

PEMANFAATAN DATA MINING UNTUK PRAKIRAAN CUACA

UTILIZATION OF DATA MINING FOR WEATHER FORECASTING

Subekti Mujiasih

Pusat Meteorologi Penerbangan dan Maritim BMKG, Jl. Angkasa I/2 Jakarta Pusat 10720

Email : subekti.mujiasih@bmkg.go.id

ABSTRACT

Weather forecasting process needs many weather components, big data size and forecaster experience. They cause the accuracy and rapid of forecasting were not well-fulfilled. In order to solve this problem, the research of prediction model was done by using Association Rule and Classification (C4.5, Classification Tree and Random Forest) methods. The input of model production were wind speed, cloud cover, dew point temperature and temperature from 9 marine stations in 2009. The input for testing the resulted model was synoptic data of Tanjung Priok Marine Station since 2002-2010. The result shows the accuracy of C4.5 is highest than others. Accuracy of C4.5 and Association Rule are about 68.5%, and 60.9%, respectively. Thus, the appropriate prediction model is the C4.5. Dominant weather component of C4.5 are cloud cover, dew point temperature and temperature.

Keywords: data mining, association rule, classification tree, random forest, weather

ABSTRAK

Proses prakiraan cuaca memerlukan banyak komponen data cuaca, jumlah data yang besar serta kemampuan prakirawan. Hal ini menyebabkan ketepatan dan kecepatan prakiraan kurang terpenuhi. Untuk memecahkan masalah tersebut, telah dilakukan penelitian model prediksi menggunakan beberapa teknik data mining yakni Association Rule, C4.5, Classification dan Random Forest. Data masukan adalah data sinoptik 9 stasiun maritim tahun 2009. Data masukan tersebut terdiri dari kecepatan angin, tutupan awan, suhu udara dan suhu titik embun. Data untuk pengujian model adalah data sinoptik Stasiun Meteorologi Maritim Tanjung Priok sejak tahun 2002 hingga 2010. Dari serangkaian pembuatan, pemilihan dan pengujian model, hasil penelitian menunjukkan Association Rule mempunyai tingkat akurasi 60.9%, sedangkan C4.5 mempunyai tingkat akurasi 68.5%. Dengan demikian model prediksi yang dipilih adalah model prediksi C4.5. Komponen cuaca yang dominan memungkinkan terjadinya hujan adalah suhu udara, suhu titik embun, dan tutupan awan.

Kata kunci: data mining, association rule, classification tree, random forest, cuaca

Naskah masuk : 3 Agustus 2011

Naskah diterima : 13 September 2011

I. PENDAHULUAN

1.1. Latar Belakang

Data mining atau sering disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran *data mining* ini bisa dipakai untuk membantu pengambilan keputusan di masa depan. Pengembangan KDD ini menyebabkan penggunaan *pattern recognition* semakin berkurang karena telah menjadi bagian data mining¹⁾. Metode ini merupakan gabungan 4 (empat) disiplin ilmu yakni statistik, visualisasi, *database*, dan *machine learning*¹⁾. Adapun *machine learning* adalah suatu area dalam *artificial intelligence* atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik pemrograman berdasarkan pembelajaran masa lalu dan bersinggungan dengan ilmu statistik kadang juga optimasi.

Kajian mengenai data mining untuk prakiraan cuaca telah banyak dilakukan. Pemilihan teknik *Data mining* menggunakan *Association Rule* dengan algoritma *Apriori* menunjukkan hasil yang lebih baik dalam hal kebenaran, proses komputasi, dan terminasi²⁾. Metode *data mining* lainnya yakni *Random Forest* memiliki kemampuan memprediksi turbulensi dan formasi tornado di wilayah benua Amerika³⁾ dan kejadian badai dalam satu jam pertama di setiap piksel data⁴⁾. Selain itu, metode *Clustering* yang merupakan teknik data mining, juga diuji untuk mendeteksi badai di wilayah Amerika⁵⁾. Tidak terbatas hanya pada faktor keakuratan, kecepatan proses prediksi cuaca dapat ditingkatkan dengan menggabungkan kemampuan teknik *data mining* SVM (*Support Vector Machine*) dan arsitektur komputasi yang berbasis *Service Oriented Architecture*⁶⁾. Tidak kalah penting adalah kemampuan *self-organizing data mining* yang diuji melalui *the enhanced Group Method of Data Handling* (e-GMDH), dimana telah dimanfaatkan untuk prakiraan parameter cuaca seperti suhu, curah hujan bulanan dan tekanan udara harian⁷⁾. Teknik data mining lain yakni *Fuzzy Association Rule* juga telah dilakukan untuk memprediksi curah hujan Moonson di wilayah India⁸⁾. Beberapa kajian menggunakan beberapa teknik data mining sekaligus. Pertama, kajian prakiraan hujan jangka pendek menggunakan *Outliner Analysis*,

