TOWARDS FORECASTING REVENUE COLLECTION USING MULTI LAYER PERCEPTRON (MLP)

A Thesis Submitted to the Faculty of Information Technology in Partial Fulfillment of the Requirements for the Degree
Master of Science (Information Technology)
Universiti Utara Malaysia

by

Mohd Afandi Md Amin

JABATAN HAL EHWAL AKADEMIK
(*Department of Academic Affairs*)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

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

MOHD. AFANDI MD. AMIN

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

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

TOWARDS FORECASTING REVENUE COLLECTION USING MULTI LAYER PERCEPTRON (MLP)

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 filed is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **MRS. ZAHARIN MARZUKI**

Tandatangan
(Signature) : _____ Tarikh (Date): 26/06/04

Nama Penyelia Kedua
(Name of 2nd Supervisor): **MR. WAN HUSSAIN WAN ISHAK**

Tandatangan
(Signature) : _____ Tarikh (Date): 26/6/04

PERMISSION OF USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti 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(s) or, in their 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 shall 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. Requests for permission to copy or to make other use of materials in this thesis, in whole or in part should be addressed to:

Dean of Faculty of Information Technology
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman

ABSTRAK

Fungsi utama Jabatan Kastam DiRaja Malaysia (KDRM) adalah memungut hasil melalui Cukai Tak Langsung disamping memberi kemudahan dan galakan kepada masyarakat perindustrian dan pedagang selaras dengan peruntukan undang-undang semasa bagi mencapai visi "*menjadikan perkhidmatan kastam dikenali dan dihormati serta bertaraf dunia*". Setiap tahun KDRM telah mensasarkan pungutan hasil tahunan selaras dengan Perancangan Strategik Jabatan 2001-2005. Sasaran pungutan hasil telah menjadi agenda yang penting kepada semua Pengarah Kastam Negeri di mana setiap negeri perlu menetapkan sasaran pungutan berdasarkan kemampuan dari segi sumber percukaian, pelesenan dan prestasi semasa.

Peramalan pungutan hasil menggunakan model statistik asas telah digunapakai oleh KDRM mulai tahun 2002 melalui penggunaan perisian lengkap peramalan *Forecast Pro Versi 4.0.* Kepentingan peramalan bagi jabatan ialah ianya dapat menyediakan satu kaedah memantau prestasi pungutan dari masa ke semasa dan mengambil langkah-langkah yang berkesan bagi memastikan sasaran pungutan dapat dicapai seperti yang telah ditetapkan.Ini adalah kerana prestasi pungutan hasil bergantung kepada beberapa factor seperti ekonomi, politik ,dasar kerajaan dan persekitaran perniagaan yang sentiasa berubah-ubah.

Kajian ini lebih menjurus kepada penerokaan kaedah baru menggunakan Rangkaian Neural Buatan ke arah peramalan pungutan hasil.Data bagi tujuan latihan telah diperolehi daripada rekod Penyata Pungutan Hasil yang disediakan oleh Bahagian Perkhidmatan Teknik, Jabatan kastam DiRaja Malaysia, Alor Setar, Kedah. Sebanyak 1 727 data telah digunakan dalam kajian ini yang terdiri daripada 7 jenis cukai dan duti serta pungutan hasil bukan cukai, bagi tempoh 159 bulan berturutan bagi tempoh 1 Januari,1990 hingga 30 Mac,2004.Rangkaian Neural Buatan melalui model perambatan balik, *Multi Layer Perceptron (MLP)* telah dilatih dengan data berkenaan bagi membangunkan model peramalan pungutan hasil. Adalah diharapkan hasil kajian ini akan dapat membantu KDRM bagi membangunkan kaedah peramalan pungutan hasil yang lengkap pada masa hadapan.

Hasil kajian ini telah membuktikan kesahihan dan kebolehpercayaan Rangkaian Neural Buatan ke arah peramalan pungutan hasil KDRM. Pencapaian daripada pembangunan model perambatan balik telah memberi ketepatan melebihi 92 peratus. Rangkaian Neural Buatan didapati dapat mengambil rekod data ke arah peramalan pungutan hasil serta menjadikan peramalan lebih pantas dan senang digunakan. Selain dari itu, kajian ini juga dapat menambahkan bidang kajian berkaitan Rangkaian Neural Buatan sedia ada di Fakulti Teknologi Maklumat, Universiti Utara Malaysia.

