



JURNAL IPTEK

MEDIA KOMUNIKASI TEKNOLOGI

homepage URL : ejurnal.itats.ac.id/index.php/iptek



Prakiraan Tinggi Gelombang Air Laut Menggunakan Data Mining

Luky Agus Hermanto¹

Jurusan Teknik Informatika, FTIf, Institut Teknologi Adhi Tama Surabaya¹

INFORMASI ARTIKEL

Jurnal IPTEK – Volume 22
Nomer 1, Mei 2018

ISSN:1411-7010
e-ISSN:2477-507X

Halaman:
37 – 44

DOI:
<https://dx.doi.org/10.31284/j.iptek.2018.v22i1.232>

EMAIL

lukyagushermanto74@itats.ac.id

RIWAYAT ARTIKEL

Tanggal diterima :
12 April 2018

Tanggal diterbit :
31 Mei 2018

PENERBIT

LPPM- Institut Teknologi
Adhi Tama Surabaya

Jurnal IPTEK by LPPM-ITATS is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

ABSTRACT

Weather forecasting needs many weathers component, big data and also forecaster experiences. They cause accuracy and rapid of forecasting were not satisfied. To answer this problem, the study to forecast model was completed by using technique in data mining such as association rule, classification and Random Forest. This research used data from Cilacap maritime station since August 2012 until August 2016. The data contains date, time, wind speed, wind direction, current direction, current speed and wave speed. Data testing taken randomly from existing data. The result shows that the accuracy of association rule and classification tree are about 79% and 88%, respectively.

Keywords: Data mining; Association rule; Classification tree; Random Forest; Ocean waves

ABSTRAK

Melakukan prakiraan cuaca memerlukan banyak komponen data cuaca, *record* dalam jumlah yang besar, serta kemampuan pelaku prakiraan. Keadaan ini mengakibatkan keakuratan dan kecepatan prakiraan menjadi kurang terpenuhi ketika kesimpulan diambil. Untuk mengatasi masalah tersebut, dilakukan penelitian pemodelan prediksi menggunakan teknik yang ada dalam konsep penambangan data, *association rule*, klasifikasi, serta Random Forest. Penelitian ini menggunakan data dari stasiun pengamatan maritim Cilacap mulai Agustus 2012 sampai dengan Agustus 2016. Data tersebut terdiri atas tanggal, waktu, kecepatan angin, arah angin, arah arus, kecepatan arus, arah gelombang, dan kecepatan gelombang. Data pengujian adalah sebagian data yang diambil secara acak dari keseluruhan data yang digunakan. Dari pengujian model, didapatkan bahwa Association Rule menghasilkan akurasi 79%, sedangkan Classification Tree menghasilkan akurasi 88%.

Kata kunci: Penambangan data; Association rule; Classification tree; Random Forest; Gelombang laut

PENDAHULUAN

Kegiatan yang meliputi proses pengumpulan, pemakaian data rekaman untuk menemukan pola keteraturan, atau hubungan data dalam ukuran (volume) yang sangat besar disebut *data mining* atau *knowledge discovery in database (KDD)*. Hasil dari *data mining* ini dapat digunakan untuk membantu pengambilan keputusan untuk masa (waktu) yang akan datang/berikutnya. Perkembangan KDD mengakibatkan penggunaan metode *pattern recognition* menjadi semakin menurun karena telah menjadi bahasan dalam data mining [1]. Metode ini menggabungkan empat keilmuan: statistik, visualisasi, *database*, serta *machine learning* [1]. *Machine learning* merupakan

suatu bidang dalam *artificial intelligence* atau disebut kecerdasan buatan yang terhubung dengan pengembangan teknik pemrograman berdasarkan pembelajaran rekam data dan ilmu statistik dan bisa juga dimanfaatkan dalam proses optimasi.