Clustering, *Prediction*, *Classification*, dan *Association Rule* telah diterapkan untuk analisis pola data meteorologi di wilayah Jalur Gaza⁹⁾. Kedua, teknik *Decision Tree*(C4.5), *Artificial Neural Network* (ANN), dan *Support Vector Machine* (SVM) telah dilakukan untuk wilayah Thailand. C4.5 digunakan untuk memprediksi status hujan atau tidak hujan. ANN digunakan untuk memprediksi jumlah hujan. SVM digunakan untuk mengklasifikasikan jumlah hujan berdasarkan tiga kelas yakni tidak hujan, hujan ringan, hujan lebat¹⁰⁾.

BMKG memiliki sekitar 10 stasiun meteorologi maritim dan 3 (tiga) stasiun yang diperbantukan untuk memberikan pelayanan meteorologi maritim. Sebagian besar stasiun tersebut melakukan pengamatan sinoptik dan sebagian diantaranya memberikan pelayanan analisa dan prakiraan cuaca maritim. Data pengamatan ini sangat penting untuk melihat karakteristik cuaca setempat dan pembuatan informasi prakiraan beberapa hari ke depan. Sementara itu dalam proses pembuatan informasi prakiraan cuaca, terdapat beberapa kendala. Pertama, sulitnya membuat informasi prakiraan karena melibatkan banyak sumber data seperti data pengamatan, data model aplikasi cuaca, data gambar kondisi awan dari satelit, data kondisi awan dari radar. Kedua, prakiraan cuaca maritim umumnya mengandalkan kemampuan dari prakirawan, sehingga interpretasi yang dihasilkan bisa berbeda antar prakirawan satu dengan yang lain karena bergantung dari pengalaman masing-masing. Perbedaan interpretasi dapat membingungkan pengguna yang pada akhirnya berpeluang menurunkan kualitas informasi yang disampaikan. Berdasarkan masalah tersebut penulis bermaksud melakukan kajian model prakiraan untuk memperoleh model yang sesuai sehingga memudahkan proses analisa dan prakiraan cuaca maritim.

1.2. Tujuan

Tujuan dari penelitian ini adalah untuk memperoleh model prakiraan yang sesuai agar memudahkan proses analisa dan prakiraan cuaca.

1.3. Teori

Beberapa teknik *data mining* yang cukup dikenal antara lain *Association Rule* dan

Classification. Metode *Classification* bisa dibagi lagi menjadi C4.5, *classification tree* dan *Random Forest*. Ketiga teknik tersebut masing-masing mempunyai kelebihan dan kekurangan. Rincian mengenai bagaimana teknik-teknik bekerja dapat dijelaskan sebagai berikut :

- a. *Association Rule* digunakan untuk memeriksa semua kemungkinan hubungan **if-then** antar item dan memilih hanya yang paling mungkin sebagai indikator dari hubungan ketergantungan antar *item*. Indikator-indikator penting yang diperlukan adalah *support*, *confidence* dan *lift ratio* dimana hubungan ketiganya adalah sebagai berikut :

