UNIVERSITI PUTRA MALAYSIA

APPLICATION OF DATA MINING TECHNIQUES FOR ECONOMIC EVALUATION OF AIR POLLUTION IMPACT AND CONTROL

IING LUKMAN

FPAS 2007 7

APPLICATION OF DATA MINING TECHNIQUES FOR ECONOMIC EVALUATION OF AIR POLLUTION IMPACT AND CONTROL

IING LUKMAN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

August 2007

To all my family,

My teachers, IPB, UPM,

and Tanah Air……Indonesia

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of requirement for the degree of Doctor of Philosophy

APPLICATION OF DATA MINING TECHNIQUES FOR ECONOMIC EVALUATION OF AIR POLLUTION IMPACT AND CONTROL

By

IING LUKMAN

August 2007

Chairman: Associate Professor Mohd Nasir Hassan, PhD

Faculty : Environmental Studies

In this research we examine aspects of the interdependence between economic development and the use of environmental and natural resources assets from global data published by United Nations. For that purpose, we use data mining techniques.

Data mining techniques applied in this thesis were: 1) Group method of data handling (GMDH), originally from engineering, introducing principles of evolution - inheritance, mutation and selection - for generating a network structure systematically to develop the automatic model, synthesis, and its validation; 2) The weighted least square (WLS) and step wise regression were also applied for some cases; 3) The classification-based association rules were applied.

Data sets for this research consist of two sets integration data of air quality data and macroeconomic data of the cross-country data of World Development Indicator 2003 (WDI 2003), and from www.nationmaster.com. The results from www.nationmaster.com were as follows: the corruption index was strongly related to the urban SO_2 concentration. The corruption index along with NO_x emission has big contribution to the debt. Debt is the debt of the home country to the foreign country or external debt or foreign debt.

The result from WDI 2003 shows that the mortality rate of children under five years old depended on sanitation and water facilities obtained from GMDH results. However, the results from stepwise regression shows that mortality rate was dependent on annual deforestation, particulate matter, nationally protected area, but the big contribution was from annual deforestation.

Based on GMDH, new Gross National Income (GNI) formula was found. Previously GNI was known as Gross National Product (GNP). It was different from the common formula of GNP. The formula or equation model of urban SO_2 concentration was also found through the GMDH algorithms. The results were then compared to WLS and Stepwise regression.

The debt was found by GMDH to be dependent on the corruption index as well as urban SO_2 concentration. Corruption index along with NO_x emission were related to debt.

Results from weighted least square using SAS software showed that the corruption index was significant to the concentration of urban SO_2.

Results from classification rules of the WDI 2003 data showed that the more energy imports net from foreign country was associated with the smaller in adjusted net saving in home country. Energy imports net were calculated as energy use in oil equivalents. This indicated that if the energy imports net was higher, then the adjusted net saving was small, and then CO_2 emissions was small also. Thus, to reduce global warming in home country, a country can import energy from foreign country. According to the result from association rules on nationamaster.com data there were indication that corruption index was related with higher urban SO_2 concentration, and inflation.

Results from association rules of item sets shows that the urban SO_2 always follows the direction of corruption index. In addition, if any country wants to reduce the urban SO_2 concentration, more works can be conducted on controlling corruption index than controlling SO_2 emission per populated area.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

APLIKASI TEKNIK PERLOMBONGAN DATA UNTUK PENILAIAN EKONOMI BAGI IMPAK DAN KAWALAN PENCEMARAN UDARA

Oleh

IING LUKMAN

Ogos 2007

Pengerusi: **Profesor Madya Mohd Nasir Hassan, PhD**

Fakulti : **Pengajian Alam Sekitar**

Dalam kajian ini, aspek saling bergantung antara pembangunan ekonomi dengan penggunaan harta alam sekitar dan sumber semulajadi daripada data global yang diterbitkan oleh Bangsa-bangsa Bersatu dikaji. Bagi tujuan itu, kami gunakan kaedah-kaedah perlombongan data.