Telah banyak kajian *data mining* dilakukan untuk prakiraan cuaca. *Data mining* fokus pada Association Rule yaitu menggunakan algoritma Apriori memperlihatkan hasil yang didapatkan lebih baik dalam hal kesesuaian, proses komputasi, dan terminasi [2]. *Data mining* lainnya yang digunakan yaitu Random Forest, mempunyai kemampuan melakukan prediksi turbulensi serta susunan tornado yang terjadi di Benua Amerika [3] dan terjadinya badai di dalam waktu sekitar satu jam pertama pada tiap sel data [4]. Metode yang lain yaitu *clustering*, juga merupakan teknik *data mining*, digunakan untuk deteksi badai di wilayah Amerika [5]. Tidak hanya pada penilaian faktor akurasi, kecepatan proses prakiraan cuaca dapat lebih meningkat dengan menggunakan gabungan kemampuan metode yang ada dalam *data mining* seperti SVM (Support Vector Machine) dan struktur komputasi yang berdasar *service-oriented architecture* [6]. Kemampuan yang paling penting adalah *self-organizing data mining* yang dapat dibuktikan melalui *enhanced Group Method of Data Handling* (e-GMDH) yang sudah diimplementasikan untuk memperkirakan variabel cuaca seperti suhu, jumlah/tinggi curah hujan dalam satu bulan, dan tekanan udara harian [7]. *Data mining* dengan teknik lain yaitu Fuzzy Association Rule juga telah dipergunakan untuk memperkirakan curah hujan Moonson di kawasan India [8]. Beberapa penelitian memanfaatkan metode-metode yang ada dalam *data mining* secara bersamaan. Pertama, penelitian prediksi curah hujan jangka pendek menggunakan metode *outlier analysis*, *clustering prediction*, *classification*, dan Association Rule telah diaplikasikan untuk menganalisis pola data meteorologi di area Jalur Gaza [10]. Kedua, metode Decision Tree (C4.5), Artificial Neural Network (ANN), dan Support Vector Machine (SVM) sudah dilakukan di wilayah Thailand. C4.5 digunakan untuk memperkirakan status, terjadi turun hujan atau tidak. ANN dimanfaatkan untuk memperkirakan jumlah curah hujan. Penggunaan SVM pada klasifikasi jumlah curah hujan, dikelompokkan berdasarkan tiga kelompok yakni tidak turun hujan, hujan ringan, dan hujan lebat [10].

Di bawah BMKG, Indonesia memiliki sekitar 10 stasiun pengamatan meteorologi maritim dan ada 3 stasiun yang diperuntukkan memberikan layanan meteorologi maritim. Sebagian besar stasiun tersebut beroperasi dengan melakukan pantauan sinoptik, yang sebagian lainnya memberikan layanan analisis serta prediksi cuaca maritim. Data hasil dari pengamatan penting untuk memantau karakteristik cuaca setempat dan penyusunan informasi prediksi untuk tempo beberapa hari kemudian. Di sisi lain, dalam proses pembuatan informasi prediksi cuaca, didapatkan beberapa kendala. Pertama, kesulitan menyusun informasi prediksi karena harus melibatkan tidak sedikit sumber data, yaitu data pengamatan, data model implementasi cuaca, data berupa citra keadaan awan dari satelit, dan data keadaan awan dari radar. Kedua, prediksi cuaca maritim pada umumnya bertumpu pada kemampuan dari prakirawan. Jadi, interpretasi yang didapatkan bisa berbeda antarprakirawan karena bergantung pada pengalaman masing-masing. Perbedaan interpretasi yang terjadi dapat memberikan keadaan yang tidak pasti kepada pengguna yang nanti akhirnya menciptakan peluang dalam menurunkan kelas informasi yang bisa disampaikan. Berdasarkan masalah tersebut, penelitian ini bertujuan untuk melakukan kajian bagaimana model prakiraan yang sesuai, yang pada akhirnya memudahkan dalam proses analisis serta prediksi cuaca maritim, khususnya prakiraan tinggi ombak di Laut Selatan atau Samudera Hindia.