Asosiasi antara X dan Y, Jika X maka Y, dinotasikan dalam bentuk X'Y

$$\text{Support (X'Y)} = P(X \text{ dan } Y) = P(X) * P(Y) \quad (1)^2$$

$$\begin{aligned} \text{Conf(X} \rightarrow \text{Y)} &= \text{supp(X} \rightarrow \text{Y)} / \text{supp(X)} \\ &= P(X \text{ and } Y) / P(X) \\ &= P(Y | X) \end{aligned} \quad (2)^2$$

$$\begin{aligned} \text{Lift(X} \rightarrow \text{Y)} &= \text{conf(X} \rightarrow \text{Y)} / \text{supp(Y)} \\ &= P(X \text{ and } Y) / (P(X)P(Y)) \end{aligned} \quad (3)^2$$

Keterangan :

X = bagian "jika", Y = bagian "maka"

P(X) = Peluang munculnya item dalam X

P(Y) = Peluang munculnya item dalam Y

P(X dan Y) = Peluang munculnya item baik dalam X maupun dalam Y

$$P(X) = \text{freq(X)} / |D|$$

$$P(Y) = \text{freq(Y)} / |D| \quad \text{freq(X)} = \text{frekwensi munculnya item dalam X}$$

$$\text{freq(Y)} = \text{frekwensi munculnya item dalam Y}$$

$$|D| = \text{Jumlah transaksi dalam database}$$

Support, *confidence* dan *Lift Ratio* adalah (%)

Nilai yang tinggi untuk *consequent* menunjukkan suatu aturan asosiasi yang kuat.

Adapun Nilai *lift ratio* > 1 menunjukkan adanya manfaat dari aturan tersebut, atau dengan kata lain semakin tinggi nilai *lift ratio*, maka lebih semakin kuat asosiasinya.

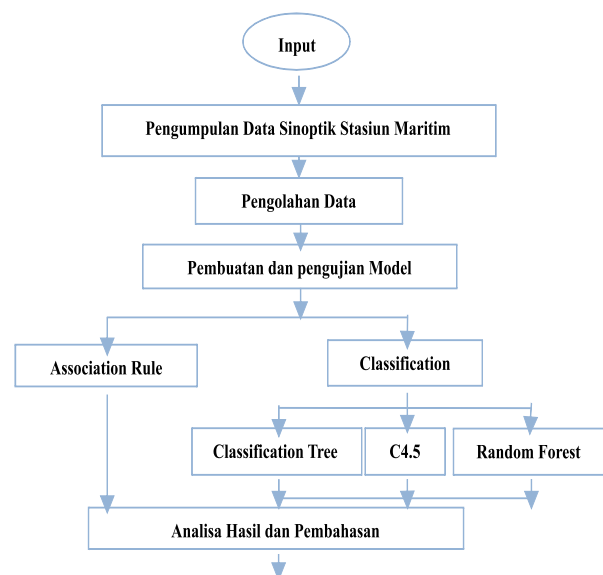
- b. *Classification* adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, yang diperlukan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui¹¹⁾. Model itu bisa berupa aturan "jika-maka", berupa *decision tree*, formula matematis atau *neural network*. *Decision tree* dimana keluarannya bernilai diskrit adalah salah satu

metode *classification* yang paling mudah untuk diinterpretasikan. Algoritma *decision tree* yang paling terkenal adalah C4.5¹²⁾ Proses klasifikasi biasanya dibagi menjadi dua fase yakni *learning* dan *test*. Pada fase *learning*, sebagian data yang telah diketahui kelas datanya diumpungkan untuk membentuk model perkiraan. Kemudian pada tahapan *test*, model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi maka model ini dapat dipakai untuk memprediksi kelas data yang belum diketahui.

Saat ini banyak sekali data mining tool dikembangkan oleh lembaga riset, universitas atau perusahaan yang bergerak di bidang teknologi informasi. Pada penelitian ini penulis menggunakan perangkat lunak *Orange Ailab* sebagai *tool data mining*¹³⁾. *Orange Ailab* adalah perangkat lunak *open source* yang memungkinkan pengguna yang tidak memahami sedikitpun tentang pemrograman dapat melakukan visualisasi dan analisis data. Fitur-fitur yang dimiliki diantaranya *scatterplots*, *bar charts*, *trees*, *dendrograms*, *networks* dan *heatmaps*.

II. METODE PENELITIAN

Metode yang digunakan dapat dilihat dalam diagram alir (Gambar 1). Dari gambar tersebut, data yang diperoleh kemudian diolah, diuji, dan dianalisis.



