

## Penggunaan Kaedah Statistik Di Dalam Pengelompokan Data Kajucaaca Bagi Meramal Taburan Hujan

Rozilawati Dollah @ Md. Zain<sup>1</sup>, Mohd Noor Md Sap<sup>2</sup>,  
Mahadi Bahari<sup>3</sup>, Aryati Bakri<sup>4</sup>  
Jabatan Sistem Maklumat  
Fakulti Sains Komputer & Sistem Maklumat  
Universiti Teknologi Malaysia  
81310 Skudai, Johor  
{zeela<sup>1</sup>,mohdnoor<sup>2</sup>,mahadi<sup>3</sup>,aryati<sup>4</sup>@fsksm.utm.my}

### Abstrak

Meramal cuaca merupakan satu tugas yang kompleks dan sukar bagi pihak Jabatan Perkhidmatan Kajucaaca Malaysia (JPKM) ekoran pertambahan data kajucaaca di dalam pangkalan data. Di antara parameter-parameter data kajucaaca yang disimpan, taburan hujan merupakan parameter yang paling sukar diramal kerana ia melibatkan pelbagai parameter data kajucaaca seperti taburan hujan, suhu, kelajuan angin, tekanan, kelembapan dan sebagainya. Salah satu cara yang boleh diaplikasikan untuk meramal taburan hujan adalah melalui pengelompokan parameter data kajucaaca dengan menggunakan kaedah statistik. Kajian ini memberikan penumpuan kepada operasi pengelompokan data kajucaaca di mana hasil dari pengelompokan ini akan digunakan sebagai input untuk meramal taburan hujan. Hasil kajian ini menunjukkan pengelompokan data kajucaaca dengan menggunakan kaedah statistik dapat membantu pihak peramal di JPKM untuk melakukan peramalan taburan hujan. Di samping itu, hasil kajian juga telah mendapati bahawa penggunaan kelompok data kajucaaca yang berlainan sebagai data input kepada proses peramalan taburan hujan telah memberikan prestasi peramalan yang lebih baik berbanding dengan penggunaan kelompok data kajucaaca yang sama.

*Katakunci:* taburan hujan, pengelompokan, penganalisaan data, kaedah statistik, peramalan

### 1.0 Pengenalan

Peramalan cuaca merupakan salah satu daripada masalah yang paling mencabar di dalam dunia sejak lebih setengah abad yang lalu. Di Malaysia, tugas memantau situasi perubahan cuaca dan mengeluarkan kenyataan, nasihat dan amaran cuaca merupakan antara peranan yang dilakukan oleh Jabatan Perkhidmatan Kajucaaca Malaysia (JPKM) (JPKM, 2000). Jabatan ini memerlukan maklumat data kajucaaca yang terdiri daripada parameter-parameter seperti data taburan hujan, suhu, angin, kelembapan, sejatan dan sinaran suria (Sani, 1984) yang tepat dari setiap stesennya untuk membuat jangkaan atau peramalan yang terperinci tentang perubahan cuaca sebelum ianya dihebahkan kepada pihak umum. Namun begitu, dari semasa ke semasa, JKPM menghadapi masalah untuk melakukan proses penganalisaan ke atas jumlah data kajucaaca yang kian meningkat memandangkan ianya dilakukan secara manual oleh beberapa orang peramal. Tambahan pula, disebabkan oleh nilai praktikal peramalan cuaca di dalam kajucaaca, ia juga merupakan masalah peramalan siri masa yang “unbiased” di dalam penyelidikan saintifik (Liu dan Lee, 1999).

Di dalam kajian ini, penyelidik telah memilih data taburan hujan sebagai parameter kaji cuaca disebabkan ianya melibatkan pelbagai pembolehubah yang saling berhubungkait dengan cara yang rumit. Mengikut Liu dan Lee (1999), di antara parameter-parameter kajucaaca, taburan hujan merupakan parameter yang paling sukar diramal. Kebanyakan perhubungan di dalam taburan hujan adalah menggambarkan hubungan ruang dan dinamik yang tidak linear (Chen dan Takagi, 1993). Data taburan hujan juga merupakan salah satu daripada elemen atau ciri yang penting dan menarik di dalam iklim yang sering dibincang dan diperkatakan. Hal ini disebabkan data taburan hujan bagi sesuatu tempat akan memberikan pengaruh kepada bidang pertanian dan juga menimbulkan masalah-masalah kepada manusia, makhluk-makhluk lain serta tanaman-tanaman, iaitu bencana alam seperti banjir, kemarau dan sebagainya (Arakawa, 1969).

