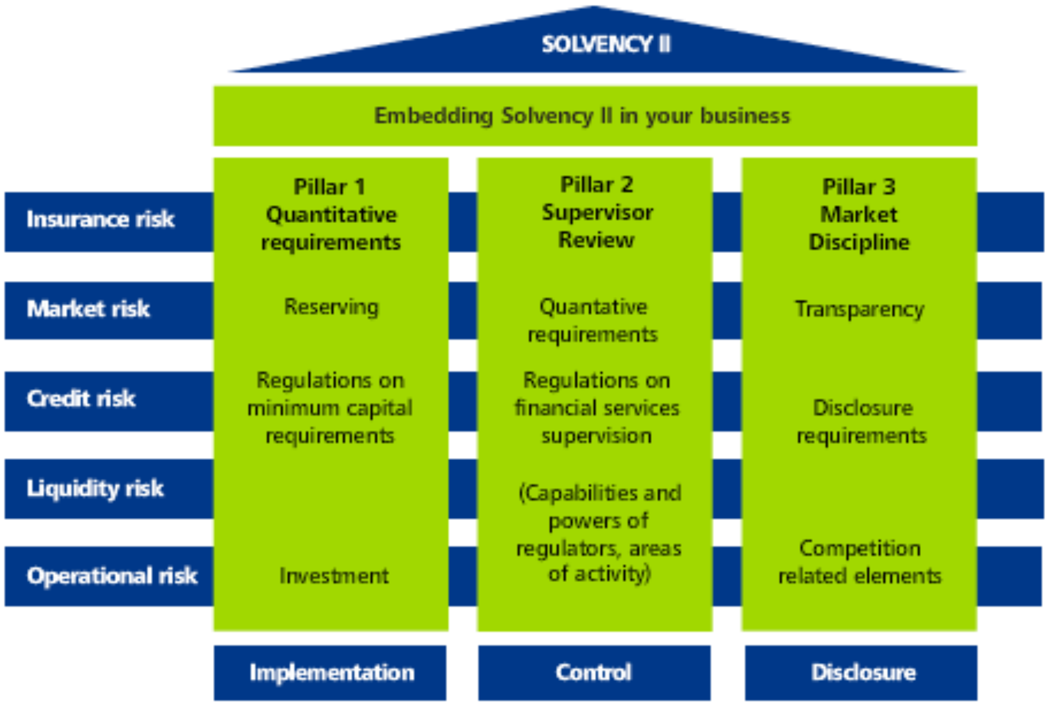

Figure 1.1: Basel II Framework

In the aftermath of the global financial crisis in 2008 and 2009, BCBS proposed Basel III in response. The proposal revises capital standards, stronger capital definition and provided a newer framework for liquidity risk [7][8]. These changes are expected to strengthen the capital requirement, which enables financial institution to absorb unexpected losses arising from financial and economic stress [8][12]. This in turn would reduce the risk of spillover effect of financial sector to a country's economy, which happened in the late 2000s crisis, and increase the resiliency of banks [5][8].

The adoption of the accord is mainly intended for the G10 and G20s [4][8]. However many countries are adopting a version of the Basel accord because it provides best practices in the areas of risk management, risk measurements and capital allocation [4]. As much as 88 non-G10 countries intend or are currently implementing Basel II [12]. In Malaysia, Bank Negara Malaysia (BNM) currently mandates financial institutions to adopt Basel II [9]. However, BNM does consider to progressively adopt Basel III, as shown in Table 1 below, beginning 2013 and to be completed by 2019 [10][11].

Table 1.1: Progressive Adoption of Basel III in Malaysia

Date	Areas of Basel III framework to be implemented
June 2012	Report leverage ratio position calculated based on Basel III regulations to BNM
2012 – 2015	Increase Tier 1 capital ratio from 4% to 6% and common equity Tier-1 capital ration to 4.5%
2015 – 2018	Consider enhancing liquidity coverage ratio and implement stable funding ratio, which banks are expected to meet the 3% leverage level
2016 – 2019	Asserting 2.5% capital conservation buffer over and above the regulatory requirement (BNM also considers to introduce countercyclical capital buffer between 0-2.5% of RWA)
Date has yet to be decided	Asserting additional loss-absorbency requirement ranging from 1% to 3.5% common equity

1.2.2 Data and Basel Accord Implementation

Basel accords are based on sophisticated risk assessments models that pose various technical challenges during implementation both for the banks and banks supervisor. Challenges can vary from the limited expertise, banking culture that is less aware of risks, tight implementation schedule, lack of data availability, less developed risk management systems to high implementation costs [12] [14][17].