Gambar 1. Metode Penelitian

2.1. Sumber Data

Data sinoptik berasal dari pengamatan 9 (sembilan) stasiun meteorologi maritim tahun 2009 berukuran (2571 raw data)¹⁴⁾. Data Sinoptik adalah data pengamatan cuaca permukaan yang dikirim dari stasiun-stasiun pengamatan cuaca di seluruh Indonesia setiap tiga jam. Data tersebut meliputi suhu udara, jumlah curah hujan, arah dan kecepatan angin, tekanan udara dan sebagainya. Data yang dikirim tersebut berupa sandi sinoptik dalam bentuk text file. Format data sinoptik mengikuti aturan WMO (*World Meteorological Organization*)¹⁵⁾. Dalam penelitian ini yang digunakan adalah data sinoptik seksi 0 dan seksi 1 yang meliputi kecepatan angin, suhu udara, suhu titik embun, tutupan awan, dan curah hujan.

2.2. Pemilihan Perangkat lunak dan teknik Data mining

Perangkat lunak yang dipilih adalah *Orange Ailab* dengan menggunakan metode *Association Rule*, *Classification Tree*, dan *Random Forest*. Perangkat lunak ini dipilih dengan alasan selain gratis, penulis ingin menguji kehandalan perangkat lunak ini dalam melakukan data mining. Metode *Association Rule* yang termasuk metode *learning unsupervised*¹⁶⁾ diujicobakan pada seluruh komponen data sinoptik yakni tanggal, stasiun, arah angin, kecepatan angin, suhu udara, suhu titik embun, tutupan awan, dan curah hujan. Sedangkan metode *Classification Tree* dan *Random Forest* termasuk metode *supervised* diujicobakan pada 4 komponen data sinoptik yakni suhu udara, suhu titik embun, tutupan awan, dan kecepatan angin untuk memprediksi cuaca di suatu wilayah.

2.3. Pengolahan Data

Data sinoptik yang telah dikumpulkan, diterjemahkan ke dalam nilai yang bukan bentuk kode yakni kecepatan angin, suhu titik embun, suhu udara, dan kondisi cuaca. Kelompok data sinoptik yang diterjemahkan adalah kelompok data seksi 1 yang digaris bawahi seperti berikut.

Format Sandi :

Seksi 0

M_iM_iM_iM_i { (D.....D****) atau (A1bwnbnnb) }
YYGGiw { Iiiii* atau 99L_aL_aL_a Q_cL_oL_oL_o**** }
MMMULaULo*** h_oh_oh_oi_m***

Seksi 1

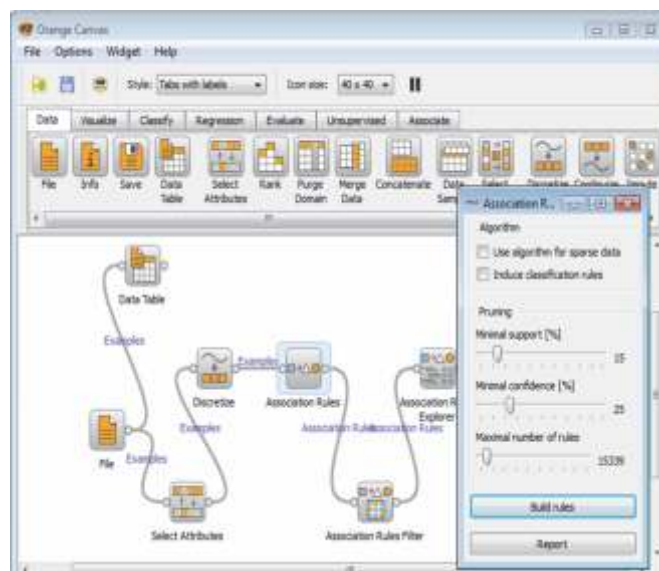
iriXhVV Nddff (00fff) 1S_nTTT { 2S_nT_dT_dT_d
(or 29UUU) } 3P_dP_dP_d{ 4PPPP (or 4a₃hhh)}
5appp 6R R R t_R { 7w w W₁W₂ (or
7wawaW_{a1}W_{a2}) } 8NhcLcMCH 9GGgg

Untuk data tanggal H, kondisi cuaca hujan/tidak hujan dan curah hujan diambil dari data tanggal H+1, karena nilai curah hujan merupakan nilai akumulasi selama 24 jam terakhir. Tahap berikutnya adalah pembuatan model dengan menggunakan data yang telah diterjemahkan dan menguji kehandalannya.

