

Penerapan Metode Association Rule Menggunakan Algoritma Apriori pada Simulasi Prediksi Hujan Wilayah Kota Bandung

Mohamad Fauzy¹, Kemas Rahmat Saleh W², Ibnu Asror³

^{1,2,3}Fakultas Informatika Telkom University, Bandung

¹mrojio12@gmail.com, ²bagindokemas@telkomuniversity.ac.id, ³ibn@telkomuniversity.ac.id

Abstrak

Prakiraan cuaca saat ini telah menjadi satu hal yang dibutuhkan bagi banyak orang di dunia. Dalam memprediksi hujan pengolahan data cuaca merupakan hal yang penting. Namun permasalahannya, data cuaca yang semakin hari semakin bertambah menyebabkan penumpukan data sehingga pengolahan data tersebut perlu penanganan lebih lanjut. Oleh karena itu pemanfaatan data mining digunakan untuk menyelesaikan masalah ini. *Association rule mining* adalah salah satu metode data mining yang dapat mengidentifikasi hubungan kesamaan antar *item*. Pada tugas akhir ini akan mencoba menerapkan *Association rule mining* dengan harapan aturan-aturan asosiatif yang dihasilkan dapat menjadi acuan dalam memprakirakan cuaca.

Tugas akhir ini dilakukan dengan tiga tahapan utama yaitu : 1) melakukan analisa pola frekuensi tinggi menggunakan algoritma apriori; 2) pembentukan aturan asosiasi (*association rule*); 3) uji kekuatan *rule* yang terbentuk dengan menghitung *lift ratio* pada masing-masing *rule*. Dataset yang digunakan adalah data klimatologi yang diambil dari BMKG stasiun geofisika kelas 1 Bandung. Hasil akhir dari tugas akhir ini berupa aturan-aturan asosiasi (*association rules*) dimana aturan-aturan ini dapat dijadikan sebagai acuan dalam memprediksi cuaca hujan atau tidak hujan untuk satu hari kedepan.

Kata kunci : *data mining, association rule, apriori, prediksi hujan*

Abstract

Weather forecast today has become a necessary thing for many people in the world. In predicting rain weather data processing is essential. But the problem, weather data that is increasingly growing cause the accumulation of data so that the data processing needs further treatment. Therefore, the use of data mining is used to solve this problem. Association rule mining is one of data mining methods that can identify similarity relationships between items. In this final project will try to implement the Association rule mining in hopes of associative rules generated can become a reference in weather forecasting.

The final task is performed by three main stages, namely: 1) to analyze high frequency patterns using algorithms priori; 2) the establishment of an association rule (association rule); 3) test the strength of the rule which is formed by calculating the ratio elevator on each rule. The dataset used is the climatological data taken from BMKG station 1st class geophysical Bandung. The end result of this thesis in the form of rules of association (association rules) in which these rules can be used as a reference in predicting the weather is rain or not rain for the next day.

Keywords: *data mining, association rule, apriori, rain forecast*

1. Pendahuluan

Prakiraan cuaca saat ini telah menjadi satu hal yang dibutuhkan bagi banyak orang di dunia. Dalam memprediksi hujan pengolahan data cuaca merupakan hal yang penting. Namun permasalahannya, data cuaca bertambah setiap hari dan prakiraan cuaca dibutuhkan untuk keesokan harinya. Oleh karena itu dibutuhkan teknik data mining yang cepat namun akurat untuk menyelesaikan masalah ini. Salah satu teknik yang dapat diterapkan yaitu *Association rule mining*.

Association rule mining adalah salah satu metode data mining yang dapat mengidentifikasi hubungan kesamaan antar *item*. Dalam metode ini dibutuhkan algoritma untuk mencari kandidat aturan

asosiasi. Salah satu algoritma yang biasa dipakai adalah algoritma apriori. Kelebihan asosiasi rule dengan apriori ini adalah lebih sederhana dan dapat menangani data yang besar. Sedangkan algoritma lainnya memiliki kelemahan dalam penggunaan memori saat jumlah data besar, tentunya berpengaruh terhadap banyaknya *item* yang diproses. Penting tidaknya aturan asosiasi dapat diketahui dengan 2 parameter, *minimum support* (prosentase kombinasi item dalam database) dan *minimum confidence* (kuatnya hubungan antar item dalam aturan asosiatif), keduanya ditentukan oleh *user*[1]. Penggunaan apriori ditujukan untuk menemukan asosiasi rule dalam simulasi prediksi hujan ini yaitu menambang keterhubungan antara *item-item* yang terkandung dalam data cuaca.

