

PENGELOMPOKAN DATA KAJICUACA BAGI PERAMALAN TABURAN HUJAN

Rozilawati Dollah @ Md. Zain dan Mohd Noor Md Sap
Fakulti Sains Komputer dan Sistem Maklumat,
Universiti Teknologi Malaysia. Skudai,
81310 Johor Bahru, JOHOR.

E-mail : zeela@fsksm.utm.my, mohdnoor@fsksm.utm.my

ABSTRAK

Penganalisan data kajicuaca merupakan satu tugas yang sangat penting tetapi sukar dan amat mencabar kepada pihak Jabatan Perkhidmatan Kajicuaca Malaysia (JPKM) ekoran penambahan jumlah data kajicuaca dari masa ke semasa. Hasil daripada penganalisan tersebut akan digunakan untuk membuat keputusan dan juga peramalan taburan hujan pada masa akan datang. Di dalam proses peramalan taburan hujan, peramal mestilah mengenalpasti parameter kajicuaca manakah yang memberikan pengaruh yang besar kepada ketepatan atau prestasi peramalan taburan hujan. Salah satu cara untuk kenalpasti parameter tersebut ialah dengan melakukan pengelompokan ke atas data kajicuaca tersebut. Oleh yang demikian, kertas kerja ini membincangkan tentang perbandingan di antara dua teknik pengelompokan iaitu teknik peraturan kesatuan dan kaedah statistik di dalam mengelompokkan data kajicuaca bagi tujuan peramalan taburan hujan. Hasil kajian telah mendapati bahawa teknik peraturan kesatuan adalah lebih sesuai untuk digunakan di dalam pengelompokan data kajicuaca berbanding kaedah statistik. Selain daripada itu, penggunaan parameter data kajicuaca yang berada di dalam kelompok yang berlainan didapati telah memberikan prestasi peramalan taburan hujan yang lebih baik berbanding penggunaan parameter data kajicuaca yang berada di dalam kelompok yang sama.

Katakunci : Pengelompokan, Perlombongan data, Peramalan, Taburan hujan dan Kajicuaca

1.0 Pengenalan

Penganalisan data kajicuaca merupakan satu tugas yang sukar dan mencabar kepada pihak Jabatan Perkhidmatan Kajicuaca Malaysia (JPKM). Ini kerana berlakunya pertambahan jumlah data kajicuaca yang dicerap dan disimpan di dalam pangkalan data JPKM. Penganalisan data kajicuaca memainkan peranan yang penting di dalam melakukan peramalan cuaca pada masa akan datang (Rozilawati *et. al*, 2003a; Rozilawati *et. al*, 2003b). Keputusan penganalisan data kajicuaca yang baik akan dapat membantu peramal di JPKM membuat jangkaan atau peramalan yang terperinci tentang perubahan cuaca sebelum mengeluarkan sebarang kenyataan, nasihat atau amaran cuaca.

Pertambahan jumlah data dan jenis data kajicuaca yang dikumpul dan disimpan menyebabkan peralatan konvensional yang digunakan untuk melakukan penganalisan data ini tidak mencukupi untuk melakukan pengekstrakan maklumat berguna yang diperlukan daripada pangkalan data tersebut (Choenni dan Siebes, 1996). Oleh yang demikian, satu cara untuk mengekstrak pengetahuan secara automatik daripada sejumlah data yang besar perlu diterokai untuk menggantikan kaedah konvensional yang digunakan bagi menampung peningkatan jumlah data kajicuaca yang disimpan di JPKM.

Sehubungan dengan itu, satu teknologi baru yang dinamakan perlombongan data merupakan salah satu teknik yang dapat membantu pihak JPKM di dalam menangani masalah penganalisan data kajicuaca yang banyak. Teknik perlombongan data boleh mengekstrak maklumat ramalan yang tersembunyi di dalam pangkalan data yang besar dengan mengecam corak atau paten data dan juga perhubungan di antara data-data. Ini kerana perlombongan data boleh mengenalpasti tren dan corak data dalam lautan maklumat dengan melakukan penjelajahan dan analisis data secara sistematik ke atas jumlah data yang banyak. Hasil daripada proses pengecaman dan perhubungan antara data kajicuaca ini akan digunakan untuk membantu di dalam proses membuat keputusan atau peramalan taburan hujan.

Terdapat pelbagai cara yang boleh digunakan untuk mengekstrak data kajicuaca daripada pangkalan data JPKM. Di antaranya ialah pengelompokan, peramalan, pengkelasan dan sebagainya, di mana hasil daripada pengekstrakan ini boleh digunakan di dalam proses membuat jangkaan atau peramalan tentang perubahan cuaca. Oleh yang demikian, kertas kerja ini akan membincangkan tentang operasi pengelompokan data kajicuaca menggunakan teknik perlombongan data bagi tujuan meramal taburan hujan. Pengelompokan merupakan cara terbaik untuk mencari kumpulan item yang mempunyai persamaan. Ini kerana pengelompokan merupakan penyusunan data yang mempunyai persamaan ke dalam satu kelas atau kelompok yang sama (Sarjon dan Mohd Noor, 2000).

Kandungan kertas kerja ini terbahagi kepada enam seksyen utama. Seksyen pertama merupakan pengenalan kepada kajian manakala latarbelakang masalah akan diperjelaskan dengan terperinci di dalam seksyen dua. 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. Seterusnya, dalam seksyen lima pula diperincikan tentang perbincangan bagi hasil keputusan setiap eksperimen yang telah dijalankan dan kertas kerja ini diakhiri dengan seksyen enam, iaitu kesimpulan bagi tajuk yang telah dibincangkan.

2.0 Latarbelakang Masalah

Pihak Jabatan Perkhidmatan Kajicuaca Malaysia (JPKM) memainkan peranan yang penting di dalam memantau situasi perubahan cuaca dan mengeluarkan kenyataan, nasihat dan amaran cuaca bila keadaan memerlukan. Pihak JPKM akan membuat pemantauan bagi kesemua stesen pencerapan kajicuaca di seluruh Malaysia. Sehubungan dengan itu, setiap stesen pencerapan ini akan memancarkan keadaan sebenar yang dicerap ke semua pejabat ramalan dan dipamerkan setiap jam (JPKM, 2002).

Secara tidak langsung, bilangan stesen yang banyak ini telah menyebabkan peningkatan jumlah data kajicuaca yang disimpan di JPKM. Pertambahan jumlah data ini telah menimbulkan kesukaran kepada pihak JPKM untuk melakukan proses penganalisan data kajicuaca termasuklah bagi tujuan melakukan peramalan cuaca (Rozilawati *et. al.*, 2003a; Rozilawati *et. al.*, 2003b). Sepertimana yang kita ketahui, maklumat yang tepat amatlah diperlukan bagi membantu pihak JPKM membuat jangkaan atau peramalan yang terperinci tentang perubahan cuaca sebelum mengeluarkan sebarang kenyataan, nasihat atau amaran cuaca.

Peramalan cuaca merupakan salah satu daripada masalah yang paling mencabar di dalam dunia sejak lebih setengah abad yang lalu. Ini bukan hanya disebabkan oleh nilai praktikalnya di dalam kajicuaca, tetapi ia juga merupakan masalah peramalan siri masa yang “unbiased” di dalam

penyelidikan saintifik (Liu, dan Lee, 1999). Justeru itu, pelbagai kaedah peramalan telah dibangunkan sejak beberapa tahun lalu dengan menggunakan rangkaian neural.

Di antara parameter-parameter kajicuaca, taburan hujan merupakan parameter yang paling sukar diramal (Liu dan Lee, 1999). Ini kerana peramalan hujan adalah merupakan satu masalah yang kompleks dan sukar kerana ia melibatkan pelbagai pembolehubah di mana ia saling berhubungkait dengan cara yang rumit. Kebanyakan perhubungannya adalah menggambarkan hubungan ruang dan dinamik yang tidak linear (Chen dan Takagi, 1993).

Terdapat banyak parameter data kajicuaca yang memainkan peranan di dalam melakukan peramalan taburan hujan. Di antara parameter-parameter data kajicuaca yang dicerap oleh pihak JPKM ialah angin, taburan hujan, taburan suhu, kelembapan bandingan, sejatan, cahaya dan sinaran matahari (Sani, 1984 & JPKM, 2002). Oleh yang demikian, kajian perlu dilakukan untuk mengenalpasti parameter manakah yang memberikan pengaruh yang besar ke atas peramalan taburan hujan. Salah satu cara yang boleh digunakan untuk mengenalpasti parameter tersebut ialah dengan cara mengelompokkan parameter-parameter data kajicuaca tersebut.

Pada masa ini, proses peramalan taburan hujan di Malaysia secara keseluruhannya dilakukan secara manual oleh beberapa orang peramal. Selepas data-data kajicuaca diperolehi sama ada dari radar, satelit ataupun lain-lain sumber seperti kaptan kapal, pelantar-pelantar minyak dan sebagainya, ia akan dianalisa sekaligus. Kemudian carta dan graf akan diplot berdasarkan data-data tersebut. Seterusnya, peramalan hujan dilakukan berdasarkan kepada corak cuaca dari carta atau graf hujan dan angin. Selain daripada itu, bantuan pengalaman dan pengetahuan dari peramal-peramal juga turut menyumbang kepada peramalan hujan (JPKM, 2002). Tugas ini memakan masa yang lama dan merupakan satu tugas yang agak sukar dilakukan.