Terdapat pelbagai kajian yang telah dilakukan oleh penyelidik ke atas peramalan atau penganggaran taburan hujan sebelum ini. Namun, kebanyakan kajian contohnya kajian yang dijalankan oleh Liu dan Lee (1999), McCullagh et. al (1999) serta Chen dan Takagi (1993) telah menggunakan teknik perlombongan data rangkaian neural pintar untuk melakukan operasi pengkelasan terhadap data taburan hujan tersebut. Sungguhpun begitu, dalam kajian ini, penyelidik memfokus kepada operasi pengelompokan di mana hasil dari pengelompokan ini akan digunakan untuk melakukan peramalan taburan hujan. Teknik ini dipilih memandangkan terdapat banyak kajian yang melibatkan pengelompokan telah berjaya meningkatkan hasil

peramalan yang dilakukan. Sebagai contoh, kajian yang dijalankan oleh Sarjon dan Mohd Noor (2000), mendapati pengelompokan dapat meningkatkan peramalan yang dilakukan. Penyelidik beranggapan, jika kaedah ini berjaya dilakukan ke atas data kajicuaca di JKPM, ianya akan dapat membantu jabatan ini untuk meramal taburan hujan.

Kandungan kertas kerja ini terbahagi kepada lima seksyen utama. Seksyen pertama merupakan pengenalan kepada kajian manakala operasi pengelompokan ke atas data kajicuaca akan diperjelaskan dengan terperinci di dalam seksyen dua nanti di samping penggunaan kaedah statistik sebagai teknik pengelompokan. Seksyen tiga akan menerangkan mengenai metodologi yang digunakan di dalam kajian ini. Seksyen empat pula mengupas mengenai beberapa siri eksperimen dan analisis keputusan yang dijalankan. Kajian lanjutan akan diperbincangkan dalam seksyen lima dan kertas kerja ini diakhiri dengan kesimpulan bagi tajuk yang telah dibincangkan.

## **2.0 Teknik Pengelompokan**

Pengelompokan merupakan kaedah untuk menghimpunkan vektor-vektor yang hampir sama atau mempunyai persamaan antara satu sama lain berdasarkan ciri-ciri datanya (Kaski, 1997). Secara amnya pengelompokan ialah memecahkan data-data kepada data yang mempunyai ciri-ciri yang sama atau hampir sama. Masalah pengelompokan data berlaku di dalam pengecaman corak, statistik, pembelajaran tanpa seliaan, rangkaian neural, perlombongan data, pembelajaran mesin dan pelbagai bidang saintifik yang lain (Hofmann dan Buhmann, 1997). Dalam kertas kerja ini, kami akan menerangkan mengenai kaedah statistik.

### **2.1 Kaedah Statistik**

Statistik merupakan fungsi yang memerihalkan ciri sesuatu sampel seperti min, sisihan piawai, kadaran dan sebagainya. Kaedah ini dapat menolong pihak yang melakukan ujikaji dalam membentangkan data dengan lebih jelas dan dalam bentuk yang mudah difahami. Statistik adalah merupakan kaedah-kaedah dan teori yang digunakan untuk memahami data. Data merupakan cerapan (biasanya berbentuk numerik) yang mewakili ciri-ciri tertentu sesuatu yang dikaji.

Terdapat banyak perisian statistik komersial dalam pasaran sekarang seperti Statistical Packages for Social Sciences SPSS, MINITAB, dan SAS. Perisian hampan elektronik seperti Excel dan kalkulator graf TI-83 juga dapat digunakan untuk tujuan analisis statistik (Hong, 2000).

Kaedah statistik adalah popular terutamanya di dalam kajian yang melibatkan analisis data dan peramalan data siri masa. Terdapat pelbagai jenis model untuk peramalan menggunakan teknik siri masa, di antaranya ialah model purata bergerak, model autoregrasi, model haluan linear, model gabungan purata bergerak terkamir autoregrasi dan sebagainya (Yeong, 1996 dan Rosdiyah, 1999). Banyak kajian yang dijalankan telah menggunakan kaedah statistik ini untuk membina model peramalan yang melibatkan data siri masa. Di antaranya ialah kajian yang dilakukan oleh Rosdiyah (1999) dan Yeong (1996).