Pada tugas akhir ini, akan melakukan penelitian dengan membuat sistem yang menerapkan metode *association rule mining* sebagai alat bantu prakiraan cuaca. Data klimatologi dalam satu hari dianggap sebagai satu transaksi. Unsur-unsur yang terdapat di dalam data yaitu suhu, rata kelembapan, rata kecepatan angin, arah angin dan curah hujan masing-masing dianggap sebuah *item*. Pendekatan yang dilakukan adalah mencari kesamaan-kesamaan antar *item* pada tiap transaksi. Kemudian dibentuk aturan asosiatif berdasarkan kesamaan-kesamaan tersebut. Aturan asosiatif dijadikan sebagai acuan dalam prakiraan cuaca pada penelitian ini.

Tugas akhir ini dilakukan dengan tiga tahapan utama yaitu : 1) melakukan analisa pola frekuensi tinggi menggunakan algoritma apriori; 2) pembentukan aturan asosiasi (*association rule*); 3) uji kekuatan *rule* yang terbentuk dengan menghitung lift ratio pada masing-masing *rule*. *Dataset* yang digunakan adalah data klimatologi yang diambil dari BMKG stasiun geofisika kelas 1 Bandung. Hasil akhir dari tugas akhir ini berupa aturan-aturan asosiasi (*association rules*) dimana aturan-aturan ini dapat dijadikan sebagai acuan dalam memprediksi cuaca hujan atau tidak hujan untuk satu hari kedepan.

2. Landasan Teori

2.1 Association Rule Mining

Analisis asosiasi atau *association rule mining* adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari salah satu teknik data mining lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien, yaitu analisis pola frekuensi tinggi (*frequent pattern mining*) [5].

Secara umum *association rule* mempunyai bentuk : LHS \Rightarrow RHS dimana LHS dan RHS tersebut adalah himpunan item; jika setiap item-item dalam LHS terdapat dalam transaksi maka item-item dalam RHS juga terdapat dalam transaksi.

Aturan asosiasi biasanya dinyatakan dalam bentuk[5]:

$$\{A,B\} \Rightarrow \{C\} \text{ (support = 10\%, confidence = 50\%)}$$

2.2 Support

Support dari suatu *association rule* adalah presentasi kombinasi *item* tersebut dalam database, dimana jika mempunyai *item* A dan *item* B maka support adalah proporsi dari transaksi dalam database yang mengandung A dan B. Rumus untuk menghitung nilai support dari dua *item* tersebut adalah sebagai berikut[5][6]:

$$\text{Support}(A, B) = P(A \cap B)$$

$$\text{Support}(A, B) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi}}$$

2.3 Confidence

Confidence dari *association rule* adalah ukuran ketepatan suatu *rule*, yaitu presentasi transaksi dalam database yang mengandung A dan mengandung B. Dengan adanya *confidence* kita dapat mengukur kuatnya hubungan antar-*item* dalam *association rule*. Rumus untuk menghitung nilai *confidence* dari dua *item* tersebut adalah sebagai berikut[5][6]:

$$\text{Confidence} = P(B | A)$$

$$\text{Confidence} = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi mengandung A}}$$

2.4 Algoritma Apriori

Algoritma apriori adalah langkah untuk proses menemukan *frequent-itemset* dengan melakukan iterasi pada data. Dimana *itemset* adalah himpunan *item-item* yang berada di dalam himpunan yang diolah oleh sistem, sedangkan *frequent-itemset* menunjukkan *itemset* yang memiliki frekuensi kemunculan lebih dari nilai minimum yang telah ditentukan (ϕ). Pada iterasi ke-*k*, semua *itemset* yang ditemukan yang memiliki *k* *item* disebut *k-itemset*. Setiap iterasi terdiri dari dua tahap yaitu pembangkitan kandidat dan pembangkitan *rule*.