Di Amerika Syarikat (USA) pula, peramalan taburan hujan dilakukan dengan menggunakan alat *Weather Surveillance Radar 1988 Doppler (WSR-88D)*. Algoritma WSR-88D hanya menggunakan teknik-teknik empirikal yang lemah di dalam penganggaran dan ini merupakan salah satu sebab berlakunya ketidaktepatan peramalan (Fulton *et. al*, 1998). Sehubungan dengan itu, data taburan hujan telah dipilih untuk diramal di dalam kajian ini. Selain daripada itu, taburan hujan juga merupakan salah satu daripada elemen atau ciri yang penting dan menarik di dalam iklim yang sering dibincang dan diperkatakan. Ini kerana 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).

Berdasarkan kepada beberapa kajian tentang peramalan atau penganggaran taburan hujan yang telah dijalankan oleh beberapa penyelidik sebelum ini, didapati kebanyakan kajian, antaranya kajian yang telah dijalankan oleh Chen dan Takagi (1993), McCullagh *et.al* (1999) serta kajian oleh Liu dan Lee (1999), telah menggunakan salah satu daripada teknik perlombongan data, iaitu teknik rangkaian neural pintar untuk melakukan operasi pengkelasan terhadap data taburan hujan tersebut bagi tujuan peramalan atau penganggaran taburan hujan. Walaubagaimanapun, masih terdapat kekurangan di dalam kajian yang telah dijalankan. Di antaranya ialah berlakunya *overlappes* di dalam set data semasa proses latihan. Sehubungan dengan itu, kajian yang telah mereka jalankan ini adalah tidak sesuai untuk digunakan bagi data input yang terlalu banyak.

Terdapat pelbagai cara yang boleh digunakan untuk mengekstrak data kajicuaca yang dikehendaki daripada pangkalan data. Di antaranya ialah pengelompokan, pengkelasan dan sebagainya, di mana hasil daripada pengekstrakan ini boleh digunakan di dalam proses membuat jangkaan atau peramalan tentang perubahan cuaca. Oleh yang demikian, kertas kerja ini

membincangkan tentang operasi pengelompokan data kajicuaca di mana hasil daripada pengelompokan ini akan digunakan untuk melakukan peramalan taburan hujan. Operasi pengelompokan data kajicuaca bagi tujuan peramalan taburan hujan telah dipilih di dalam kajian ini kerana terdapat banyak kajian yang melibatkan pengelompokan telah berjaya meningkatkan prestasi peramalan yang dilakukan. Contohnya, kajian yang telah dilakukan oleh Sarjon dan Mohd Noor (2000), mendapati pengelompokan dapat meningkatkan peramalan yang dilakukan.

Selain daripada itu, operasi pengelompokan juga akan membahagikan satu pangkalan data kepada beberapa kelompok yang mempunyai ciri atau sifat yang sama. Ini bermaksud atribut-atribut data kajicuaca yang mempunyai ciri yang sama akan dikumpulkan ke dalam satu kelompok. Bagi melaksanakan kajian ini, dua teknik pengelompokan telah dipilih untuk melakukan pengelompokan data kajicuaca iaitu teknik peraturan kesatuan dan kaedah statistik. Teknik peraturan kesatuan telah dipilih untuk mengelompokkan data kajicuaca ini kerana teknik peraturan kesatuan ini dapat mencari perkaitan di antara koleksi item-item atau atribut-atribut dalam pangkalan data kajicuaca (Rozilawati *et. al*, 2003a). Manakala kaedah statistik pula dipilih kerana ia dapat mencari persamaan di antara atribut-atribut di dalam pangkalan data kajicuaca bagi tujuan pengelompokan dengan cara mengira atau mengukur jarak di antara atribut-atribut tersebut (Rozilawati *et. al*, 2003b).

Untuk memilih atau menentukan teknik pengelompokan manakah yang lebih baik di antara kaedah statistik dan teknik peraturan kesatuan, beberapa eksperimen yang melibatkan kelompok-kelompok data kajicuaca yang dihasilkan oleh kedua-dua teknik ini digunakan sebagai data input kepada proses peramalan taburan hujan. Hasil keputusan peramalan taburan hujan ini kemudiannya digunakan sebagai perbandingan dan penganalisan dibuat berdasarkan kepada prestasi peramalan, nilai ralat RMS dan juga pekali korelasi yang diperolehi.

3.0 Metodologi

Perlaksanaan kajian ini adalah melibatkan beberapa aktiviti. Antara aktiviti-aktiviti yang terlibat di dalam perlaksanaan kajian ini adalah seperti berikut :-

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

Jadual 1 : Sampel data kajicuaca

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
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
...
...
250	8/11/00	9:00	222.5	0.0		0.0

- ii) **Mengira nilai purata dan sisihan piawai bagi data kajicuaca** : pengiraan nilai purata dan sisihan piawai bagi data kajicuaca 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-atribut sampel data kajicuaca tersebut dengan menggunakan formula-formula berikut :

$$\bar{X} = \frac{\sum X_i}{n} \quad \dots \text{ untuk pengiraan nilai purata}$$

di mana;

X_i = setiap nilai data individu

n = jumlah bilangan data

... (1)

dan untuk menghitung nilai sisihan piawai;

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

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \mu^2}$$

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

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

σ ialah sisihan piawai populasi
 x ialah cerapan dalam set data populasi
 μ ialah min untuk set data populasi
 Σ ialah hasil tambah
 N ialah jumlah bilangan cerapan

- 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. Pengiraan persamaan ini dilakukan dengan menggunakan formula jarak Euclidean seperti berikut (Korkmaz, 1997; Sarjon dan Mohd Noor, 2000).

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

di mana ; ... (2)

$$\begin{aligned} d_{ij} &= \text{jarak di antara kes } i \text{ dan } j \\ x_{ik} &= \text{nilai pembolehubah } k_{th} \text{ bagi kes } i_{th} \\ x_{jk} &= \text{nilai pembolehubah } k_{th} \text{ bagi kes } j_{th} \end{aligned}$$

- 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 :

$$d(c_i, c_j) = \min(\{d(\gamma_k, \gamma_l) \mid \gamma_k \subseteq c_i \text{ and } \gamma_l \subseteq c_j\})$$

di mana ; ... (3)

$$\begin{aligned} d(c_i, c_j) &= \text{jarak antara kelompok di antara dua kelompok tunggal} \\ c_i &= (\gamma_k) \text{ dan } c_j = (\gamma_l) \end{aligned}$$

- v) **Pengkelasan data kajicuaca** : tujuan pengkelasan data kajicuaca dilakukan ialah untuk memudahkan pengiraan nilai *support* dan *confidence*. Bagi tujuan ini, kaedah kuantitatif yang diperkenalkan oleh Fact, (1987) telah digunakan untuk membina pengagihan kelas. **Jadual 2** di bawah menunjukkan data kajicuaca yang telah dikelaskan.

Jadual 2 : Sampel data kajicuaca yang dikelaskan

ID	Tarikh	Masa	Windvane	Kelembapan	Taburan Hujan
1	8/1/00	0:00	W1	H1		RF1
2	8/1/00	1:00	W2	H1		RF1
3	8/1/00	2:00	W2	H1		RF1
4	8/1/00	3:00	W2	H1		RF1
5	8/1/00	4:00	W2	H1		RF1
6	8/1/00	5:00	W1	H1		RF1
7	8/1/00	6:00	W2	H1		RF1
8	8/1/00	7:00	W2	H1		RF1
9	8/1/00	8:00	W2	H1		RF1
10	8/1/00	9:00	W2	H1		RF1
11	8/1/00	10:00	W2	H1		RF1
12	8/1/00	11:00	W2	H1		RF1
13	8/1/00	12:00	W2	H1		RF1
14	8/1/00	13:00	W2	H1		RF1
15	8/1/00	14:00	W1	H1		RF1
16	8/1/00	15:00	W1	H1		RF1
17	8/1/00	16:00	W1	H1		RF1
...
...
250	8/11/00	9:00	W2	H1		RF1

- vi) **Mengira nilai support dan confidence bagi data kajicuaca** : pengiraan nilai *support* dan *confidence* bagi data-data kajicuaca dilaksanakan bagi tujuan mencari set item yang *frequent*. Ia dilakukan dengan menggunakan teknik peraturan kesatuan. Teknik peraturan kesatuan ini bertujuan untuk mencari item (atribut) yang dikatakan mempunyai perkaitan dengan item (atribut) yang lain dalam satu pangkalan data pada masa yang sama. Berdasarkan kepada sampel data kajicuaca dalam **Jadual 2** di atas, pengiraan nilai *support* dan *confidence* dilakukan ke atas data tersebut. Faktor *support* adalah parameter yang menunjukkan peraturan kesatuan yang telah dikesan dari keseluruhan set data di dalam satu transaksi. Manakala faktor *confidence* pula merupakan darjah yang mengatakan bahawa peraturan tersebut adalah betul bagi setiap rekod. Nilainya adalah dikira seperti berikut (Cheng, 1998) :

$$\begin{aligned}
 \text{Support}(X \Rightarrow Y) &= \text{Kebarangkalian}(X \cup Y) \\
 \text{Confidence}(X \Rightarrow Y) &= \text{Kebarangkalian}(Y / X)
 \end{aligned}
 \dots (4)$$

Di mana, kebarangkalian $(X \cup Y)$ merupakan kebarangkalian kedua-dua X dan Y berlaku di dalam pangkalan data dan kebarangkalian (Y / X) pula merupakan kebarangkalian bersyarat bahawa Y berlaku bila X berlaku.

