

# **TELECOMMUNICATION SUBSCRIPTION FRAUD DETECTION USING NEURAL NETWORK**

**A thesis submitted to the Faculty of Information Technology in partial  
fulfillment of the requirement for the degree  
Master of Science (Intelligent System)  
Universiti Utara Malaysia**

**By  
NURATIKHAH SHAHRIN**

**© Nuratikhah Shahrin, 2008.  
All Rights Reserved.**

710  
210  
13



**KOLEJ SASTERA DAN SAINS**  
**(College of Arts and Sciences)**  
**Universiti Utara Malaysia**

**PERAKUAN KERJA KERTAS PROJEK**  
**(Certificate of Project Paper)**

Saya, yang bertandatangan, memperakukan bahawa  
*(I, the undersigned, certify that)*

**NURATIKAH SHAHRIN**

calon untuk Ijazah  
*(candidate for the degree of)* **MSc. (Intelligent System)**

telah mengemukakan kertas projek yang bertajuk  
*(has presented his/ her project paper of the following title)*

**TELECOMMUNICATION SUBSCRIPTION FRAUD**  
**DETECTION USING NEURAL NETWORK**

seperti yang tercatat di muka surat tajuk dan kulit kertas projek  
*(as it appears on the title page and front cover of project paper)*

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan  
dan meliputi bidang ilmu dengan memuaskan.  
*(that the project paper acceptable in form and content, and that a satisfactory  
knowledge of the field is covered by the project paper).*

Nama Penyelia Utama  
*(Name of Main Supervisor):* **ASSOC. PROF. FADZILAH SIRAJ**

Tandatangan  
*(Signature)*

:

Tarikh (Date) :

27/06/08

Nama Penyelia Kedua  
*(Name of 2<sup>nd</sup> Supervisor):* **MISS NOORAINI YUSOFF**

Tandatangan  
*(Signature)*

:

Tarikh (Date) :

17/06/08

## **PERMISSION TO USE**

In presenting this thesis in partial fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor, in her absence, by the Dean of the Faculty of Information Technology. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain should not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Request for permission to copy or to make use of material in this thesis, in whole or in part should be addressed to:

**Dean of the Faculty of Information Technology**

**Universiti Utara Malaysia**

**06010 UUM Sintok**

**Kedah Darul Aman**

## **ABSTRACT**

### **(Bahasa Malaysia)**

*Setiap tahun, penipuan telekomunikasi menelan kos berbilion dolar daripada keuntungan pelaburan. Jenayah penipuan telekomunikasi kebiasaannya dimotivasikan oleh wang atau sesuatu keadaan yg terdesak tanpa melibatkan nama, atau kedua-duanya. Melalui penggunaan internet, komuniti penyalahgunaan menjadi semakin berkolaborasi, dan akibatnya, mereka menjadi lebih bijak. Oleh itu, aplikasi pengesan penipuan perlu menjadi lebih sofistikated supaya selaras dengan penjenayah. Sebuah sistem untuk menghalang langganan penipuan dalam talian tetap dengan impak yang tinggi ke atas penghantar jarak jauh telah dicadangkan. Sistem tersebut meliputi teknik rangkaian neural, menggunakan kaedah 'Multilayer Perceptron' (MLP). Sejumlah 158 sampel data dari Telekom Malaysia Bhd yang telah dikumpulkan, dilatih dan diuji dengan menggunakan model. Modul ramalan membenarkan pengenalpastian penipuan yang berpotensi ketika proses pendaftaran. Analisis data menunjukkan perkaitan yang kuat antara pemalar masukan yang terdiri daripada persamaan, faktor kepastian pengguna, penunjuk kes dan matlamat. Hasil kajian menunjukkan 78% ketepatan ramalan berjaya diperolehi. Daripada keputusan yang diperolehi, dapat disimpulkan bahawa rangkaian neural berpotensi untuk digunakan sebagai pengesan penipuan dalam telemounikasi.*