2.4. Pembuatan Model

2.4.1. Metode Association Rule

Pada pengujian ini dipilih *support* = 15% dan *confidence* = 25%, *maximum rule* = 15399, dengan alasan nilai *support* sebesar 15% akan memperkecil jumlah *rule* yang terjadi dan menghemat memori komputer. Adapun *confidence* sebesar 25% adalah untuk memperkuat asosiasi antar *rule* sehingga *rule* yang dihasilkan dapat lebih bermanfaat untuk prediksi. Selain itu, untuk menentukan kuatnya hubungan antara komponen cuaca digunakan *Lift ratio* > 1 seperti pada Gambar 2.

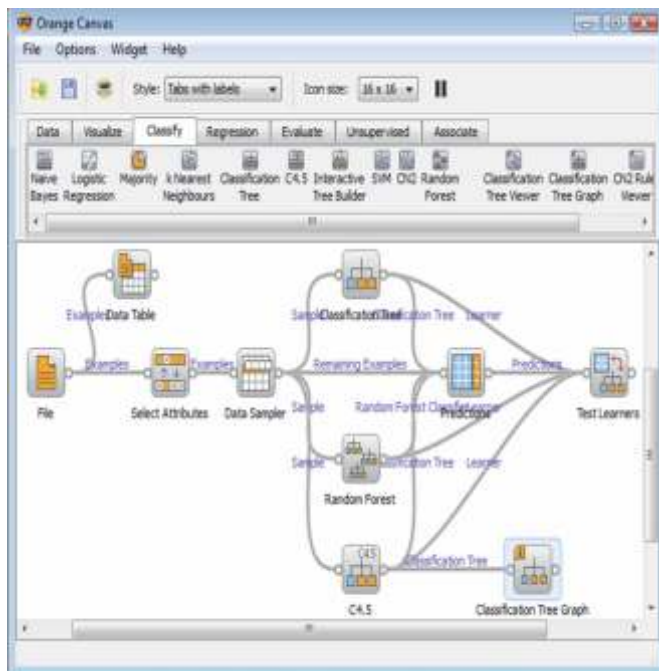


Gambar 2. Proses Association Rule pada Orange Ailab

2.4.2. Metode C4.5, Classification Tree, dan Random Forest

Komponen data sinoptik yang menjadi masukan adalah suhu udara, suhu titik embun,

perbedaan suhu udara dan suhu titik embun serta kecepatan angin. Metode yang digunakan untuk membuat model dengan masukan komponen data cuaca tersebut adalah C4.5, *Classification Tree*, dan *Random Forest*. Model keluaran masing-masing metode tersebut diuji dengan sebagian data masukan untuk melihat kehandalan model. Setelah itu, hasilnya dibandingkan untuk mendapatkan akurasi yang tertinggi, dan memutuskan model prediksi yang paling baik. Proses pembuatan model menggunakan perangkat lunak Orange tampak seperti pada Gambar 3.



Gambar 3. Proses *classification* pada Orange Ailab

III. HASIL DAN PEMBAHASAN

3.1. Association Rule

Hasil *running Orange Ailab* dengan masukan seluruh komponen cuaca sinoptik dengan *support* = 15% dan *confidence* = 25%, *maximum Rule* = 15399 menghasilkan 42 kelompok *rule* seperti tampak pada Gambar 4. Dari kelompok tersebut, penulis melakukan penyaringan dengan beberapa tahap. Pertama, penulis memilih *rule-rule* dengan bentuk asosiasi jika "komponen cuaca seperti suhu udara atau suhu titik embun atau tutupan awan atau kecepatan angin mempunyai nilai tertentu" maka "kondisi hujan yang terjadi bagaimana". Kedua, memilih *rule* dengan nilai *lift rationnya* lebih besar dari satu (*Lift ratio* > 1) seperti pada Gambar 4. Ketiga, jika terdapat *rule* yang sama, maka dipilih