However, the degree of these challenges varies across developing and developed countries, as shown in Figure 2 below, because banks in developed countries are more risk-aware and have more sophisticated risk measurement management processes and systems compared to its developing counterparts [4][15] [16][17]. In addition, a bank that adopts the Internal Ratings Based (IRB) approach rather than the standardized approach in order to comply with the Pillar 1 requirements faces greater complexities due to the complicated models and calculation methods it imposed [15].

In implementing the Pillar 1, data constraints are common for banks across the globe. An enhance data infrastructure is required to allow banks to extract required information across the various systems in a bank [17][14]. This also includes having the capability to obtain a single and consolidated view of borrowers or group of borrowers that would enable effective monitoring of borrowers and segmentation of exposures [14] [18]. Accomplishing a single view would require systems to be fully integrated between borrowers, exposures and collateral management systems [18]. These requirements are feat given the complex nature of banking systems, as shown in Figure 1.2 below.

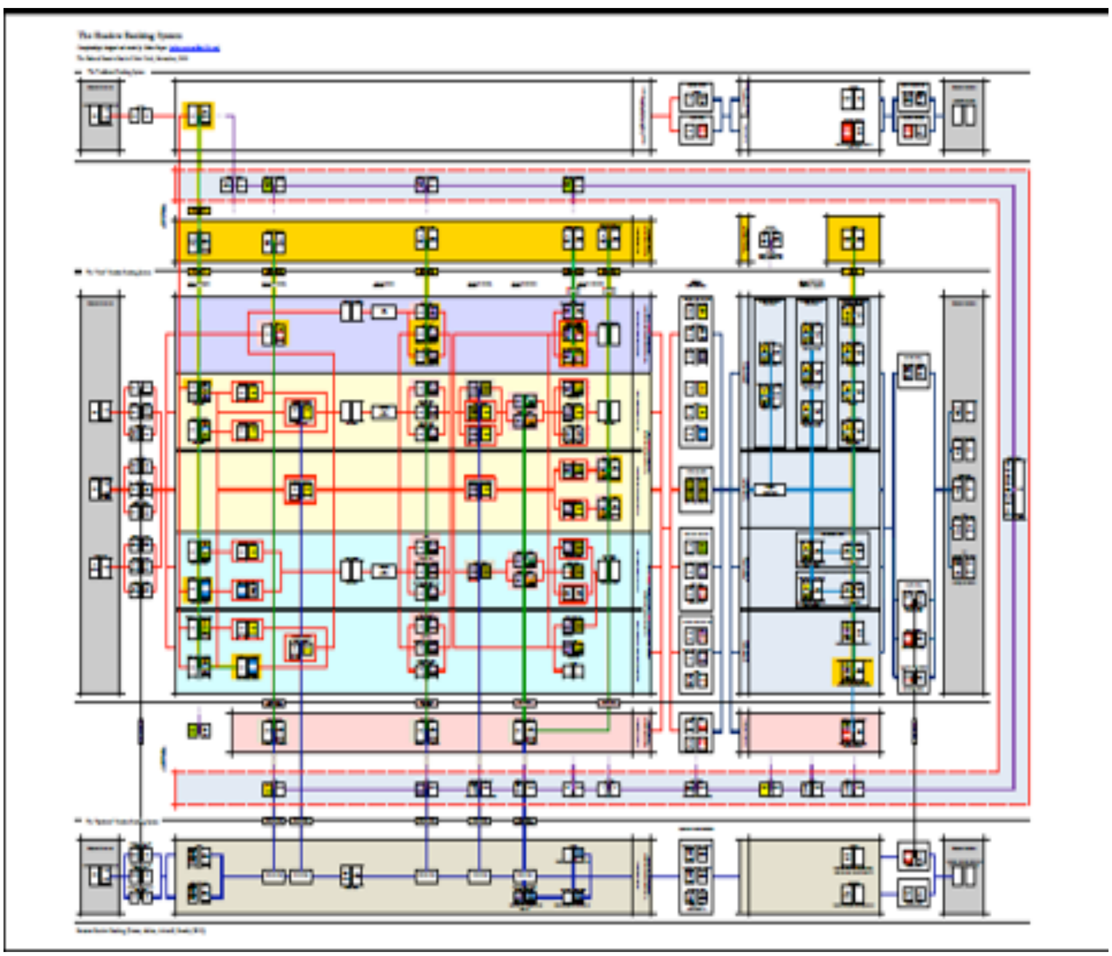

Figure 1.2: The complexity of banking systems [Federal New York Reserve]

In addition to system requirements, various types of data are required and depending on the banks business practices these data might not be available in the banks systems. Data gap analysis addresses this problem by identifying the required information a bank needs to conform to the Basel accord [12]. The types of data that a bank may require include the granular details of borrowers' financial statements, collateral, liquidity and borrowers details [12] [18]. To obtain an adequately robust model for calculations, storing historical data for a full economic cycle data, which is typically a minimum period of 10 years, is also advisable [18].