- vii) **Mendapatkan set item yang frequent** : nilai *support* dan *confidence* yang diperolehi kemudian digunakan untuk mengasingkan set item yang *frequent* dan set item yang tidak *frequent*. Ia ditentukan dengan membandingkan nilai *support* dan *confidence* yang diperolehi dengan nilai minimum *support* dan juga nilai minimum *confidence*. Nilai minimum *support* dan nilai minimum *confidence* di dalam kajian

ini ialah 0.1. Oleh yang demikian, bagi data kajicuaca yang melebihi atau sama dengan nilai minimum *support* dan minimum *confidence*, maka ia akan menjadi atribut atau set item yang *frequent*. Selain daripada itu, untuk menjanakan set item yang *frequent* ini, algoritma Apriori yang diperkenalkan oleh Agrawal dan Srikant juga boleh digunakan (Sarjon dan Mohd Noor, 2000). Perincian bagi algoritma Apriori ini boleh dirujuk pada **Lampiran B**.

- viii) **Mengelompokkan set item** : selepas dibuat perbandingan di antara nilai *support* dan *confidence* dengan nilai minimum masing-masing, set item yang *frequent* akan diperolehi. Atribut atau set item yang *frequent* akan dikelompokkan ke dalam kelompok atau kumpulan yang sama manakala set item yang tidak *frequent* pula dikelompokkan ke dalam satu kelompok yang berlainan. Bagi tujuan pengujian teknik pengelompokan, hanya kelompok set item yang *frequent* sahaja yang akan digunakan sebagai data input kepada proses peramalan taburan hujan.
- ix) **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 teknik peraturan kesatuan dan juga kaedah statistik menjadi input kepada proses peramalan tersebut. Sehubungan dengan itu, beberapa eksperimen yang melibatkan kelompok-kelompok data kajicuaca yang menggunakan teknik pengelompokan peraturan kesatuan dan kaedah statistik telah dilaksanakan. Ianya bertujuan untuk membuat perbandingan prestasi peramalan taburan hujan di antara dua teknik pengelompokan data kajicuaca yang telah digunakan.
- x) **Penganalisan dan perbandingan hasil peramalan** : keputusan peramalan taburan hujan digunakan untuk penganalisan dan perbandingan di antara dua teknik pengelompokan tersebut. Ianya bertujuan untuk melihat sejauhmanakah ketepatan dan keberkesanan di antara teknik peraturan kesatuan dan kaedah statistik di dalam mengelompokkan data kajicuaca. Perbandingan 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 juga pekali korelasi digunakan bagi tujuan perbandingan keberkesanan dan ketepatan peramalan.

4.0 Eksperimen

Pengelompokan ialah membahagikan satu pangkalan data kepada beberapa kelompok yang mempunyai ciri atau sifat yang sama. Apabila atribut data yang mempunyai ciri yang sama telah dikumpulkan ke dalam satu kelompok, ini bermakna atribut data yang berada di dalam kelompok yang lain mempunyai ciri yang berbeza daripada mereka. Atribut-atribut di dalam satu kelompok mestilah mempunyai persamaan yang boleh diukur menggunakan kaedah tertentu, contohnya mengira jarak di antara atribut-atribut tersebut.

Terdapat pelbagai cara untuk mengelompokkan data kajicuaca di dalam satu pangkalan data. Di antaranya ialah dengan menggunakan kaedah statistik, teknik perlombongan data dan sebagainya. Sehubungan dengan itu, kajian ini telah dilaksanakan dengan menggunakan dua teknik pengelompokan yang berbeza iaitu kaedah statistik dan juga teknik peraturan kesatuan, iaitu salah satu teknik dalam perlombongan data.

Kaedah statistik ini telah digunakan untuk mencari persamaan di antara atribut-atribut di dalam pangkalan data kajicuaca bagi tujuan pengelompokan. Ia dilakukan dengan cara mengira atau mengukur jarak di antara atribut-atribut tersebut. Pengukuran ini dilakukan dengan menggunakan nilai purata dan sisihan piawai bagi sampel data kajicuaca yang digunakan. **Jadual 3** di bawah menunjukkan nilai purata dan sisihan piawai bagi setiap atribut data kajicuaca yang digunakan di dalam kajian ini (Rozilawati *et. al*, 2003b).

Jadual 3 : 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 digunakan untuk mengira persamaan di antara atribut data kajicuaca. Pengiraan persamaan ini dilakukan untuk mengira jarak di antara atribut-atribut tersebut. Pengiraan jarak ini adalah bertujuan untuk mengenalpasti atribut data kajicuaca manakah 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. Pengiraan persamaan dilakukan dengan menggunakan formula jarak Euclidean. Selepas pengiraan jarak di antara atribut data kajicuaca dilakukan, atribut-atribut tersebut dikelompokkan kepada beberapa kelompok berdasarkan kepada jarak yang diperolehi. Bagi tujuan pengelompokan ini, perbandingan jarak di antara atribut dibuat. Jarak antara atribut yang rendah menunjukkan persamaan yang tinggi atau kuat dan begitulah sebaliknya. Kelompok ini ditakrifkan sebagai jarak yang terdekat di antara dua kelompok. **Jadual 4** di bawah menunjukkan hasil keputusan pengelompokan atribut data kajicuaca menggunakan kaedah statistik (Rozilawati *et. al*, 2003b).

Jadual 4 : 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 telah dihasilkan. Kelompok-kelompok ini kemudiannya digunakan sebagai data input kepada proses peramalan taburan hujan.

Manakala teknik peraturan kesatuan pula akan mencari perkaitan di antara koleksi item-item atau atribut-atribut dalam pangkalan data kajicuaca. Perkaitan ini diperolehi dengan cara mencari set item yang *frequent*. Ia dilakukan dengan mengira nilai *support* dan juga nilai *confidence* bagi set item atau atribut yang terdapat di dalam pangkalan data kajicuaca. Faktor *support* dan *confidence*

merupakan dua parameter yang digunakan untuk menentukan kesahihan dan keutamaan peraturan di antara atribut-atribut yang dikaitkan. Sebelum data kajicuaca dikelompokkan, ia perlu dikelaskan terlebih dahulu. Tujuan pengelasan ini dilakukan ialah untuk memudahkan pengiraan nilai *support* dan *confidence* dilakukan. **Jadual 5** di bawah menunjukkan sampel data kajicuaca yang telah dikelaskan di dalam kajian ini.

Jadual 5 : Sampel data kajicuaca yang dikelaskan

ID	Tarikh	Masa	Windvane	Kelembapan	Taburan Hujan
1	8/1/00	0:00	W1	H1		RF1
2	8/1/00	1:00	W2	H1		RF1
3	8/1/00	2:00	W2	H1		RF1
4	8/1/00	3:00	W2	H1		RF1
5	8/1/00	4:00	W2	H1		RF1
6	8/1/00	5:00	W1	H1		RF1
7	8/1/00	6:00	W2	H1		RF1
8	8/1/00	7:00	W2	H1		RF1
9	8/1/00	8:00	W2	H1		RF1
10	8/1/00	9:00	W2	H1		RF1
11	8/1/00	10:00	W2	H1		RF1
12	8/1/00	11:00	W2	H1		RF1
13	8/1/00	12:00	W2	H1		RF1
14	8/1/00	13:00	W2	H1		RF1
15	8/1/00	14:00	W1	H1		RF1
16	8/1/00	15:00	W1	H1		RF1
17	8/1/00	16:00	W1	H1		RF1
...
...
250	8/11/00	9:00	W2	H1		RF1

Di mana ;

- W = *Windvane*,
- H = *Kelembapan*,
- E = *Energy*,
- T = *Suhu*,
- TE = *Tekanan*,
- R = *Radiation*,
- WS = *Kelajuan Angin*, dan
- RF = *Taburan Hujan*.

Berdasarkan kepada jadual sampel data kajicuaca di atas, pengiraan nilai *support* dan *confidence* dilakukan ke atas data tersebut. Hasil daripada pengiraan nilai *support* dan *confidence* yang diperolehi bagi setiap atribut data kajicuaca yang digunakan di dalam kajian ini ditunjukkan oleh **Jadual 6** di bawah (Rozilawati *et. al*, 2003a);

Jadual 6 : Nilai *Support* dan *Confidence* bagi data kajicuaca

Bil	Set Item	<i>Support</i>	<i>Confidence</i>
1.	W1, H1	0.64	1.00
2.	W1, H6	0.00	0.00
3.	W2, H1	0.36	0.99
4.	W2, H6	0.01	0.01
5.	W1, E1	0.48	0.75
6.	W1, E2	0.06	0.10
7.	W1, E3	0.01	0.006
8.	W1, E4	0.03	0.044
...
...
690.	R10, WS9	0.00	0.00

Nilai *support* dan *confidence* ini digunakan untuk mengasingkan set item yang *frequent* dan set item yang tidak *frequent*. Ia ditentukan dengan membandingkan nilai *support* dan *confidence* yang diperolehi dengan nilai minimum *support* dan juga nilai minimum *confidence*. Nilai minimum *support* di dalam kajian ini ialah 0.1 manakala nilai minimum *confidence* pula ialah 0.1. Selepas dibuat perbandingan di antara nilai *support* dan *confidence* dengan nilai minimum masing-masing, set item yang *frequent* akan diperolehi. **Jadual 7** pula menunjukkan sebahagian daripada set item yang *frequent*. **Lampiran C** menunjukkan senarai set item *frequent* yang diperolehi.