## **ABSTRACT**

### **(English)**

*Telecommunications fraud costs carriers billions of revenue dollars annually. The telecom fraud criminal is often ingenious and is typically motivated by money, anonymity, or both. Through use of the Internet, the fraud community is becoming more collaborative, and as a result, more ingenious. Because of this, fraud detection applications must become more sophisticated to keep pace with the criminals. A system to prevent subscription fraud in fixed telecommunications with high impact on long-distance carriers is proposed. The system consists of a neural network technique, using Multilayer Perceptron (MLP). A total of 158 data samples from Telecom Malaysia Bhd were collected trained and tested using model. The prediction module allows identifying potential fraudulent customers at the time of subscription. The analysis of the data shows a reasonably strong correlation between the input variable, which consist of severity, user certainty factor, indicator case and target. The result shows that 78% of prediction accuracy has been obtained. From the result that has been produce, neural network has a potential to be used for detecting fraud in telecommunication.*

## **ACKNOWLEDGEMENT**

Firstly, I would like to extend my thanks and gratitude to Telecom Bhd who have give their cooperation during this research, where they have fully support in all the information I needs for this research.

Next, I would like to extend my thanks to University Utara Malaysia (UUM) for facilities and resources provided. Without the support of the facilities from UUM this research may not be come true so smoothly.

Besides that, I would like to take this opportunity to express my thanks and gratitude to my Supervisor, Associate Professor Fadzilah Siraj, for her guidance, wonderful support and efforts in assisting me carrying out these research and thesis to become a reality.

Last but not least, my gratitude also goes to my beloved family members who have done great job of supporting me throughout this research. They have been very understanding and offered lots of encouragement when the system and research I was working on did not seem to work the way it should be.

To those I have not mentioned above, but had ever helped me, I really appreciate it and felt very thankful for all your helps. Thanks to all of you.

## TABLE OF CONTENTS

<b>PERMISSION TO USE</b>	ii
<b>ABSTRACT (Bahasa Malaysia)</b>	iii
<b>ABSTRACT (English)</b>	iv
<b>ACKNOWLEDGEMENT</b>	v
<b>TABLE OF CONTENT</b>	vi
<b>LIST OF TABLES</b>	vii
<b>LIST OF FIGURES</b>	viii
<b>CHAPTER 1 : INTRODUCTION</b>	
1.1 Background	1
1.2 Problem Statement	3
1.3 Objectives	5
1.4 Scope of the study	5
1.5 Significance of the study	5
1.6 Organization of the report	6
<b>CHAPTER 2 : LITERATURE REVIEWS</b>	
2.1 Overview	7
2.2 Fraud Detection Using Neural Network	10
2.3 Fraud Detection Using Other Techniques	12
2.4 Summary	13
<b>CHAPTER 3 : METHODOLOGY</b>	
3.1 Overview of Methodology	14
3.1.1 Awareness of Problem	16
3.1.2 Suggestion	16
3.1.2.1 Layer 1	17
3.1.2.2 Layer 2	18
Layer 3	18
3.1.3 Development	18
3.1.3.1 Data Acquisition	19
3.1.3.2 Data Description and Selection	19
3.1.3.3 System Development	22
3.1.4 Evaluation	34
3.1.5 Conclusion	34
3.2 Summary	34

## **CHAPTER 4 : RESULT AND DISCUSSION**

4.1	Neural Network Detection System	35
4.2	Preprocessing Data Result	38
4.3	Training Data Result	40
4.4	Prediction Data Result	42
4.4.1	Test Case 1	43
4.4.2	Test Case 2	44

## **CHAPTER 5 : CONCLUSION**

5.1	Conclusion	45
5.2	Future Enhancement	46

<b>REFERENCES</b>	<b>47</b>
-------------------	-----------

<b>APPENDIX: System User Manual</b>	<b>54</b>
-------------------------------------	-----------



## **LIST OF TABLES**

Table 3.1	Sample of Datasets	20
Table 3.2	Description of each Attribute	21
Table 3.3	Sample of Datasets after Attributes Selection and Classifying the telecommunication fraud into different classes.	22

## LIST OF FIGURES

Figure 3.1	General Methodology of Design Research (Vaishnavi & Kuechler, 2005)	15
Figure 3.2	Flow of Neural Network Architecture	17
Figure 3.3	Neural Network Module Process	23
Figure 3.4	Neural Network Structure Diagram	23
Figure 3.5	Normalization Sequence	24
Figure 3.6	Normalization Algorithm	25
Figure 3.7	Backpropagation Algorithm	27
Figure 3.8	Flowchart of the training process	30
Figure 3.9	Feedforward Algorithm	32
Figure 3.10	Flowchart of the prediction process	33
Figure 4.1	Preprocessing Interface	36
Figure 4.2	Training Interface	37
Figure 4.3	Prediction Interface	38
Figure 4.4	Raw Data	39
Figure 4.5	Result of Normalization	40
Figure 4.6	Data Partition for Training	41
Figure 4.7	Result of Training	42
Figure 4.8	Result of Non Fraud Cases	43
Figure 4.9	Result of Fraud Cases	44

