

PENERAPAN DISCRETIZE BY FREQUENCY DALAM MENINGKATKAN AKURASI ALGORITMA C4.5 DALAM MEMPREDIKSI CUACA PADA JALUR PANTURA TEGAL-PEKALONGAN-SEMARANG

Kurnia Adi Cahyanto¹, Esti Mulyani², Fachrul Pralienka Bani Muhamad³

^{1,2,3}Politeknik Negeri Indramayu

Email: ¹kurnia@polindra.ac.id, ²estimulyani@polindra.ac.id, ³fachrul.pbm@polindra.ac.id

Abstrak

Cuaca merupakan suatu hal yang sangat vital dalam kehidupan manusia sehari-hari. Hampir seluruh aktifitas yang dilakukan manusia di dunia ini tidak dapat lepas dari keadaan cuaca yang berlangsung. Oleh karena itu, pengamatan cuaca di sini sangatlah diperlukan untuk melakukan prediksi tentang cuaca. Jalur pantura adalah salah satu jalur yang sangat penting bagi pulau Jawa, khususnya jalur pantura Tegal-Pekalongan-Semarang, sehingga dibutuhkan semacam prediksi cuaca untuk mengetahui intensitas hujan di jalur ini. Algoritma C4.5 yang selama ini banyak digunakan untuk melakukan klasifikasi sekaligus prediksi terbukti relatif akurat dalam melakukan proses prediksi ini. Namun, di lain sisi dengan berkembangnya ilmu pengetahuan teknologi komputasi, maka nilai akurasi ini masih dapat ditingkatkan, salah satunya dengan menggunakan *discretize by frequency* dalam langkah *preprocessing data*. Hasil akurasi menunjukkan, bahwa sebelum dilakukan *discretize by frequency* dalam langkah *preprocessing data* nilai akurasi C4.5 adalah 72,99% dan setelah dilakukan *discretize by frequency* dalam langkah *preprocessing data* nilai akurasi C4.5 meningkat menjadi 86,13%.

Kata Kunci: prediksi cuaca, pantura, algoritma C4.5, *discretize by frequency* dan akurasi

Abstract

Weather is a very vital thing in everyday human life. Almost all activities carried out by humans in this world cannot be separated from the prevailing weather conditions. Therefore, observing the weather here is very necessary to make predictions about the weather. The north coast path is one of the most important routes for Java, especially the Tegal-Pekalongan-Semarang pantura line, so it needs some kind of weather prediction to find out the intensity of rain on this route. The C4.5 algorithm which has been widely used to make classifications and predictions has proven to be relatively accurate in carrying out this prediction process. However, on the other hand with the development of computational technology science, the value of this accuracy can still be improved, one of them is by using discretize by frequency in the data preprocessing step. Accuracy results show that before discretize by frequency in the preprocessing step the data accuracy value of C4.5 is 72.99% and after discretize by frequency in the preprocessing step the data accuracy value C4.5 increases to 86.13%.

Keywords: weather prediction, pantura, C4.5 algorithm, *discretize by frequency* and accuracy

I. PENDAHULUAN

Cuaca sangat berhubungan dengan kegiatan atau aktifitas sehari-hari manusia. Dalam hal ini kelancaran pekerjaan ataupun kegiatan lainnya juga dapat saja dipengaruhi oleh cuaca (Novisnky, Stevy and Bobanto, 2017).

Dalam hal transportasi, prediksi cuaca juga sangat penting, khususnya di pesisir. Sehingga ketidaktepatan prediksi cuaca dapat menyebabkan ketidaklancaran jalur

transportasi khususnya di jalur pantura Tegal-Pekalongan-Semarang.

Bahkan, kondisi cuaca diyakini mempengaruhi berbagai aspek yang dapat mempengaruhi keselamatan jalan, yaitu keputusan apakah atau tidak untuk melakukan perjalanan atau pilihan moda transportasi (Bijleveld and Churchill, 2009).