To ensure the delivery of good model performance, sound data management processes and procedures are critical to ensure data collected are of good quality. A sound data management includes the practice of data governance, metadata, data standardization, reference data unification, data archiving and data stewardship [13] [18]. And a good quality data includes its availability, completeness, accuracy and consistency across the banks information systems [12][13] [19]. Ensuring high data quality and integrity standards that are observed at all times requires clear lines of authority and accountability, which can only be achieved when quality is a business culture acknowledge both by the top management and business units. [18]

1.2.3 Foundation for Capital Requirement in Basel

In financial institutions, capital provisions provide buffers against bank loses, protect creditors in the event banks fails and curbing the culture of excessive risk taking and shirking by shareholders and top managers [3][20]. Banks losses occur constantly, losing capital and interest in the form of borrowers defaulting from obligations. This type of loss is called expected losses (EL) and financial institutions view this loss as a cost component of doing business and are covered by the bank through the use of provisioning and pricing policies. The actual value of EL is not possible known but it can be reasonably estimated based on business experience.

In addition to EL, less frequently, financial institutions experience unexpected losses (UL), which are of higher value than EL, as shown in Figure 1.3. Examples of UL are operations loss due to mishandling of payment operations due to bank mergers, massive speculating trading or lack of internal control. The exact timing of UL occurrence and the severity is not known and when the severity is considerable, it is possible that the market cannot sufficiently cover the losses [20]. In this case capital is required to cover the losses.

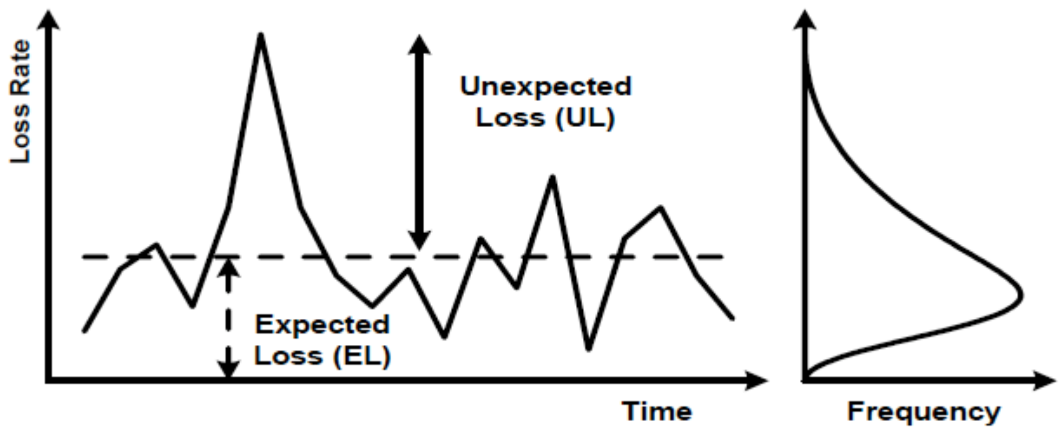

Figure 1.3: The relationship between EL and UL and their frequency of occurrence [20]

Even though holding capital would allow banks to absorb losses, minimizing the capital provision would allow banks to use its capital resources for profitable investments. This conflicting role of capital needs to be carefully balanced. The Basel accord only provides a framework for the minimum values of capital provision for UL a bank should hold, which is 8% of RWA, calculated based on formula 1.1 and 1.2 [20]. Probability of default (PD), exposure at default (EAD) and loss given default (LGD) are the random factors of EL experienced by banks, which represent the unknown variable of actual loss rate, and exact number of defaults in a given year exact amount outstanding.

$$RWA = K * 12.5 * EAD \qquad (1.1)$$

Standard normal distribution (N) applied to threshold and conservative value of systematic factor

Inverse of the standard normal distribution (G) applied to PD to derive default threshold

