

**PENGELOMPOKAN DATA KAJI CUACA MENGGUNAKAN TEKNIK  
PENGELOMPOKAN HIERARKI AGGLOMERATIVE BAGI PERAMALAN  
TABURAN HUJAN.**

Mahadi Bahari<sup>1</sup>, Rozilawati Dollah @ Md. Zain<sup>2</sup>,  
Mohd Noor Md Sap<sup>3</sup>, Mohamad Fahmi Mohamad Adini<sup>4</sup>

Fakulti Sains Komputer dan Sistem Maklumat

Universiti Teknologi Malaysia

81310 Skudai, Johor.

Tel : 07-5576160 ext. <sup>1</sup>34207, <sup>2</sup>32425, <sup>3</sup>32419

Fax : 07-5565044

E-Mel : {<sup>1</sup>mahadi, <sup>2</sup>zeela, <sup>3</sup>mohdnoor}@fsksm.utm.my,  
<sup>4</sup>fahmiadini@gmail.com}

**Abstrak**

Kertas kerja ini melaporkan penggunaan teknik pengelompokan hierarki *Agglomerative* bagi melakukan peramalan taburan hujan. Tujuan utama kajian ini adalah untuk melihat keberkesanan serta prestasi algoritma yang terdapat di dalam teknik pengelompokan hierarki. Kertas kerja ini bermula dengan penerangan ke atas pengelompokan hierarki yang memfokus kepada algoritma *Single Link*, *Average Link* dan *Complete Link*. Melalui penggunaan algoritma-algoritma tersebut, kelompok dihasilkan berdasarkan pembentukan susunan skema pengelompokan dengan mengurangkan jumlah kelompok bagi setiap proses. Kelompok yang dihasilkan, diperolehi daripada gabungan kelompok-kelompok yang terhampir (sama) kepada satu kelompok. Kelompok-kelompok yang dihasilkan melalui ketiga-tiga algoritma tersebut akan digunakan sebagai input bagi melakukan peramalan taburan hujan. Langkah-langkah yang terlibat di dalam proses pengelompokan ini akan diterangkan dengan lebih jelas di dalam bahagian metodologi kajian. Seterusnya, kertas kerja ini akan menerangkan mengenai eksperimen yang dilakukan ke atas kelompok-kelompok yang dihasilkan dengan menggunakan ketiga-tiga algoritma di atas. Pengukuran prestasi pengelompokan dibuat berdasarkan hasil pengelompokan ialah nilai ralat min punca kuasa dua (RMS) dan nilai pekali kolerasi yang dihasilkan di dalam setiap eksperimen yang telah dijalankan. Hasil kajian menunjukkan bahawa peramalan taburan hujan yang terbaik diperolehi melalui penggunaan algoritma *Complete-Link*.

*Kata kunci:* Perlombongan data, Pengelompokan, Pengelompokan Hierarki *Agglomerative* dan Peramalan taburan hujan.

## 1.0 Pengenalan

Peramalan cuaca merupakan suatu proses atau kerja yang penting dan rumit dalam sesebuah negara. Keadaan cuaca boleh mempengaruhi keadaan persekitaran seperti keadaan muka bumi, aliran sungai dan bencana alam. Faktor-faktor ini juga memberi kesan kepada keadaan semasa sesebuah negara seperti keadaan ekonomi serta kependudukan masyarakat. Jabatan Perkhidmatan Kaji cuaca Malaysia (JPKM) merupakan agensi terpenting yang memainkan peranan utama dalam mengawal serta menguruskan dan membuat pemerhatian tentang keadaan cuaca di Malaysia.

Elemen yang utama dalam peramalan cuaca ialah peramalan taburan hujan, di mana hujan merupakan satu elemen yang rumit serta kompleks untuk diramal kerana ianya mempunyai kepelbagaian pemboleh ubah yang wujud serta mempunyai kaitan yang kompleks antara satu sama lain (Chen and Takagi, 1993; Ultsch and Guimareas, 1996). Berdasarkan JPKM, pemboleh ubah-pemboleh ubah atau atribut-atribut yang terlibat dengan peramalan taburan hujan ialah *windvane*, *humidity*, *energy*, *temperature*, *tension*, *radiation* dan *windspeed*. Kepelbagaian pemboleh ubah ini menyebabkan tugas peramalan menjadi kompleks.

Sehubungan dengan itu, satu penyelesaian telah dikaji bagi mengatasi masalah peramalan taburan hujan ini. Antara kaedah penyelesaian yang telah dikenalpasti ialah teknik perlombongan data iaitu dengan mengaplikasikan kaedah pengelompokan ke atas data yang ingin dikaji. Matlamat utama kajian yang dijalankan ialah untuk mengurangkan bilangan pemboleh ubah yang boleh digunakan untuk melakukan peramalan taburan hujan. Kaedah yang digunakan ialah dengan menggunakan teknik pengelompokan hierarchical, di mana perkaitan antara pemboleh ubah akan dikaji bagi membolehkan penggunaan pemboleh ubah-pemboleh ubah yang tertentu sahaja dalam membuat sesuatu peramalan.