Jadual 7 : Kelompok set item yang *frequent*

Kelompok	Set Item
Kelompok 1	W1, H1
Kelompok 2	W2, H1
Kelompok 3	W1, E1
Kelompok 4	W2, E1
Kelompok 5	W1, T2
Kelompok 6	W1, T3
Kelompok 7	W1, T4
Kelompok 8	W2, T1
Kelompok 9	W1, TE1
Kelompok 10	W2, TE1
..	...
Kelompok 47	R2, WS1

Berdasarkan kepada **Jadual 7** di atas, set item yang *frequent* ini dikelompokkan dan ia menjadi data input kepada proses peramalan taburan hujan. Manakala set item yang tidak *frequent* tidak digunakan bagi tujuan pengujian peramalan taburan hujan di dalam kajian ini.

Bagi tujuan pengujian teknik pengelompokan manakah yang lebih baik di antara teknik peraturan kesatuan dan kaedah statistik, beberapa eksperimen pengujian terhadap keputusan peramalan taburan hujan yang telah dihasilkan. Pengujian ini dilakukan dengan menggunakan kelompok-kelompok data kajicuaca yang dihasilkan dengan menggunakan kedua-dua teknik

pengelompokan tersebut. Kelompok-kelompok ini digunakan sebagai data input kepada pakej NeuNetPro bagi melakukan peramalan taburan hujan.

Secara amnya, eksperimen peramalan taburan hujan ini dibahagikan kepada dua bahagian utama, iaitu;

- a) Pengelompokan data kajicuaca menggunakan kaedah statistik
 - melibatkan kelompok data kajicuaca yang sama
 - melibatkan kelompok data kajicuaca yang berlainan
- b) pengelompokan data kajicuaca menggunakan teknik peraturan kesatuan
 - melibatkan kelompok data kajicuaca yang sama
 - melibatkan kelompok data kajicuaca yang berlainan

Bagi tujuan pengujian peramalan taburan hujan, kelompok-kelompok data kajicuaca yang dihasilkan oleh kaedah statistik akan digunakan sebagai data input kepada pakej NeuNetPro. Di dalam proses peramalan, data taburan hujan digunakan sebagai nilai yang diramal. Pengujian peramalan ini dilaksanakan dengan melakukan beberapa eksperimen 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 (Rozilawati *et. al.*, 2003b) 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 8(a), (b) dan (c) berikut menunjukkan keputusan peramalan taburan hujan yang dihasilkan di dalam tiga eksperimen di atas.

Jadual 8(a) : Keputusan peramalan taburan hujan bagi atribut di dalam Kelompok 1

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
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
8/1/00	3:00	0	-0.02	0.02
8/1/00	4:00	0	-0.02	0.02
8/1/00	5:00	0	0.14	0.14
8/1/00	6:00	0	0.35	0.35
8/1/00	7:00	2.5	0.34	2.16
8/1/00	8:00	0.5	0.19	0.31
8/1/00	9:00	0	-0.02	0.02
8/1/00	10:00	0	0.08	0.08
8/1/00	11:00	0	0.24	0.24
8/1/00	12:00	0	0.19	0.19

Jadual 8(a) : Keputusan peramalan taburan hujan bagi atribut di dalam Kelompok 1 (Sambungan)

8/1/00	13:00	0	0.54	0.54
8/1/00	14:00	0	0.57	0.57
...
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 pada beberapa hari pertama. Walaubagaimapun, nilai perbezaannya kemudian semakin meningkat dan agak tidak sekata.

Jadual 8(b) : Keputusan peramalan taburan hujan bagi atribut di dalam Kelompok 2

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
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
8/1/00	3:00	0	0.24	0.24
8/1/00	4:00	0	0.23	0.23
8/1/00	5:00	0	0.23	0.23
8/1/00	6:00	0	0.23	0.23
8/1/00	7:00	2.5	0.21	2.29
8/1/00	8:00	0.5	0.21	0.29
8/1/00	9:00	0	0.22	0.22
8/1/00	10:00	0	0.23	0.23
8/1/00	11:00	0	0.22	0.22
8/1/00	12:00	0	0.21	0.21
8/1/00	13:00	0	0.21	0.21
8/1/00	14:00	0	0.21	0.21
...
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 8(c) : Keputusan peramalan taburan hujan bagi atribut di dalam Kelompok 3

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
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
8/1/00	3:00	0	0.11	0.11
8/1/00	4:00	0	0.12	0.12
8/1/00	5:00	0	0.10	0.10

Jadual 8(c) : Keputusan peramalan taburan hujan bagi atribut di dalam Kelompok 3 (Sambungan)

8/1/00	6:00	0	0.11	0.11
8/1/00	7:00	2.5	0.13	2.37
8/1/00	8:00	0.5	0.17	0.33
8/1/00	9:00	0	0.19	0.19
8/1/00	10:00	0	0.21	0.21
8/1/00	11:00	0	0.20	0.20
8/1/00	12:00	0	0.15	0.15
8/1/00	13:00	0	0.25	0.25
8/1/00	14:00	0	0.25	0.25
...
8/11/00	9:00	0	0.24	0.24

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 kajicua yang berlainan juga telah dilaksanakan. Di antara eksperimen-eksperimen tersebut (Rozilawati *et. al*, 2003b) 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 8(d): Keputusan peramalan taburan hujan menggunakan atribut *Windvane* dan Suhu

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
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
8/1/00	3:00	0	-0.12	0.12
8/1/00	4:00	0	-0.09	0.09
8/1/00	5:00	0	0.02	0.02
8/1/00	6:00	0	0.06	0.06
8/1/00	7:00	2.5	0.06	2.44
8/1/00	8:00	0.5	0.03	0.47
8/1/00	9:00	0	-0.02	0.02
8/1/00	10:00	0	0.01	0.01

Jadual 8(d): Keputusan peramalan taburan hujan menggunakan atribut *Windvane* dan Suhu (Sambungan)

8/1/00	11:00	0	0.30	0.30
8/1/00	12:00	0	0.33	0.33
8/1/00	13:00	0	0.44	0.44
8/1/00	14:00	0	0.07	0.07
...
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 8(e) : Keputusan peramalan taburan hujan menggunakan atribut *Energy* dan Suhu

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
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
8/1/00	3:00	0	-0.09	0.09
8/1/00	4:00	0	-0.08	0.08
8/1/00	5:00	0	-0.01	0.01
8/1/00	6:00	0	0.01	0.01
8/1/00	7:00	2.5	0.02	2.48
8/1/00	8:00	0.5	0.24	0.26
8/1/00	9:00	0	0.21	0.21
8/1/00	10:00	0	0.20	0.20
8/1/00	11:00	0	0.41	0.41
8/1/00	12:00	0	0.36	0.36
8/1/00	13:00	0	-0.24	0.24
8/1/00	14:00	0	-0.05	0.05
...
8/11/00	9:00	0	0.26	0.26

Daripada dua eksperimen yang dijalankan di bawah, eksperimen kelima memberikan prestasi peramalan yang lebih baik daripada eksperimen keempat. Ini dibuktikan oleh nilai ralat RMSnya yang lebih kecil dan juga nilai pekali korelasinya yang lebih besar daripada eksperimen keempat.

Manakala bagi tujuan peramalan taburan hujan yang menggunakan teknik peraturan kesatuan untuk mengelompokkan data kajuca, sebanyak tiga eksperimen yang menggunakan atribut daripada kelompok data kajuca yang sama sebagai data input telah dijalankan (Rozilawati *et. al.*, 2003a), iaitu;

- c) Eksperimen A - melibatkan atribut kelembapan dan suhu, iaitu (H1, T3).
- d) Eksperimen B - melibatkan atribut kelajuan angin dan tekanan, iaitu (WS1,TE1).
- e) Eksperimen C - melibatkan atribut windvane dan energy, iaitu (W2,E1).

Jadual 9(a), (b) dan (c) berikut menunjukkan hasil peramalan taburan hujan bagi ketiga-tiga eksperimen-eksperimen yang telah dilaksanakan.

Jadual 9(a): Keputusan peramalan taburan hujan bagi Eksperimen A

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
8/1/00	11:00	0	0.22	0.22
8/1/00	17:00	1.0	0.18	0.18
8/1/00	18:00	0	-0.09	0.09
8/1/00	19:00	0	0.04	0.04
8/2/00	10:00	0	0.44	0.44
8/2/00	14:00	0	0.34	0.34
8/2/00	15:00	0	-0.11	0.11
8/2/00	16:00	0	0.09	0.09
8/2/00	17:00	0	0.04	0.04
8/2/00	18:00	0	0.10	0.10
8/3/00	0:00	0	0.81	0.81
8/3/00	1:00	0	0.17	0.17
8/3/00	2:00	0	-0.20	0.20
8/3/00	3:00	0	-0.37	0.37
8/3/00	4:00	0	-0.27	0.27
8/3/00	5:00	0	0.03	0.03
8/3/00	6:00	0	0.08	0.08
8/4/00	0:00	0	-0.15	0.15
8/9/00	9:00	0.5	0.21	0.29
8/10/00	19:00	0	0.02	0.02
8/10/00	20:00	0	0.13	0.13
8/11/00	2:00	0	0.02	0.02
8/11/00	3:00	0	0.03	0.03
8/11/00	4:00	0	0.06	0.06
...
8/11/00	8:00	0	-0.02	0.02

Daripada jadual di atas, didapati perbezaan di antara nilai sebenar taburan hujan dan nilai yang diramal adalah tidak sekata.