Inverse of the standard normal distribution (G) applied to confidence level to derive conservative value of systematic factor

$$\text{Capital requirement (K)} = [LGD * N[(1 - R)^{-0.5} * G(PD) + (R / (1 - R))^{0.5} * G(0.999)] - PD * LGD] * (1 - 1.5 \times b(PD))^{-1} \times (1 + (M - 2.5) * b(PD)) \quad (1.2)$$

1.3 Problem Statement

Delta Bank Bhd (DB)[1] is a financial institution that operates commercial banking, investment banking and insurance banking to both consumers and businesses in Malaysia. Regulated under BNM, DB is required to conform to the Basel II requirements both for the purpose of capital adequacy and improvement in risk management. Since its implementation, DB has constantly provided more than the required minimum RWCR of 8% and has set a higher capital ratio in preparation to implement Basel III requirements and actively practice risk management as a strategic function of the company.

However, like any other financial institution, data constraints remain a challenging factor in implementation of Basel, especially data quality. Table 2 summarizes the condition of quality that feeds into the risk management system used to calculate capital provision. It is important to note that data quality is hard to quantify because identifying error requires knowledge of the true nature of the data. Given the massive amount of data (millions of data) feeding into the analytics system, it is impossible to constantly determine the exact nature of data. The table below only provides an adequate reflection of the condition of the data for known error based on rejected data or flagged error. It is expected that the actual percentage is considerably lower.

---

[1] Although DB is a fictitious name, the research was conducted at an actual company

Table 1.2: The percentage of good quality data entering the risk management system

Function	Average % of Good Data
Deposit	99.2
Loan	92.4
Customer	81.4
Collateral	76.3
Treasury	93.4
Hire Purchase	87.1
Credit Card	82.5
Insurance	95.6

The effects of poor quality data onto regulatory capital can be significant. For example, when data is missing the analytics system resorts to using conservative estimates, which can potentially avoid the uplift of regulatory capital. In addition, misclassification of assets due incorrect counterparty details, missing ratings or missing product details can also results in retaining potential provision, which is not reflective of the actual risk. This form of over-estimation of capital requirement can hinder the use of capital for more profitable activities. For DB bank this unnecessary reservation due to poor data quality can freeze millions of Ringgits.

To address the quality issue of the data, DB has established data governance and quality programs, which helps to improve front-end systems. However, data undergoes various transformations along the way before it reaches the risk management system and this gap of poor quality data within the information chain has yet to be addressed by DB. Nevertheless, for the risk management system itself, a team of 20 business analysts and a team of 5 IT specialists maintain it. About 45% of the resource time and effort is spent on addressing data quality issues, where 20% is spent on detection, 60% on root cause analysis and 20% on rectification.

In addition, it is not unusual for quality issues detected in the risk management system to be trickled down the other departments. This is because the risk management system is the end user of data from front-end systems, data warehouse and operational data warehouses. This adds the workload of departments responsible for these systems. In addition, because of priority differences between departments, often the analysts will result to interim solution to improve the quality of data in the risk management system. This leads to the complexity of the information system; not just the risk management system but also the bank's itself.

1.4 Research Question

The research objective is to improve the quality of capital provisioning by providing a method that can identify unusual and unexpected behaviors in the output of the calculation of the analytical system, namely RWA of each portfolio in the system.

i. What are the anomaly detection methods that can be used for this application?
ii. How effective is the selected anomaly detection method in detecting anomaly in a live data stream?
iii. How effective is the selected anomaly detection method in detecting data anomalies during data profiling?
iv. What are the qualitative advantage and disadvantage of selected detection method?
v. How would the selected detection method be implemented as a part of existing data quality management program?

1.5 Scope

The risk management system is a system that contains millions of portfolios. This, to limit the scope of the research, the study will focus on the top 26 portfolio that the bank constantly analyzed. This project also focuses on reducing the time analysts spend on detection, through the use of identifying anomalies in quality of the output of the risk management system through the use of automation.

In addition, the scope of this study focuses on the RWA figure. The reason for selecting this variable is because even though regulatory framework has changed, RWA remains an important and dependent variable for determining capital provision. The RWA provides a common measure for bank's risk, ensure the capital allocated is commensurate with the risks and has the potential to highlight where destabilizing asset class bubbles are arising [21]. In addition, RWA also provides a figure to assess the strengths of banks and a figure, which policy makers, banks, and financial institution supervisors can refer to in order to provide solutions in the event of financial crisis [21].

