

Model Optimasi Prediksi dengan Model Association Rule Best Time Series(ARBT) Pada Data Mining Time Series

Relita Buaton¹, Muhammad Zarlis², Herman Mawengkang³, Syahril Efendi⁴

¹Graduate Program of Computer Science, Department of Computer Science, Universitas Sumatera Utara, Medan, Indonesia

^{2,4}Department of Computer Science, Universitas Sumatera Utara, Medan, Indonesia

³Department of Mathematics, Universitas Sumatera Utara, Medan, Indonesia

Email: ¹bbcbuaton@gmail.com, ²m.zarlis@usu.ac.id, ³mawengkang@usu.ac.id, ⁴syahril1@usu.ac.id

Abstrak—Era globalisasi dan Industri 4.0 yang sangat mempengaruhi pola tatanan kehidupan manusia yang berkaitan erat dengan perubahan lingkungan dan berdampak pada beberapa tahun belakangan ini terjadi banyak fenomena alam, bencana alam, banjir, gempa, perubahan suhu, peningkatan permukaan air laut dan bahkan telah terjadi gangguan listrik hampir setengah Pulau Jawa dan bencana paling terkini yang masih melanda dunia dan masih belum ditemukan vaksinnnya adalah virus corona. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB) sebanyak 1.978 bencana alam dan non alam terjadi di Indonesia hingga tahun 2020 diantaranya bencana banjir dengan 748 peristiwa, bencana puting beliung dengan 527 kasus, tanah longsor 370 kasus dan kebakaran hutan dan lahan dengan 270 kasus mengakibatkan mengungsi mencapai angka 3,9 juta jiwa dan meninggal 273, 422 orang luka-luka dan 12 orang hilang. Tingginya angka kejadian bencana alam dan non alam yang terjadi di hampir seluruh penjuru bumi yang diyakini memiliki hubungan dan korelasi yang kuat antar kejadian dan peristiwa maka perlu dicari sebuah formula yang mendalam dengan kajian komputasi untuk menemukan model yang dapat digunakan untuk mengetahui korelasi antar kejadian serta mampu memprediksi peristiwa yang akan terjadi setiap terjadi peristiwa atau bencana. Maka bagaimana menemukan model komputasi untuk mengetahui korelasi antar bencana yang terjadi dan bagaimana mengoptimalkan prediksi bencana yang akan terjadi di masa yang akan datang serta melakukan pengujian dan simulasi model komputasi dengan metode ARBT(Association Rule Best Time Series). Penelitian ini menghasilkan model komputasi dengan metode ARBT yang mampu menghasilkan model komputasi untuk forecasting dengan variable dan bobot yang teruji berdasarkan pengembangan metode RBT dan association rule, ditemukan model dan simulasi hipotesa forecasting suhu udara dengan metode ARBT dengan multi variable untuk mendapatkan tingkat akurasi yang maksimal.

Kata Kunci: Optimasi Prediksi, Metode ARBT

1. PENDAHULUAN

Seiring dengan perkembangan zaman dan teknologi serta peradaban manusia yang memiliki korelasi kuat dengan perubahan pola kehidupan masyarakat saat ini. Khususnya dampak perubahan global atau global warning yang sangat berpengaruh dengan terganggunya keseimbangan dalam segala aspek khususnya dampak lingkungan. Beberapa tahun belakangan ini terjadi banyak fenomena alam, bencana alam, banjir, gempa, perubahan suhu, peningkatan permukaan air laut dan bahkan beberapa tahun belakangan ini telah terjadi gangguan listrik hampir setengah pulau Jawa dan bencana paling terkini yang masih melanda dunia dan masih belum ditemukan vaksinnnya adalah Virus Corona. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB) mencatat sebanyak 1.978 bencana alam dan non alam terjadi di Indonesia hingga tahun 2020, tren bencana di Indonesia tahun 2020 secara umum terus meningkat yang sifatnya hidro meteorology diantaranya bencana banjir menempati peringkat pertama dengan 748 peristiwa, bencana puting beliung dengan 527 kasus, tanah longsor 370 kasus dan kebakaran hutan dan lahan dengan 270 kasus mengakibatkan mengungsi mencapai angka 3,9 juta jiwa dan meninggal 273, 422 orang luka-luka dan 12 hilang, dan bencana saat ini yang masih sangat membutuhkan perhatian dunia adalah bencana yang masuk dalam kategori bencana non alam di Indonesia yaitu virus corona. Berdasarkan data peristiwa, setiap terjadi bencana alam dan non alam, juga memakan banyak korban, semua itu terjadi karena keterbatasan kemampuan prediksi dimasa yang akan datang, pihak instansi terkait tidak mampu memprediksi apa yang akan terjadi dimasa yang akan datang sehingga banyak bencana dan fenomen yang memakan korban. Seandainya hal tersebut dapat diprediksi maka tentu dapat diantisipasi dan meminimize korban. Oleh sebab itu dibutuhkan prediksi yang akurat sehingga dapat diketahui dan diantisipasi sebelum peristiwawa terjadi dan dapat meminimalkan kerugian serta dampak negatifnya.