Pada tahap pembangkitan kandidat (*candidate generation*) dimana himpunan semua *frequent(k-1)-itemset* F_{k-1} yang digunakan pada pass ke-(*k-1*) digunakan untuk membangkitkan kandidat *itemset* C_k . Prosedur pembangkitan kandidat menjamin bahwa C_k adalah superset dari himpunan. Kemudian di-scan dalam tahap perhitungan *support* (*support counting*). Pada akhir pass C_k diperiksa untuk menentukan kandidat mana yang muncul, menghasilkan F_k . Perhitungan *support* berakhir ketika C_k atau C_{k+1} kosong.

Pada tahap membangkitkan *rule*, akan dibangkitkan lebih dahulu *candidate rule*. *Candidate rule* berisi semua kemungkinan *rule* yang memiliki *support* > *minimum support* karena imputan *candidate rule* adalah *frequent-itemset*. Kemudian *candidate rule* akan di-join dengan table F untuk menemukan *support antecedent*. *Confidence rule* dihitung dengan cara membandingkan *support rule* dengan *support antecedent rule*. Hanya *rule* yang mempunyai *confidence* > *minimum confidence* yang disimpan dalam table *rule* (table R)[6].

2.5 Lift Ratio

Lift ratio adalah suatu ukuran untuk mengetahui kekuatan aturan asosiasi (*association rule*) yang telah terbentuk. Nilai *lift ratio* biasanya digunakan sebagai penentu apakah aturan asosiasi valid atau

tidak valid. Untuk menghitung *lift ratio* digunakan rumus sebagai berikut:

$$Lift\ Ratio = \frac{Confidence\ (A,B)}{Benchmark\ Confidence\ (A,B)}$$

Untuk mendapatkan nilai *benchmark confidence* sendiri dapat dihitung menggunakan rumus sebagai berikut:

$$Benchmark\ Confidence = \frac{Nc}{N}$$

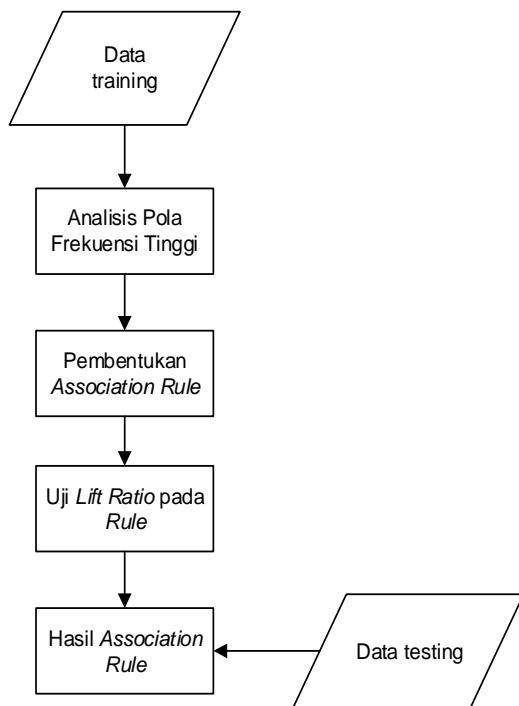
Keterangan:

- Nc = jumlah transaksi dengan item yang menjadi consequent

- N = jumlah transaksi basis data

3. Analisis dan Perancangan Sistem

Gambar 1. Perancangan Umum Sistem



Skenario keseluruhan sistem :

1. Data training dan data testing adalah data yang sudah melalui proses preprocessing secara manual.
2. Dengan menggunakan algoritma apriori akan dilakukan training terhadap itemset yang terdapat di dalam data klimatologi yaitu dengan membentuk kandidat kandidat itemset atau biasa disebut dengan k-itemset.
3. Setiap k-itemset yang terbentuk sebelumnya kemudian dilakukan perhitungan nilai support

untuk menghitung jumlah transaksi yang memuat semua item di dalam k-itemset tersebut. K-itemset yang nilai supportnya lebih tinggi dari nilai minimum support kemudian dijadikan kandidat sebagai pembentukan aturan asosiasi (association rule).