Yeong (1996) di dalam kajiannya telah membuat perbandingan di antara model-model yang dinyatakan di atas dengan menggunakan pakej Statistica dan SAS (Sistem Analisis Statistik) bagi tujuan menganalisa dan meramal kadar hujan. Hasil kajian menunjukkan model yang dihasilkan dengan menggunakan pakej Statistica menghasilkan peramalan yang lebih jitu. Manakala kajian yang dilakukan oleh Rosdiyah (1999) pula telah membuat peramalan taburan hujan menggunakan model '*Transfer Function*'.

### 3.0 Metodologi Kajian

Terdapat pelbagai cara untuk mengelompokkan data kajiucua di dalam satu pangkalan data. Di antaranya ialah dengan menggunakan kaedah statistik. Kaedah statistik ini digunakan untuk mencari persamaan di antara atribut-atribut di dalam pangkalan data kajiucua bagi tujuan pengelompokan. Oleh yang demikian, pelaksanaan kajian ini adalah melibatkan beberapa aktiviti, di antaranya ialah;

- i) **Mengumpul dan menganalisa data kajiucua** : data yang digunakan di dalam kajian ini adalah terdiri daripada data kajiucua (setiap jam) dari bulan September 1993 hingga bulan Februari 2001, yang diperolehi daripada stesen pencerapan Kluang (MPOB). Jadual 1 di bawah menunjukkan sample data kajiucua. Di antara atribut data kajiucua yang digunakan dalam kajian ini ialah *windvane*, kelembapan, *energy*, suhu, tekanan, *radiation*, kelajuan angin dan taburan hujan. Lampiran A menunjukkan sebahagian daripada sampel data kajiucua yang digunakan.

**Jadual 1 : Sampel data kajiucua**

ID	Tarikh	Masa	Windvane	Kelembapan	.....	Taburan Hujan
1	8/1/00	0:00	197.0	0.0		0.0
2	8/1/00	1:00	201.2	0.0		0.0
3	8/1/00	2:00	206.5	0.0		0.0
...	...	...	...	...		...
...	...	...	...	...		...
250	8/11/00	9:00	222.5	0.0		0.0

- ii) **Mengira nilai purata dan sisihan piawai bagi data kajiucua** : berdasarkan kepada jadual sampel data kajiucua di atas, pengiraan nilai purata dan sisihan piawai bagi data kajiucua dilaksanakan dengan menggunakan kaedah statistik. Purata bagi suatu set data adalah merupakan hasil tambah semua nilai data dibahagi oleh jumlah bilangan data. Manakala sisihan piawai pula mengukur serakan dengan menunjukkan secara purata jarak cerapan-cerapan dari purata (min). Mengikut kaedah statistik, pengiraan nilai purata dan sisihan piawai ini boleh dilakukan ke atas atribut sampel data kajiucua tersebut dengan menggunakan formula-formula berikut dan **Jadual 2** di bawah menunjukkan nilai purata dan sisihan piawai bagi setiap atribut data kajiucua.

untuk pengiraan nilai purata;

$$\bar{X} = \frac{\sum Xi}{n} \quad \dots (1)$$

di mana;

$X_i$  = setiap nilai data individu  
 $n$  = jumlah bilangan data

dan untuk menghitung nilai sisihan piawai;

$$s = \sqrt{\frac{\sum x^2 - (\sum x)^2/n}{n - 1}} \quad \sigma = \sqrt{\frac{\sum x^2}{N} - \mu^2}$$

atau  $s = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n - 1}} \quad \dots (2)$

$s$  ialah sisihan piawai sampel  
 $x$  ialah cerapan dalam set data sampel  
 $\bar{x}$  ialah min untuk set data sampel  
 $\sum$  ialah hasil tambah  
 $n$  ialah jumlah bilangan cerapan

$\sigma$  ialah sisihan piawai populasi  
 $x$  ialah cerapan dalam set data populasi  
 $\mu$  ialah min untuk set data populasi  
 $\sum$  ialah hasil tambah  
 $N$  ialah jumlah bilangan cerapan

**Jadual 2 : Nilai purata dan sisihan piawai bagi sampel data kajicuaca**

Atribut	Purata	Sisihan Piawai
<i>Windvane</i>	182	101
Kelajuan Angin	0.019	0.095
Kelembapan	4.7	5.8
Suhu	0.3	8.5
<i>Energy</i>	87	170
Tekanan	19	49
<i>Radiation</i>	139	280

Nilai purata dan sisihan piawai ini akan digunakan untuk mengira persamaan di antara atribut data kajicuaca.