Tingginya angka kejadian bencana alam dan non alam yang terjadi di hampir seluruh penjuru bumi yang diyakini memiliki hubungan dan korelasi yang kuat antar kejadian dan peristiwa maka perlu dibangun sebuah formula yang mendalam dengan kajian komputasi untuk menemukan model yang dapat digunakan untuk mengetahui korelasi antar kejadian serta mampu memprediksi peristiwa yang akan terjadi setiap terjadi peristiwa atau bencana, bagaimana menemukan model komputasi untuk mengetahui korelasi antar bencana yang terjadi serta mengoptimalkan prediksi yang akan terjadi dimasa yang akan datang.

2. METODOLOGI PENELITIAN

2.1 Data Mining Time Series

Salah satu bidang ilmu pengetahuan yang digunakan untuk mengoptimalkan prediksi adalah data mining. Data mining dapat digunakan membuat prediksi berdasarkan data time series. Data mining adalah proses analitis tentang penemuan pengetahuan dalam kumpulan data dengan jumlah yang besar dan kompleks, data mining merupakan disiplin ilmu yang

berada pada interseksi statistika dan ilmu komputer. Lebih tepatnya lagi, data mining merupakan hasil hibridisasi dari statistika, ilmu komputer, kecerdasan buatan dan *machine learning* [1]. Forecasting menjadi salah satu topik pembicaraan penting dalam komunitas *data mining*. Kebanyakan penelitian yang muncul dalam *literature* berkaitan dengan objek diskrit [2], misalnya pada penekanan prediksi, *database queries* [3], peramalan cuaca, prediksi prestasi akademik, dan lain-lain. Dalam statistika, Deret Berkala (*Time Series*) merupakan salah satu topik yang selalu dikaitkan dengan peramalan melalui sederetan data yang bergantung pada periode waktu. Data deret berkala sering terjadi dalam aplikasi bisnis dan dalam sains. Suatu deret berkala adalah kumpulan pengamatan yang dibuat secara kronologis. Data dari deret berkala memiliki karakteristik seperti berukuran besar, berdimensi tinggi dan dimutakhirkan secara kontinu. Karakteristik selanjutnya adalah sifat numerik dan kontinu dari data selalu dipandang sebagai suatu kesatuan bukan numerik individu. Karena itu, tidak seperti database tradisional dimana pencarian similaritas dilakukan berdasarkan *matching*, pencarian similaritas dalam data deret berkala dilakukan berdasarkan pendekatan.

2.2 Metode RBT(Rule Best Time Series)

Penemuan aturan pada dasarnya merupakan inti dari penemuan pengetahuan (knowledge discovery) melalui data mining. Dengan kata lain, ingin diindikasikan relasi antara variable, siklus atau pola deret berkala yang secara tipikal muncul dalam database sangat besar, model yang ditemukan untuk mengoptimalkan penemuan rule dalam time series data mining model RBT(Rule Best Time Series)

$$RBT = \sum_{i \in T} \sum_{j \in W} W_{ij} \sum_{k \in S} C_{jk} + \sum_{i \in \alpha} \sum_{j \in \beta} R_{ij} + \sum_{i \in \mu} \sum_{j \in \rho} B_{ij} \quad (1)$$