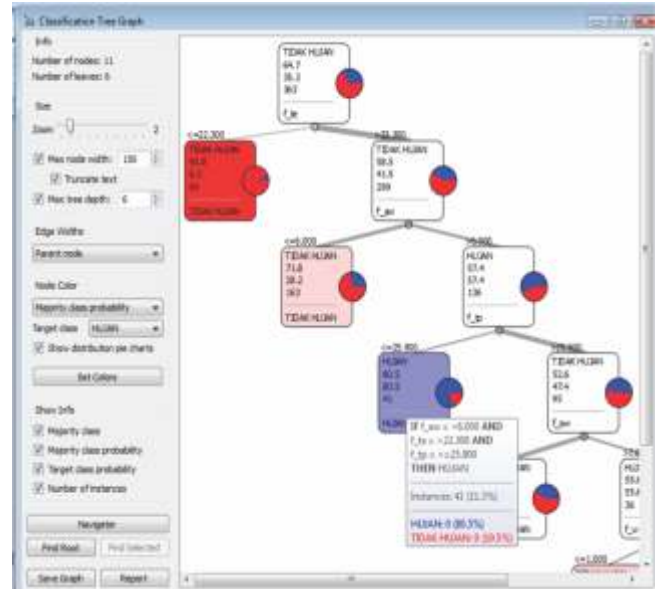
lift rationnya yang paling tinggi. Hasil penyaringan didapatkan satu model dengan bentuk : **Jika suhu udara > 23.9 dan suhu titik embun 23.6 maka terjadi hujan** dengan nilai *support* = 0.307, *confidence* = 0.613, *lift ratio* = 1.248. Pengujian dilakukan dengan menggunakan data sinoptik stasiun Meteorologi Tanjung Priok sejak tahun 2002 hingga 2010. Setelah dilakukan pengujian ternyata tingkat akurasinya sebesar 60.9%

Gambar 4. Hasil Metode Association Rule

3.2. C4.5, Classification Tree, dan Random Forest

Pembuatan model menggunakan teknik C4.5, *Classification Tree* dan *Random Forest* menghasilkan Gambar-5 dan 6. Gambar-5 menunjukkan perbandingan kondisi cuaca riil, hasil keluaran dari masing-masing metode. Sebagian data menunjukkan kesesuaian hasil prediksi C4.5 dengan kondisi cuaca riil. Pada Gambar-6 pemilihan sampling dilakukan dengan *cross validation* dimana *number of folds* adalah 5, *repeat training set* data 5 kali, dengan ukuran data yang dijadikan *training test* adalah sebesar 20% dari total data. Hasil tampak pada Gambar-6, dari ketiga teknik terlihat bahwa tingkat akurasi C4.5 paling tinggi yakni sebesar 69.5% dibandingkan dengan metode *Random Forest* yakni 64.6% dan *Classification Tree* sebesar 64.4%. Dengan demikian model yang digunakan untuk pengujian adalah model keluaran C4.5. Gambar-7 menunjukkan banyaknya model prediksi yang terjadi dari 4 komponen data cuaca tersebut. Namun keluaran yang dihasilkan model umumnya tidak hujan, maka menurut pendapat penulis model ini lebih tepat sebagai "model prediksi tidak hujan".

Kemungkinan ini disebabkan interval data yang relatif pendek, sehingga untuk memperbaiki model diperlukan interval data yang cukup panjang, paling tidak antara 5 - 10 tahun. Gambar 7 menunjukkan salah satu model prediksi yang melibatkan lebih dari satu komponen cuaca. Dimana jika diketahui tutupan awan sebesar 6 suhu titik embun lebih besar dari 22.3 dan suhu udara lebih kecil dari 25.9 maka model tersebut memprediksikan cuaca berpeluang hujan. Selain itu Gambar-7 juga menunjukkan sejumlah model yang dihasilkan metode C4.5 yang mungkin digunakan. Pengujian dilakukan dengan menggunakan data sinoptik stasiun Meteorologi Tanjung Priok sejak tahun 2002 hingga 2010. Setelah dilakukan pengujian ternyata tingkat akurasi sebesar 68.5%



