

THE UNIVERSITY OF THE SOUTH PACIFIC
SCHOOL OF DISTANCE EDUCATION

TECHNOLOGY AND THE ENVIRONMENT

ENVIRONMENTAL SCIENCE AND TECHNOLOGY
TECHNOLOGY AND THE ENVIRONMENT

**PERLOMBONGAN DATA : KAJIAN TERHADAP TEKNIK-TEKNIK
PENORMALAN DATA DI DALAM PENUKARAN DATA**

MOHD. FIRDAUS BIN ABD. RAHMAN

**Tesis Ini Dikemukakan Bagi
Memenuhi Sebahagian Daripada Syarat Untuk
Memperolehi Sarjana Muda Teknologi Maklumat
Kolej Universiti Sains dan Teknologi Malaysia (KUSTEM)**

1100043479



**JABATAN SAINS KOMPUTER
FAKULTI SAINS DAN TEKNOLOGI**

**PENGAKUAN DAN PENGESAHAN LAPORAN
PROJEK PENYELIDIKAN II**

Adalah ini diakui dan disahkan bahawa laporan penyelidikan bertajuk:

**Perlombongan Data : Kajian Terhadap Teknik-teknik Penormalan Di Dalam Penukaran
Data**

Oleh **Mohd. Firdaus bin Abd. Rahman**, No.Matrik **UK 6410** telah diperiksa dan semua pembedaan yang disarankan telah dilakukan. Laporan ini dikemukakan kepada Jabatan Sains Komputer sebagai memenuhi sebahagian daripada keperluan memperoleh **Ijazah Sarjana Muda Teknologi Maklumat (Kejuruteraan Perisian)**, Fakulti Sains dan Teknologi, Kolej Universiti Sains dan Teknologi Malaysia.

Disahkan oleh :

Penyelia Utama

Pn. Zuriana Abu Bakar

Cop Rasmi:

ZURIANA ABU BAKAR
Pensyarah
Jabatan Sains Komputer
Fakulti Sains dan Teknologi
Kolej Universiti Sains dan Teknologi Malaysia
(KUSTEM)
21030 Kuala Terengganu.

Tarikh:

04.04.05

Ketua Jabatan Sains Komputer

P.M Dr. Mustafa Mat Deris

Cop Rasmi:

Prof. Madya Dr. Mustafa Mat Deris
Ketua
Jabatan Sains Komputer
Fakulti Sains & Teknologi
Kolej Universiti Sains dan Teknologi Malaysia
21030 Kuala Terengganu.

Tarikh:

04.04.05

PENGAKUAN

Dengan ini saya mengakui bahawa segala karya ini adalah hasil kerja saya sendiri kecuali nukilan dan ringkasan yang tiap-tiap satunya telah saya jelaskan sumbernya.

.....
MOHD. FIRDAUS BIN ABD. RAHMAN

4 APRIL 2005

Disahkan oleh :



.....
P.M Dr. Mustafa Mat Deris
Ketua
Jabatan Sains Komputer
Fakulti Sains dan Teknologi
Kolej Universiti Sains dan Teknologi Malaysia



.....
Pn. Zuriana Abu Bakar
Pensyarah
Jabatan Sains Komputer
Fakulti Sains dan Teknologi
Kolej Universiti Sains dan Teknologi Malaysia

PENGHARGAAN

Syukur ke hadrat Ilahi dengan limpah kurniaNya, saya berjaya menyiapkan kertas saranan projek yang dikehendaki tepat pada masanya. Ingin juga saya menyampaikan ribuan terima kasih kepada Pn. Zuriana binti Abu Bakar selaku penyelia projek yang telah banyak meluangkan masa dengan membantu memberikan pandangan, komen serta tunjuk ajar bagi menjayakan kajian ini.

Setinggi-tinggi penghargaan juga kepada kedua ibu bapa saya yang galak memberikan sokongan yang tidak pernah putus serta berkat doa kalian akhirnya segala tugas dan cabaran sebagai pelajar Kolej Universiti Sains & Teknologi Malaysia berjaya ditempuhi

Ribuan terima kasih juga kepada rakan-rakan seperjuangan yang turut sama menjayakan kajian ini samada langsung ataupun tidak langsung. Akhir sekali, terima kasih kepada semua pihak yang membantu di dalam menyiapkan kajian ini. Segala jasa baik, bantuan dan sokongan yang diberikan oleh semua pihak amatlah dihargai.

DATA MINING : A STUDY ON NORMALIZATION TECHNIQUES IN DATA TRANSFORMATION

ABSTRACT

Data normalization is the one of the most important element in KDD process (Knowledge Discovery in Databases). The objective of normalization is to ensure the training data will be in standardized format. The purpose of this research is to give a view or detail briefing of normalization techniques and the differentiate of it. The techniques that involve in this research are Decimal scaling, min-max normalization and normalization using Z-score. This technique also will be a analyst in purpose of term and condition that will make all of the normalization an efficient technique. The data for this research will be a numerical data of air pollution and tornado. This data is a gift from Alam Sekitar Malaysia (ASMA) and the tornado taken from the KDD site. This research will give the differentiate of each technique and will answer the question on what technique is the best. This research will also produce a new fact on data normalization for purpose on data mining.

ABSTRAK

Penormalan data adalah salah satu elemen penting di dalam proses KDD (*Knowledge Discovery in Databases*). Objektif penormalan data adalah untuk memastikan data yang terlibat berada di dalam format yang boleh diproses. Tujuan kajian ini dijalankan adalah untuk memberikan gambaran yang terperinci tentang teknik penormalan serta perbezaan di antaranya. Kaedah penormalan yang terlibat di dalam kajian ini adalah penormalan *Decimal scaling*, penormalan *Min-max* dan penormalan *Z-skor*. Data yang digunakan di dalam kajian ini adalah berbentuk numerik pencemaran udara dan tornado. Data tersebut diperolehi daripada sumbangan pihak Alam Sekitar Malaysia (ASMA) dan data tornado daripada laman web KDD sendiri. Kajian ini akan memberikan perbezaan terhadap kaedah penormalan yang dikaji serta memberikan jawapan terhadap kaedah yang terbaik. Kajian ini juga telah menghasilkan penemuan-penemuan yang berguna dan membawa manfaat, terutama kepada perlombongan data.