Dengan mengoptimalkan data time series data mining melalui optimasi penemuan rule dengan analisis cluster sequence time series data mining, ditemukannya sebuah model baru dalam penemuan pengetahuan baru pada data time series dengan pendekatan diskritisasi deret berkala sliding window yang disebut dengan model RBT(Rule Best Time series). Dengan menggunakan MAD, MSD dan MAPE dihasilkan tingkat akurasi 87% dan persentase error sebesar 4,7% akurasi model sangat baik dan akurat. Untuk menghasilkan dan optimalisasi pengetahuan dari sekumpulan data juga dapat menggunakan *Clustering time series* digunakan pada berbagai bidang. Selain itu, *clustering time series* juga diterapkan pada salah satu bidang dalam domain *clustering time series* yaitu pengenalan pola dengan menggunakan *sequence time series* [4].

Data *time series* T satu dimensi dengan panjang n direpresentasikan dengan

$$T = x_1, \dots, x_n \quad (2)$$

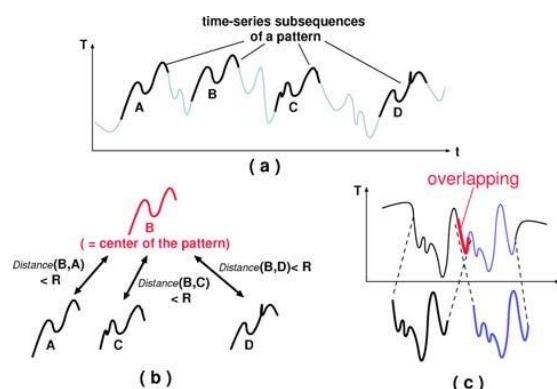
Masing masing x_i adalah variabel bernilai real

Data *time series* T_m m dimensi dengan panjang n adalah *sequence*/urutan m set variabel bernilai real, direpresentasikan dengan

$$T_m = (x_{11}, \dots, x_{m1}), \dots, (x_{1n}, \dots, x_{mn}) \quad (3)$$

Time series subsequence(TSS) C dengan panjang q bagian dari T, direpresentasikan dengan

$$C_{p,q} = x_p, \dots, x_{p+q-1} (q \leq n, 1 \leq p \leq n - q + 1) \quad (4)$$

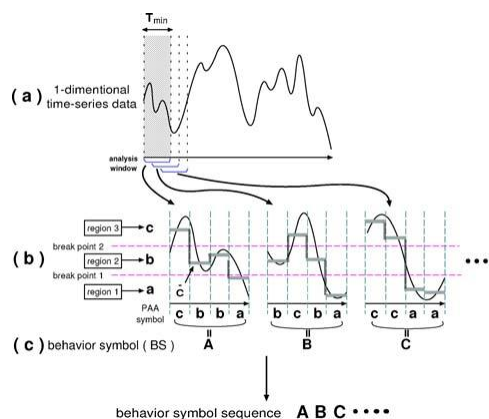


Gambar 1. Batasan Pola (a) Batasan perilaku, (b) Batasan jarak (c) Batasan yang Tidak Tumpang Tindih [5]

2.3 Algoritma Penemuan Motif

Algoritma penemuan motif secara dinamis untuk mendeteksi motif dari data *time series* satu dimensi. Beberapa algoritma yang digunakan adalah *Minimum description length*(MDL), *Akaike's information criterion*(AIC), *Bayesian Information criterion*(BIC). AIC menghasilkan model terbaik berdasarkan kemampuan prediksi, dalam hal ini AIC menemukan pola yang sering muncul dan bukan prediksi *time series*. BIC menghasilkan model terbaik berdasarkan teorema bayes dan MDL menghasilkan model terbaik dengan meminimalkan deskripsi panjang keseluruhan kumpulan data. MDL adalah model terbaik untuk meringkas atau merangkum data times series.