Kaedah pelombongan data yang digunakan dalam tesis ini adalah: 1) Kaedah kumpulan dalam pengendalian data (GMDH) berasal dari bidang kejuruteraan yang memperkenalkan prinsip evolusi –warisan, mutasi dan pilihan – untuk menjana satu struktur rangkaian yang teratur dalam pembinaan model automatik, sintesis and pengesahannya; 2) Regresi kuasa-dua berpemberat (WLS) dan regresi langkah cerdik juga diaplikasikan untuk sesetengah kes; 3) Peraturan asosiasi berasaskan klasifikasi digunakan.

Set data untuk kajian ini terdiri daripada dua set integrasi data dari data kualiti udara dan data makroekonomi dari kumpulan data serata dunia dari penunjuk pembinaan dunia 2003 (WDI 2003), dan juga dari www.nationmaster.com. Keputusan dari www.nationmaster.com adalah seperti berikut: indeks rasuah berkait rapat dengan kandungan urban $S0_2$. Indeks rasuah berserta kepulan NO_x memiliki sumbangan yang besar kepada hutang. Hutang bermakna hutang sesebuah negara sendiri ke negara asing atau hutang ke dunia luar atau hutang ke negara luar.

Keputusan dari WDI 2003 menunjukkan bahawa kadar kematian kanak-kanak dibawah umur lima tahun bergantung kepada kebersihan dan fasiliti air yang diperolehi dari keputusan GMDH. Walau bagaimanapun, keputusan dari regresi langkah cerdik menunjukkan kadar kematian bergantung kepada kepupusan setiap tahun kawasan perhutanan, jirim debu halus, kawasan lindungan negara, tetapi sumbangan besar yalah dari kepupusan setiap tahun kawasan perhutanan.

Berasaskan pada GMDH, kemudiannya formula pendapatan kotor kebangsaan (GNI) yang baru telah ditemui. Dahulunya GNI dikenali sebagai hasil kotor kebangsaan (GNP). Ianya berbeza dengan formula biasa hasil kotor kebangsaan. Formula atau persamaan model kandungan urban SO_2 juga ditemui menggunakan algoritma GMDH. Keputusannya kemudian diperbandingkan dengan keputusan dari WLS dan regresi langkah cerdik.

Hutang ditemui oleh GMDH sebagai bergantung pada indeks rasuah sepertimana kandungan urban SO_2. Index rasuah bersama dengan kepulan NO_x berkait kepada hutang. Keputusan dari WLS menggunakan perisian SAS menunjukkan bahawa indeks rasuah penting kepada kandungan urban SO_2.

Keputusan dari peraturan klasifikasi daripada data WDI 2003 menunjukan bahawa lebih banyaknya impot bersih tenaga dari negara asing berkaitan dengan simpanan bersih terubahsuai yang kecil di negara sendiri. Impot tenaga bersih dicongak sebagai penggunaan tenaga yang disamakan dengan penggunaan minyak. Ini bermakna bahawa jika impot bersih tenaga adalah tinggi, maka simpanan bersih terubahsuai adalah kecil, dan kepulan CO_2 pun kecil. Maka, untuk mengurangkan kepanasan global di negara sendiri, sesebuah negara boleh mengimpot tenaga dari negara asing. Berasaskan pada keputusan dari peraturan asosiasi pada data nationmaster.com ada tanda bahawa indeks rasuah berkait dengan kandungan urban SO_2 dan juga inflasi.

Keputusan dari peraturan asosiasi daripada set-set item menunjukkan bahawa urban SO_2 selalu mengikut hala indeks rasuah. Dengan demikian, jika sesebuah negara ingin mengurangkan kadar kandungan urban SO_2, lebih banyak kerja yg boleh dilaksanakan dengan mengawal indeks rasuah daripada dengan mengawal kepulan SO_2 pada setiap kawasan penduduk.