TINJAUAN PUSTAKA

Ada banyak metode dalam *data mining* yang cukup dikenal antara lain Association Rule dan *classification*. Metode *classification* dapat dibagi lagi yaitu C4.5, Classification Tree, dan Random Forest. Ketiga metode tersebut memiliki kelebihan serta kekurangan. Rincian kerja dari teknik-teknik tersebut bisa diuraikan sebagai berikut.

- a. Association Rule dipakai untuk melihat segala kemungkinan bentuk hubungan *if-then* di antara *item* serta memilih hanya hubungan yang paling memungkinkan sebagai penanda dari relasi ketergantungan di antara *item*. Indikator-indikator yang dianggap bernilai penting yang dibutuhkan yaitu *support*, *confidence*, dan *lift ratio*. Hubungan dari ketiganya seperti berikut.

Asosiasi antara X dan Y, Jika X maka Y, disimbolkan $X \rightarrow Y$

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

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

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

dengan

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 Y

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

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

freq(X) = frekuensi munculnya *item* dalam X

freq(Y) = frekuensi munculnya *item* dalam Y

$|D|$ = banyaknya transaksi di *database support, confidence, dan lift ratio* adalah (%)

Consequent dengan nilai yang tinggi memperlihatkan aturan asosiasi bernilai kuat. Adapun nilai pada *lift ratio* > 1 memperlihatkan adanya fungsi yang bisa diambil dari metode tersebut. Atau bisa dengan pernyataan lain, semakin tinggi level *lift ratio*, semakin kuat hubungan asosiasinya.

- b. *Classification* merupakan suatu proses untuk dapat menemukan fungsi atau model yang akan membedakan atau menjelaskan hal tersebut sebagai konsep atau kelas data yang diperlukan dalam memperkirakan objek tersebut dari kelas mana yang labelnya belum diketahui [11]. Model tersebut dapat berbentuk aturan *jika-maka, decision tree*, atau formula matematis ataupun Neural Network. *Decision tree*, keluarannya akan bernilai diskret, merupakan salah satu metode *classification* yang paling mudah untuk diinterpretasikan. Algoritma dari *decision tree* yang paling dikenal adalah C4.5 [12].

Proses klasifikasi pada umumnya terbagi atas dua fase yaitu *learning* (pembelajaran) dan *testing* (pengujian). Pada proses *learning*, diambil sebagian data yang telah diketahui sejak awal kelas datanya, dimasukkan untuk membentuk perkiraan model. Pada tahap *testing*, model yang terbentuk akan diujikan dengan mengambil sebagian data yang lain untuk dapat mengetahui ketepatan model tersebut. Apabila ketepatannya memenuhi, maka model tersebut dapat dipakai untuk memprediksi kelompok (kelas) data yang belum diketahui.

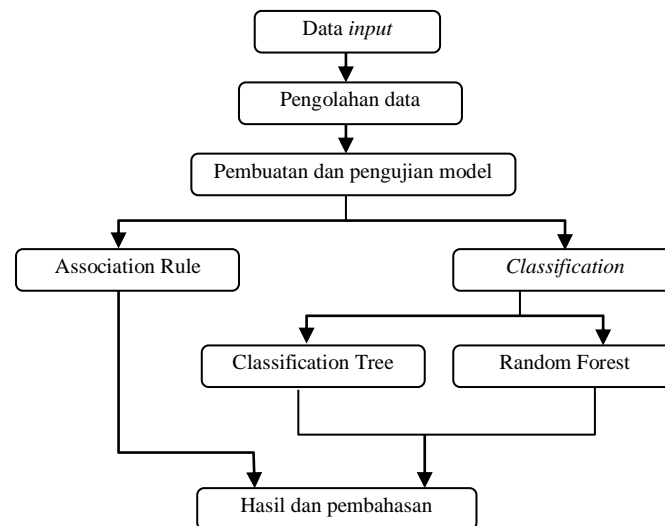
Dewasa ini, banyak *tool data mining* yang telah dikembangkan oleh lembaga riset, perguruan tinggi, atau perusahaan di bidang teknologi informasi. Perangkat lunak Orange Ailab pada penelitian ini digunakan sebagai *tool* untuk *data mining* [13]. Orange Ailab adalah perangkat lunak *open source* yang memberikan pengguna kemudahan dalam memvisualisasi serta menganalisis data. Fitur-fitur yang dimiliki antara lain *scatterplot, bar chart, tree, dendrogram, network, dan heat map*.