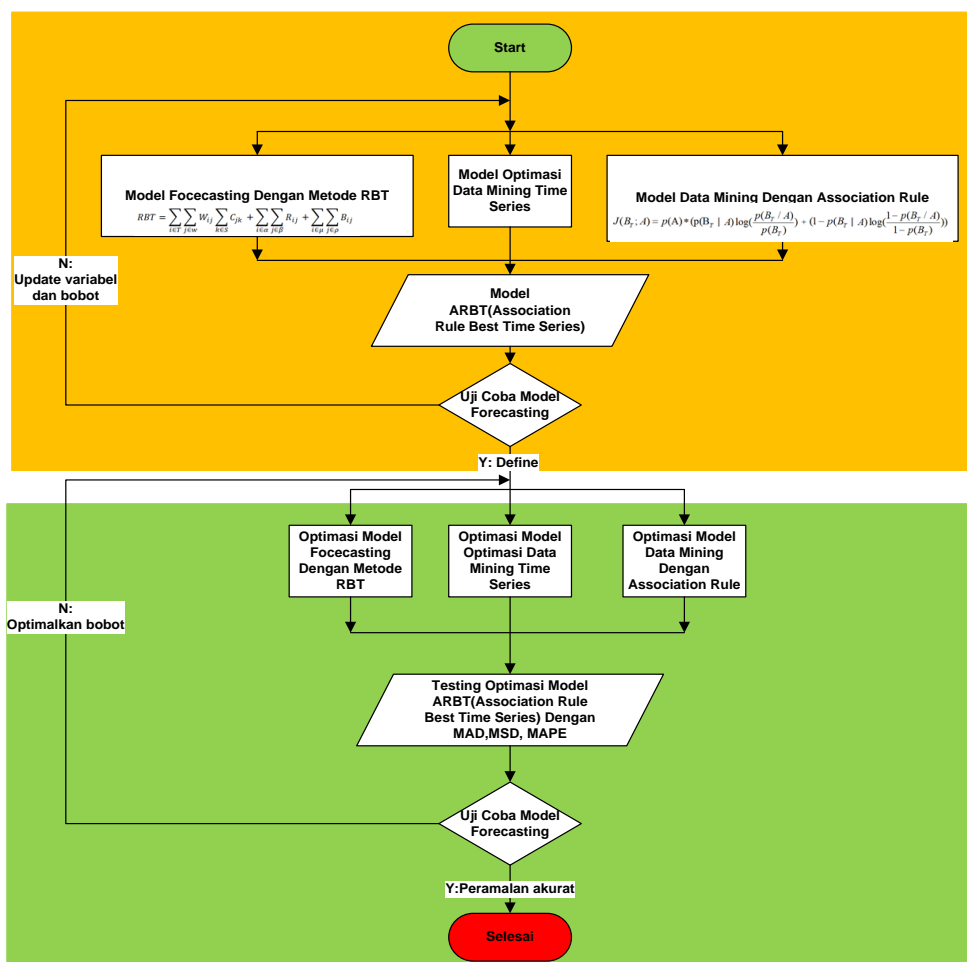
1. Mengubah *time series* menjadi urutan simbol
2. Ekstraksi pola dengan Prinsip MDL memiliki kelemahan yakni pola yang sama hampir tidak muncul dalam deret waktu dan terdapat *noise* dalam *time series*.



Gambar 2. Visualisasi algoritma mengubah *time series* menjadi urutan simbol[6]

Beberapa penelitian tentang forecasting telah dilakukan tetapi masih belum akurat dan memiliki kelemahan, secara umum prediksi yang digunakan hanya menggunakan dua variabel yakni jumlah dan waktu, dalam hal ini pada waktu t akan terdapat sejumlah x(contoh pada tahun 2022 akan ada bencana banjir sebesar 20 peristiwa di pulau Jawa). Prediksi ini cenderung meleset karena hanya didasarkan penekanan pola jumlah yang terjadi pada tahun-tahun sebelumnya. Maka pada penelitian ini akan dilakukan pengembangan penelitian yaitu dengan menambah variabel dan mencari korelasi atau hubungan beberapa variabel sehingga diperoleh hasil prediksi yang lebih akurat dengan model yang ditemukan. Contoh(pada tahun 2020 akan mengalami peningkatan jumlah bencana alam sebanyak 200 kejadian, prediksi ini diperkuat karena adanya korelasi kuat peningkatan perambahan hutan dan alih fungsi hutan dengan kecenderungan perubahan lingkungan yang berpengaruh juga dengan peningkatan intensitas hujan dan longsor di beberapa daerah). Dari contoh prediksi tersebut hasil semakin akurat karena adanya beberapa faktor pendukung dan adanya korelasi kuat antar variabel tersebut.