Namun begitu sudah terdapat beberapa kajian terperinci yang telah dijalankan ke atas peramalan taburan hujan menggunakan teknik perlombongan data. Kebanyakan kajian yang dilakukan menumpukan kepada penggunaan rangkaian neural sebagai asas peramalan taburan hujan. Dalam kajian yang dilakukan oleh Diyankov et al (1992) berkaitan penggunaan rangkaian neural dalam peramalan cuaca telah menggunakan teknik pengesanan imej di mana ia telah menghasilkan corak imej bagi cuaca sepanjang tahun yang boleh digunakan dalam penentuan peramalan cuaca. Menurut kajian beliau, kajian itu telah mencapai kebarangkalian yang hampir tepat kepada data sebenar (91%). Begitu juga kajian yang dijalankan oleh Chen dan Takagi (1993), di mana kajian mereka tertumpu kepada penggunaan imej satelit dalam

meramal hujan menggunakan rangkaian neural. Penganalisisan yang dijalankan telah menghasilkan ketepatan klasifikasi terhadap pengujian data iaitu 90.45%.

## 2.0 Teknik Pengelompokan Hierarki Agglomerative

Pengelompokan merupakan salah satu kaedah yang digunakan di dalam perlombongan data. Tujuan utama teknik pengelompokan ialah untuk mengumpulkan elemen-elemen yang mempunyai persamaan ke dalam satu kelompok, di mana setiap kelompok mempunyai perbezaan antara satu sama lain. Pengelompokan membolehkan sesuatu corak atau susunan objek yang berkaitan dapat dikenalpasti. Ini membolehkan sebarang persamaan atau perbezaan di dalam satu set data diketahui. Teknik pengelompokan dibahagikan kepada dua jenis utama iaitu *partitional* dan hierarki. Dalam kajian ini, kami menumpukan kepada pengelompokan hierarki dalam menghasilkan kelompok yang dikehendaki. Pengelompokan hierarki dilakukan dengan menentukan persamaan di antara kelompok untuk digabungkan atau dipecahkan bagi menghasilkan struktur hierarki berdasarkan kepada matriks *proximity*. Hasil daripada pengelompokan hierarki selalunya digambarkan melalui pepohon perduaan atau dendogram (Xu dan Wunsch, 2005). Penggabungan atau pemisahan kelompok dilakukan berdasarkan algoritma di dalam pengelompokan hierarki yang digunakan.

Sejak kebelakangan ini, terdapat pelbagai kajian yang telah dijalankan menggunakan teknik pengelompokan hierarki. Namun begitu, tidak terdapat kajian yang menggunakan teknik pengelompokan hierarki ke atas data kaji cuaca dalam peramalan taburan hujan. Contohnya dalam kajian yang dilakukan oleh Zhao dan Karypis (2002) telah menggunakan teknik pengelompokan hierarki dalam mengelompokkan perkataan-perkataan di dalam sesuatu dokumen. Kajian yang dilakukan oleh Szymkowiak et al (2001) juga menggunakan teknik pengelompokan hierarki dalam menentukan vektor setiap perkataan di dalam sesuatu maklumat *e-mail*. Kajian-kajian yang dijalankan oleh penyelidik sebelum ini hanya menumpukan kepada pengelompokan data berdasarkan set data berbentuk dokumen dan bukan berdasarkan data kaji cuaca. Oleh sebab itu, penyelidikan ini cuba mengaplikasikan teknik pengelompokan data hierarki ke atas data kaji cuaca bagi tujuan peramalan taburan hujan.

Pengelompokan Hierarki dibahagikan kepada dua algoritma iaitu *Agglomerative (bottom-up)* dan *Divisive (top-down)*. Dalam algoritma *Agglomerative*, kelompok dihasilkan melalui pembentukan susunan skema pengelompokan dengan mengurangkan jumlah kelompok pada setiap langkah pengelompokan. Kelompok yang dihasilkan diperolehi

daripada gabungan kelompok-kelompok yang terhampir (sama) kepada satu kelompok. Manakala dalam algoritma *Divisive*, kelompok dijana dengan menghasilkan susunan skema pengelompokan melalui penambahan bilangan kelompok bagi setiap langkah pengelompokan. Kelompok yang dihasilkan adalah dengan memisahkan kelompok kepada dua (Haldiki et al, 2001).

Namun begitu, skop kajian ini hanya tertumpu kepada algoritma *Agglomerative* sebagai kaedah pengelompokan yang akan diguna pakai ke atas set data kajian. Ini memandangkan, bagi setiap kelompok dengan  $n$  objek pada algoritma *Divisive*, terdapat kemungkinan sebanyak  $2^{N-1} - 1$  dua-subset pembahagian perlu dilakukan (Xu dan Wunsch, 2005). Justeru itu, pengelompokan *divisive* tidak banyak digunakan secara praktiknya.

Algoritma *Agglomerative* dibahagikan kepada tiga jenis iaitu algoritma *Single Link*, algoritma *Average Link* dan algoritma *Complete Link*. Setiap algoritma ini mempunyai perbezaan dari segi pengiraan jarak di antara kelompok-kelompok bagi pembentukan sesuatu kelompok. Menurut Lin (1999), algoritma *Agglomerative* dimulakan dengan  $n$  kelompok dan setiap kelompok mengandungi satu set data bagi atribut yang berkaitan. Selepas itu, setiap kelompok akan digabungkan berdasarkan persamaan ciri di antara kelompok sehingga kelompok dikelompokkan mengikut yang dikehendaki atau sehingga pembentukan satu kelompok. Secara amnya, algoritma *Agglomerative* boleh diringkaskan dengan mengikuti prosidur seperti berikut:

- (i) Bermula dengan  $n$  kelompok dan satu sampel setiap kelompok.
- (ii) Cari persamaan antara kelompok dan gabungkan kelompok yang sama.
- (iii) Ulang langkah (ii) sehingga kelompok yang dikehendaki atau satu kelompok.