METODE

Metode yang digunakan pada penelitian ini dijelaskan dalam diagram pada Gambar 1. Pada Gambar 1 tersebut, data yang didapatkan kemudian diolah, diuji, selanjutnya dianalisis.

Sumber Data

Data sinoptik berasal dari stasiun pengamatan meteorologi maritim P15 Cilacap mulai Agustus 2012 sampai dengan Agustus 2016, berukuran 35.137 *raw data*. Data tersebut meliputi tanggal, waktu, kecepatan udara, arah angin, arah arus, arah gelombang, dan kecepatan arus.



Gambar 1. Metode penelitian

Pemilihan Perangkat Lunak dan Metode dalam *Data Mining*

Perangkat lunak yang dipergunakan adalah Orange Ailab yang menggunakan metode Association Rule, Classification Tree, dan Random Forest. Perangkat lunak ini dipilih karena bersifat *open source*. Association Rule yang termasuk metode *unsupervised learning* [14] diujicobakan pada seluruh komponen data yang sinoptik, yakni waktu, kecepatan udara, arah angin, arah arus, arah gelombang, dan kecepatan arus. Sedangkan penggunaan Classification Tree serta Random Forest yang merupakan metode *supervised* akan diujicobakan pada bagian komponen data yang sinoptik yakni waktu, kecepatan udara, arah angin arah arus, arah gelombang, dan kecepatan arus.

Pengolahan Data

Data yang dikumpulkan berupa data sinoptik, dipetakan ke dalam nilai yang bisa diolah oleh *tool*, terutama pada kategori *output*, yaitu tinggi gelombang berdasar pada 3 kelompok: kategori rendah, sedang, dan tinggi.

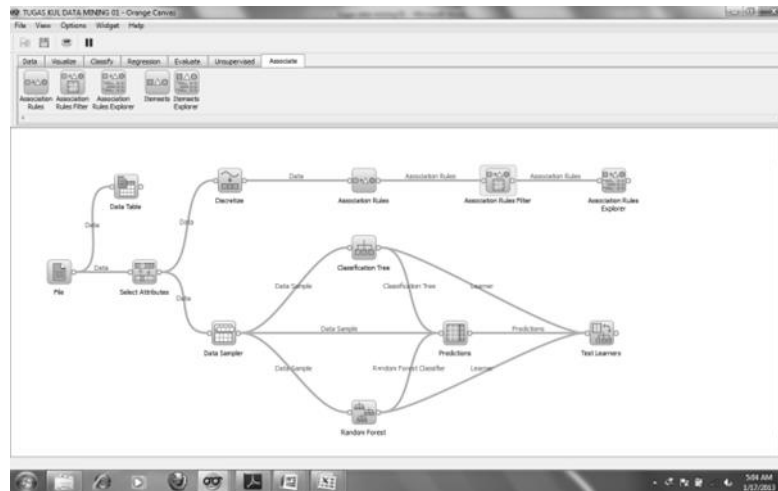
Pembuatan Model

Metode Association Rule

Penelitian ini, pada proses pengujiannya, memilih *support* sebesar 15%, *confidence* 25%, serta *maximum rule* sebanyak 15.500. Dengan dasar nilai *support* sebesar 15% akan berakibat menjadi lebih kecil jumlah *rule* yang akan terjadi dan menghemat memori komputer. Adapun nilai *confidence* sebesar 25% itu adalah untuk lebih memperkuat hubungan (asosiasi) antaraturan (*rule*) sehingga aturan yang akan dihasilkan dapat bernilai lebih bermanfaat untuk melakukan prediksi. Untuk menghasilkan kuatnya nilai hubungan antarkomponen cuaca, akan digunakan nilai *lift ratio* > 1 sesuai Gambar 2.