4. Dilakukan perhitungan *confidence* dari tiap k-itemset untuk menentukan apakah kandidat tersebut dapat dijadikan sebagai aturan asosiasi (association rule) atau tidak.
5. Rule yang terbentuk di evaluasi kekuatannya dengan cara uji lift ratio. Nilai lift ratio dari sebuah rule didapatkan melalui perbandingan *confidence* rule tersebut dengan *benchmark confidence*.
6. Selanjutnya dikeluarkan hasil akhir aturan asosiasi dengan format jika (k-itemset) maka hujan atau tidak hujan. Nilai k-itemset tergantung berdasarkan k-itemset yang terbentuk.

4. Pengujian

4.1 Tujuan Pengujian

Tujuan dari pengujian yang dilakukan adalah sebagai berikut ini:

1. Menganalisa pengaruh *minimum support* terhadap jumlah *frequent itemset* yang dibangkitkan algoritma apriori.
2. Menganalisa pengaruh *minimum confidence* terhadap jumlah *rule* yang dihasilkan berdasarkan *frequent itemset*.
3. Menganalisa pengaruh *minimum support* dan *minimum confidence* terhadap akurasi dari sistem yang telah dibangun.
4. Mengetahui kekuatan dari tiap *rule* yang dihasilkan.

4.2 Dataset

Dataset yang digunakan oleh sistem yang dibangun berupa *file* bertipe excel yang berisikan data cuaca yang diambil dari BMKG stasiun geofisika kelas 1 Bandung. Penginputan data ke *file* excel dilakukan secara manual karena data yang di dapat dari BMKG merupakan data fisik. Lalu data akan masuk ke proses *data preprocessing* yang akan menghasilkan dataset yang akan dipakai oleh sistem.

5. Analisis Hasil Pengujian

5.1 Analisis Hasil Pengaruh *Minimum Support* dan *Minimum Confidence* terhadap *Frequent Itemset*

Tabel 1. *frequent itemset*

mincof\minsup	10%	20%	30%	40%	50%
50%	73	40	19	17	7
60%	73	40	19	17	7
70%	73	40	19	17	7

Berdasarkan tabel 1, dapat dilihat jumlah *frequent itemset* terbanyak dibangkitkan oleh *minimum support* terkecil dalam pengujian ini yaitu 10%, sedangkan *frequent itemset* paling sedikit dibangkitkan oleh *minimum support* sebesar 50%. Hal ini dikarenakan nilai *minimum support* adalah nilai acuan dalam perhitungan sebuah *itemset*. Hanya *itemset* yang nilai supportnya sama atau lebih besar dari nilai *minimum support* saja yang dijadikan sebagai *frequent itemset*. Sedangkan *minimum confidence* tidak memiliki pengaruh dalam pembangkitan *frequent itemset* karena tidak terjadi perhitungan *confidence* pada tahap ini. Maka dapat disimpulkan bahwa semakin tinggi *minimum support* yang digunakan maka akan semakin sedikit jumlah *frequent itemset* yang dibangkitkan.

5.2 Analisis Hasil Pengaruh *Minimum Support* *Minimum* dan *Confidence* terhadap *Rule*

Tabel 2. Jumlah *rule*

mincof\minsup	10%	20%	30%	40%	50%
50%	9	5	2	2	0
60%	3	1	0	0	0
70%	2	0	0	0	0

Berdasarkan tabel 2, dapat dilihat jumlah aturan asosiatif yang terbentuk terbanyak dihasilkan oleh *minimum support* 10% dan *minimum confidence* 50% yaitu sebanyak 9 *rule*. Pada setiap penambahan nilai *minimum support* dan *minimum confidence* yang diujikan terlihat *rule* yang terbentuk semakin berkurang. Penurunan ini dipengaruhi oleh jumlah *frequent itemset* yang dibangkitkan pada pengujian sebelumnya. Pada *minimum support* 50% dapat dilihat tidak terbentuk satu pun aturan asosiatif, hal ini diakibatkan nilai *confidence* dari *frequent itemset* yang dibangkitkan tidak ada yang memenuhi syarat *minimum confidence* yang diujikan. Oleh karena itu, dapat dilihat bahwa *minimum confidence* berpengaruh pada pembentukan *rule*. Pada pengujian ini dapat disimpulkan semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka *rule* yang dihasilkan akan semakin sedikit.