Pengiraan persamaan atau jarak di antara kelompok ditentukan dengan menentukan pasangan kelompok yang sepatutnya dikelompokkan. Pengiraan ini berbeza mengikut jenis algoritma *Agglomerative* yang digunakan. Bagi algoritma *Single Link*, pengiraan jarak di antara kelompok dilakukan dengan mencari jarak yang terdekat di antara sampel dalam satu kelompok dan sampel di dalam kelompok yang lain.

Formula jarak algoritma *Single Link* (Cimiano et al, 2001; Lin, 1999):

$$d(C_i, C_j) = \underset{a \in C_i, b \in C_j}{\text{mind}}(a, b) \quad \dots (1)$$

di mana  $d(C_i, C_j)$  merupakan fungsi jarak yang dibincangkan tadi di dalam matriks proximity. Manakala di dalam algoritma *Average Link*, pengiraan jarak di antara kelompok dilakukan dengan menentukan jarak purata di antara sampel di dalam kelompok atau sampel dengan

kelompok lain. Formula jarak algoritma *Average Link* (Cimiano et al, 2001; Lin, 1999) adalah:

$$d(C_i, C_j) = \frac{1}{n_{c_i} n_{c_j}} \sum_{a \in C_i, b \in C_j} d(a, b) \quad \dots (2)$$

di mana  $n_{c_i}$  merupakan jumlah kelompok bagi kelompok  $C_i$ . Pengiraan jarak di dalam algoritma *Complete Link* pula diperolehi dengan menentukan jarak yang terpanjang di antara sampel di dalam satu kelompok dan sampel di dalam kelompok yang lain. Formula jarak algoritma *Complete Link* (Cimiano et al, 2001; Lin, 1999) adalah:

$$d(C_i, C_j) = \max_{a \in C_i, b \in C_j} d(a, b) \quad \dots (3)$$

### 3.0 Metodologi

Perlaksanaan kajian ini telah dilakukan mengikut beberapa aktiviti. Di antara aktiviti-aktiviti yang terlibat ialah mengumpul dan menganalisis data kaji cuaca, mengira jarak Euclidean, membangunkan atur cara, mengelompokkan set data kaji cuaca, melakukan pengujian kelompok data kaji cuaca, menganalisis keputusan pengujian dan akhir sekali ialah membuat perbandingan keputusan pengujian yang telah dihasilkan.

Bagi menjayakan perlaksanaan kajian ini, sebanyak 500 set data kaji cuaca, bermula dari 1 September 2000 sehingga 21 September 2000 telah dikumpul. Data tersebut telah diperolehi daripada Jabatan Perkhidmatan Kaji cuaca Malaysia. Terdapat lapan pemboleh ubah atau atribut utama data kaji cuaca yang telah diguna pakai di dalam perlaksanaan kajian ini, di antaranya ialah *windvane*, *humidity*, *energy*, *temp*, *tension*, *radiation*, *windspeed* dan *rainfall*.

Selepas data kaji cuaca tersebut dianalisis, pengiraan jarak Euclidean pula dilakukan. Proses ini dijalankan untuk mendapatkan persamaan di antara atribut-atribut data kaji cuaca tersebut untuk tujuan mengelompokkan atribut-atribut yang berkaitan. Formula yang digunakan untuk melakukan pengiraan jarak *Euclidean* ialah :

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad \dots (4)$$

di mana  $x = (x_1, x_2, \dots, x_n)$  dan  $y = (y_1, y_2, \dots, y_n)$

Berdasarkan pengiraan jarak Euclidean yang telah dihasilkan, algoritma pengelompokan hierarki bagi *single link*, *average link* dan *complete link* dibangunkan menggunakan bahasa pengaturcaraan C. Dengan menggunakan atur cara yang telah dibangunkan, atribut-atribut data kaji cuaca (kecuali atribut *rainfall*) dikelompokkan kepada beberapa kelompok iaitu kelompok 2, 3, 4, 5 dan 6. Proses ini dilakukan berulang-ulang kali dengan menggunakan beberapa set data yang terdiri daripada 100 set data, 200 set data, 300 set data, 400 set data dan 500 set data kaji cuaca.

Setelah pengelompokan dijalankan ke atas set data kaji cuaca, setiap kelompok yang terhasil diuji dengan menggunakan Pakej NeuNet Pro 2.3. Pengujian ini telah dibahagikan kepada dua kategori, iaitu eksperimen yang melibatkan atribut-atribut dari kelompok yang sama dan juga eksperimen yang melibatkan atribut-atribut dari kelompok yang berlainan. Di dalam pengujian ini, atribut-atribut dari kelompok yang sama dan atribut-atribut dari kelompok yang berlainan digunakan sebagai data input untuk melakukan proses peramalan taburan hujan. Pengujian kelompok ini dilakukan untuk membuat penganalisan dan perbandingan prestasi peramalan taburan hujan di antara set-set eksperimen yang dijalankan.

Penganalisan kemudian dilakukan ke atas hasil pengujian tersebut untuk menentukan prestasi teknik pengelompokan menggunakan algoritma pengelompokan hierarki. Penganalisan ini dilakukan dengan melihat nilai perbezaan di antara nilai sebenar taburan hujan serta nilai ramalan taburan hujan yang dihasilkan. Berdasarkan kepada perbezaan nilai tersebut, nilai ralat min punca kuasa dua (RMS) dan pekali korelasi akan dihasilkan. Sehubungan dengan itu, nilai RMS dan pekali korelasi ini digunakan sebagai perbandingan bagi menentukan keberkesanan teknik pengelompokan hierarki.