# CHAPTER 1

## INTRODUCTION

In this chapter, the first section explains the background of the study that mainly involves the fraud detection and fraud detection technique using Neural Network. Next, the research problem is presented, followed by the objectives, significance of the project and scope for the study. As a conclusion for this chapter, the organization of this study is presented.

### **1.1 Background**

Telecommunication fraud is increasing dramatically each year resulting in loss of a large amount of euros worldwide. The history of telecommunications crime, including several types of fraudulent activities, was reviewed by Collins (1999a, 1999b, 2000). Fraud detection involves identifying fraud as quickly as possible once it has been committed.

The contents of  
the thesis is for  
internal user  
only

## REFERENCES

- Barghava, B., Zhong, Y., & Lu, Y. (2003). Fraud formalization and detection. In Y. Kambayashi, M. Mohania, W. Wöb (Eds.), *Data Warehousing and Knowledge Discovery, 5th Int. Conference, DaWak 2003, Prague, Czech Republic* (pp. 330-339). LCNS Vol. 2737, Springer-Verlag.
- Bolton, R.J., & Hand, D.J. (2002). Statistical fraud detection: A review. *Statistical Science*, 17(3), 235-249.
- Burge, P., & Shawe-Taylor, J. (2001). An unsupervised neural network approach to profiling the behavior of mobile phone users for use in fraud detection. *Journal of Parallel and Distributed Computing*, 61, 915-925.
- Cahill, M. H., Lambert, D., Pinheiro, J.C., & Sun, D. X. (2002). Detecting fraud in the real world. In J. Abello, P. M. Pardalos, M. G. C. Resende (Eds.), *Handbook of Massive Data Sets*, (pp. 913-930). Dordrecht: Kluwer Academic Publishers.
- Chan, P. K., Fan, W., Prodromidis, A. L., & Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems*, 14(6), 67-74.
- Collins, M. (1999a). Telecommunications Crime – Part 1. *Computer & Security*, 18, 577-586.

Collins, M. (1999b). Telecommunications Crime – Part 2. *Computer & Security*, 18, 683-692.

Collins, M. (2000). Telecommunications Crime – Part 3. *Computer & Security*, 19, 141-148.

Cortes, C., Pregibon, D., & Volinsky, C. (2001). Communities of interest. In F. Hoffman, D. J. Hand, N. Adams, G. Guimaraes (Eds.), *Advanced in Intelligent Data Analysis, 4<sup>th</sup> Int. Conference IDA 2001, Cascais, Portugal* (pp. 105-114). LCNS Vol. 2189, Springer-Verlag.

Cortes, C., Pregibon, D., & Volinsky, C. (2003). Computational methods for dynamics graphs. *Journal of Computational and Graphical Statistics*, 12(4), 950-970.

Daskalaki, S., Kopanas, I., Goudara, M., & Avouris, N. (2003). Data mining for decision support on customer insolvency in telecommunications business. *European Journal of Operational Research*, 145, 239-255.

Ezawa, K.J., & Norton, S.W. (1996). Constructing Bayesian networks to predict uncollectible telecommunications accounts. *IEEE Expert*, 11(5), 45-51.

Fawcett, T., & Provost, F. (1997a). Combining data mining and machine learning for effective fraud detection. In *AI Approaches to Fraud Detection & Risk Management*, Workshop Technical Report WS-97-07 (pp. 14-19). AAAI Press, Menlo Park, California.

Fawcett, T., & Provost, F. (1997b). Adaptive fraud detection. *Data Min. Knowl. Disc.*, 1(3), 291-316.

Fawcett, T. (2002a). Case study: Adaptive fraud detection. In W. Klösgen, & J. Zytkow (eds.), *Handbook of Data Mining and Knowledge Discovery* (section H1.2.1). Oxford U. Press, UK.