- iii) **Mengira persamaan di antara data kajicuaca** : proses mengira persamaan di antara atribut-atribut data kajicuaca dilakukan dengan mengira jarak di antara atribut-atribut tersebut dengan menggunakan nilai purata dan sisihan piawai yang telah diperolehi. Pengiraan jarak ini bertujuan untuk mengenalpasti atribut-atribut yang mempunyai persamaan yang kuat dan lemah bagi tujuan pengelompokan. Persamaan di antara atribut-atribut tersebut dinilai berdasarkan kepada jarak yang diperolehi, di mana jarak yang terkecil dikira sebagai atribut yang mempunyai persamaan yang tinggi dan sebaliknya. Pengiraan persamaan ini dilakukan dengan menggunakan formula jarak Euclidean seperti berikut (Korkmaz, 1997; Sarjon dan Mohd Noor, 2000).

$$d_{ij} = \sqrt{(x_{ik} - x_{jk})^2} \quad \dots (3)$$

di mana ;

- $d_{ij}$  = jarak di antara kes i dan j
- $x_{ik}$  = nilai pembolehubah  $k_{th}$  bagi kes  $i_{th}$
- $x_{jk}$  = nilai pembolehubah  $k_{th}$  bagi kes  $j_{th}$

- iv) **Mengelompokkan data kajicuaca berdasarkan persamaan** : data kajicuaca akan dikelompokkan kepada kumpulan masing-masing berdasarkan kepada hasil persamaan yang diperolehi daripada pengiraan jarak di antara atribut-atribut data kajicuaca. Persamaan di antara atribut-atribut dinilai berdasarkan kepada jarak yang diperolehi, di mana jarak yang terkecil dikira sebagai atribut yang mempunyai persamaan yang tinggi dan kuat dan begitulah sebaliknya. Kelompok ini ditakrifkan sebagai jarak yang terdekat di antara dua kelompok mengikut persamaan berikut dan Jadual 3 di bawah menunjukkan keputusan pengelompokan atribut data kajicuaca.

$$d(c_i, c_j) = \min(\{d.f(\gamma k, \gamma l) \mid \gamma k \subseteq c_i \text{ and } \gamma l \subseteq c_j\}) \quad \dots (4)$$

di mana ;

- $d(c_i, c_j)$  = jarak antara kelompok di antara dua kelompok tunggal
- $c_i = (\gamma k)$  dan  $c_j = (\gamma l)$

**Jadual 3 : Pengelompokan atribut data kajicuaca**

Kelompok	Penerangan
Kelompok 1	{ Kelembapan, suhu }
Kelompok 2	{ Kelajuan angin, Tekanan }
Kelompok 3	{ <i>Windvane</i> , <i>Energy</i> }
Kelompok 4	{ <i>Radiation</i> }

Daripada perbandingan jarak di antara atribut yang telah dilakukan, didapati 4 kategori kelompok yang telah dihasilkan. Kelompok-kelompok ini akan digunakan sebagai data input kepada proses peramalan taburan hujan.

- v) **Peramalan taburan hujan** : pengujian peramalan taburan hujan ini dilakukan dengan menggunakan program NeuNet Pro versi 2.3, di mana data kajicuaca yang telah dikelompokkan dengan menggunakan kaedah statistik akan menjadi input kepada proses peramalan tersebut. Sehubungan dengan itu, beberapa eksperimen yang melibatkan kelompok-kelompok data kajicuaca yang menggunakan kaedah statistik telah dilaksanakan. Ianya bertujuan untuk membuat perbandingan prestasi peramalan taburan hujan di antara kelompok data kajicuaca yang sama dan juga kelompok data kajicuaca yang berlainan.
- vi) **Penganalisaan dan perbandingan hasil peramalan** : keputusan peramalan taburan hujan telah digunakan untuk membuat penganalisaan dan perbandingan di antara kelompok data kajicuaca yang sama dengan kelompok data kajicuaca yang berlainan. Ianya bertujuan untuk melihat sejauhmanakah ketepatan dan keberkesanan kaedah statistik di dalam mengelompokkan data kajicuaca. Di samping itu, ianya juga bertujuan untuk melihat kesan penggunaan kelompok data kajicuaca manakah yang memberikan pengaruh kepada prestasi peramalan taburan hujan. Perbandingan ini dilakukan dengan mencari perbezaan di antara nilai taburan hujan yang sebenar dan nilai taburan hujan yang diramal. Selain daripada itu, nilai ralat min punca kuasa dua (RMS) dan pekali korelasi juga turut digunakan bagi tujuan perbandingan keberkesanan dan ketepatan peramalan.