#### **4.0 Eksperimen**

Di dalam kajian yang dijalankan, proses pengelompokan menggunakan algoritma *Agglomerative* telah dilakukan ke atas 500 set data kaji cuaca. Proses pengelompokan ini telah dibahagikan kepada tiga bahagian utama iaitu pengelompokan menggunakan algoritma *Single Link*, algoritma *Average Link* dan algoritma *Complete Link*. Eksperimen ini dijalankan berulang-ulang kali dengan mengelompokkan 100 set data, 200 set data, 300 set data, 400 set data dan 500 set data kaji cuaca kepada 2 kelompok, 3 kelompok, 4 kelompok, 5 kelompok dan 6 kelompok.

Selepas pengelompokan dijalankan ke atas set-set data tadi, pengujian kelompok pula dilakukan bagi setiap kelompok data kaji cuaca yang telah dihasilkan. Ia melibatkan pengujian peramalan taburan hujan dengan menggunakan atribut-atribut dari kelompok data kaji cuaca yang sama dan juga kelompok data kaji cuaca yang berlainan. Pengujian ini dilakukan dengan menggunakan aplikasi NeuNet Pro 2.3. Keputusan peramalan taburan hujan yang diperolehi digunakan untuk membuat penganalisan dan perbandingan.

#### 4.1 Algoritma *Single Link*

Berdasarkan eksperimen yang telah dilakukan, Jadual 1 berikut adalah merupakan hasil pengelompokan data kaji cuaca dengan menggunakan algoritma *Single Link*.

**Jadual 1 : Hasil pengelompokan data kaji cuaca menggunakan algoritma *Single Link***

Bil. Kelompok	100 Set Data	200 Set Data	300 Set Data	400 Set Data	500 Set Data
2	( a ) ( b, c, d, e, f, g )	( a, b, d, e, g ) ( c, f )	( a, b, d, e, g ) ( c, f )	( a, b, d, e, g ) ( c, f )	( a, b, d, e, g ) ( c, f )
3	( a ) ( b, d, e, g ) ( c, f )	( a ) ( b, d, e, g ) ( c, f )	( a ) ( b, d, e, g ) ( c, f )	( a ) ( b, d, e, g ) ( c, f )	( a ) ( b, d, e, g ) ( c, f )
4	( a ) ( b, d, e, g ) ( c ) ( f )	( a ) ( b, d, e, g ) ( c ) ( f )	( a ) ( b, d, e, g ) ( c ) ( f )	( a ) ( b, d, e, g ) ( c ) ( f )	( a ) ( b, d, e, g ) ( c ) ( f )
5	( a ) ( b, d, g ) ( c ) ( e ) ( f )	( a ) ( b, d, g ) ( c ) ( e ) ( f )	( a ) ( b, d, g ) ( c ) ( e ) ( f )	( a ) ( b, d, g ) ( c ) ( e ) ( f )	( a ) ( b, d, g ) ( c ) ( e ) ( f )
6	( a ) ( b, g ) ( c ) ( d ) ( e ) ( f )	( a ) ( b, g ) ( c ) ( d ) ( e ) ( f )	( a ) ( b, g ) ( c ) ( d ) ( e ) ( f )	( a ) ( b ) ( c ) ( d, g ) ( e ) ( f )	( a ) ( b ) ( c ) ( d, g ) ( e ) ( f )

a – windvane, b – humidity, c – energy, d – temp, e – tension, f – radiation, g – windspeed

Setelah kelompok-kelompok data kaji cuaca dikenalpasti, pengujian kelompok pula dilakukan. Proses pengujian ini melibatkan penggunaan atribut-atribut data kaji cuaca (kecuali atribut *rainfall*) sebagai data input untuk melakukan proses peramalan taburan hujan. Pengujian ini telah dibahagikan kepada dua kategori, iaitu pengujian bagi kelompok data kaji

cuaca yang sama dan juga pengujian bagi kelompok data kaji cuaca yang berlainan. Oleh yang demikian, atribut-atribut dari kelompok yang sama dan juga atribut-atribut dari kelompok yang berlainan telah digunakan sebagai data input kepada proses peramalan tersebut. Bagi setiap kategori pengujian, terdapat beberapa eksperimen yang telah dijalankan.

Setiap eksperimen tersebut melibatkan atribut-atribut bagi 2 kelompok, 3 kelompok, 4 kelompok, 5 kelompok dan 6 kelompok data kaji cuaca dan juga menggunakan 100 set data, 200 set data, 300 set data, 400 set data dan 500 set data kaji cuaca. Pengujian bagi setiap kategori kelompok dan set data dilakukan beberapa kali dan nilai purata bagi nilai ralat min punca kuasa dua (RMS) dan pekali korelasi telah diambil dan digunakan untuk tujuan perbandingan. Jadual 2(a) dan 2(b) menunjukkan hasil peramalan taburan hujan yang dilakukan ke atas kelompok atribut data kaji cuaca yang sama dan juga kelompok atribut data kaji cuaca yang berlainan.