Metode Classification Tree dan Random Forest

Metode yang dipakai untuk membangun model dengan input komponen data cuaca adalah metode Classification Tree dan Random Forest. *Output* model masing-masing metode tersebut akan diujicobakan dengan sebagian data *input* untuk melihat kemampuan model. Hasilnya setelah itu dibandingkan untuk mendapatkan nilai akurasi tertinggi dan memutuskan model nilai prediksi yang terbaik.



Gambar 2. Pemodelan dengan Orange Ailab

HASIL DAN PEMBAHASAN

Association Rule

Orange Ailab yang menggunakan input seluruh komponen data pengamatan dengan *support* 15%, *confidence* 25%, dengan *maximum rule* sebanyak 15.500 menghasilkan 29 kelompok *rule* seperti terdapat pada Gambar 3. Dari kelompok tersebut, dapat dilakukan penyaringan dengan dua tahap. Pertama, memilih aturan (*rule*) dengan bentuk hubungan jika “komponen seperti waktu, kecepatan angin, arah angin, arah arus, arah gelombang, kecepatan arus atau mempunyai nilai tertentu” maka “kondisi tinggi gelombang laut yang akan terjadi bagaimana”. Kedua, memilih *rule* yang memiliki nilai *lift ratio* lebih besar dari satu (*lift ratio* > 1).

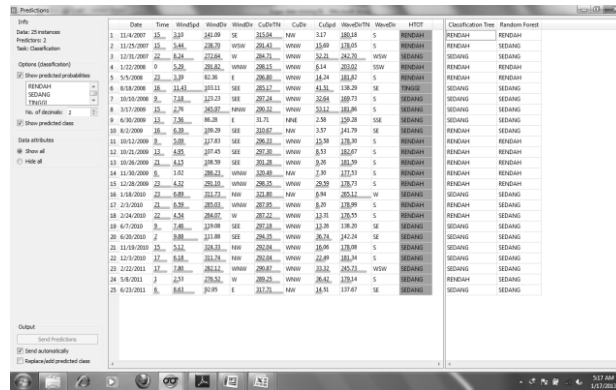
Supp	Conf	Lift	Lev	Strg	Cev	Antecedent	Consequent
0.290	0.308	0.801	-0.032	1.190	0.520	HTOT=SEDANG	CuDir=WNW
0.202	0.621	1.003	0.001	1.904	0.325	D_CuSpd=(8.55, 20.54]	CuDir=WNW
0.287	0.690	1.114	0.029	1.490	0.416	HTOT=RENDAH	CuDir=WNW
0.200	0.737	1.190	0.032	2.277	0.272	WaveDir=5	CuDir=WNW
0.324	0.736	1.189	0.051	1.495	0.441	WaveDir=5	CuDir=WNW
0.150	0.775	1.251	0.030	3.196	0.194	D_CuSpd=(8.55, 20.54] ...	CuDir=WNW
0.150	0.341	1.688	0.061	0.458	0.441	WaveDir=5	CuDir=WNW D_CuSpd=(8.55, 20.54]
0.200	0.455	1.586	0.074	0.650	0.441	WaveDir=5	CuDir=WNW HTOT=RENDAH
0.150	0.462	1.423	0.045	0.988	0.325	D_CuSpd=(8.55, 20.54]	CuDir=WNW WaveDir=5
0.200	0.482	1.486	0.066	0.780	0.416	HTOT=RENDAH	CuDir=WNW WaveDir=5
0.202	0.326	1.003	0.001	0.525	0.619	CuDir=WNW	D_CuSpd=(8.55, 20.54]
0.168	0.403	1.240	0.032	0.782	0.416	HTOT=RENDAH	D_CuSpd=(8.55, 20.54]
0.194	0.440	1.352	0.050	0.738	0.441	WaveDir=5	D_CuSpd=(8.55, 20.54]
0.150	0.463	1.423	0.045	1.002	0.324	CuDir=WNW WaveDir=5	D_CuSpd=(8.55, 20.54]
0.287	0.463	1.114	0.029	0.671	0.619	CuDir=WNW	HTOT=RENDAH
0.168	0.515	1.240	0.032	1.278	0.325	D_CuSpd=(8.55, 20.54]	HTOT=RENDAH
0.200	0.618	1.486	0.066	1.281	0.324	CuDir=WNW WaveDir=5	HTOT=RENDAH
0.272	0.617	1.485	0.089	0.943	0.411	WaveDir=5	HTOT=RENDAH
0.163	0.369	0.700	-0.067	1.180	0.441	WaveDir=5	HTOT=SEDANG
0.290	0.469	0.901	-0.032	0.840	0.619	CuDir=WNW	HTOT=SEDANG
0.166	0.701	1.348	0.043	2.203	0.236	WindDir=SEE	HTOT=SEDANG
0.163	0.313	0.709	-0.067	0.847	0.520	HTOT=SEDANG	WaveDir=5
0.324	0.524	1.189	0.051	0.712	0.619	CuDir=WNW	WaveDir=5
0.194	0.596	1.352	0.050	1.356	0.325	D_CuSpd=(8.55, 20.54]	WaveDir=5
0.272	0.654	1.485	0.089	1.090	0.416	HTOT=RENDAH	WaveDir=5