Umumnya, manusia akan mengharapkan efek negatif hujan dan suhu ekstrim pada permintaan transportasi. Hal ini terutama berlaku untuk perjalanan yang dibuat untuk tujuan rekreasi, karena mereka dapat dengan mudah

dijadwal ulang atau dibatalkan (Ommeren, Koetse and Sabir, 2010).

Oleh karena itu, kebutuhan akan akurasi prediksi cuaca juga diharapkan akan sangat membantu manusia dalam aktifitas keseharian yang membutuhkan informasi yang lebih tepat dalam hal prakiraan cuaca.

Selain itu, keadaan cuaca sendiri juga juga dipengaruhi beberapa parameter membuat objek dari cuaca itu sendiri memiliki banyak variabel dari data-data lampau yang akan sangat mungkin dimanfaatkan keberadaanya untuk membuat sebuah prakiraan cuaca beberapa hari ke depan (Cipta, 2014). Sementara itu dalam proses pembuatan informasi prakiraan cuaca ini terdapat beberapa kendala.

Dengan pesatnya teknologi yang berkembang saat ini, prakiraan cuaca dapat dilakukan dengan membangun sebuah sistem penunjang keputusan berbasis komputasi yang mengimplementasikan algoritma klasifikasi data mining (DM) atau disebut juga *Knowledge Discovery in Databases* (KDD) (Subhan and Fanani, 2017). Setiap harinya Badan Meteorologi dan Geofisika (BMKG) melakukan pengukuran terhadap faktor-faktor cuaca. Ketersediaan data yang melimpah tentang faktor-faktor cuaca tersebut akan mendukung penggalian informasi untuk prakiraan cuaca menggunakan data mining. Data mining mampu memberikan suatu prakiraan kondisi cuaca yang akan terjadi (T.P and Indah Sari, 2018).

Algoritma C4.5 adalah salah satu teknik data mining berbasis klasifikasi yang sering digunakan untuk melakukan prediksi dari beberapa data yang pada umumnya bersifat binominal (Novandya, Adhika., Oktria, 2017).

Pada penelitian ini akan dilakukan analisis terhadap penerapan diskretisasi untuk meningkatkan akurasi klasifikasi algoritma C4.5 dalam memprediksi cuaca pada jalur pantura tegal-pekalongan-semarang.

II. METODE

Penelitian ini menggunakan metode yang terdiri atas beberapa subbab yaitu jenis dan sumber data, metode pengumpulan data, metode pengembangan sistem, dan kerangka pikir penelitian yang digunakan dalam menyelesaikan masalah penelitian.

Jenis dan Sumber Data

Jenis data yang digunakan untuk penelitian ini adalah data sekunder. Data sekunder adalah data yang telah diolah lebih lanjut dan disajikan dengan baik oleh pihak pengumpul data primer atau pihak lain (Ratnasari, 2017).

Metode Pengumpulan Data

1. Studi Pustaka

Digunakan sumber pustaka yang relevan untuk mengumpulkan informasi yang diperlukan dalam penelitian ini, yaitu penelitian terkait dan sumber pustaka yang berupa buku, jurnal, prosiding seminar nasional, skripsi maupun tesis.

Setelah sumber pustaka terkumpul, dilanjutkan dengan penelaahan dari sumber pustaka tersebut, yang pada akhirnya sumber pustaka itu dijadikan landasan untuk menganalisis permasalahan.

2. Pengumpulan Data

Menggunakan data curah hujan harian BMKG dalam kurun waktu kurang lebih 10 tahun, yaitu dari tahun 2008-2018 sebagai *data training* dan data curah hujan harian BMKG dalam kurun waktu Mei 2018 sampai dengan Juli 2018 sebagai *data testing*.

3. Analisis dan Perancangan Sistem

Menganalisis data yang telah dikumpulkan melalui beberapa langkah-langkah dalam *data warehousing* dan sistem yang akan digunakan untuk penyelesaian masalah.

4. Implementasi Sistem

Pengimplementasian algoritma C4.5 untuk data-data yang sudah diperoleh yang ke dalam rancangan sistem yang telah dibuat.