**Jadual 2(a) : Keputusan peramalan taburan hujan bagi kelompok berlainan**

KELOMPOK	100 Set Data		200 Set Data		300 Set Data		400 Set Data		500 Set Data	
	RMS	CC	RMS	CC	RMS	CC	RMS	CC	RMS	CC
2 Kelompok	0.95	0.07	0.85	0.27	0.88	0.17	0.96	0.12	0.95	0.09
3 Kelompok	0.93	0.15	0.81	0.39	0.85	0.26	0.96	0.16	0.94	0.16
4 Kelompok	0.88	0.37	0.79	0.46	0.85	0.23	0.92	0.33	0.92	0.28
5 Kelompok	0.91	0.22	0.78	0.49	0.78	0.49	0.89	0.25	0.90	0.34
6 Kelompok	0.84	0.38	0.68	0.64	0.74	0.56	0.79	0.67	0.78	0.56

Berdasarkan hasil pengujian bagi kelompok data kaji cuaca berlainan yang telah dilaksanakan, maka penganalisan terhadap keputusan peramalan taburan hujan telah dibuat. Secara keseluruhannya, didapati kesemua eksperimen yang melibatkan set data 100, 200, 300, 400 dan 500 menunjukkan bahawa hasil atau prestasi peramalan semakin meningkat, selari dengan bilangan atribut-atribut data kaji cuaca yang digunakan di dalam pengujian tersebut. Atau dengan kata lain, semakin banyak bilangan atribut data kaji cuaca yang digunakan sebagai data input kepada proses peramalan taburan hujan tersebut, semakin baik prestasi peramalan yang dihasilkan. Ini dibuktikan oleh nilai RMS dan juga pekali korelasi yang telah dihasilkan di dalam pelaksanaan eksperimen yang berkaitan.

Selain daripada itu, hasil eksperimen juga menunjukkan keputusan yang dihasilkan oleh eksperimen yang melibatkan set data kaji cuaca 200 memberikan keputusan peramalan yang terbaik berbanding set data kaji cuaca yang lain bagi semua kelompok 2, 3, 4, 5 dan 6



yang dijalankan. Keadaan ini mungkin juga disebabkan oleh kewujudan bilangan data melampau yang banyak di dalam set data 300, 400 dan 500.

**Jadual 2(b) : Keputusan peramalan taburan hujan bagi kelompok sama**

KELOMPOK	100 Set Data		200 Set Data		300 Set Data		400 Set Data		500 Set Data	
	RMS	CC	RMS	CC	RMS	CC	RMS	CC	RMS	CC
2 Kelompok	0.87	0.29	0.82	0.29	0.77	0.36	0.88	0.30	0.90	0.28
3 Kelompok	0.94	0.12	0.89	0.08	0.84	0.19	0.85	0.21	0.95	0.13
4 Kelompok	0.70	0.07	0.67	0.06	0.63	0.14	0.77	0.16	0.69	0.14
5 Kelompok	0.95	0.03	0.86	0.15	0.86	0.14	1.03	0.07	0.96	0.08
6 Kelompok	0.95	0.03	0.90	0.06	0.88	0.10	0.93	0.03	0.80	0.06

Manakala bagi eksperimen yang melibatkan penggunaan atribut-atribut dari kelompok data kaji cuaca yang sama pula menunjukkan keputusan peramalan taburan hujan yang berbeza. Berdasarkan kepada keputusan pada Jadual 2(b) di atas, maka dapatlah dirumuskan bahawa keputusan peramalan yang dihasilkan oleh 4 kelompok memberikan prestasi yang terbaik berbanding dengan empat eksperimen yang lain. Prestasi ini boleh dilihat pada nilai RMS dan juga nilai pekali korelasi yang dihasilkan oleh eksperimen yang melibatkan penggunaan atribut-atribut data kaji cuaca dari 4 kelompok di dalam kesemua eksperimen sama ada yang melibatkan 100 data set, 200 data set, 300 data set, 400 data set, mahupun 500 data set. Di samping itu, keputusan peramalan taburan hujan yang ditunjukkan oleh eksperimen yang melibatkan 300 set data kaji cuaca juga memberikan keputusan prestasi peramalan yang lebih baik jika dibandingkan dengan prestasi peramalan taburan hujan yang melibatkan set data yang lain.

Oleh yang demikian, secara keseluruhannya bolehlah dirumuskan bahawa keputusan peramalan taburan hujan yang dihasilkan oleh eksperimen yang melibatkan penggunaan atribut dari kelompok data kaji cuaca yang berlainan menunjukkan prestasi peramalan taburan hujan yang lebih baik berbanding dengan prestasi peramalan taburan hujan yang dihasilkan oleh eksperimen yang melibatkan penggunaan atribut dari kelompok data kaji cuaca yang sama. Ini dibuktikan oleh keputusan peramalan taburan hujan yang dihasilkan di dalam kelima-lima eksperimen yang melibatkan bilangan set data kaji cuaca yang berbeza.

Selain daripada itu, keputusan peramalan bagi eksperimen kelompok berlainan menunjukkan semakin banyak bilangan atribut data kaji cuaca yang digunakan di dalam pengujian tersebut, semakin tinggi prestasi peramalan yang dihasilkan. Ini dapat dilihat pada nilai RMS yang semakin berkurang dan juga nilai pekali korelasi yang semakin meningkat.

## 5.2 Algoritma *Average Link*

Proses pengelompokan telah dilaksanakan ke atas set data kaji cuaca dengan menggunakan algoritma *Average Link*. Jadual 3 menunjukkan hasil pengelompokan data kaji cuaca tersebut.