ABSTRACT

Royal Customs Department Malaysia (RCDM) has the main function to collect revenue through indirect taxes, instead of giving facilitation and encourage industrialization and trades according to current provisions by laws and regulations towards vision *"to be recognized respected and world class services"*. RCDM was setting revenue collection target yearly along with Strategic Planning 2001-2005. The revenue collection target has become an important agenda for every State Director of Customs through nationwide. Every state has to set their collection target based on the capability and tax resources, licensees and current performance.

Forecasting revenue collection used statistical fundamental model has been used by RCDM by the year 2002 via deployment completed forecasting software that recognized as Forecast Pro Version 4.0. The significant of revenue forecasting is it provided a method to monitor collection performance and to take effective planning in order to ensure that revenue collection target can be achieved. This is as a result of the revenue collection performance was depends on various factors such as economics, politics, government policy and business environment that always been changing.

This study is more on a new exploration technique using Artificial Neural Network (ANN) towards forecasting revenue collection of RCDM. The data sets were gathered from Monthly Revenue Return that provided and allowed to be used by Technique Division, RCDM, Alor Setar,Kedah. The data sets comprises of 1 727 data that composed by 7 types of the duty and tax with non tax revenue from 159 successive month starting from 1st January, 1990 to 30th Mac, 2004. ANN with back propagation model, MLP has been used to train data sets in order to develop forecasting model revenue collection. Hopefully this study can be assists RCDM to develop a complete forecasting revenue collection tools using ANN for future enhancement.

This study has proven the capability and reliability of ANN towards forecasting revenue collection. The results from BP model have proven the accuracy of forecasting is more than 92 percent. The ANN was found that can feed the data towards forecasting revenue collection. Thus, making it faster and easy to use. On the other hand, this study also adds more study domain related to current ANN applications for Faculty of Information Technology, Northern University of Malaysia.

ACKNOWLEDGEMENTS

Many thanks to the Most Gracious and Most Merciful, Allah for giving me the patience and consistency in completing this study. I dedicated this thesis to my beloved wife, *Nurul Huda* and kids; *Nur Syafiqah, Muhammad Hanif Syauqi* and *Nur Iman* for giving support , loves and prayers in completing my degree.

I wish to acknowledge gratefully the Universiti Utara Malaysia, Sintok, Kedah Darul Aman for providing the facilities for this research undertaking. The same goes true to both my supervisors Puan Zaharin Marzuki and Encik Wan Hussain Wan Ishak whose tiring continually supervise me in time, all the way to its final reproduction.

I am grateful to Jabatan Perkhidmatan Awam, Malaysia (JPA) for the scholarships, study leave and allowances that approved for my further study. I also acknowledge Royal Customs Department Malaysia (RCDM) especially RCDM, Kedah for allowing the access of revenue collection records. This study would not have become visible without the cooperation and support from Kedah State Director of Customs, especially Senior Assistant Director, Tn. Hj. Mutalib Embi and Senior Customs Officer, Puan Norhayati Zakaria from Technique Division, for providing the preliminary materials and data. The contributions to the success of this study are very much venerated.

Special appreciation to my parents for their loves and prayers. Sincere thanks to collegues,sharing experiences through the years.

Mohd Afandi Md Amin
June 2004

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

Revenue refers to government income due to taxation. In Malaysia there are two types of taxation has been implemented; direct taxation and indirect taxation that was organized by different authority. Royal Customs Department Malaysia (RCDM) is obliged to enforce indirect taxation, whereby Inland Revenue Board (IRB) concerned in direct taxation that involved income tax collection for every taxable person subject to the current provisions.

RCDM was placed under The Ministry of Finance to implement the laws and regulations regarding indirect taxation. An indirect tax has several categories or resources such as Import Duty, Export Duty, Sales Tax, Service Tax and Excise Duty. By the year 2003, indirect taxes share 25.4% of Federal Government Revenue (FGR). Whereby, non-tax revenue share 7.5%, Excise Duty and Sales Tax contributed about 5.5% and 9.5% respectively for FGR (*Malaysia Economics Report 2003/2004*).

This study is an exploration of the use of Data Mining technique, especially Neural Network (NN) to build a forecasting (predicting) model for revenue collection. Generally, data mining sometimes called data or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful

The contents of the thesis is for internal user only

BIBLIOGRAPHY

Abdul-Kareem, S., Baba, S., Zubairi, Y. Z., Prasad, U., & A.Wahid, M. I. (2001). *Back Propagation Neural Network For Medical Prognosis: A Comparison Of Different Training Algorithms*. Retrieved April 27, 2004, from World Wide Web: http://www.sat.ait.ac.th/ej-sat/articles/3.1/sarmeem.pdf

Abidi, S. S. R. (1998). *Neural Networks: Their Efficacy Towards The Malaysia IT Environment*. Retrieved May 13, 2004, from World Wide

Aleksander, I., & Morton, H. (1991). *An ntroduction to Neural Computing*. London: Chapman & Hall.