Jadual 9(b): Keputusan peramalan taburan hujan bagi Eksperimen B

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
8/1/00	0:00	0	0.02	0.02
8/1/00	1:00	0	0.02	0.02
8/1/00	2:00	0	-0.13	0.13
8/1/00	3:00	0	0.13	0.13
8/1/00	4:00	0	0.02	0.02

Jadual 9(b): Keputusan peramalan taburan hujan bagi Eksperimen B

8/1/00	5:00	0	0.02	0.02
8/1/00	6:00	0	0.12	0.12
8/1/00	7:00	2.5	0.14	1.36
8/1/00	8:00	0.5	0.14	0.36
8/1/00	9:00	0	0.21	0.21
8/1/00	12:00	0	0.14	0.14
8/1/00	13:00	0	0.22	0.22
8/1/00	14:00	0	0.14	0.14
8/1/00	15:00	0	0.18	0.18
8/4/00	21:00	0.5	0.22	0.28
8/4/00	22:00	0	0.14	0.14
8/4/00	23:0	0	0.22	0.22
...
8/11/00	9:00	0	0.22	0.22

Jadual di atas menunjukkan keputusan peramalan taburan hujan yang menggunakan kelompok atribut data kajicuaca yang sama iaitu atribut kelajuan angin dan tekanan.

Jadual 9(c): Keputusan peramalan taburan hujan bagi Eksperimen C

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
8/1/00	1:00	0	0.01	0.01
8/1/00	2:00	0	0.03	0.03
8/1/00	3:00	0	0.05	0.05
8/1/00	4:00	0	0.09	0.09
8/1/00	6:00	0	0.07	0.07
8/1/00	7:00	2.5	0.11	2.39
8/1/00	8:00	0.5	1.01	0.21
8/1/00	9:00	0	1.09	1.09
8/1/00	10:00	0	1.09	1.09
8/1/00	11:00	0	1.07	1.07
8/1/00	12:00	0	1.03	1.03
8/1/00	17:00	1	0.95	0.05
8/1/00	23:00	0	0.11	0.11
8/2/00	0:00	0	0.11	0.11
8/2/00	1:00	0	0.11	0.11
8/2/00	2:00	0	0.11	0.11
8/2/00	3:00	0	0.11	0.11
8/2/00	4:00	0	0.11	0.11
...
8/11/00	0:00	0	-0.03	0.03

Jadual 9(c) menunjukkan keputusan peramalan taburan hujan yang menggunakan kelompok atribut data kajicuaca yang sama bagi eksperimen C, di mana ia melibatkan atribut *windvane* dan *energy*.

Bagi peramalan taburan hujan dengan menggunakan atribut daripada kelompok data kajicuaca yang berlainan sebagai data input pula, sebanyak dua eksperimen yang berkaitan telah dijalankan (Rozilawati *et. al*, 2003a), iaitu;

- a) Eksperimen D - melibatkan atribut dari kelompok yang berlainan iaitu *windvane* dan suhu (W2, T3).
- b) Eksperimen E - melibatkan atribut dari kelompok berlainan, iaitu atribut *energy* dan suhu (E1, T4).

Jadual 9(d) dan **9(e)** di bawah menunjukkan hasil peramalan taburan hujan bagi kedua-dua eksperimen yang telah dilakukan.

Jadual 9(d): Keputusan peramalan taburan hujan bagi Eksperimen D

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
81/00	11:00	0	0.27	0.27
8/1/00	17:00	1	0.80	0.20
8/3/00	3:00	0	0.03	0.03
8/5/00	0:00	0	-0.01	0.01
8/5/00	1:00	0	-0.01	0.01
8/5/00	2:00	0	-0.01	0.01
8/6/00	20:00	0	-0.08	0.08
8/6/00	21:00	0	0.14	0.14
8/6/00	22:00	0	-0.11	0.11
8/7/00	19:00	0	-0.07	0.07
8/8/00	19:00	0	0.01	0.01
8/8/00	20:00	0	0.01	0.01
8/8/00	21:00	0	0.02	0.02
88/00	22:00	0	0.06	0.06
...
8/11/00	7:00	0	-0.07	0.07

Jadual di atas menunjukkan hasil peramalan taburan hujan menggunakan kelompok atribut data kajicuaca yang berlainan, iaitu atribut *windvane* dan suhu.

Jadual 9(e): Keputusan peramalan taburan hujan bagi Eksperimen E

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
8/1/00	12:00	0	1.37	1.37
8/1/00	16:00	1.5	0.94	0.56
8/3/00	11:00	5	1.35	3.65
8/3/00	12:00	1	0.65	0.35
8/3/00	14:00	0	1.36	1.36

Jadual 9(e): Keputusan peramalan taburan hujan bagi Eksperimen E (Sambungan)

8/3/00	15:00	0	0.28	0.28
8/4/00	22:00	0	-0.04	0.04
8/4/00	23:00	0	0.06	0.06
8/5/00	18:00	0	1.41	1.41
8/6/00	17:00	1	0.45	0.55
8/6/00	18:00	0	0.72	0.72
8/6/00	19:00	0	0.33	0.33
8/7/00	16:00	0	0.49	0.49
8/8/00	16:00	0	0.19	0.19
8/8/00	17:00	0	0.85	0.85
...
8/10/00	15:00	0.5	0.78	0.28

Jadual 9(e) pula menunjukkan keputusan peramalan taburan hujan yang menggunakan kelompok atribut data kajicuaca yang berlainan, di mana atribut *energy* dan suhu telah digunakan sebagai data input kepada proses peramalan tersebut.

5.0 Perbincangan

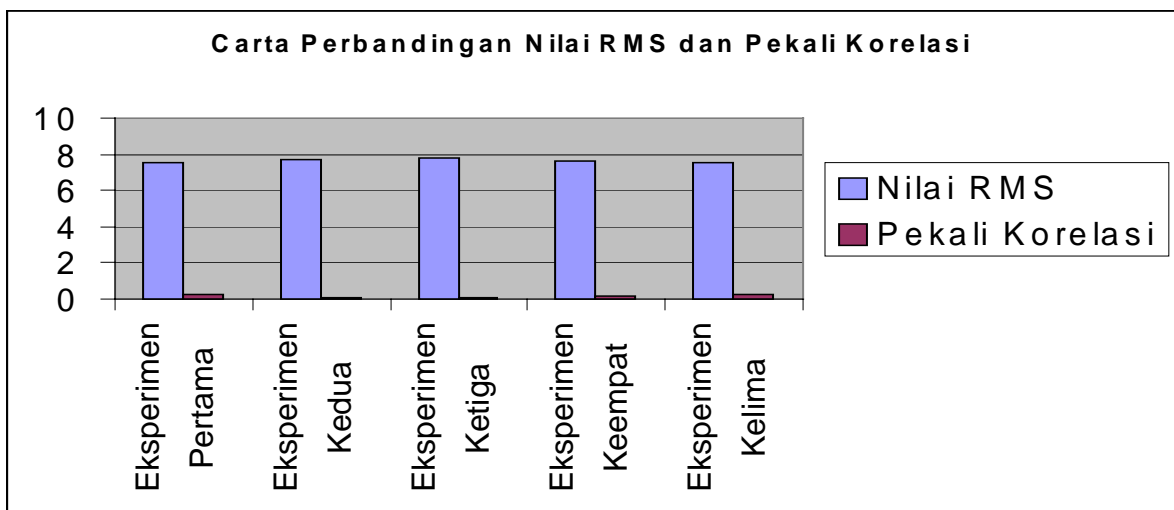
Eksperimen peramalan taburan hujan yang telah dijalankan adalah bertujuan untuk membuat perbandingan di antara teknik pengelompokan statistik dengan teknik pengelompokan peraturan kesatuan. Ianya adalah untuk menentukan teknik pengelompokan yang manakah yang terbaik di antara teknik peraturan kesatuan dan kaedah statistik dalam mengelompokkan data kajicuaca. Selain daripada itu, perbandingan peramalan taburan hujan dengan menggunakan kelompok data kajicuaca yang sama dan kelompok data kajicuaca yang berlainan juga dilaksanakan bagi setiap teknik pengelompokan statistik dan peraturan kesatuan. Perbandingan di antara dua teknik pengelompokan tersebut dilakukan berdasarkan kepada dua faktor yang digunakan untuk mengukur prestasi peramalan taburan hujan iaitu nilai ralat min punca kuasa dua (RMS) dan nilai pekali korelasi (*correlation coefficient*).

Daripada keputusan lima eksperimen peramalan taburan hujan yang telah dijalankan dengan menggunakan kaedah statistik untuk mengelompokkan data kajicuaca, secara keseluruhannya bolehlah dikatakan prestasi peramalan yang dilakukan menggunakan atribut di dalam kelompok yang berlainan memberikan prestasi peramalan yang lebih baik jika dibandingkan dengan kelompok yang sama. Walaubagaimanapun, 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 10** dan **Rajah 1** di bawah menunjukkan pengukuran keputusan peramalan taburan hujan yang telah diperolehi (Rozilawati *et. al*, 2003b).