**Jadual 3 : Hasil pengelompokan data kaji cuaca menggunakan algoritma *Average Link***

Bil. Kelompok	100 Set Data	200 Set Data	300 Set Data	400 Set Data	500 Set Data
2	( a,c,d,e,f,g ) ( b )	( a,c,d,e,f,g ) ( b )	( a,c,d,e,f,g ) ( b )	( a,c,d,e,f,g ) ( b )	( a,c,d,e,f,g ) ( b )
3	( a ) ( b ) ( c,d,e,f,g )	( a ) ( b,c, d, e, f ) ( g )	( a, c, d, e, g ) ( b ) ( f )	( a, c, d, e, g ) ( b ) ( f )	( a, c, d, e, g ) ( b ) ( f )
4	( a ) ( b ) ( c,d,f,g ) ( e )	( a ) ( b,c,d, f ) ( e ) ( g )	( a, c, d, e ) ( b ) ( f ) ( g )	( a, c, d, e ) ( b ) ( f ) ( g )	( a, c, d, e ) ( b ) ( f ) ( g )
5	( a ) ( b ) ( c,f,g ) ( d ) ( e )	( a ) ( b ) ( c, d, f ) ( e ) ( g )	( a, e, d ) ( b ) ( e ) ( f ) ( g )	( a, e, d ) ( b ) ( e ) ( f ) ( g )	( a, e, d ) ( b ) ( e ) ( f ) ( g )
6	( a ) ( b ) ( c,g ) ( d ) ( e ) ( f )	( a ) ( b ) ( c, d ) ( e ) ( f ) ( g )	( a ) ( b ) ( c, d ) ( e ) ( f ) ( g )	( a ) ( b ) ( c, d ) ( e ) ( f ) ( g )	( a ) ( b ) ( c, d ) ( e ) ( f ) ( g )

a – *windvane*, b – *humidity*, c – *energy*, d – *temp*, e – *tension*, f – *radiation*, g – *windspeed*

Berdasarkan kelompok-kelompok data kaji cuaca yang telah dihasilkan dengan menggunakan algoritma *Average Link*, pengujian kelompok dilaksanakan dengan menjalankan beberapa eksperimen. Di antaranya melibatkan penggunaan atribut dari kelompok yang sama dan juga dari kelompok yang berlainan. Di dalam pelaksanaan eksperimen tersebut, atribut data kaji cuaca (atribut selain *rainfall*) telah digunakan sebagai data input kepada proses pengujian, iaitu proses peramalan taburan hujan. Keputusan yang dihasilkan di dalam setiap eskperimen tersebut, akan dicatatkan dan nilai purata bagi nilai ralat min punca kuasa dua (RMS) dan juga nilai pekali korelasi digunakan untuk perbandingan keputusan. Jadual 4(a) dan 4(b) menunjukkan hasil purata bagi nilai RMS dan pekali korelasi yang dihasilkan oleh setiap eksperimen yang dijalankan.

**Jadual 4(a) : Keputusan peramalan taburan hujan bagi kelompok berlainan**

KELOMPOK	100 Set Data		200 Set Data		300 Set Data		400 Set Data		500 Set Data	
	RMS	CC	RMS	CC	RMS	CC	RMS	CC	RMS	CC
2 Kelompok	0.94	0.09	0.87	0.24	0.74	0.15	1.09	0.14	0.96	0.10
3 Kelompok	0.94	0.13	0.83	0.35	0.86	0.23	0.94	0.26	0.78	0.15
4 Kelompok	0.95	0.09	0.85	0.30	0.74	0.54	0.86	0.62	0.81	0.53
5 Kelompok	0.89	0.33	0.70	0.62	0.68	0.64	0.77	0.71	0.74	0.63
6 Kelompok	0.94	0.17	0.58	0.76	0.75	0.55	0.76	0.73	0.76	0.61

Jadual 4(a) di atas menunjukkan keputusan peramalan taburan hujan bagi kelompok berlainan. Bagi eksperimen yang melibatkan penggunaan atribut dari kelompok data kaji cuaca yang berlainan sebagai data input kepada pengujian, didapati keputusan peramalan taburan hujan yang dihasilkan memberikan prestasi peramalan yang semakin baik atau meningkat selari dengan pertambahan bilangan atribut data kaji cuaca yang digunakan. Ini ditunjukkan oleh kesemua eksperimen yang melibatkan penggunaan data set (kecuali 100 set data menunjukkan prestasi yang tidak sekata). Jadual 4(a) juga menunjukkan prestasi peramalan taburan hujan yang dihasilkan di dalam eksperimen yang melibatkan penggunaan 400 set data memberikan prestasi yang terbaik berbanding dengan eksperimen yang lain.

**Jadual 4(b) : Keputusan peramalan taburan hujan bagi kelompok sama**

KELOMPOK	100 Set Data		200 Set Data		300 Set Data		400 Set Data		500 Set Data	
	RMS	CC	RMS	CC	RMS	CC	RMS	CC	RMS	CC
2 Kelompok	0.92	0.04	0.74	0.43	0.81	0.29	0.97	0.34	0.84	0.34
3 Kelompok	0.92	0.16	0.88	0.13	0.84	0.18	0.55	0.67	0.95	0.13
4 Kelompok	0.94	0.10	0.89	0.11	0.88	0.11	1.11	0.05	0.97	0.04
5 Kelompok	0.95	0.05	0.87	0.15	0.89	0.09	1.11	0.24	0.97	0.01
6 Kelompok	0.95	0.03	0.89	0.08	0.89	0.07	1.11	0.06	0.97	0.01

Manakala Jadual 4(b) pula menunjukkan hasil peramalan taburan hujan yang dilakukan ke atas kelompok atribut data kaji cuaca yang sama. Berdasarkan kepada keputusan peramalan taburan hujan yang ditunjukkan di dalam jadual di atas, didapati semakin banyak bilangan atribut data kaji cuaca yang digunakan sebagai data input kepada proses pengujian, semakin menurun prestasi peramalan taburan hujan yang dihasilkan. Prestasi ini ditunjukkan oleh nilai RMS yang semakin tinggi dan juga nilai pekali korelasi

yang semakin menghampiri 0 di dalam kesemua eksperimen yang melibatkan 100 set data, 200 set data, 300 set data, 400 set data dan 500 set data kaji cuaca.