5.3 Analisis Hasil Pengaruh *Minimum Support* *Minimum* dan *Confidence* terhadap akurasi

Tabel 3. Akurasi data training

mincof\minsup	10%	20%	30%	40%	50%
50%	76.78	76.71	49.73	49.73	0
60%	42.67	26.71	0	0	0
70%	15.96	0	0	0	0

Berdasarkan tabel 3 di atas, akurasi tertinggi diperoleh sebesar 76.78% dengan *minimum support* 10% dan *minimum confidence* 50%. Sedangkan pada *minimum support* 50%, dan yang bernilai akurasi 0 lainnya tidak dapat dihitung besarnya akurasi karena tidak ada *rule* yang terbentuk. Dari grafik diatas dapat disimpulkan semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka akurasi yang dihasilkan akan semakin kecil. Hal ini terjadi karena *rule-rule* yang dihasilkan semakin kuat keterkaitan antar *item*-nya. Sehingga pada saat data training diuji kedalam *rule* dan dicocokkan kembali dengan data nilai kebenarannya semakin berkurang. Nilai tersebut di prosentasekan dan dijadikan nilai akurasi dari data training. Dalam pengujian akurasi data training dapat dilihat sistem berfungsi optimal pada *minimum support* sebesar 20% dan *minimum confidence* sebesar 50% yang menghasilkan akurasi sebesar 76.71%.

Tabel 4. Akurasi data testing

mincof\minsup	10%	20%	30%	40%	50%
50%	75.89	75.89	50.41	50.41	0
60%	30.69	24.11	0	0	0
70%	6.58	0	0	0	0

Berdasarkan tabel 4 di atas, akurasi tertinggi diperoleh sebesar 75.89% terdapat pada dua *minimum support* berbeda yaitu *minimum support* 10% dan *minimum support* 20% dengan *minimum confidence* 50%. Akurasi berikutnya juga diperoleh pada dua *minimum support* yang berbeda yaitu *minimum support* 30% dan *minimum support* 40% dengan *minimum confidence* 50% sebesar 50.41%. Sedangkan pada *minimum support* 50%, dan yang bernilai akurasi 0 lainnya tidak dapat dihitung besarnya akurasi karena tidak ada *rule* yang terbentuk. Dari grafik diatas dapat disimpulkan semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka akurasi yang dihasilkan akan semakin kecil. Hal ini terjadi karena *rule-rule* yang dihasilkan semakin kuat keterkaitan antar *item*-nya. Sehingga pada saat data testing diuji kedalam *rule* dan dicocokkan kembali dengan data nilai kebenarannya semakin berkurang. Nilai tersebut di prosentasekan dan dijadikan nilai akurasi dari data testing. Dalam pengujian akurasi data testing dapat dilihat sistem berfungsi optimal pada *minimum support* sebesar 20% dan *minimum confidence* sebesar 50% yang menghasilkan akurasi sebesar 75.89%.

5.4 Analisis Hasil *Lift Ratio* terhadap Hasil *Rule*

Tabel 5. *Lift ratio* tiap *rule*

minimum support	minimum confidence	aturan asosiatif	confidence	lift ratio

10%	50%	Agak Panas(temperature) E(windDirection) -> tidak hujan	73.08	1.46
10%	50%	E(windDirection) -> tidak hujan	72.81	1.45
10%	50%	W(windDirection) -> hujan	62.9	1.26
10%	50%	Agak Basah(humidity) Agak Kencang(windSpeed) -> tidak hujan	53.41	1.06
10%	50%	Agak Kencang(windSpeed) -> tidak hujan	53.38	1.06
10%	50%	Agak Panas(temperature) Pelan(windSpeed) -> tidak hujan	51.94	1.03
10%	50%	Pelan(windSpeed) -> tidak hujan	50.88	1.01
10%	50%	Agak Basah(humidity) -> tidak hujan	50.66	1.01
10%	50%	Agak Panas(temperature) -> tidak hujan	50.62	1.01
20%	50%	W(windDirection) -> hujan	62.9	1.26
20%	50%	Agak Panas(temperature) Pelan(windSpeed) -> tidak hujan	51.94	1.03
20%	50%	Pelan(windSpeed) -> tidak hujan	50.88	1.01
20%	50%	Agak Basah(humidity) -> tidak hujan	50.66	1.01
20%	50%	Agak Panas(temperature) -> tidak hujan	50.62	1.01
30%	50%	Agak Basah(humidity) -> tidak hujan	50.66	1.01
30%	50%	Agak Panas(temperature) -> tidak hujan	50.62	1.01
40%	50%	Agak Basah(humidity) -> tidak hujan	50.66	1.01
40%	50%	Agak Panas(temperature) -> tidak hujan	50.62	1.01