Untuk mencapai hasil forecasting yang optimal dilakukan kerangka penelitian sebagai berikut:



Gambar 3. Kerangka Penelitian

Penelitian dilakukan dengan mengembangkan hasil penelitian terdahulu yakni mengembangkan metode RBT dengan kemampuan memprediksi data time series untuk 2 variabel. Metode RBT dikorelasikan dengan metode asosiasi pada data mining time series sehingga dihasilkan model ARBT(Association Rule Best Time Series), Diskritisasi deret berkala secara pengelompokan (clustering)

$$D(s) = a_{j(1)} \cdot a_{j(2)} \cdots a_{j(n-w+1)} \quad (5)$$

Setiap simbol a_k menyajikan suatu bentuk dasar dan yang ingin diperoleh adalah aturan penemuan yang mencakup pola yang terbentuk dari bentuk-bentuk dasar. Berikutnya penentuan deret berkala

$$d(\hat{x}, \hat{y}) = (\sum_i (x_i - y_i)^2)^{1/2} \quad (6)$$

Sebagai *metric* dalam pengelompokkan. *Metric* lainnya mencakup *metric* umum L_p yang didefinisikan dengan

$$L_p(\hat{x}, \hat{y}) = (\sum_i (x_i - y_i)^2)^{1/p} \quad (7)$$

model yang dihasilkan diverifikasi dan validasi sesuai dengan target bobot dan variable. Kemudian dilanjutkan dengan mengoptimalkan metode ARBT, metode dikembangkan dengan multi variable dan bobot untuk mencapai target peramalan dengan tingkat akurasi yang tinggi. Tingkat akurasi diperoleh dengan pengujian MAD, MSD dan MAPE. Terdapat beragam *metric* dapat dipakai untuk meranking aturan, antara lain dengan metode data envelopment analysis [7], dengan nilai sensitivitas [8]. Dalam penelitian ini dipakai ukuran-J, yang didefinisikan sebagai berikut:

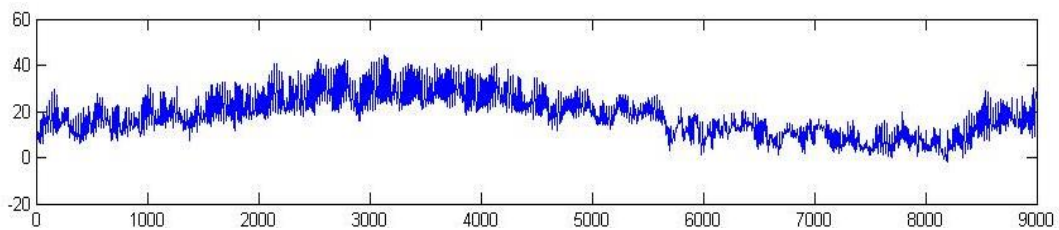
$$J(B_T; A) = p(A) * (p(B_T | A) \log(\frac{p(B_T | A)}{p(B_T)}) + (1 - p(B_T | A)) \log(\frac{1 - p(B_T | A)}{1 - p(B_T)})) \quad (8)$$

Dimana, dalam konteks aturan sikuen, $p(A)$ adalah probabilitas keterjadian symbol A dilokasi acak dalam sikuen, $p(B_T)$ adalah probabilitas keterjadian sekurang-kurangnya satu B dalam *window* yang terpilih secara acak dengan durasi t dan $p(B_T | A)$ adalah probabilitas keterjadian paling sedikit satu B dalam *window* terpilih acak dengan durasi T dengan diketahui bahwa *window* telah mengandung A . Jadi suku pertama dari bentuk ukuran ini yaitu $p(A)$ merupakan bias terhadap aturan yang sering terjadi. Suku kedua dikenal sebagai *cross-entropy*, yaitu informasi yang diperoleh (atau derajat kejutan) untuk berpindah dari probabilitas prior $p(B_T)$ ke probabilitas posterior $p(B_T | A)$.