Jadual 10 : Keputusan nilai RMS dan pekali korelasi

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

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

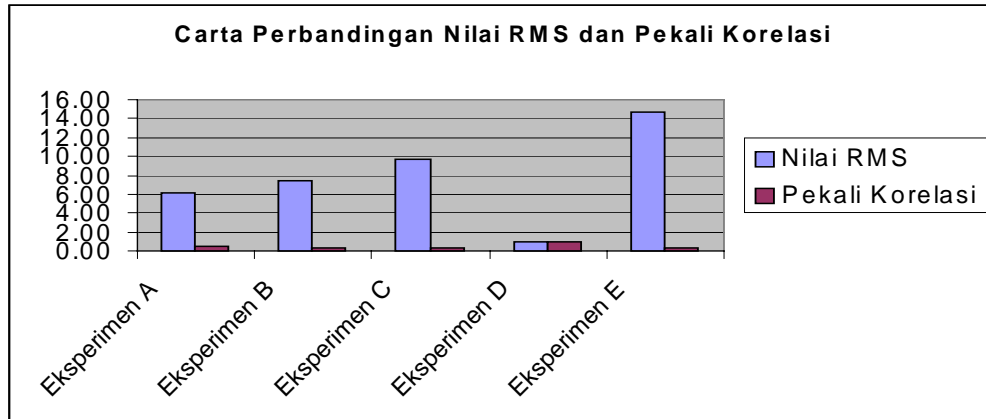


Rajah 1 : Perbandingan keputusan peramalan bagi setiap eksperimen

Selain daripada itu, penganalisan dan perbandingan juga telah dibuat ke atas hasil peramalan taburan hujan yang dilakukan terhadap data kajicuaca yang telah dikelompokkan dengan menggunakan teknik peraturan kesatuan. Tujuan perbandingan ini dilakukan ialah untuk melihat atau menilai sejauhmanakah prestasi hasil peramalan taburan hujan berdasarkan penggunaan atribut data kajicuaca yang dikelompokkan dengan menggunakan teknik peraturan kesatuan dan juga kesan penggunaan kelompok data kajicuaca yang sama dan juga kelompok data kajicuaca yang berlainan ke atas hasil peramalan taburan hujan. Sehubungan dengan itu, lima lagi eksperimen yang melibatkan penggunaan kelompok data kajicuaca yang sama dan juga kelompok data kajicuaca yang berlainan telah dilakukan. **Jadual 11** di bawah menunjukkan keputusan nilai RMS dan juga nilai pekali korelasi tersebut (Rozilawati *et. al*, 2003a).

Jadual 11 : Keputusan nilai RMS dan pekali korelasi

Eksperimen	Nilai RMS	Pekali Korelasi
Eksperimen A	6.15	0.52
Eksperimen B	7.46	0.38
Eksperimen C	9.65	0.39
Eksperimen D	1.02	0.91
Eksperimen E	14.71	0.29



Rajah 2 : Perbandingan keputusan peramalan bagi setiap eksperimen

Rajah 2 di atas menunjukkan nilai RMS yang dihasilkan di dalam eksperimen E jauh lebih tinggi berbanding dengan eksperimen-eksperimen yang lain. Manakala nilai pekali korelasi dalam eksperimen D pula memberikan nilai yang tertinggi di mana ia menghampiri nilai 1. Ini menunjukkan bahawa prestasi peramalan yang dihasilkan oleh eksperimen D merupakan yang terbaik jika dibandingkan dengan eksperimen yang lain. Oleh yang demikian, dapatlah dirumuskan bahawa penggunaan atribut di dalam kelompok yang berlainan menghasilkan keputusan peramalan taburan hujan yang lebih tepat berbanding dengan penggunaan kelompok yang sama. Ini boleh dibuktikan daripada nilai pekali korelasi di dalam eksperimen D yang menghampiri nilai 1 berbanding dengan eksperimen peramalan taburan hujan menggunakan kelompok yang sama sebagai data input kepada peramalan taburan hujan.

Berdasarkan kepada keputusan eksperimen-eksperimen yang telah dijalankan, perbandingan pengelompokan data kajiucua menggunakan teknik peraturan kesatuan dan kaedah statistik dilakukan, di mana ia dibahagikan kepada dua, iaitu;

- a) Kesan kelompok data kajiucua sebagai data input kepada proses peramalan taburan hujan.
- b) Kesan teknik pengelompokan data kajiucua.

Penganalisaan dan perbandingan kesan kelompok data kajiucua ke atas prestasi peramalan taburan hujan ini dibahagikan kepada dua iaitu kesan kelompok yang sama dan kesan kelompok yang berlainan.

5.1 Perbandingan kesan kelompok sama

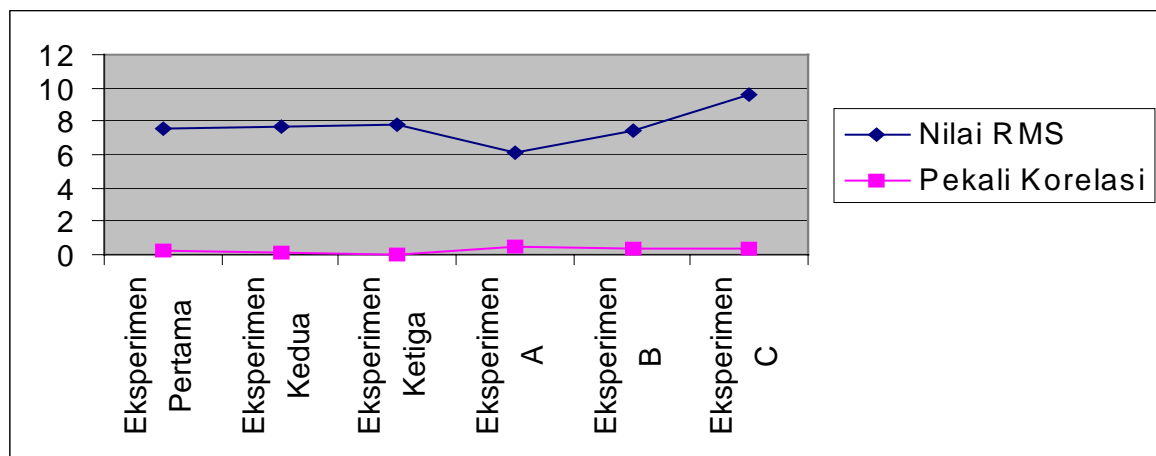
Penganalisaan dan perbandingan bagi kesan kelompok yang sama ini dilakukan ke atas hasil eksperimen peramalan yang menggunakan teknik pengelompokan peraturan kesatuan dan statistik. Penganalisaan dan perbandingan ini dilakukan untuk melihat kesan kelompok yang sama tetapi menggunakan dua teknik pengelompokan yang berlainan. Ianya bertujuan untuk melihat teknik pengelompokan manakah yang lebih baik.

Bagi membandingkan kesan penggunaan kelompok yang sama sebagai data input kepada proses peramalan taburan hujan, keputusan eksperimen yang diperolehi daripada enam eksperimen iaitu eksperimen pertama, kedua dan ketiga (menggunakan teknik pengelompokan statistik) serta eksperimen A, B dan C (menggunakan teknik pengelompokan peraturan kesatuan) dibandingkan. Keputusan perbandingan nilai RMS dan pekali korelasi bagi keenam-enam eksperimen tersebut ditunjukkan di dalam **Jadual 12** di bawah (Rozilawati *et. al.*, 2003a; Rozilawati *et. al.*, 2003b).

Jadual 12 : Perbandingan nilai RMS dan pekali korelasi bagi kelompok sama

Teknik Pengelompokan	Eksperimen yang dijalankan	Nilai RMS	Pekali Korelasi
Statistik	Eksperimen Pertama	7.58	0.23
Statistik	Eksperimen Kedua	7.73	0.11
Statistik	Eksperimen Ketiga	7.77	0.05
Peraturan Kesatuan	Eksperimen A	6.15	0.52
Peraturan Kesatuan	Eksperimen B	7.46	0.38
Peraturan Kesatuan	Eksperimen C	9.65	0.39

Berdasarkan jadual di atas, graf diplotkan bagi tujuan perbandingan di antara dua teknik pengelompokan data kajicuaa tersebut.



Rajah 3 : Perbandingan nilai RMS dan pekali korelasi bagi kelompok sama

Daripada jadual dan graf yang diplotkan di atas, didapati nilai RMS yang diperolehi daripada teknik pengelompokan statistik adalah lebih kurang sama di dalam ketiga-tiga eksperimen yang dijalankan, manakalan nilai RMS yang diperolehi daripada teknik pengelompokan peraturan

kesatuan agak tidak sekata. Secara puratanya bolehlah dikatakan bahawa nilai RMS yang dihasilkan daripada teknik pengelompokan peraturan kesatuan lebih rendah daripada teknik pengelompokan statistik. Ini ditunjukkan oleh nilai RMS yang dihasilkan di dalam eksperimen A dan B.

Selain daripada itu, pekali korelasi yang diperolehi daripada teknik pengelompokan statistik juga lebih rendah jika dibandingkan dengan nilai pekali korelasi yang dihasilkan oleh teknik pengelompokan peraturan kesatuan. Oleh yang demikian, daripada hasil eksperimen yang diperolehi menunjukkan peramalan taburan hujan ke atas kelompok data kajicuaca yang sama menggunakan teknik pengelompokan peraturan kesatuan menghasilkan prestasi peramalan taburan hujan yang lebih baik berbanding teknik pengelompokan statistik.