1.6 Significance

The main significance of this research is that the financial institution would have an appropriate method to improve the quality of the RWA, which could not be addressed effectively using existing methods available in the bank. The outcome of this research would help analyst to detect anomalies automatically and faster, subsequently, reducing the resources spent on ensuring RWA and data quality. On a corporate level and in the long-term, a better quality data will help the financial institution to make well-informed risk management decisions.

1.7 Organization

The remainder of this report is organized as follows. In the next chapter, a literature review of a selection of papers dealing with anomaly detection and its application in data quality control and assurance is given. Hereafter, chapter 3 explains the methodology of this research. Thereafter the results of the research will be analyzed and discussed in the subsequent chapter 4. The final chapter, chapter 5, concludes the paper with a summary, implication, limitation and future work of the research.

1.8 Conclusion

The Basel Accord, Basel I, Basel II and Basel III were drafted to improve the safety and the soundness of the global banking system by enhancing the risk measurement and capital adequacy within banks. Many countries are adopting the Basel Accord and Malaysia is currently adopting Basel II with the initiative to slowly adapt Basel III. Under the Basel Accord, banks are required to set aside capital provision to cover losses from UL. Because UL is unknown, it is estimated by exaggerating the EL of assets defaulting. And according to the accord, at least 8.5% of this exaggerated figure, the RWA, should be set aside.

To determine the capital provision requirement of the Basel Accord is a technical challenge. There are various challenges, which are dependant on the bank’s risk culture, country and methods used for modeling its risk under the Basel Accord. However, data constraint is one of the challenges faced by all banks, particularly in the area of information system, data availability and data quality. In the case of DB bank, data constraints, particularly data quality, is a major challenge in conforming the Basel requirement.

For DB bank, the effect of poor quality data causes the models to use conservative estimates, which leads to retaining potential provision for profitable investments that are not reflective of the actual risks faced by the bank. About 45% of the analysts and IT personal of the system, which are used to model the risk for Basel, are used to address data quality issue. Estimated of this 45%, 20% is used for anomaly detection, 60% for analyzing and 20% for rectifying at their end. It is not abnormal for issues to be trickled down to other departments, where they are also required to analyzed data and rectify related issues at their respective end.

The scope of the project is to improve the time spent on addressing data quality issues, by expediting and automating the process of detection. In addition, the scope of this project is limited to the top assets portfolio of the bank. The research questions of the project includes: the quantitative and qualitative effectiveness of the detection method for two scenarios: data profiling and live stream, and the suggestion of implementing the detection method in the context of the bank's data quality management initiatives.

REFERENCES

1. Bank Negara Malaysia, Prudential financial policy department. (2009). Risk-weighted capital adequacy framework (basel i) (BNM/RH/GL 001 - 21)
2. International Monetary Fund, (n.d.). Global financial stability report - capital to assets ratio (5). Retrieved from website: http://data.worldbank.org/indicator/FB.BNK.CAPA.ZS/countries
3. Tarullo, D. (2008). Banking on basel: the future of international financial regulation. (p. 29). Washington DC: Peterson Instituter for International Economics.
4. Delloite, (2011). Understanding the framework - adopting the basel II accord in asia pacific. Retrieved from website: www.deloitte.com/assets/Dcom../02720_Basel_II_Adopting_SCRN.pdf
5. Flach, J. (2009, June). The basel ii framework and the supervisory assessment of banks' internal risk management. Risk taking and risk transfer of financial market, Jena, Germany.

7. Delloite Touche LLP. (2011, Oct). Basel ii & credit risk management: risk data and modeling consideration. SAS analytics 2011 conference, Orlando, USA.
8. KPMG, (2011). Basel iii: issues and implication. Retrieved from website: www.kpmg.com/../en/IssuesAndInsights/../basell-III-issues-implications.pdf
9. Bank Negara Malaysia, Prudential financial policy department. (2012). Capital adequacy framework (basel ii - risk weighted assets) (BNM/RH/GL 001- 22)
10. Bank Negara Malaysia, Prudential financial policy department, Islamic banking and takaful department. (2011). Notification on the implementation of basel iii (BNM/RH/NT 007-25)
11. Financial Daily. (2011, Dec 23). Basel iii for malaysian banks. The Edge. Retrieved from http://www.theedgemalaysia.com/highlights/198319-basel-iii-for-malaysian-banks.html
12. Accenture, (2011). Basel iii and its consequences - confronting a new regulatory environment (ACC11-0719 / 02-2470)
13. Delloite Touche LLP. (2011, Oct. Key data considerations for risk programs.