3. ANALISA DAN PEMBAHASAN

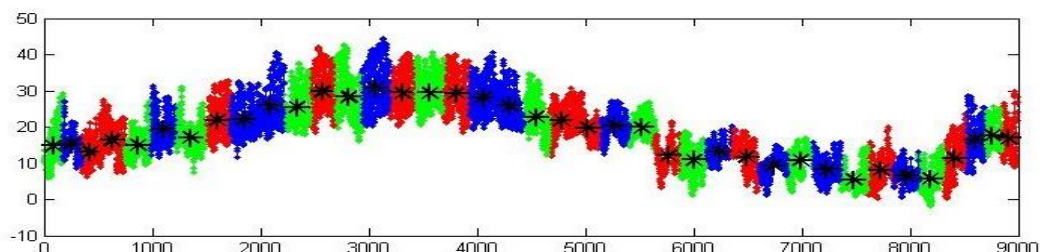
3.1 Analisa Time Series

Data pengujian berisi respons temperatur dalam satuan derajat celsius dari perangkat multisensor gas yang digunakan di lapangan di kota Italia. Rata-rata respons per jam dicatat bersama dengan referensi konsentrasi gas dari penganalisa bersertifikat yang dapat diunduh melalui <http://archive.ics.uci.edu/ml/datasets/Air+Quality>. Dataset berisi 9358 respons rata-rata per jam dari berbagai 5 sensor kimia oksida logam yang tertanam dalam Perangkat Multisensor Kimia Kualitas Udara. Perangkat itu terletak di lapangan di area yang sangat tercemar, di permukaan jalan, di dalam kota Italia. Data direkam dari Maret 2004 hingga Maret 2005 (satu tahun) mewakili rekaman terpanjang yang tersedia secara bebas dari respons perangkat sensor kimia kualitas udara yang digunakan di lapangan.



Gambar 4. Grafik *time series* temperatur

Gambar 4 menunjukkan hasil plot data *time series* respons temperatur dalam satuan derajat celsius, terlihat jelas bahwa grafiknya tidak linear terjadi perubahan bentuk seiring berjalannya waktu setiap bulan. Secara visual sangat sulit dianalisis jika menggunakan analisis pola, sulit mengidentifikasi aturan yang terkandung didalamnya dan potensial menarik sebagai pendukung keputusan atau peramalan. Langkah berikutnya segmentasi grafik menjadi beberapa *sliding window* yang disebut dengan *sub sequence time series*, menunjukkan bahwa, grafik dibagi menjadi beberapa *window*, dalam kasus ini dibagi dalam 13 *window* dan selanjutnya *window* tersebut akan menjadi pusat analisis



Gambar 5. Grafik *time series* periode bulanan yang terdiri dari beberapa *window*

Gambar 5 menunjukkan hasil analisis untuk setiap *window*, dari s dibentuk semua *window* (subsikuen) s_1, \dots, s_{n-w+1} dengan lebar w dimana $s_i = (x_i, \dots, x_{i+w-1})$, s terdiri dari 13 *window* dengan lebar $w = 1$ bulan, dengan menggunakan persamaan jarak euclid $d(\hat{x}, \hat{y}) = (\sum_i (x_i - y_i)^2)^{1/2}$ setiap *window* menghasilkan titik-titik yang diperoleh melalui perhitungan jarak *similarity*, titik tersebut merupakan hasil *similarity* terhadap perubahan suhu untuk setiap *window* seiring dengan terjadinya perubahan waktu. Setiap *window* menghasilkan 3 titik pusat *similarity*, informasi masing-masing analisis *window* disajikan pada tabel 1.