Gambar 3. Hasil metode Associate Rule

Classification Tree dan Random Forest

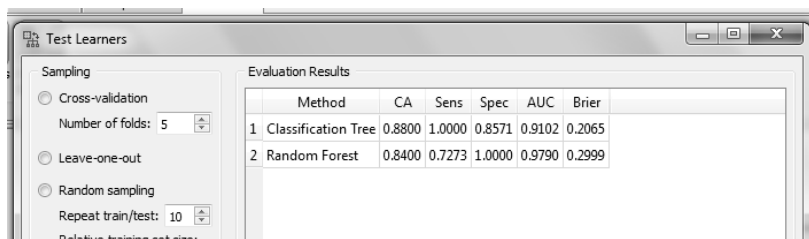
Pemodelan menggunakan teknik C4.5, Classification Tree, dan Random Forest menghasilkan Gambar 4 dan 5. Gambar 4 menunjukkan perbandingan hasil kondisi cuaca nyata (riil) dengan keluaran hasil setiap metode. Sebagian data masih menunjukkan adanya kesesuaian dari hasil prediksi teknik C4.5 dibandingkan dengan keadaan cuaca riil. Pada Gambar 5, dengan *cross validation*, pemilihan sampel dilakukan dengan *number of folds* sebesar 5, *repeat training set* data sebesar 5 kali, dan ukuran data yang dijadikan *training test* dari total data adalah sebesar 20%.

Pada Gambar 5 tampak dari kedua teknik bahwa tingkat akurasi Classification Tree lebih tinggi yaitu sebesar 88%, sedangkan dengan Random Forest yakni 84%.



Model	Accuracy
Classification Tree	88%
Random Forest	84%

Gambar 4. Perbandingan hasil Classification Tree dan Random Forest



Method	CA	Sens	Spec	AUC	Brier
1 Classification Tree	0.8800	1.0000	0.8571	0.9102	0.2065
2 Random Forest	0.8400	0.7273	1.0000	0.9790	0.2999

Gambar 5. Perbandingan akurasi Classification Tree dan Random Forest

Berdasar hasil pembuatan serta pengujian model, penggunaan teknik Association Rule dan *classification*, maka dari eksperimen ini dapat disampaikan bahwa model prediksi yang bisa dihasilkan dari kedua metode yang digunakan tersebut menunjukkan tidak adanya perbedaan dalam ukuran yang besar, baik dari sisi komponen penyusun cuaca ataupun dari nilai batas syarat setiap komponen cuaca. Hasil Association Rule menunjukkan faktor penentu tinggi gelombang adalah kecepatan arus laut serta arah gelombang.