14. Oracle, (2009). Key challenges and best practices for Basel II implementation
15. Egan, B. (2007, March). Meeting the challenges of the implementation of Basel ii. Risk management Australia 2007 conference.
16. Gottschalk, R., & Griffith-Jones, S. UK Department for International Development, (2006). Review of basel ii implementation in low-income countries
17. Deloitte Touche Tohmatsu, (2005). From framework to execution effective planning and implementation of the Basel II accord in asia pacific (AM_HO_02720_05.05)
18. Bank Negara Malaysia, (2005). Bank negara malaysia annual report 2005. Retrieved from website: 03/06/03 (BN)
19. KPMG, (2012). Basel infrastructure survey 2012.
20. Basel Committee on Banking Supervision, (2005). An explanatory note on the basel ii irb risk weight functions. Retrieved from website: http://www.bis.org/bcbs/irbriskweight.pdf
21. Sonali, D., & Amadou, N. R. S. International Monetary Fund, (2012). How risky are banks' risk weighted assets? evidence from the financial crisis (WP/12/36)
22. Auer, M., Pfoestl, G. V., Tuerk, B., & Berg, C. Accenture, (2012). Basel iii rwa optimization going beyond compliance (11-5420 / 12-3631 SL)
23. Hay, G. (2011, Nov 24). Measuring risks at europe's banks. The New York Times. Retrieved from http://www.nytimes.com/2011/11/25/business/global/measuring-risks-at-europes-banks.html?_r=0
24. Nadal, J., & Serrano, J. Accenture, (2012). The new importance of risk-weighted assets across europe (12-1190)
25. Chen, T. X., Meyer, M., anapathi, N., Liu, S., & Cirella, J. Research Triangle Institute International, (2011). Improving data quality in relational databases: Overcoming functional entanglements (doi:10.3768/rtipress.2011.op.0004.1105)RTI Press.
26. Zaidi, J. (2012, Feb 1). Proactive and reactive techniques to address information quality challenges head on. The Data Administration Newsletter,
27. Grubbs, F. (1969). Procedures for detecting outlying observations in samples.

28. Laurikkala, J., Juhola, M., and Kentala, E.: 2000, 'Informal Identification of Outliers in Medical Data'. In: Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology IDAMAP-2000 Berlin, 22 August. Organized as a workshop of the 14th European Conference on Artificial Intelligence ECAI-2000.
29. Rosner, B. "On the Detection of Many Outliers." Technometrics 17 (1975): 221–227.
30. Dixon, W. J. "Analysis of Extreme Values." The Annals of Mathematical Statistics 21, 4 (December1950): 488-506.
31. Dixon, W. J. "Ratios Involving Extreme Values." The Annals of Mathematical Statistics 22, 1 (March 1951): 68-78.x
32. Vapnik, V. N. 1995. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA.
33. Box, George & Jenkins, Gwilym. Time Series Analysis: Forecasting and Control. Holden-Day, 1970.
34. Kasunic, M., Mc Curley, J., Goldenson, D., & Zubrow, D. (2011). An investigation of techniques for detecting data anomalies in earned value management data. Department of Defense, USA.
35. Florac, William A. & Carleton, Anita D. Measuring the Software Process: Statistical Process Control for Software Process Improvement. Addison-Wesley Professional, 1999.
36. Wheeler, Donald J. & David S. Chambers. Understanding Statistical Process Control, 3rd ed.SPC Press, 2010.
37. Crowder, S. V. "Design of Exponentially Weighted Moving Average Schemes." Journal of Quality Technology 21 (1989):155-162.
38. Montgomery, D. C. (2012). Introduction to statistical quality control. (7th ed.). USA: Wiley
39. Lucas, J.M (1982). Combined Shewhart-CUSUM quality control scheme, Journal of Quality Technology. Vol 14 (2), pp 51-59
40. Shewhart, Walter A. Economic Control of Quality of Manufactured Product. Van Nostrand, 1931.
41. Ramaswamy, S., Rastogi, R., and Shim, K.: 2000, 'Efficient Algorithms for Mining Outliers from Large Data Sets'. In: Proceedings of the ACM