ACKNOWLEDGEMENTS

All praises do to Allah SWT, Lord of the universe. Only by His grace and mercy this thesis can be completed.

I wish to express my sincere thanks to my supervisor Associate Prof. Dr. Mohd Nasir Hassan for his invaluable guidance, support and continuous encouragement throughout the course of the project.

My gratitude also goes to the member of my supervisory committee, Assoc. Prof. Dr. Noor Akma Ibrahim for unending helps during my study, starting from giving me guidance, great advice, until trying to find me some alternative resource fund for my living costs, and providing much time for the methodology discussion, and checking the whole content of this thesis draft. My gratitude also goes to another supervisory committee Assoc. Prof. Dr. Md Nasir Sulaiman for his useful suggestions and helpful comments during the preparation of this thesis.

I also wish to express my sincere gratitude to Dr. Gregory Alexander Ivakhnenko from National Institute for Strategic Studies of the Council of National Security and Defense of Ukraine, Kiev, Ukraine for sending me the GMDH Software and its source codes, and also to Dr Bing Liu from National University of Singapore for obtaining the CBA Software.

Grateful acknowledgement is made to Malaysian Government, Universiti Putra Malaysia, and Department of Mathematics for giving me a room for doing research.

Financial supports from Malaysian Government under the IRPA programme during my time of study, financial supports from my siblings (especially Ali Abdurrahman SH, MH and Drs Aziz Taufik Hirzi, MSi) and from best friend Ir. Slamet Soedarsono, MSc of BAPPENAS Jakarta, and from excellent friend Ir. Elfajri Hanif, Jakarta, and from Uni Sjarkiah Muhammad Yasie, Kelantan, all are gratefully acknowledged.

Special thanks are extended to other members of the academic and technical staff of the Mathematics Department, Institut Penyelidikan Matematik (INSPEM), and Environmental Studies Faculty (especially Assoc. Prof. Dr. M Pauzi Zakaria), students and friends who helped me in every way possible and providing a congenial and enthusiastic atmosphere in the laboratory.

Acknowledgement is also extended to Indonesian Student Association, for giving me moral supports, the sweet friendship that made life easier during my difficult-time in Malaysia, especially to Dollaris Riauaty Suhadi, Abdul Kudus, and Aris Slamet Widodo .

I wish to express my deepest gratitude to my late parents, brothers and sisters for their prayers, continuous moral support and unending encouragement.

Last but not least, I wish especially to acknowledge my beloved wife, Dr Maria Viva Rini, and my dearest sons (Tanukh and Fatih) and daughters (Haifa, Putri, and Ghaida) for their love, support, patience and understanding.

I certify that an Examination Committee met on 20 April 2007 to conduct the final examination of Iing Lukman on his Doctor of Philosophy thesis entitled "Data Mining for Economic Evaluation of Air Pollution Impacts and Controls" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

Ahmad Makmum Abdullah, PhD
Associate Professor
Department of Environmental Management
Faculty of Environmental Studies
Universiti Putra Malaysia
(Chaiman)

Latifah Binti Abd Manaf, PhD
Lecturer (Assistant Professor)
Department of Environmental Sciences
Faculty of Environmental Studies
Universiti Putra Malaysia
(Internal Examiner)

Isa Bin Daud, PhD
Associate Professor
Department of Mathematics
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Abdul Razak Bin Hamdan, PhD
Professor
Department of System and Management Sciences
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
(External Examiner)

_____

HASANAH MOHD. GHAZALI, PhD
Professor / Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee are as follows:

Mohd Nasir Hassan, PhD
Associate Professor
Faculty of Environmental Studies
Universiti Putra Malaysia
(Chairman)

Noor Akma Ibrahim, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Md Nasir Sulaiman, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

AINI IDERIS, PhD
Professor /Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 9 August 2007

TABLE OF CONTENTS

LIST OF TABLES