KESIMPULAN

Untuk lebih memenuhi tuntutan kecepatan dan akurasi prakiraan, model prediksi yang bisa dibangun dan dimanfaatkan selanjutnya yaitu model keluaran dari Classification Tree dengan tingkat akurasi 88%. Komponen penyusun cuaca yang dominan (memungkinkan) terjadinya tinggi ombak adalah kecepatan arus laut serta arah gelombang.

DAFTAR PUSTAKA

- [1] B. Santosa. *Data mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Edisi Pertama. Yogyakarta: Graha Ilmu, 2007.
- [2] S. Nandagopal, S. Karthik, and V.P. Arunachala. "Mining of Meteorological Data Using Modified Apriori Algorithm," *European Journal of Scientific Research*, vol. 47, no. 2, pp. 295-308, 2010.
- [3] A. Mcgovern, et al. "Understanding Severe Weather Processes through Spatiotemporal Relational Random Forest," *Proceedings of Conference on Intelligent Data Understanding*, pp. 213-227.
- [4] J. K. Williams, D. A. Ahijevych, C. J. Kessinger, T. R. Saxen, M. Steiner, and S. Dettling. "A Machine Learning Approach to Finding Weather Regimes and Skillful Predictor Combinations for Short-term Storm Forecasting," 2008. Available: <http://nldr.library.ucar.edu/repository/collections/OSGC-000-000-003-270>.

- [5] X. Li, et al. "Real-Time Storm Detection and Weather Forecast Activation through Data Mining and Events Processing," *Journal of Earth Science Informatics*, vol. 1, no. 2, pp. 49-57, 2008.
- [6] C. B. C. Latha, S. Paul, E. Kirubakaran, and Sathianarayanan. "A Service Oriented Architecture for Weather Forecasting Using Data Mining," *International Journal of Advanced Networking and Applications*, vol. 2, no. 2, pp. 608-613, 2010.
- [7] G. C. Onwubolu, et al. "Self-Organizing Data Mining for Weather Forecasting," *Proceedings of IADIS European Conference Data Mining*, pp. 81-88, 2007.
- [8] C.T. Dhanya and N. Kumar. "Data Mining for Evolving Fuzzy Association Rules for Predicting Monsoon Rainfall of India," *Journal of Intelligent System*, vol. 18, no. 3, pp. 193-209, 2009.
- [9] S. N. Kohail and A. M. El-Halees. "Implementation of Data Mining Techniques for Meteorological Data Analysis (A case study for Gaza Strip)," *International Journal of Informatics and Communication Technology Research*, vol. 1, no. 3, pp. 96-100, 2011.
- [10] L. Ingsrisawang, S. Ingsrisawang, S. Somchit, P. Aungsuratana, and W. Khantiyanan. "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*, pp. 248-253, 2008.
- [11] M. Hahsler, "A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules," Feb. 9, 2018. [Online]. Available: http://michael.hahsler.net/research/association_rules/measures.html
- [12] Iqbal. "Penerapan Data mining di Badan Meteorologi dan Geofisika untuk Memprediksi Cuaca di Jakarta," MTI Thesis, Fakultas Ilmu Komputer, Univeristas Indonesia, 2007.
- [13] Orange Data Mining. *Education in Data Science*. [Online]. Available: <https://orange.biolab.si/features/interactive-data-visualization>.
- [14] Turban, et al. *Decision Support and Business Intelligence Systems*, 2009.

Halaman ini sengaja dikosongkan