#### 4.0 Eksperimen

Bagi tujuan pengujian kaedah statistik sebagai teknik pengelompokan, beberapa eksperimen pengujian terhadap keputusan peramalan taburan hujan yang dihasilkan telah dijalankan. Pengujian ini dilakukan dengan menggunakan kelompok-kelompok data kajicuaca yang dihasilkan seperti di dalam Jadual 3 di atas. Kelompok-kelompok ini digunakan sebagai data input kepada pakej NeuNetPro bagi melakukan peramalan taburan hujan. Di dalam proses peramalan, data taburan hujan digunakan sebagai nilai yang diramal. Pengujian peramalan ini dilaksanakan dengan melakukan beberapa eskperimen yang menggunakan atribut data kajicuaca yang berada di dalam kelompok yang sama dan juga atribut daripada kelompok yang berlainan. Ianya bertujuan untuk melihat ketepatan hasil peramalan.

Sehubungan dengan itu, tiga eksperimen yang melibatkan kelompok data kajicuaca yang sama telah dilaksanakan, iaitu;

- (a) Eksperimen Pertama – terdiri daripada kelompok 1 yang mempunyai atribut kelembapan dan suhu,
- (b) Eksperimen Kedua - iaitu kelompok 2 yang terdiri daripada atribut kelajuan angin dan tekanan, dan
- (c) Eksperimen Ketiga – terdiri daripada kelompok 3 yang mempunyai atribut *windvane* dan *energy*.

Jadual 4(a), (b) dan (c) berikut menunjukkan keputusan peramalan taburan hujan yang dihasilkan di dalam tiga eksperimen di atas.

**Jadual 4(a) : Keputusan peramalan taburan hujan bagi atribut di dalam Kelompok 1**

Tarikh	Masa	Taburan Hujan (mm <sup>3</sup> )	Nilai yang Diramal (mm <sup>3</sup> )	Perbezaan (mm <sup>3</sup> )
8/1/00	0:00	0	-0.03	0.03
8/1/00	1:00	0	-0.02	0.02
8/1/00	2:00	0	-0.02	0.02
...	...	...	...	...
8/11/00	9:00	0	0.22	0.22

Perbezaan di antara nilai sebenar taburan hujan dengan nilai yang diramal agak sekata dan rendah.

**Jadual 4(b) : Keputusan peramalan taburan hujan bagi atribut di dalam Kelompok 2**

Tarikh	Masa	Taburan Hujan (mm <sup>3</sup> )	Nilai yang Diramal (mm <sup>3</sup> )	Perbezaan (mm <sup>3</sup> )
8/1/00	0:00	0	0.23	0.23
8/1/00	1:00	0	0.23	0.23
8/1/00	2:00	0	0.24	0.24
...	...	...	...	...
8/11/00	9:00	0	0.21	0.21

Secara keseluruhannya, perbezaan di antara nilai yang diramal dengan nilai sebenar taburan hujan agak sekata.

**Jadual 4(c) : Keputusan peramalan taburan hujan bagi atribut di dalam Kelompok 3**

Tarikh	Masa	Taburan Hujan (mm <sup>3</sup> )	Nilai yang Diramal (mm <sup>3</sup> )	Perbezaan (mm <sup>3</sup> )
8/1/00	0:00	0	0.11	0.11
8/1/00	1:00	0	0.11	0.11
8/1/00	2:00	0	0.11	0.11
...	...	...	...	...
8/11/00	9:00	0	0.24	0.24

Perbezaan di antara nilai yang diramal dengan nilai sebenar taburan agak sekata dan lebih rendah berbanding dengan eksperimen 2. Daripada ketiga-tiga eksperimen yang dijalankan di atas, didapati prestasi peramalan taburan hujan yang dilaksanakan di dalam eksperimen pertama adalah lebih baik berbanding dengan dua eksperimen yang lain. Ini diperolehi daripada nilai ralatnya yang lebih rendah dan juga nilai pekali korelasinya yang lebih tinggi.