Berdasarkan tabel 5 diatas *lift ratio* untuk semua *rule* yang berhasil dibentuk memiliki nilai lebih besar dari 1 (*lift ratio* >1). Hal ini menunjukkan bahwa semua *rule* tersebut bersifat kuat dan valid

untuk digunakan sebagai acuan dalam memprediksi hujan.

6. Kesimpulan dan Saran

6.1 Kesimpulan

Berdasarkan pengujian yang dilakukan dalam Tugas Akhir ini, dapat disimpulkan bahwa :

1. Metode *association rule mining* dengan menggunakan algoritma apriori dapat diterapkan pada sistem simulasi prediksi hujan.
2. Semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka semakin sedikit jumlah *frequent itemset* dan *rule* yang terbentuk serta akurasi semakin berkurang.
3. Semua *rule* yang dihasilkan pada penelitian ini memiliki nilai *lift ratio* lebih dari 1.00 sehingga dapat digunakan sebagai acuan dalam memprediksi hujan.

6.2 Saran

Saran yang diperlukan untuk pengembangan sistem lebih lanjut adalah sebagai berikut :

1. Kategori yang dilakukan pada penelitian ini adalah hujan dan tidak hujan. Untuk kedepannya dapat dilakukan pengkategorian yang lebih luas sehingga dapat diketahui intensitas hujan yang turun.
2. Untuk selanjutnya dapat menggunakan dataset yang jauh lebih besar dari sekarang, misalnya data yang diolah sebanyak 10 tahun keatas.
3. Diharapkan untuk penelitian selanjutnya dapat menerapkan algoritma pembangkitan *frequent itemset* yang berbeda, seperti *FP-Growth*.

Daftar Pustaka:

- [1] Han, Jiawei., Kamber, Micheline.(2009). Data mining: Concepts and Techniques. San Francisco, CA. Morgan Kauffman.
- [2] Santosa, Budi.(2007). Data mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu-Bisnis.Edisi Pertama.
- [3] The Parallel Computer Center. 1995. Data Mining : An Introduction. Student Note. The Queen's University of Belfast. <http://www.pcc.qub.ac.uk>
- [4] Mujiasih, Subekti. (2011). Pemanfaatan Data Mining Untuk Prakiraan Cuaca. Pusat Meteorologi Penerbangan dan Maritim BMKG Jakarta.
- [5] Kusrini, Luthfi., Emha Taufiq.(2009). Algoritma Data Mining. Yogyakarta: Andi.
- [6] Kusumo, Dana S., Bijaksana, M. Arief., Darmantoro, Dhinta.(2003). Data Mining dengan Algoritma Apriori pada RDBMS Oracle. Bandung; Sekolah Tinggi Teknologi Telkom.

- [7] Indrawan, Bani P., Maharani, Warih., K, Angelina P.(2003). Analisis Sentimen Berdasarkan Fitur Produk Menggunakan Opinion Lexicon dan Wordnet. Bandung; Institut Teknologi Telkom.
- [8] Chengqi, Zhang., Shichao, Zhang.,(2002). Association Rule Mining : Models and Algorithms (Lecture Notes in Computer Science).
- [9] S.Nandagopal., S.Karthik., V.P.Arunachalam(2010). Mining of Meteorological Data Using Modified Apriori Algorithm. EuroJournals Publishing, Inc. <http://www.eurojournals.com/ejsr.htm>
- [10] F.M, Tria., Hidayati, Hetti., Effendy, Veronikha.(2014). Simulasi Klasifikasi Hujan Wilayah Kota Bandung dengan Metode Decision Tree Menggunakan Algoritma C4.5. Bandung; Universitas Telkom.
- [11] Witten, Ian H & Eibe Frank, 2005, *Data Mining : Practical Machine Learning Tools and Technique*, University of Waikato, Morgan Kaufmann Publisher.