Gambar 7. Classification graph dari C4.5

Berdasarkan hasil pembuatan dan pengujian model pada teknik *Association rule* dan *Classification* (Gambar 7), maka penulis berpendapat bahwa model prediksi yang dihasilkan dari dua metode tersebut tidak memiliki perbedaan yang besar baik dari komponen cuaca yang menyusunnya ataupun dari nilai syarat batas masing-masing komponen cuaca. Hasil *Association rule* menunjukkan faktor penentu hujan atau tidak hujan adalah suhu udara dan suhu titik embun. Sedangkan *Classification* yang diwakili C4.5 menunjukkan faktor penentu hujan atau tidak hujan adalah suhu udara, suhu titik embun, dan tutupan awan. Namun interval syarat batasnya tidak jauh berbeda. Menurut metode *Association Rule*, nilai syarat batas suhu udara ≥ 23.9 dan suhu titik embun ≥ 23.6 . Pada metode C4.5, syarat terjadinya hujan jika suhu udara ≤ 25.9 , suhu titik embun ≥ 22.3 ditambah tutupan awan ≥ 6 . Dengan membandingkan tingkat akurasi, *Association Rule* mempunyai tingkat akurasi 60.9%, sedangkan C.45 mempunyai tingkat akurasi 68.5%. Dengan demikian model prediksi yang disarankan adalah model prediksi C4.5. Tingkat akurasi sebesar 68.5% ini sebenarnya masih mengandung resiko apakah prediksi cuaca sesuai dengan kenyataan. Untuk meningkatkan tingkat akurasi tampaknya diperlukan data seluruh stasiun meteorologi maritim yang memiliki interval 5 - 10 tahun.

	f _{aw}	f _{te}	f _{tp}	f _{ws}	f _{ch}	Random Forest	C4.5	Classification Tr
60	5.0	22.4	21.8	3.0	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
61	8.0	23.0	27.3	0.0	TIDAK HUJAN	HUJAN	TIDAK HUJAN	TIDAK HUJAN
62	3.0	24.2	27.8	3.0	HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
63	7.0	24.3	27.1	0.0	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
64	1.0	18.6	23.8	8.0	HUJAN	HUJAN	TIDAK HUJAN	TIDAK HUJAN
65	6.0	24.0	27.8	8.0	HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
66	7.0	22.7	27.2	8.0	HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
67	7.0	22.1	27.6	0.0	HUJAN	TIDAK HUJAN	TIDAK HUJAN	HUJAN
68	7.0	24.1	27.0	5.0	HUJAN	TIDAK HUJAN	TIDAK HUJAN	HUJAN
69	3.0	21.7	26.5	4.0	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
70	6.0	22.0	27.2	0.0	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
71	3.0	21.6	26.4	4.0	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
72	1.0	11.6	22.0	0.0	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
73	7.0	21.8	28.0	4.0	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
74	6.0	23.4	27.2	7.0	HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
75	7.0	23.0	26.2	0.0	HUJAN	TIDAK HUJAN	TIDAK HUJAN	TIDAK HUJAN
76	8.0	21.5	25.8	2.0	HUJAN	TIDAK HUJAN	HUJAN	TIDAK HUJAN

Gambar5. Perbandingan hasil metode C4.5, Classification Tree, dan Random Forest

Evaluation Results						
	Method	CA	Sens	Spec	AUC	Brier
1	Classification Tree	0.6444	0.4785	0.7348	0.6360	0.5661
2	Random Forest	0.6458	0.4902	0.7306	0.6545	0.4766
3	C4.5	0.6954	0.3340	0.8924	0.6485	0.4190

Gambar 6. Perbandingan akurasi metode C4.5, Classification tree, dan Random Forest

IV. KESIMPULAN

Untuk memenuhi kecepatan dan ketepatan prakiraan, diperoleh model prediksi yang dapat digunakan selanjutnya adalah model keluaran C4.5 dengan tingkat akurasi 68.5%. Komponen cuaca yang dominan memungkinkan terjadinya hujan adalah suhu udara ≤ 25.9 , suhu titik embun ≥ 22.3 , dan tutupan awan ≥ 6 . Pada kajian model selanjutnya, diperlukan data dengan interval 5 - 10 tahun untuk memperbaiki akurasi model.