Amir, A. F. (2001). *Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results*. Retrieved Mac 23, 2004, from World Wide Web: http://www.alumni.caltech.edu/~amir/bankr.pdf

Amstrong, J. S. (1983). *Strategic Planning and Forecasting Fundamental*. Retrieved Mac 28,2004, from World Wide Web: http://citeseer.ist.psu.edu/armstrong83strategic.html

Baharom, S., & Shang, W. S. (1996). *Neural Networks Approach in Predicting KLSE Composite Index*: Jabatan Matematik,UKM.

Berson, A., & Smith, S. J. (1997). *Data Warehousing, Data Mining and OLAP*: McGraw-Hill Companies.

Bigus, J. P. (1996). *Data Mining With Neural Networks: Solving Business Problems From Application Development To Decision Support*: McGraw-Hill.

Blum, A. (1992). *Neural Networks in C++.An Object-Oriented Framework for Building Connectionist Systems*: John Wiley & Sons, Inc.

Boon, T. H. (2003). *Forecasting Techniques in Economics and Business*: Faculty of Economics and Management, Universiti Putra Malaysia.

Brockett, P., Cooper, W., Golden, L., & Piktakong, U. (1994). A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency. *Journal of Risk and Insurance, Vol. 61*(3), pp. 402-424.

Clarkson, T. (1999). *Applications of Neural Networks in Telecommunications*. Retrieved April 30, 2004, from World Wide Web: http://www.erudit.de/erudit/events/tc-d/16_04_99/1_clarkson.pdf

Dajani, R. E., Miquel,M.,Rubel,P.,Maison-Blanche,P.,Fayn,J. and Forlini,M-C. (2001). *A Neural Network Approach for Predicting and Modeling the Dynamical Behavior of Cardiac Ventricular Repolarisation.* Retrieved October 9, 2003, from World Wide Web: http://cmbi.bjmu.edu.cn/2001/medinfo_2001/Papers/Ch6/Eldajani.pdf

Deo, M. C., & Naidu, S. (1999). Real Time Wave Forecasting Using Neural Networks. *Ocean Engineering, 26*, 191-203.

Do, A. Q., & Grudintski, G. (1992). A Neural Network Approach to Residential Property Appraisal. *The Real Estate Appraiser, Vol. 58*(3), pp. 38–45.

Dorffner. (1999). *Internal Report for NeuroNet.* Retrieved March 11,2004, from World Wide Web: http://www.kcl.ac.uk/neuronet

Dzeroski, S., & Lavrac, N. (2001). *Relational Data Mining*: Springer-Verlag Heidelberg,Germany.

Fausett, L. (1994). *Fundamental of Neural Networks :Architecture, Algorithm and Applications*. Englewood Cliff: Prentice Hall.

Fayyad, U., Piatetsky-Shapiro ,G. and Smyth ,P. (1997). *From Data Mining to Knowledge Discovery in Databases*. Retrieved May 10, 2004, from World Wide Web: http://www.kdnuggets.com/gpspubs/aimag-kdd-overview -1996-Fayyad.pdf

Franses, P., & Griensven, K. (1998). Forecasting Exchange Rates Using Neural Networks for Technical Trading Rules. *Studies in Non Linear Dynamics and Econometrics Quarterly Journal, 2*(4).

Glorfeld, L. W., & Hardgrave, B. C. (1996). An Improve Method for Developing Neural Networks: The Case of Evaluating Commercial Loan Creditworthiness. *Computer Operation Research, 23*(10), 933-944.

Gonzalez, S. (2000). *Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models*. Retrieved March 12, 2004, from World Wide Web: http://ann.martinsewell.com/Gonz00.pdf

Hamad, D., & Betrouni, M. (1995). *Artificial Neural Networks for Nonlinear Projection and Exploration Data Analysis.* Paper presented at the International Conference, Ales,France.

Haykin, S. (1994). *Neural Networks. A Comprehensive Foundation.* New York: Macmillan College Publishing Company.

Heikkonen, J., & Lampinen, J. (1999). *Building Industrial Applications with Neural Networks*. Retrieved March 11, 2004, from World Wide Web: http://www.erudit.de/erudit/events/esit99/12642_P.pdf.

Hernandez-Aquirre, A., Koutsougeras, C., & Chris, B. (2002). Sample Complexity For Function Learning Tasks Through Linear Neural Networks. *International Journal on Artificial Intelligence Tools, 11*(4), 499-511.