Manakala dua lagi eksperimen yang melibatkan peramalan taburan hujan menggunakan kelompok data kajiucua yang berlainan juga telah dilaksanakan. Di antara eksperimen-eksperimen tersebut ialah;

- (a) Eksperimen Keempat – terdiri daripada kelompok berlainan, iaitu atribut *windvane* dan suhu.
- (b) Eksperimen Kelima – mengandungi dua atribut daripada kelompok yang berlainan, iaitu atribut *energy* dan suhu.

Keputusan peramalan taburan hujan bagi kedua-dua eksperimen berikut adalah seperti di dalam jadual di bawah;

**Jadual 4(d): Keputusan peramalan taburan hujan menggunakan atribut *Windvane* dan Suhu**

Tarikh	Masa	Taburan Hujan (mm <sup>3</sup> )	Nilai yang Diramal (mm <sup>3</sup> )	Perbezaan (mm <sup>3</sup> )
8/1/00	0:00	0	-0.09	0.09
8/1/00	1:00	0	-0.10	0.10
8/1/00	2:00	0	-0.10	0.10
...	...	...	...	...
8/11/00	9:00	0	0.36	0.36

Perbezaan di antara nilai yang diramal dengan nilai sebenar taburan hujan adalah tidak sekata.

**Jadual 4(e) : Keputusan peramalan taburan hujan menggunakan atribut *Energy* dan Suhu**

Tarikh	Masa	Taburan Hujan (mm <sup>3</sup> )	Nilai yang Diramal (mm <sup>3</sup> )	Perbezaan (mm <sup>3</sup> )
8/1/00	0:00	0	-0.08	0.08
8/1/00	1:00	0	-0.09	0.09
8/1/00	2:00	0	-0.09	0.09
...	...	...	...	...
8/11/00	9:00	0	0.26	0.26

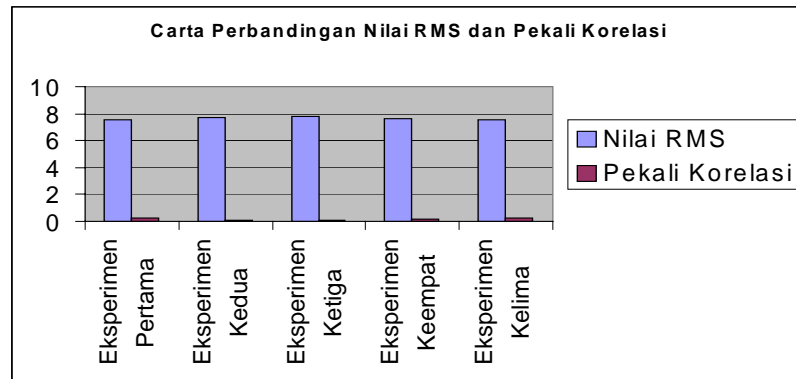
Perbezaan di antara nilai sebenar dengan nilai yang diramal adalah tidak sekata. Daripada keputusan lima eksperimen peramalan taburan hujan yang telah dijalankan dengan menggunakan kaedah statistik untuk mengelompokkan data kajiucaca, secara keseluruhannya bolehlah dikatakan prestasi peramalan yang dilakukan menggunakan atribut di dalam kelompok yang berlainan memberikan prestasi peramalan yang lebih baik daripada kelompok yang sama.

Walaupun bagaimanapun, berdasarkan carta perbandingan keputusan peramalan taburan hujan di bawah, didapati nilai RMS yang dihasilkan daripada kelima-lima eksperimen agak sekata tetapi nilai pekali korelasi bagi setiap eksperimen tersebut adalah tidak sekatan. Ini dibuktikan oleh hasil kajian di mana nilai RMS yang lebih kecil dan juga nilai pekali korelasinya yang lebih besar. Jadual 5 dan Rajah 1 di bawah menunjukkan pengukuran keputusan peramalan taburan hujan yang telah diperolehi.

**Jadual 5 : Keputusan nilai RMS dan pekali korelasi**

Eksperimen	Nilai RMS	Pekali Korelasi
<b>Eksperimen Pertama</b>	7.58	0.23
<b>Eksperimen Kedua</b>	7.73	0.11
<b>Eksperimen Ketiga</b>	7.77	0.05
<b>Eksperimen Keempat</b>	7.65	0.18
<b>Eksperimen Kelima</b>	7.51	0.26

Rajah 1 di bawah menunjukkan perbandingan di antara keputusan pengukuran prestasi peramalan taburan hujan yang diperolehi.