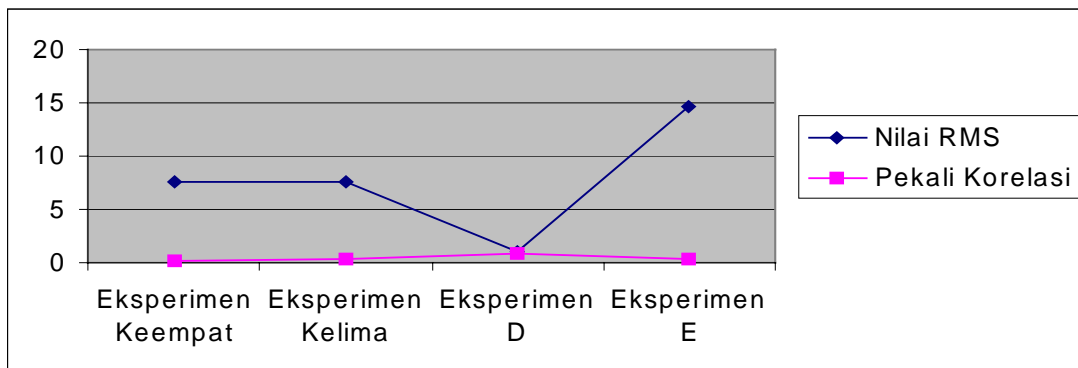
5.2 Perbandingan kesan kelompok berlainan

Penganalisaan dan perbandingan bagi kesan kelompok yang berlainan ini dilakukan ke atas hasil eksperimen peramalan taburan hujan yang menggunakan teknik pengelompokan peraturan kesatuan dan statistik. Bagi membandingkan kesan penggunaan kelompok data kajicuaca yang berlainan sebagai data input kepada proses peramalan taburan hujan, keputusan eksperimen yang diperolehi daripada empat eksperimen iaitu eksperimen keempat dan kelima (menggunakan kaedah pengelompokan statistik) serta eksperimen D dan E (menggunakan kaedah pengelompokan peraturan kesatuan) dibandingkan. Keputusan perbandingan nilai RMS dan pekali korelasi bagi semua eksperimen tersebut ditunjukkan di dalam **Jadual 13** di bawah (Rozilawati *et. al*, 2003a; Rozilawati *et. al*, 2003b).

Jadual 13 : Perbandingan nilai RMS dan pekali korelasi bagi kelompok berlainan

Teknik Pengelompokan	Eksperimen yang dijalankan	Nilai RMS	Pekali Korelasi
Statistik	Eksperimen Keempat	7.65	0.18
Statistik	Eksperimen Kelima	7.51	0.26
Peraturan Kesatuan	Eksperimen D	1.02	0.91
Peraturan Kesatuan	Eksperimen E	14.71	0.29

Berdasarkan jadual di atas, graf diplotkan bagi tujuan perbandingan nilai RMS dan pekali korelasi di antara dua teknik pengelompokan data kajicuaca tersebut.



Rajah 4 : Perbandingan nilai RMS dan pekali korelasi bagi kelompok berlainan

Berdasarkan jadual dan graf yang diplotkan di atas, didapati nilai RMS yang diperolehi daripada teknik pengelompokan statistik adalah lebih kurang sama di antara dua eksperimen yang telah dijalankan. Manakala nilai RMS yang dihasilkan oleh peramalan menggunakan teknik pengelompokan peraturan kesatuan pula adalah tidak sekata, di mana nilai RMS di dalam eksperimen E jauh lebih tinggi daripada nilai RMS di dalam eksperimen-eksperimen lain. Namun begitu, pekali korelasi yang dihasilkan daripada peramalan menggunakan kelompok berlainan yang dihasilkan oleh teknik pengelompokan peraturan kesatuan adalah lebih tinggi berbanding dengan nilai pekali korelasi yang dihasilkan di dalam peramalan menggunakan kelompok berlainan yang diperolehi daripada teknik pengelompokan statistik.

Daripada keempat-empat eksperimen yang dijalankan di atas, didapati prestasi peramalan taburan hujan ke atas kelompok data kajicuaca berlainan yang menggunakan teknik pengelompokan peraturan kesatuan adalah lebih baik jika dibandingkan dengan prestasi peramalan taburan hujan ke atas kelompok data kajicuaca berlainan yang diperolehi daripada teknik pengelompokan statistik. Oleh yang demikian, maka dapatlah diputuskan bahawa bagi melakukan peramalan taburan hujan menggunakan kelompok data kajicuaca yang berbeza, teknik pengelompokan peraturan kesatuan adalah lebih baik berbanding dengan teknik pengelompokan statistik.

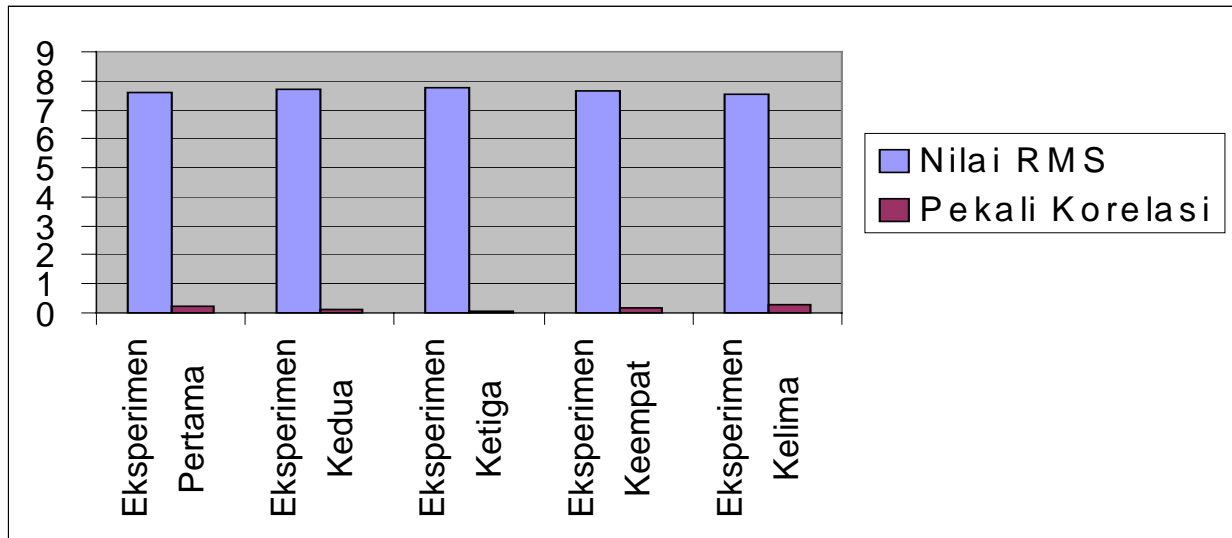
5.3 Perbandingan kesan teknik pengelompokan statistik

Bagi tujuan membandingkan kesan teknik pengelompokan statistik (bagi kedua-dua kelompok sama dan berlainan) ke atas prestasi peramalan taburan hujan, keputusan eksperimen yang diperolehi daripada lima eksperimen yang dilaksanakan telah dibandingkan. Di antara eksperimen-eksperimen tersebut ialah eksperimen pertama, kedua, ketiga, keempat dan kelima. Penganalisaan dan perbandingan ini bertujuan untuk melihat kelompok manakah (di antara kelompok sama dan kelompok berlainan yang dihasilkan oleh teknik pengelompokan statistik) yang lebih baik untuk digunakan sebagai data input di dalam peramalan taburan hujan. **Jadual 14** berikut menunjukkan keputusan perbandingan nilai RMS dan pekali korelasi bagi kelima-lima eksperimen yang telah dijalankan.

Jadual 14 : Perbandingan nilai RMS dan pekali korelasi bagi teknik pengelompokan statistik

Kelompok	Eksperimen yang dijalankan	Nilai RMS	Pekali Korelasi
Kelompok Sama	Eksperimen Pertama	7.58	0.23
Kelompok Sama	Eksperimen Kedua	7.73	0.11
Kelompok Sama	Eksperimen Ketiga	7.77	0.05
Kelompok Berlainan	Eksperimen Keempat	7.65	0.18
Kelompok Berlainan	Eksperimen Kelima	7.51	0.26

Berdasarkan kepada jadual di atas, carta diplotkan untuk perbandingan di antara nilai RMS dan pekali korelasi yang diperolehi daripada eksperimen yang telah dijalankan.



Rajah 5 : Perbandingan nilai RMS dan pekali korelasi bagi teknik pengelompokan statistik

Daripada jadual dan graf yang dihasilkan di atas, didapati nilai RMS yang diperolehi daripada kelompok yang sama dan kelompok berlainan adalah lebih kurang sama. Walaubagaimanapun, nilai RMS yang dihasilkan daripada peramalan taburan hujan menggunakan kelompok data kajicuaca yang berlainan adalah lebih rendah berbanding dengan nilai RMS yang dihasilkan daripada peramalan taburan hujan menggunakan kelompok sama. Tetapi nilai pekali korelasi yang diperolehi daripada hasil peramalan taburan hujan menggunakan kelompok yang berlainan adalah lebih tinggi daripada nilai pekali korelasi yang dihasilkan menggunakan kelompok yang sama. Oleh yang demikian, didapati prestasi peramalan taburan hujan menggunakan kelompok data kajicuaca yang berlainan adalah lebih baik berbanding dengan kelompok yang sama bagi teknik pengelompokan statistik.