42. Wettschereck, D.: 1994, ‘A study of distance-based machine learning algorithms’.Ph.D. thesis, Department of Computer Science, Oregon State University,Corvallis.
43. Tang, J., Chen, Z., Fu, A. W.-C., and Cheung, D.: 2002, ‘A Robust Outlier DetectionScheme for Large Data Sets’. Submitted .
44. Knorr, E. M. and Ng, R. T. 1999. Finding intensional knowledge of distance-based outliers. InThe VLDB Journal. 211-222.
45. Bolton, R. J. and Hand, D. J.: 2001, ‘Unsupervised Profiling Methods for Fraud Detection’. In: Credit Scoring and Credit Control VII, Edinburgh, UK, 5-7 Sept. Bradley, P. S., Fayyad, U. M., and Mangasarian, O. L.: 1999, ‘Mathematical programming for data mining: formulations and challenges’. INFORMS Journal on Computing 11(3), 217–238.
46. Bolton, R. J. and Hand, D. J.: 2001, ‘Unsupervised Profiling Methods for Fraud Detection’. In: Credit Scoring and Credit Control VII, Edinburgh, UK, 5-7 Sept. Bradley, P. S., Fayyad, U. M., and Mangasarian, O. L.: 1999, ‘Mathematical programming for data mining: formulations and challenges’. INFORMS Journal on Computing 11(3), 217–238.
47. Allan, J., Carbonell, J., Doddington, G., Yamron, J., and Yang, Y.: 1998, ‘Topic detection and tracking pilot study: Final report’. In: Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop.
48. John, G. H.: 1995, ‘Robust Decision Trees: Removing Outliers from Databases’. In:Proceedings of the First International Conference on Knowledge Discovery and Data Mining. Menlo Park, CA, pp. 174–179, AAAI Press.
49. Skalak, D. B. and Rissland, E. L.: 1990, ‘Inductive Learning in a Mixed Paradigm Setting’. In: Proceedings of the Eighth National Conference on Artificial Intelligence, Boston, MA. pp. 840–847.
50. Tax, D. M. J., Ypma, A., and Duin, R. P. W.: 1999, ‘Support Vector Data Description Applied to Machine Vibration Analysis’. In: Proceedings of ASCI’99, Heijen, Netherlands.
51. DeCoste, D. and Levine, M. B.: 2000, ‘Automated Event Detection in Space Instruments: A Case Study Using IPEX-2 Data and Support Vector Machines’.In: Proceedings of the SPIE Conference on Astronomical

52. Nairac, A., Townsend, N., Carr, R., King, S., Cowley, P., and Tarassenko, L.: 1999, ‘A system for the analysis of jet system vibration data’. Integrated ComputerAided Engineering 6(1), 53 – 65.
53. Bishop, C. M.: 1994, ‘Novelty detection and Neural Network validation’. In: Proceedingsof the IEE Conference on Vision, Image and Signal Processing. pp.217–222.
54. Fawcett, T. and Provost, F. J.: 1999, ‘Activity Monitoring: Noticing Interesting Changes in Behavior’. In: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 53–62.
55. Siaterlis, C. and Maglaris, B. 2004. Towards multisensor data fusion for dos detection. In Proceedings of the 2004 ACM symposium on Applied computing. ACM Press, 439-446.
56. Ester, M., Kriegel, H-P., and Xu, X.: 1996, ‘A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise’. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, pp. 226–231. AAAI Press
57. Zhang, T., Ramakrishnan, R., and Livny, M.: 1996, ‘BIRCH: An Efficient Data Clustering Method for Very Large Databases’. In: H. V. Jagadish and I. S. Mumick (Eds.): Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Quebec, Canada, June 4-6, 1996. pp. 103–114, ACM Press.
58. Kohonen, T.: 1997, ‘Self-Organizing Maps’ volume 2. Springer-Verlag, Heidelberg.
59. Ng, R. T. and Han, J.: 1994, ‘Efficient and Effective Clustering Methods for Spatial Data Mining’. In: Proceedings of the 20th International Conference on Very Large Data Bases, September 12–15, 1994, Santiago, Chile. pp. 144–155, Morgan Kaufmann Publishers.
60. Lazaveric, A., Ertoz L, Kumar V, Ozgur A, & Srivastava J (2003). A comprarative study of anomaly detection schemes in network intrusion detection. In Proceedings of the 7th UNISEX security Symposium.
61. Johnson, R. (1992), Applied Multivariate Statistical Analysis. Prentice Hall.
62. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A

63. Hodge, V.J and Austin J. (2004) A survey outlier detection methodologies. Kluwer Academic Publishers, Netherlands.
64. Yongro, P. (2005). A statistical process control approach for network intrusion detection. Georgia Institute of Technology.
65. Tan, P.N. Steinbach, M., Kumar, V. (2005) Introduction to Data Mining. Addision-Wesley.