Oleh yang demikian, dapatlah dirumuskan bahawa dengan menggunakan algoritma *Average Link* untuk mengelompokkan data kaji cuaca, didapati keputusan peramalan taburan hujan yang dihasilkan oleh kelompok yang berlainan memberikan prestasi yang lebih baik berbanding dengan keputusan peramalan taburan hujan yang dihasilkan oleh kelompok yang sama.

### 5.3 Algoritma *Complete Link*

Jadual 5 menunjukkan hasil pengelompokan data kaji cuaca yang dilakukan dengan menggunakan algoritma *Complete Link*. Berdasarkan hasil pengelompokan ini, beberapa eksperimen telah dilaksanakan untuk melihat prestasi peramalan taburan hujan yang dihasilkan. Eksperimen ini dijalankan dengan melibatkan penggunaan atribut-atribut data kaji cuaca dari kelompok yang sama dan juga kelompok yang berlainan.

**Jadual 5 : Hasil pengelompokan data kaji cuaca menggunakan algoritma *Complete Link***

Bil. Kelompok	100 Set Data	200 Set Data	300 Set Data	400 Set Data	500 Set Data
2	( a, d, f, g ) ( b, c, e )	( a, d, f, g ) ( b, c, e )	( a, c, d ) ( b, e, f, g )	( a, c, d ) ( b, e, f, g )	( a, b, c, d ) ( e, f, g )
3	( a, d, f, g ) ( b ) ( c, e )	( a, d, f, g ) ( b ) ( c, e )	( a, c, d ) ( b ) ( e, f, g )	( a, c, d ) ( b ) ( e, f, g )	( a, c, d ) ( b ) ( e, f, g )
4	( a, f, g ) ( b ) ( c, e ) ( d )	( a, f, g ) ( b ) ( c, e ) ( d )	( a, c ) ( b ) ( d ) ( e, f, g )	( a, c, d ) ( b ) ( e ) ( f, g )	( a, c ) ( b ) ( d ) ( e, f, g )
5	( a, f, g ) ( b ) ( c ) ( d ) ( e )	( a, f, g ) ( b ) ( c ) ( d ) ( e )	( a, c ) ( b ) ( d ) ( e, f ) ( g )	( a, c ) ( b ) ( d ) ( e ) ( f, g )	( a, c ) ( b ) ( d ) ( e, f ) ( g )
6	( a, f ) ( b ) ( c ) ( d ) ( e ) ( g )	( a, f ) ( b ) ( c ) ( d ) ( e ) ( g )	( a ) ( b ) ( c ) ( d ) ( e, f ) ( g )	( a ) ( b ) ( c ) ( d ) ( e ) ( f, g )	( a ) ( b ) ( c ) ( d ) ( e, f ) ( g )

a – windvane, b – humidity, c – energy, d – temp, e – tension, f – radiation, g - windspeed

Jadual 6(a) menunjukkan keputusan peramalan taburan hujan yang melibatkan penggunaan atribut data kaji cuaca dari kelompok berlainan. Manakala 6(b) pula menunjukkan keputusan peramalan taburan hujan yang menggunakan atribut data kaji cuaca dari kelompok sama sebagai data input kepada eksperimen yang telah dijalankan.

**Jadual 6(a) : Keputusan peramalan taburan hujan bagi kelompok berlainan**

KELOMPOK	100 Set Data		200 Set Data		300 Set Data		400 Set Data		500 Set Data	
	RMS	CC	RMS	CC	RMS	CC	RMS	CC	RMS	CC
2 Kelompok	0.94	0.09	0.87	0.22	0.88	0.15	1.06	0.10	0.96	0.11
3 Kelompok	0.95	0.08	0.83	0.32	0.84	0.28	0.95	0.34	0.94	0.19
4 Kelompok	0.71	0.27	0.73	0.52	0.76	0.48	0.90	0.55	0.91	0.32
5 Kelompok	0.74	0.16	0.71	0.57	0.60	0.71	0.89	0.59	0.76	0.61
6 Kelompok	0.72	0.25	0.71	0.60	0.56	0.75	0.76	0.73	0.76	0.61

Bagi eksperimen yang melibatkan penggunaan atribut dari kelompok yang sama, didapati keputusan peramalan taburan hujan yang dihasilkan menunjukkan semakin banyak bilangan atribut data kaji cuaca yang digunakan sebagai data input kepada proses peramalan, semakin meningkat prestasi peramalan taburan hujan yang dihasilkan. Ini ditunjukkan oleh kesemua eksperimen yang melibatkan 100 set data, 200 set data, 300 set data, 400 set data dan 500 set data, di mana nilai RMS yang dihasilkan semakin menurun manakala nilai pekali korelasinya semakin meningkat. Selain daripada itu, didapati eksperimen yang menggunakan 200 set data, 300 set data dan 400 set data menunjukkan prestasi peramalan yang lebih baik berbanding yang lain.