V. DAFTAR PUSTAKA

- ¹⁾ Santosa, Budi. (2007). *Data mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu- Bisnis. Edisi Pertama.
- ²⁾ Nandagopal, S., Karthik, S., & Arunachalam, V.P. (2010). Mining of Meteorological Data Using Modified Apriori Algorithm. *European Journal of Scientific Research*, 47(2), 295-308.
- ³⁾ McGovern, Amy., Supinie, Timothy., John Cagne II, David., Troutman, Nathaniel., Collier, Matthew., A. Brown, Rodger., et. al. (2010). Understanding Severe Weather Processes Through Spatiotemporal Relational Random Forest. *Proceedings of Conference on Intelligent data Understanding*. 213-227.
- ⁴⁾ Williams, K., Ahijevych., Kessinger, C.J., Saxon, T.R., Steiner, M., & Dettling, S. (2008). A machine-learning approach to finding weather regimes and skillful predictor combinations for short-term storm forecasting. *Presentation/webcast*. Situs <http://nldr.library.ucar.edu/repository/collections/OSGC-000-000-003-270> diakses tanggal 10 Desember 2011
- ⁵⁾ Li, Xiang., Plale, Beth., Vijayakumar, Nithya., Ramachandran, Rahul., Graves, Sara., & Conover, Helen. (2008). Real-Time Storm Detection and Weather Forecast Activation through Data Mining and Events Processing. *Journal of Earth Science Informatics*, 1(2), 49-57.
- ⁶⁾ Christalin Latha, C. Beulah., Paul, Sujni., Kirubakaran, E., & Sathianarayanan. (2010). A Service Oriented Architecture for weather Forecasting Using Data Mining. *International Journal of Advanced Networking and Applications*. 2(2), 608-613.
- ⁷⁾ C. Onwubolu, Godfrey., Buryan, Petr., Garimella, Sitaram., Ramachandran, Visagaperuman., Buadromo, Viti., & Abraham, Ajith. (2007). Self-Organizing Data Mining For Weather Forecasting. *Proceedings of IADIS European Conference Data Mining*. 81-88.
- ⁸⁾ Dhanya, C.T., & Kumar, Nagesh. (2009). Data Mining for Evolving Fuzzy Association Rules for Predicting Monsoon Rainfall of India. *Journal of Intelligent System*. 18(3), 193-209.
- ⁹⁾ N. Kohail, Sarah., & El-Halees, Alaa M. (2011). Implementation of Data Mining Techniques for Meteorological Data Analysis (A case study for Gaza Strip). *International Journal of Informatics and Communication Technology Research*, 1(3), 96-100.
- ¹⁰⁾ Ingsrisawang, Lily., Ingsrisawang, Supawadee., Somchit, Saisuda., Aungsuratana, Prasert., & Khantiyanan, Wawarut. (2008). Machine Learning Techniques for Short-Time Rain Forecasting System in the Northeastern Part of Thailand. *International Journal of World Academy of science, engineering and Technology*. v41-43, 248-253.
- ¹¹⁾ Situs http://michael.hahsler.net/research/association_rules/measures.html diakses tanggal 18 Juni 2011
- ¹²⁾ Iqbal. (2007). *Penerapan Data mining di Badan Meteorologi dan Geofisika untuk memprediksi cuaca di Jakarta*. Thesis Program Studi Magister Teknologi Informasi, Fasilkom UI.
- ¹³⁾ Situs www.orange.com diakses tanggal 8 Desember 2010
- ¹⁴⁾ Situs www.ogimet.com diakses tanggal 01 Desember 2010
- ¹⁵⁾ Technical Documentation-471. (2001). *Guide to Marine Meteorological Services*. World Meteorological Organization.
- ¹⁶⁾ Turban et.al., (2009). *Decision Support and Business Intelligence Systems. Ninth Edition*. Pearson.