Rajah 1 : Perbandingan keputusan peramalan bagi setiap eksperimen

Berdasarkan carta perbandingan keputusan peramalan taburan hujan di atas, didapati nilai RMS yang dihasilkan daripada kelima-lima eksperimen agak sekata tetapi nilai pekali korelasi bagi setiap eksperimen tersebut adalah tidak sekata. Oleh yang demikian, secara keseluruhannya bolehlah dikatakan bahawa prestasi peramalan taburan hujan yang dilakukan dengan menggunakan atribut di dalam kelompok yang berlainan memberikan prestasi peramalan taburan hujan yang lebih baik daripada kelompok yang sama. Ini ditunjukkan oleh hasil kajian di mana nilai RMS yang lebih kecil dan juga nilai pekali korelasi yang lebih besar.

## 5.0 Perbincangan

Kajian terhadap teknik pengelompokan data kajiucua merupakan satu bidang yang menarik untuk diselidiki. Ini kerana operasi pengelompokan telah banyak digunakan di dalam pelbagai bidang terutamanya bidang pengelompokan aksara (Azah Kamilah, 1999), data tak berstruktur jenis teks (Siti Sakira, 1998), imej (Shahliza, 1999) dan sebagainya. Dan terdapat banyak teknik pengelompokan terutamanya teknik perlombongan data yang boleh digunakan untuk melakukan pengelompokan.

Eksperimen peramalan taburan hujan yang telah dijalankan di dalam kajian ini adalah bertujuan untuk melihat tahap keupayaan teknik pengelompokan dengan menggunakan kaedah statistik di dalam memberi nilai ramalan yang tepat ke atas data taburan hujan. Hasilnya menunjukkan bahawa teknik pengelompokan statistik dapat memberikan prestasi peramalan yang baik. Penyelidik juga mendapati bahawa penggunaan kelompok data kajiucua yang berlainan sebagai data input kepada proses peramalan taburan hujan telah memberikan prestasi peramalan yang lebih baik berbanding dengan kelompok data kajiucua yang sama.

Berdasarkan kepada hasil eksperimen, dapatlah dirumuskan bahawa teknik pengelompokan statistik dapat meningkatkan prestasi peramalan taburan hujan. Selain daripada itu, faktor kelompok yang digunakan sebagai data input kepada proses peramalan turut mempengaruhi keputusan peramalan. Hal ini berlaku, di mana penggunaan kelompok yang berlainan dapat memberikan prestasi peramalan yang lebih baik berbanding dengan penggunaan kelompok yang sama. Sepanjang kajian ini dilakukan, penyelidik tidak menggunakan set item data kajiucua yang tidak *frequent*. Penyelidik juga menganggap bahawa jarak di antara atribut data kajiucua yang paling kecil mempunyai ciri persamaan yang kuat dan begitulah sebaliknya. Selain daripada itu, data-data kajiucua yang digunakan dianggap bersih dan bebas dari hingar.

## 6.0 Kesimpulan

Kajian ini dijalankan untuk mengkaji teknik pengelompokan data kajiucua bagi tujuan meramal taburan hujan. Terdapat pelbagai cara yang boleh digunakan untuk melakukan pengelompokan data kajiucua, di antaranya ialah dengan menggunakan kaedah statistik, teknik perlombongan data dan sebagainya. Sehubungan dengan itu, kajian ini dilaksanakan dengan menggunakan teknik statistik untuk mengelompokkan data kajiucua bagi tujuan peramalan taburan hujan. Hasil kajian ini menunjukkan prestasi peramalan taburan hujan yang menggunakan kelompok atribut data kajiucua yang berlainan adalah lebih baik berbanding dengan menggunakan kelompok atribut data kajiucua yang sama.



## RUJUKAN

Arakawa, H. (1969). "Climates of Northern dan Eastern Asia." World Survey of Climatology, Vol.8, Amsterdam : Elsevier Publishing Company.

Azah Kamilah Muda (1999). "Pengekstrakan Ciri dan Perangkaian Aksara Taip Pelbagai Variasi : Kajian Perbandingan Antara Teknik Momen Geometric Tak Varian, Momen Regular dan Momen Legendre." Universiti Teknologi Malaysia:Tesis Sarjana.