**Jadual 6(b) : Keputusan peramalan taburan hujan bagi kelompok sama**

KELOMPOK	100 Set Data		200 Set Data		300 Set Data		400 Set Data		500 Set Data	
	RMS	CC	RMS	CC	RMS	CC	RMS	CC	RMS	CC
2 Kelompok	0.95	0.08	0.95	0.09	0.83	0.35	0.91	0.49	0.87	0.41
3 Kelompok	0.94	0.11	0.80	0.33	0.84	0.30	0.90	0.26	0.92	0.19
4 Kelompok	0.95	0.06	0.87	0.22	0.87	0.16	0.95	0.18	0.93	0.14
5 Kelompok	0.95	0.05	0.87	0.17	0.88	0.12	0.95	0.10	0.98	0.04
6 Kelompok	0.95	0.04	0.89	0.10	0.88	0.16	0.95	0.09	0.98	0.03

Berdasarkan keputusan peramalan yang dipaparkan pada Jadual 6(b) di atas, didapati prestasi peramalan yang dihasilkan semakin menurun selari dengan peningkatan bilangan atribut data kaji cuaca yang digunakan di dalam proses peramalan taburan hujan tersebut. Ini ditunjukkan oleh kesemua eksperimen yang dijalankan, yang melibatkan penggunaan 100 set data, 200 set data, 300 set data, 400 set data dan 500 set data.

Oleh yang demikian, maka bolehlah dirumuskan bahawa dengan mengelompokkan data kaji cuaca menggunakan algoritma *Complete Link*, didapati prestasi peramalan taburan hujan yang dihasilkan dengan menggunakan atribut data kaji cuaca dari kelompok berlainan adalah lebih baik jika dibandingkan dengan prestasi peramalan taburan hujan yang dihasilkan dengan menggunakan atribut data kaji cuaca dari kelompok sama. Ini dibuktikan oleh nilai RMS dan pekali korelasinya yang ditunjukkan pada Jadual 6(a) dan Jadual 6(b).

## 6.0 Kesimpulan

Berdasarkan kepada keputusan eksperimen yang telah dijalankan, maka secara keseluruhannya bolehlah dirumuskan bahawa prestasi peramalan taburan hujan yang dihasilkan oleh eksperimen yang melibatkan penggunaan atribut-atribut data kaji cuaca dari kelompok yang berlainan adalah lebih baik jika dibandingkan dengan prestasi peramalan taburan hujan yang menggunakan atribut data kaji cuaca dari kelompok yang sama. Ini ditunjukkan oleh nilai RMS eksperimennya yang semakin menghampiri 0 dan juga nilai pekali korelasinya yang semakin menghampiri 1 di dalam kesemua eksperimen yang melibatkan ketiga-tiga teknik pengelompokan iaitu algoritma *Single Link*, *Average Link* dan juga *Complete Link*.

Walau bagaimanapun, didapati algoritma *Complete Link* memberikan prestasi peramalan taburan hujan yang lebih baik berbanding dua algoritma yang lain, diikuti oleh algoritma *Single Link* dan akhir sekali *Average Link*. Sehubungan dengan itu, kajian ini boleh diteruskan dengan membuat perbandingan di antara teknik pengelompokan Hierarki *Agglomerative* dengan teknik-teknik pengelompokan yang lain, contohnya teknik pengelompokan Hierarki *Divisive* (seperti *Divisive Analysis* dan *Monothetic Analysis*) dan teknik pengelompokan *Partitional* (seperti *Square Error*, *Graph Theoretic*, *Mixture Resolving* dan *Mode Seeking*).

## Rujukan

Chen, T., and Takagi, M. (1993). Rainfall prediction of geostationary Meteorological satellite images using artificial neural network. *International Geoscience and Remote Sensing Symposium*. 3:1247-1249.

- Cimiano, P., Hotho, A. and Staab, S. (2001). Comparing conceptual, divisive and agglomerative clustering for learning taxonomies from text.
- Haldiki, M., Batistakis, Y., and Vazirgiannis, M. (2001). Clustering algorithms and validity measures. Tutorial paper, *Proceedings of SSDBM Conference*.3-22.
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data Clustering : A review. *ACM Computing Surveys (CSUR)*. 31(3):264-323.
- Lin, H. (1999). Survey and implementation of clustering algorithms. Theses. Hsinchu, Taiwan, Republic of China.
- Liu, J. N. K., and Lee, R. S. T. (1999). Rainfall forecasting from multiple point sources using neural networks. *In Proceedings of the 1999 IEEE International Conference on Systems, Man, and Cybernetics (SMC '99)*. 3:429-434.
- Malaysia Meteorological Services (2004).  
[Online] Available : <http://www.kjc.gov.my/>
- McCullagh, J., Bluff, K., and Ebert, E. (1995). A Neural network model for rainfall estimation. *The Second New Zealand International Two Stream Conference on Artificial Neural Networks and Expert Systems*. 389-392.
- McCullagh, J., Bluff, K., and Hendtlass, T. (1999). Evolving expert neural networks for meteorological rainfall estimations. *Proceedings of the International Conference on Neural Information Processing and Intelligent Information Systems IEEE (ICONIP '99)*. 2:585-590.
- Ochiai, K., Suzuki, H., Shinozawa, K., Fujii, M. and Sonehara, N. (1995). Snowfall and rainfall forecasting from weather radar images with artificial neural networks. *Proceedings of IEEE International Conference*. 2:1182-1187.
- Szymkowiak, A., Larsen, J. and Hansen, L. K. (2001). Hierarchical clustering for data mining.

Ultsch, A., and Guimareas, G. (1996). Classification and prediction of hail using self-organizing neural networks. *In Proceedings of the International Conference on Neural Networks ICNN '96*. 1622-1627.

V. Diyankov, Vladimir A. Lykov and Serge A. Terekhoff (1992). Artificial neural networks in weather forecasting.

Xu, R. and Wunsch, D. (2005). Survey of Clustering Algorithms. *IEE Transactions on Neural Networks*. Vol 16. 645-678.

Zhao, Y. and Karypis, G. (2002). Evaluation of hierarchical clustering algorithms for document datasets. (*CIKM'02*). 515-524.