Hussain, W. I. W. (2002). *Predicting Diseases Using Multi-Backpropagation.* Unpublished Master of Science (IT) Thesis, Universiti Utara Malaysia.

Ishikawa, M. (1995). Learning of Modular Structured Networks. *Artificial Intelligence, 75*, 51-52.

Jensen, H. L. (1992). Using neural networks for credit scoring. *Managerial Finance, Vol. 18*(6), pp. 15–26.

Jordan, M. I. (1995). *Why The Logistic Function? A Tutorial Discussion On Probabilities and Neural Networks.* Retrieved April 20, 2004, from World Wide Web: http://www.cs.berkeley.edu/~jordan/papaers/uai.psZ/Jordan.ps

Kawaguchi, K. (2000). *A Multithreaded Software Model For Back propagation Neural Networks Application.* University of Texas, El Paso.

Kiviluoto, K. (1998). Predicting bankruptcies with the self-organizing map. *Neurocomputing, Vol. 21*(1-3), pp. 203–224.

Kuhrer, M., & Macha, C. (2000). *Forecasting and Trading Stock Market Indices: A Comparison of the Results achieved by Neural Networks and by Technical Analysis.* Retrieved March 15, 2004, from World Wide Web: http://ifm.wu-wien.ac.at/Forschung/K%FChrer.pdf

Kuvayev, L. (1998). *Predicting Financial Markets With Neural Networks.* Retrieved March 15,2004, from World Wide Web: http://citeseer.ist.psu.edu/386387.html.

Le Cun, Y., & Drucker, H. (1992). *Improving Generalization Performance Using Double Backpropagation.* Retrieved March 13, 2004, from World Wide Web: http://yann.lecun.com/exdb/publis/pdf/drucker-lecun-92.pdf

Lee, K. C., Han, I., & Kwon, Y. (1996). Hybrid Neural Network Models for Bankruptcy Predictions. *Decision Support Systems, Vol.18*, pp. 63–72.

Li, Y., Liu, L., Yang, T., & Chiu, W. (1997). *Comparing The Performance Of Mathematical Models for Surgical Decisions on Head Injury Patients.* Retrieved October 7, 2003, from World Wide Web: http://www.amia.org/pubs/symposia/D004007.pdf.

Lisboa, P. G. J. (1992). *Neural Networks Current Applications*: Chapman & Hall.

Malaysia Economics Report 2003/2004. Kuala Lumpur: Percetakan Nasional Malaysia Berhad.

Manilla, H. (2001). *Data Mining: Selected Issues And An Environmental Study.* Retrieved December 10, 2003, from World Wide Web: http://www.matstat.umu.se/personal/oleg/dm2001/dmamm-print.pdf

Masri, A., Mohd Faidzul, N., Khairuddin, O., & Miswan, S. (2000). *The Effect of Return Functions On Individual Stock Price (KLSE) Prediction Model Using Neural Networks.* Retrieved April 17, 2004, from World Wide Web: http://www.cs.nott.ac.uk/~mxa/ReturnFunction.PDF

McMenamin, S. J. (1997): "A primer on neural networks for forecasting", *Journal of Business Forecasting*, Vol. 16, Fall 1997, p.17-22.

Moody, J., Levin, U. and Rehfuss, S. (1993): "Predicting the U.S. index of industrial production", *Neural Network World,* 3(6), p.791-794 in special issue:*Proceedings of Parallel Applications in Statistics and Economics.*

Mohd Yusof, M., & Siti Noraini, S. (2001) .Recognition of Noisy Numerals using Neural Networks. International Journal of Computer and Engineering Management. Retrieved Mac 10, 2004 from World Wide: http://www.journal.au.edu/ijcim/2001/sep01.article4.pdf

Morris, R. (1997). Predicting failure: A failure in prediction? *Accountancy, December*, pp. 152–153.

Ohno-Machado, L. (1996). *Medical Applications of Artificial Neural Networks Connectionist Model of Survival.* Standford University.

Parker, D. (1995). Learning-Logic: Casting The Cortex Of The Human Brain In Silicon. *Technical Report TR-47.Center For Computational Research In Economics And Management Science.*

Patterson, D. W. (1996). *Artificial Neural Networks.Theory and Applications.* New Jersey: Prentice Hall.