Fawcett, T., & Provost, F. (2002b). Fraud detection. In W. Klösgen, & J. Zytkow (eds.), *Handbook of Data Mining and Knowledge Discovery* (section F2). Oxford U. Press, UK.

Fawcett, T. (2003). ROC graphs: Notes and practical considerations for data mining researchers. HP Labs. Technical Report HPL-2003-4.

Ghosh, S. Reilly, D.L. Credit card fraud detection with a neural-network. *Decision Support and Knowledge-Based Systems, Proceedings of the Twenty-Seventh Hawaii International Conference*. Volume 3: 621-630, 1994.

- Hansen, J. V., McDonald, J. B., Messier, W. F., & Bell, T. B. A generalized qualitative – response model and the analysis of management fraud. *Management Science*, 42(7), 1022-1032, 1996.
- Hoath, P. (1999). What's new in telecoms fraud?. *Computer Fraud & Security*, 1999(2), 13- 19.
- Hoath, P. (1998). Telecoms fraud, the gory details. *Computer Fraud & Security*, 1998(1), 10-14.
- Hong, S. J., & Weiss, S. M. (2001). Advances in predictive models for data mining. *Pattern Recognition Letters*, 22, 55-61.
- Liao, S. H. (2005). Expert systems methodologies and applications - a decade review from 1995 to 2004. *Expert Systems with Applications*, 28, 93-103.
- Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., & Kaushansky, H. (2000) Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Trans. Neural Networks*, 11(3), 690-696.
- M. Taniguchi, M. Haft, J. Hollmen, and V. Tresp. Fraud detection in communication networks using neural and probabilistic methods. In *Proceedings of the 1998 IEEE*



*International Conference in Acoustics, Speech and Signal Pmcessing*, volume 2, pages 1241-1244,1998.

Paredes, R.D. (2005). Lessons from the deregulation transition in Chile's local telephony market. *Telecommunications Policy*, 29, 333-350.

Patterson, D.W. (1995). *Artificial neural networks: theory and applications*. Singapore: Prentice Hall.

Phua, C., Lee, V., Smith, K., & Gayler R. (2005). A comprehensive survey of data mining-based fraud detection research. Available at <http://www.bsys.monash.edu.au/people/cphua/>

Provost, F., & Fawcett, T. (1997). Analysis and visualization of classifier performance with nonuniform class and cost distributions. In *AI Approaches to Fraud Detection & Risk Management*, Workshop Technical Report WS-97-07 (pp. 57-63). AAAI Press, Menlo Park, California.

Rosset, S., Murad, U., Neumann, E., Idan, Y., & Pinkas, G. (1999). Discovery of fraud rules for telecommunications - challenges and solutions. In *Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 409-413.

Shawe-Taylor, J., Howker, K., & Burge, P. (1999). Detection of fraud in mobile telecommunications. *Information Security Technical Report*, 4(1), 16-28.

Shawe-Taylor, J., Howker, K., Gosset, P., Hyland, M., Verrelst, H., Moreau, Y., Stoermann, C., & Burge, P. (2000). Novel techniques for profiling and fraud in mobile telecommunications. In P. J. G. Lisboa, B. Edisbury, & A. Vellido (eds.), *Business Applications of Neural Networks. The State-of-the-Art of Real World Applications* (pp. 113-139). World Scientific, Singapore.

Stehmann, O. (1995). Network liberalization and developing countries: The case of Chile. *Telecommunications Policy*, 19(9), 667-684.

Stolfo, S. J., Fan, D. W., Lee W. L., Prodromidis, A. L. & Chan, P. K. (1997) Credit card fraud detection using meta-learning: issues and initial results. In *AI Approaches to Fraud Detection & Risk Management*, Workshop Technical Report WS-97-07 (pp. 83- 90). AAAI Press, Menlo Park, California.

Wong, B. K., Bodnovich, T. A., & Selvi, Y. (1997). Neural network applications in business: a review and analysis of the literature (1988-95). *Decision Support Systems*, 19, 301-320.

Wong, B. K., & Monaco, J. A. (1995). Expert system applications in business: a review and analysis of the literature (1977-1993). *Information & Management*, 29, 141-152.

Y. Moreau, B. Preneel, P. Burge, J. Shawe-Taylor, C. Stoermann, and C. Cooke (1997).  
Novel techniques for fraud detection in mobile telecommunication networks. In *ACTS  
Mobile Summit*, Grenada, Spain.