67. Stefano, C., Sansone, C & Vento, M. (2000) To reject or not to reject: that is the question- an aswer in case of neural classifiers. IEEE Transactions of Systems, Management and Cybernatics 30, 1, 84-94
68. He, Z., Xu, X. & Deng, S. (2003) Discovering cluster-based local outliers. Pattern Recognition Letters 24, 9-10, 1641-1650.
69. Rousseeuw, P.J, A. M. Leroy (1986, 2003). *Robust Regression and Outlier Detection*. Wiley.
70. Galeano, P., Pea, D., and Tsay, R. S. 2004. Outlier detection in multivariate time series via projection pursuit. Statistics and Econometrics Working Papers ws044211, Universidad Carlos III, Departamento de EstadÃ3stica y EconometrÃ3ca. Sep.
71. Western Electric Company (1956), *Statistical Quality Control handbook.* (1 ed.), Indianapolis, Indiana: Western Electric Co
72. Lu, C.W. and M.R Reynolds, Jr (1999). Ewma control charts for monitoring the mean of autocorrelated processes, Journal of Quality Technology, Vol. 31(2) pp.189-206.
73. Rigdon, S. E., Cruthis, E. N., Champ, C. W. (1994). Design strategies for individuals and moving range control charts. Journal of Quality Technology 26(4):274-287.
74. Page, E.S. (1961) Cummulative sum control charts, Technometrics, Vol. 3(1) pp 1-9
75. Roberts, S.W. (1995). Control chart tests based on geometric moving averages, Technometrics, Vol 42(1), pp 97-102
76. Bower, K. M. Minitab. (2005). Using exponentially weighted average. www.minitab.com/uploadedFiles/Shared.../ewma_control_charts.pdf

77. Berthouex, P.M, W.G Hunter and L. Palleson (1978). Monitoring sewage treatment plants: some quality control aspects, Journal of Quality Technology , Vol 10 (4), pp 139-149.
78. Runger, G.C and T.R Willemain (1996), Batch means control charts for autocorellated data. IIE Transactions. Vp; 28(6), pp 483-487.
79. Hill, D. J., Minsker, B. S., & Amir, E. (2007). Real-time Bayesian Anomaly Detection for Environmental Sensor Data. In Proceedings of the 32nd conference of IAHR.
80. Juran, J.M. (1951) Quality Control Handbook. USA-McGraw Hill
81. Karr, A. F., Sanil, A. P., & Banks, D. L. (2005). Data Quality: A Statistical Perspective (Report No. 151). Research Triangle Park, NC: National Institute of Statistical Science.
82. Hinrich, H. (2001). Statistical quality control of warehouse data. In Databases and information systems (pp. 69-84). Norwell, MA: Kluwer Academic Publisher.
83. Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A framework for analysis of data quality research. IEEE transaction on knowledge and data engineering, 7(4).
84. Wang R,Y, A Product Perspective on Total Data Quality Management," Communications of the ACM, February 1998, pp. 58-63.
85. Pipino, L. L., Lee, W. L., & Wang, R. Y. (2002, April) Data Quality Assessment. Communications of the ACM, 45, 211-217.
86. Vannan, E. (2001, Novemeber) Quality Data - An Improbable Dream? Educause Review, 56-58
87. Sandeep Savla & Mathew Ninan (2005). SETLabs Briefings. Vol. 3 No 4, Dec 2005
88. Dasu, T. (2013). Data Glitches. In Handbook Of Data Quality: Research and Practice (pp. 163-179). Berlin, Germany: Springer-Verlag
89. Gunisetti L (2011). Detection of Discordant Observations and Visualization of Data. IJCA Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET) (9):15-18, 2011. Published by Foundation of Computer
90. Hanninen, S. (2010). Applying data mining techniques to ERP system

91. Thiprungsri. S, Vaserhelyi, M.A (2011) Cluster analaysis for anomaly detection in accountring data: an audit approach. The International Journal of Digital Accounting Research Vol.11, 2011, pp 69-84.
92. Wei, W., Li, J., Cao, L., & Chen, J. (2012). Effective detection of sophisticated online banking fraud on extremely imbalanced data . In World Wide Web. Retrieved from 10.1007/s11280-012-0178-0.
93. Lohsin, David. (2006) Monitoring data quality performance using data quality metric. Informatica
94. Simon, G, ORACLE. (2008) The Emerging Field of Financial Data Quality Management. Redwood Shores CA USA
95. Xu, H., Nord, J.H, Brown, N. Nord, G.D (2002) Data quality issues in implementing an ERP. Industrial Management & Data Systems, In World Wide Web. Retrieved from 10.1108/02635570210414668