Peramunetilleke, D., & Wong, R. K. (2002). *Currency Exchange Rate Forecasting from News Headlines.* Retrieved Mac 15, 2004, from World Wide Web: http://crpit.com/confpapers/CRPITV5Peramunetilleke.pdf

Qi, M., & Yang, S. (2003). Forecasting consumer credit card adoption: What Can We Learn About The Utility Function? *International Journal of Forecasting, 17*, 71-85.

Roush, W. B., & Cravener, T. L. (1997). *Artificial Neural Network Prediction of Amino Acid Level in Feed Ingredients.* Retrieved October 9, 2003, from World Wide Web: http://www.poultryscience.org/toc/papers/97/ps97721.pdf

Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures. *Decision Sciences,, Vol. 23*, pp. 899–916.

Sarle, S. W. (1997). *Neural Network FAQ, part 1 of 7:Introduction.* Retrieved Mac 2, 2004, from World Wide Web: ftp://ftp.sas.com/pub/neural/FAQ.html

Schalkoff, R. J. (1997). *Artificial Neural Networks*: The MacGraw-Hill Companies, Inc.

Shachmurove, Y. (2001). *Applying Artificial Neural Networks to Business, Economics and Finance.* Retrieved Mac 21, 2004, from World Wide Web: http://www.econ.upenn.edu/Centers/CARESS/CARESSpdf/02-08.pdf.

Skabar, A., & Cloete, I. (2000). *Neural Networks, Financial Trading and the Efficient Markets Hypothesis.* Retrieved Mac 14, 2004, from World Wide Web: http://www.geocities.com/ francorbusetti/ChengEssay.pdf

Skapura, D. M. (1995). *Building Neural Networks.* New York: Addison-Wesley Publishing Company.

Smith, K. A., Willis, R. J., & Brooks, M. (2000). An Analysis Of Customer Retention And Insurance Claim Patterns Using Data Mining: A Case Study. *Journal Of The Operations Research Society, 51*, 532-541.

Stergiou, C., & Siganos, D. (1999). *Neural Networks.* Retrieved April,2003, from World Wide Web: http://www.neural.com

Stolorz, P., Lapedes, A., & Yuan, X. (1995). *Predicting Protein Secondary Structure Using Neural Net and Statistical Methods.* Retrieved October 10, 2003, from World Wide Web: http://netec.wustl.edu/BibEc/data/Papers/wopsafiwp95-02-014.html

Tan, C. N. W. (1999). *An Artificial Neural Networks Primer with Financial. Applications Examples in Financial Distress Predictions and Foreign Exchange Hybrid Trading System.* Retrieved April 2, 2004, from World Wide Web: http://www.smartquant.com/references/NeuralNetworks/neural28.pdf

Tkacz, G. (2001). Neural Network Forecasting Of Canadian GDP Growth. *International Journal of Forecasting, 17*, 57-69.

Tsoukalas, L. H. (1998). Neurofuzzy Approaches to Anticipation: A New Paradigm for Intelligent Systems. *IEEE Transactions On Systems, Man, And Cybernetics-Part B: Cybernetics, 8*(4).

Vellido, A., Lisboa , P. J. G., & Vaughan , J. (1999). Neural Networks In Business: A Survey Of Applications (1992-1998). *Expert Systems With Applications, 17*, 51-70.

Wei, C., Wei, I., Lorry, W., & Chein, H. L. (1996). *Forecasting the 30-year U.S. Treasury Bond with a System of Neural Networks*. Retrieved Mac 24, 2004, from World Wide Web: http://www.geocities.com/francorbusetti/ChengEssay.pdf.

Wenrui, H., Murray , C., Kraus, N., & Rosati, J. (2003). Development Of A Regional Neural Network For Coastal Water Level Predictions. *Extended Technical Note. Ocean Engineering, 30*, 2275-2295.

Werbos, P. J. (1994). *Beyond Regression: New Tools For Prediction And Analysis In The Behavioral Science.* Unpublished PhD Dissertation, Harvard University.

Wong, B. K., Bodnovich, T. A., & Selvi, Y. (1995). A Bibliography of Neural Network Business Applications Research. *Expert Systems*, August 1995.

Worzala, E., Lenk, M., & Silva, A. (1995). An Exploration of Neural Networks and its Application to Real Estate evaluation. 1. *The Journal of Real Estate Research, Vol. 10*(2), pp.185–120.

Yao, J., & Poh, H. L. (1995). *Equity Forecasting A Case Study On The KLSE Index.* Retrieved Mac 5, 2004, from World Wide Web: http://citeseer.ist.psu.edu/yao95equity.html

Zahedi, F. (1993). *Intelligent Systems for Business: Expert Systems with Neural Networks*. Belmont,USA: Wadsworth Publishing Company.