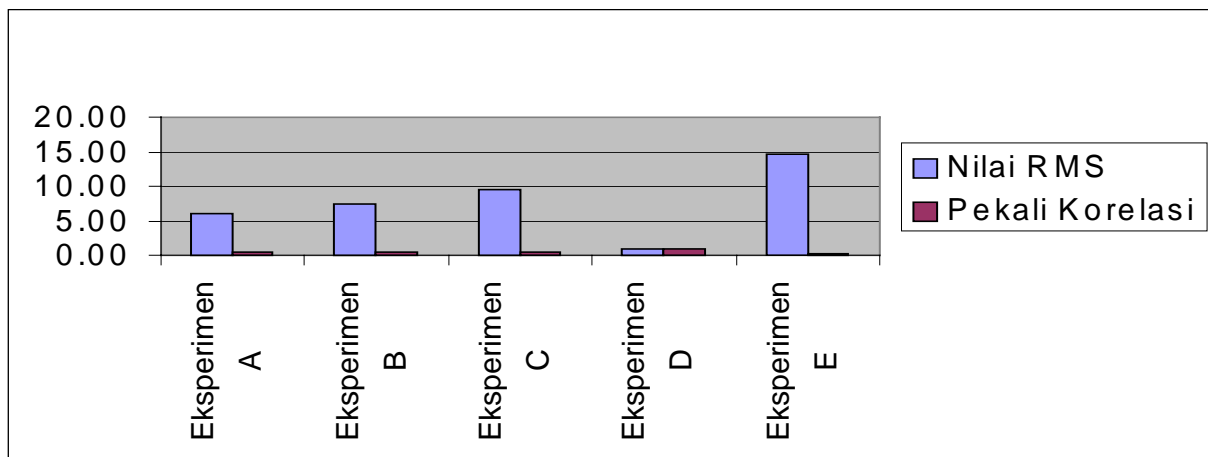
5.4 Perbandingan kesan teknik pengelompokan peraturan kesatuan

Penganalisan dan perbandingan kesan teknik pengelompokan peraturan kesatuan ke atas prestasi peramalan taburan hujan dilakukan dengan menggunakan kelompok yang sama dan berlainan sebagai data input kepada operasi peramalan taburan hujan. Bagi melihat kesan penggunaan teknik pengelompokan peraturan kesatuan ini, keputusan eksperimen yang diperolehi daripada lima eksperimen yang telah dijalankan, digunakan sebagai perbandingan. Di antara eksperimen-eksperimen tersebut ialah eksperimen A, B, C, D dan E. Keputusan nilai RMS dan pekali korelasi bagi semua eksperimen tersebut adalah seperti yang ditunjukkan oleh **Jadual 15** di bawah.

Jadual 15 : Perbandingan nilai RMS dan pekali korelasi bagi teknik pengelompokan peraturan kesatuan

Kelompok	Eksperimen yang dijalankan	Nilai RMS	Pekali Korelasi
Kelompok Sama	Eksperimen A	6.15	0.52
Kelompok Sama	Eksperimen B	7.46	0.38
Kelompok Sama	Eksperimen C	9.65	0.39
Kelompok Berlainan	Eksperimen D	1.02	0.91
Kelompok Berlainan	Eksperimen E	14.71	0.29

Berdasarkan kepada jadual di atas, graf diplotkan untuk perbandingan di antara nilai RMS dan pekali korelasi yang diperolehi daripada eksperimen yang telah dijalankan.



Rajah 6 : Perbandingan nilai RMS dan pekali korelasi bagi teknik pengelompokan peraturan kesatuan

Berdasarkan keputusan peramalan taburan hujan yang dihasilkan, didapati nilai RMS yang diperolehi daripada kelompok yang sama, iaitu eksperimen A, B dan C adalah lebih tinggi berbanding dengan nilai RMS yang diperolehi daripada kelompok berlainan, iaitu eksperimen D. Manakala nilai pekali korelasi bagi peramalan taburan hujan yang menggunakan kelompok yang berlainan adalah lebih tinggi jika dibandingkan dengan nilai pekali korelasi yang dihasilkan di dalam peramalan taburan hujan menggunakan kelompok yang sama. Sehubungan dengan itu bolehlah disimpulkan bahawa prestasi peramalan taburan hujan yang menggunakan kelompok data kajicuaa yang berlainan adalah lebih baik daripada prestasi peramalan yang menggunakan kelompok sama bagi teknik pengelompokan peraturan kesatuan. Ini kerana nilai pekali yang dihasilkan di dalam eksperimen yang menggunakan kelompok data kajicuaa yang berlainan adalah menghampiri nilai 1.

6.0 Kesimpulan

Peramalan cuaca merupakan satu tugas yang agak sukar dilakukan terutamanya di Malaysia. Ini kerana cuaca di Malaysia sering berubah-ubah. Di antara parameter-parameter data kajicuaca seperti kelembapan, suhu, tekanan, kelajuan angin, taburan hujan dan sebagainya, peramalan taburan hujan merupakan parameter yang paling sukar diramal. Peramalan taburan hujan yang dilakukan oleh pihak JPKM adalah secara manual. Berdasarkan data-data kajicuaca yang diperolehi, proses penganalisaan data kajicuaca dan seterusnya meramal taburan hujan dilakukan secara manual oleh beberapa orang peramal. Di dalam membuat peramalan taburan hujan ini, faktor bantuan seperti pengalaman peramal dan juga pengetahuan turut diambil kira. Oleh yang demikian, tugas ini memakan masa yang lama dan agak sukar dilakukan.

Sehubungan dengan itu, kajian ini telah dijalankan untuk mengkaji teknik pengelompokan data kajicuaca bagi tujuan meramal taburan hujan. Terdapat pelbagai cara yang boleh digunakan untuk melakukan pengelompokan data kajicuaca, di antaranya ialah dengan menggunakan kaedah statistik, teknik perlombongan data dan sebagainya. Kajian ini dilaksanakan dengan menggunakan kaedah statistik dan juga salah satu teknik perlombongan data iaitu teknik peraturan kesatuan untuk mengelompokkan data kajicuaca bagi tujuan peramalan taburan hujan. Bagi menentukan teknik manakah yang terbaik di antara dua teknik tersebut, beberapa eksperimen peramalan taburan hujan telah dijalankan.

Kesimpulannya, hasil kajian telah membuktikan bahawa teknik peraturan kesatuan adalah lebih baik di dalam melakukan pengelompokan data kajicuaca bagi tujuan peramalan taburan hujan berbanding dengan kaedah statistik. Selain daripada itu, kajian ini juga menunjukkan prestasi peramalan taburan hujan yang menggunakan kelompok atribut data kajicuaca yang berlainan adalah lebih baik berbanding dengan menggunakan kelompok atribut data kajicuaca yang sama bagi kedua-dua teknik pengelompokan.

RUJUKAN

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

Cheng, Shan (1998). "Statistical Approaches to Predictive Modeling in Large Databases." Simon Fraser University : Tesis Sarjana.

Choenni, R. dan Siebes, R. (1996). "Data Mining Applications." Computer Aided Decision Support in Telecommunications. UK: Chapman dan Hall.

Fulton, R., J. Breidenbach, D. J. Seo, D. Miller, dan T. O'Bannon, (1998). "The WSR-88D rainfall algorithm." Weather and Forecasting, 13, 377-395.

Hua, Zhu (1998). "On-line Analytical Mining of Association Rules." Simon Fraser University : Tesis Sarjana.

Jabatan Perkhidmatan Kajicuaca Malaysia (2002). [Online] Available : http://www.kjc.gov.my/main_m.html

- 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.
- Nieuwolt, S. (Terjemahan Jahi, J.) (1985). "Klimatologi Kawasan Tropika : Suatu Pengenalan Iklim Kawasan Garisan Lintang Rendah." Kuala Lumpur : Dewan Bahasa dan Pustaka.
- Ronkainen, P. (1998). "Attribute Similarity Event Sequence Similarity in Data Mining." Department of Computer Science : University of Helsinki.
- Rosdiyah Dahlan (1999). "Analisa Taburan Hujan Menggunakan Model Transfer Function." Universiti Teknologi Malaysia : Tesis Sarjana Muda.
- Roselina Sallehuddin (1999). "Penggunaan Model Rangkaian Neural dalam Peramalan Siri Masa Bermusim." Universiti Teknologi Malaysia : Tesis Sarjana.
- Rozilawati Dollah @ Md. Zain, Mohd Noor Md Sap, Mahadi Bahari dan Aryati Bakri (2003a). "Pengelompokan Data Kajicuaca Menggunakan Teknik Peraturan Kesatuan.;" Proceeding in The National Conference Management Science and Operations Research 2003.; 24 – 25 Jun 2003; 115 – 131.
- Rozilawati Dollah @ Md. Zain, Mohd Noor Md Sap, Mahadi Bahari dan Aryati Bakri (2003b). "Penggunaan Kaedah Statistik di dalam Pengelompokan Data Kajicuaca bagi Meramal Taburan Hujan.;" Proceeding in The Seminar Kebangsaan Sains Pemutusan 2003.; 15 – 16 Oktober 2003; 71 – 78.
- Sham Sani (1984). "Beberapa Aspek Iklim Negeri Sabah." UKM : Bangi.
- Sarjon Defit dan Mohd Noor Mad Sap (2000). " Predictive Data Mining Based On Similarity and Clustering Methods. "Jurnal Teknologi Maklumat , Fakulti Sains Komputer dan Sistem Maklumat : UTM. Jilid 12, Bil.2.
- Talava, L. (1998). "Exploring Efficient Attribute Prediction in Hierarchical Clustering." Universitat Politecnica de Catalunya.
- Yair, E.Z., Dan, R. dan Dmitry, Z. (1999). "Word Prediction and Clustering." Dept. of Computer Science. University of Illinois : Urbana.
- Yeong, Yoke Wei (1996). "Penggunaan Model Siri Masa untuk Menganalisis dan Meramal Kadar Hujan." Universiti Teknologi Malaysia : Tesis Sarjana Muda.