Tabel 1. Pola yang ditemukan pada setiap analisis *window*

Window	Bulan	Similarity	Pola Yang Ditemukan		
			Trend Suhu	Perubahan Suhu	Waktu (Hari)
I	3 (1-510)	85-14,7 °C	Suhu cenderung	Start	0
		255-15,3 °C	°C sampai dengan	Meningkat	7
		425-13,0 °C	15,3 °C	Menurun	7
II	4 (511-1230)	630-16,4 °C	Suhu cenderung	Meningkat	8
		870-15,0 °C	°C sampai dengan	Menurun	10
		1110-19,4 °C	19,4 °C	Meningkat	10
III	5 (1231-1974)	1354-17,1 °C	Suhu cenderung	Menurun	10
		1602-21,8 °C	°C sampai dengan	Meningkat	10
		1850-22,2 °C	22,2 °C	Meningkat	10
IV	6 (1975-2694)	2094-25,8 °C	Suhu cenderung	Meningkat	10
		2334-25,5 °C	°C sampai dengan	Menurun	10
		2575-29,8 °C	29,8 °C	Meningkat	10
V	7 (2695-3438)	2818-28,3 °C	Suhu cenderung	Menurun	10
		3066-30,9 °C	°C sampai dengan	Meningkat	10
		3314-29,5 °C	29,5 °C	Menurun	10
VI	8 (3439-4182)	3562-28,2 °C	Suhu cenderung	Menurun	10
		3810-29,2 °C	28,2°C sampai	Meningkat	10
		4058-28,2 °C	dengan 29,2 °C	Menurun	10
VII	9 (4183-4902)	4302-25,9 °C	Suhu cenderung	Menurun	10
		4542-22,8 °C	°C sampai dengan	Menurun	10
		4782-21,8 °C	25,9 °C	Menurun	10
VIII	10 (4903-5646)	5026-19,8 °C	Suhu cenderung	Menurun	10
		5273-20,3 °C	°C sampai dengan	Meningkat	10
		5522-19,9 °C	20,3 °C	Menurun	10
IX	11 (5647-6366)	5766-12,1 °C	Suhu cenderung	Menurun	10
		6006-11,0 °C	°C sampai dengan	Menurun	10
		6246-13,2 °C	13,2 °C	Meningkat	10
X	12 (6367-7110)	6490-11,8 °C	Suhu cenderung	Menurun	10
		6738-09,2 °C	°C sampai dengan	Menurun	10
		6987-10,8 °C	11,8 °C	Meningkat	10
XI	1 (7111-7854)	7235-8,3 °C	Suhu cenderung	Menurun	10
		7482-5,5 °C	°C sampai dengan	Menurun	10
		7730-8,1 °C	°C	Meningkat	10
XII	2 (7855-8526)	7966-6,5 °C	Suhu cenderung	Menurun	10
		8190-05,7 °C	°C sampai dengan	Menurun	10
		8414-11,4 °C	11,4 °C	Meningkat	9
XIII	3 (8527-8991)	8604-16,5 °C	Suhu cenderung	Meningkat	8
		8759-17,7 °C	°C sampai dengan	Meningkat	6
		8914-17,0 °C	17,7 °C	Menurun	7

Tabel 1 memuat informasi tentang pola yang dihasilkan berdasarkan analisa grafik pada gambar 5. Jumlah baris pada kolom *window* sesuai dengan jumlah sequen, serta kolom tahun berdasarkan lebar interval *window* yang telah ditetapkan. Masing-masing *window* memiliki 3 interval 1 bulan yang dihasilkan dari perhitungan jarak *similarity* dan setiap interval bulan memiliki data peningkatan atau penurunan suhu rata-rata, data ini diperoleh dari titik-titik *similarity* pada gambar 5. Perubahan pola terjadi seiring dengan perubahan waktu dalam satuan hari yang ditunjukkan pada kolom waktu. Terjadi peningkatan dan penurunan suhu rata-rata setiap *window*, informasi ini termasuk hal yang menarik untuk dijadikan *rule*/aturan dalam pengambilan keputusan atau prediksi dimasa yang akan datang.

3.2 Penemuan Rule Dengan Tingkat Confidence

Berdasarkan tabel 1 dihasilkan 37 *rule*/aturan, setiap *window* memiliki *rule*, format aturan yang digunakan adalah format dengan menggunakan logika proposisi yakni Jika A terjadi, maka B terjadi dalam waktu T hari. Disini A dan B merupakan

bentuk dasar, yaitu titik-titik *similarity* yang dihasilkan oleh diskritisasi pada tabel 1. Dengan menggunakan notasi matematika pada persamaan 5 dihasilkan 37 *rule*. *Rule* tersebut kemudian dicari korelasi kuat dan cenderung dengan menghitung tingkat keyakinan tinggi dengan menggunakan persamaan $c(A \overset{T}{\Rightarrow} B) = \frac{F(A,B,T)}{F(A)}$, sehingga diperoleh *rule* sebagai berikut:

Tabel 2. *Rule* dengan tingkat *confidence* berdasarkan pola yang ditemukan

No	<i>Rule</i>	<i>Confidence</i>
1	IF Suhu Meningkatkan THEN akan Menurun dalam 7 sampai dengan 10 hari hingga 7 ^o C	$= \frac{12}{16} \times 100\% = 75\%$
2	IF Suhu Menurun THEN akan Meningkatkan dalam 7 sampai dengan 10 hari hingga 31 ^o C	$= \frac{11}{21} \times 100\% = 52\%$
3	IF Suhu Menurun THEN akan Menurun dalam 10 hari hingga 5,5 ^o C	$= \frac{10}{21} \times 100\% = 48\%$
4	IF Suhu Meningkatkan THEN akan Meningkatkan dalam 10 hari hingga 25 ^o C	$= \frac{4}{16} \times 100\% = 25\%$

Untuk mendapatkan aturan yang mendekati data asli, maka berusaha memangkas aturan dengan menentukan tingkat keyakinan untuk masing-masing aturan yang diperoleh. 4 Aturan diatas merupakan hasil penggabungan interval bulan seiring dengan perubahan pola yang ditemukan dari aturan semula yakni 37 aturan. 4 aturan diurutkan berdasarkan tingkat keyakinan atau *confidence* masing-masing *rule*.

4. KESIMPULAN

Dengan mengoptimalkan data *time series data mining* melalui optimasi penemuan *rule* dengan analisis *cluster sequence time series data mining*, ditemukannya sebuah model baru dalam penemuan pengetahuan baru pada data *time series* dengan pendekatan diskritisasi deret berkala *sliding window* yang disebut dengan model RBT(*Rule Best Time series*). Similaritas deret berkala berdasarkan ukuran jarak kemiripan antara deret berkala lainnya, membuat *rule* dengan menyajikannya berdasarkan ukuran *J-Measure* sehingga menghasilkan informasi atau pengetahuan atau *trend* dan pola dalam database dari data *time series* yang sifatnya tidak tentu, yang sebelumnya tidak diketahui serta mampu menemukan *rule* yang menarik dengan tingkat persentase *confidence* dan perangkingan ketertarikan *rule*. Keputusan atau informasi tersebut dapat digunakan untuk mengambil keputusan, atau peramalan dimasa yang akan datang Hasil perangkingan dengan pengukuran *J* diurutkan berdasarkan hasil terkecil. Sebagai rangkuman hasilnya dan membandingkannya dengan hasil pengukuran *J* setiap *rule*. Berdasarkan analisa pengetahuan dari setiap window maka dapat ditentukan *rule* yang berpotensi menarik serta memangkas *rule* yang tidak menarik, *rule* yang cenderung memiliki redundansi, *rule* yang tidak memiliki redundansi, maka *best* rulanya adalah **IF** Suhu Meningkatkan **THEN** akan Menurun dalam 7 sampai dengan 10 hari hingga 5^oC.

REFERENCES

- [1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Mag.*, vol. 17, no. 3, p. 37, 1996, doi: 10.1609/aimag.v17i3.1230.
- [2] M. Shokoohi-Yekta, J. Wang, and E. Keogh, "On the Non-Trivial Generalization of Dynamic Time Warping to the Multi-Dimensional Case," *Data Min. Proceeding 2015 ...*, pp. 289–297, 2015, doi: 10.1137/1.9781611974010.33.
- [3] Y. Li, J. Lin, and T. Oates, "Visualizing Variable-Length Time Series Motifs.," *Sdm*, pp. 895–906, 2012.
- [4] S. Zolhavarieh, S. Aghabozorgi, and Y. W. Teh, "A review of subsequence time series clustering.," *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/312521.
- [5] R. Buaton *et al.*, "Decision Tree Optimization in Data Mining with Support and Confidence," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012056.
- [6] R. Buaton *et al.*, "Time Series Optimization on Data Mining," *J. Phys. Conf. Ser.*, vol. 1235, no. 1, 2019, doi: 10.1088/1742-6596/1235/1/012014.
- [7] M. Toloo and S. Nalchigar, "On Ranking Discovered Rules of Data Mining by Data Envelopment Analysis : Some New Models with Applications," no. January 2011, 2015, doi: 10.5772/13659.
- [8] H. X. Huynh, F. Guillet, T. Q. Le, and H. Briand, "Ranking objective interestingness measures with sensitivity values," vol. 24, pp. 122–132, 2008.