Chen, Tao dan Takagi, Mikio (1993). "Rainfall Prediction of Geostationary Meteorological Satellite Images Using Artificial Neural Network." Jepun : University of Tokyo.

Hofmann, T. dan Buhmann, J. (1997). "Inferring Hierarchical Clustering Structures by Deterministic Annealing." [Online] Available : <http://www-dbv.cs.uni-bonn.de>

Hong, KS. (2000). "Statistik Untuk Sains Sosial". Nota Elektronik, Universiti Malaysia Sarawak. [Online] Available : [http://www.calm.unimas.my/hong/Laman\\_Web\\_Statistik/Teks-Statistik/statistik\\_teks.htm](http://www.calm.unimas.my/hong/Laman_Web_Statistik/Teks-Statistik/statistik_teks.htm)

Jabatan Perkhidmatan Kajucaua Malaysia (2000). [Online] Available : <http://www.kjc.gov.my/monsoonb.htm>

Kaski, S. (1997). "Data Exploration Using Self-Organizing Maps."Helsinki University of Technology : Tesis Ph.D.,

Korkmaz, E. (1997). " A Method For Improving Automatic Word Categorization." The Middle East Technical University : Tesis Sarjana.

Liu, James N. K. dan Lee, Raymond S.T. (1999). "Rainfall Forecasting from Multiple Point Sources Using Neural Networks." Hong Kong : Hong Kong Polytechnic University.

McCullagh, J., Bluff, K. dan Hendtlass, T. (1999). "Evolving Expert Neural Networks for Meteorological Rainfall Estimation." IEEE.

Rosdiyah Dahlan (1999). "Analisa Taburan Hujan Menggunakan Model Transfer Function." Universiti Teknologi Malaysia : Tesis Sarjana Muda.

Siti Sakira Kamarudin (1998). "Clustering Unstructured Data From A Text Database Using Self-Organizing Maps." Universiti Teknologi Malaysia : Tesis Sarjana.

Sham Sani (1984). "Beberapa Aspek Iklim Negeri Sabah." UKM : Bangi.

Sarjon Defit dan Mohd Noor Mad Sap (2000a). " Data Mining : A Preview ." Jurnal Teknologi Maklumat , Fakulti Sains Komputer dan Sistem Maklumat : UTM. Jilid 12, Bil.1.

Sarjon Defit dan Mohd Noor Mad Sap (2000b). " Predictive Data Mining Based On Similarity and Clustering Methods. "Jurnal Teknologi Maklumat , Fakulti Sains Komputer dan Sistem Maklumat : UTM. Jilid 12, Bil.2.

Shahliza Abd Halim (1999). "Feature Extraction and Clustering of Trademark Images : A Comparison between Geometric Invariant and Zernike Moment Invariant." Universiti Teknologi Malaysia : Tesis Sarjana.

Yeong, Yoke Wei (1996). "Penggunaan Model Siri Masa untuk Menganalisis dan Meramal Kadar Hujan." Universiti Teknologi Malaysia : Tesis Sarjana Muda.

**LAMPIRAN A**

<b>ID</b>	<b>Tarikh</b>	<b>Masa</b>	<b>Windvane</b>	<b>Kelembapan</b>	<b>.....</b>	<b>Taburan Hujan</b>
1	8/1/00	0:00	197.0	0.0		0.0
2	8/1/00	1:00	201.2	0.0		0.0
3	8/1/00	2:00	206.5	0.0		0.0
4	8/1/00	3:00	235.5	0.0		0.0
5	8/1/00	4:00	293.7	0.0		0.0
6	8/1/00	5:00	143.2	0.0		0.0
7	8/1/00	6:00	201.2	0.0		0.0
8	8/1/00	7:00	343.8	0.0		2.5
9	8/1/00	8:00	261.2	0.0		0.5
10	8/1/00	9:00	304.7	0.0		0.0
11	8/1/00	10:00	302.3	0.0		0.0
12	8/1/00	11:00	286.7	7.4		0.0
13	8/1/00	12:00	270.0	8.2		0.0
14	8/1/00	13:00	298.3	8.1		0.0
15	8/1/00	14:00	106.5	7.8		0.0
16	8/1/00	15:00	185.8	0.0		0.0
17	8/1/00	16:00	192.4	5.1		1.5
...	...	...	...	...		...
...	...	...	...	...		...
250	8/11/00	9:00	222.5	0.0		0.0