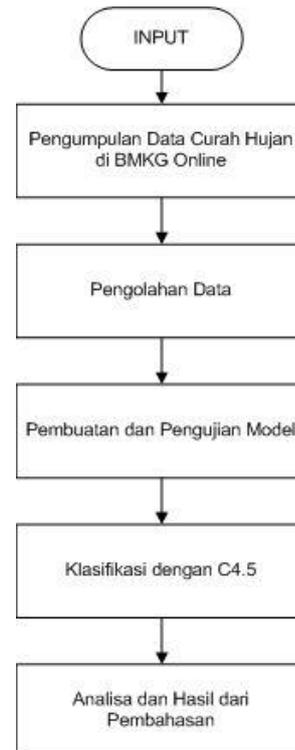
5. Pengujian dan Analisis Implementasi

Melakukan pengujian dan analisis terhadap C4.5 untuk mendapatkan tingkat akurasi yang tinggi. Kemudian menganalisis hasil yang didapat dan diprediksi agar bisa digunakan dalam kehidupan nyata.

6. Penyusunan Laporan

Tahap terakhir yang dilakukan dalam Metodologi Penelitian. Melakukan penyusunan terhadap apa yang sudah dilakukan ke dalam satu laporan tertulis sesuai dengan struktur dan alur penelitian.

Diagram Alir Metode Penelitian



Gambar 1. Skema Diagram Alir Metode Penelitian

III. HASIL DAN PEMBAHASAN

Pengumpulan Data

1. Menyeleksi dan memilih data

Dalam penelitian ini menggunakan dua *subset* data, yaitu *data training* dan *data testing*. Data yang akan digunakan diambil dari situs www.bmkg.go.id dengan *dataset* cuaca dari Data Iklim Harian 1 Januari 2009 sampai dengan 31 Desember 2017 sebagai *data training* dan Data Iklim Harian 24 Mei 2018 sampai dengan 22 Juli 2018 sebagai *data testing*. Data cuaca dapat dilihat pada Tabel 1.

Tabel 1 Data Cuaca yang Didapat dari Situs www.bmkg.go.id

Nama Stasiun	WMO ID	Tanggal	Suhu Min	Suhu Max	Suhu
Stasiun Meteorologi Tegal	96797	01/01/2009	24,9	31	27
Stasiun Klimatologi Semarang	96835	01/01/2009	25,4	30,4	27,4
Stasiun Meteorologi Maritim Tanjung Mas	96837	01/01/2009	24	31	27,2
Stasiun Meteorologi Ahmad Yani	96839	01/01/2009	24,2	31,4	27,6
Stasiun Meteorologi Cilacap	96805	02/01/2009	25,4	31,4	27,1
Stasiun Meteorologi Tegal	96797	02/01/2009	25,8	32,6	27,7
Stasiun Klimatologi Semarang	96835	02/01/2009	24,8	30	27
Stasiun Meteorologi Maritim Tanjung Mas	96837	02/01/2009	24,3	31	27,3
Stasiun Meteorologi Ahmad Yani	96839	02/01/2009	23,7	30,4	27,1
Stasiun Meteorologi Cilacap	96805	03/01/2009	24,2	31,6	27,9
Stasiun Meteorologi Tegal	96797	03/01/2009	23,8	33,8	29,1

Data pada Tabel 1 berbentuk *spreadsheet* yang diunduh pada tanggal 24 Juni 2018, dengan atribut sebagai berikut:

- Nama stasiun adalah nama tempat untuk melakukan pengamatan cuaca dan melaporkan keadaan cuaca maupun iklim pada waktu tertentu.
- WMO ID adalah nomor identitas yang diberikan oleh World Meteorological Organization sebagai lembaga dunia yang menjadi organisasi Perserikatan Bangsa-Bangsa untuk menangani laporan cuaca dan iklim sedunia.
- Tanggal adalah waktu dimana telah dilakukan pengamatan. Dalam hal ini data yang diambil dilakukan secara harian.
- Suhu minimum adalah suhu yang terendah ketika dilakukan pengukuran suhu di hari tersebut, satuan dalam derajat Celcius (°C).

- Suhu maksimum adalah suhu yang tertinggi ketika dilakukan pengukuran suhu di hari tersebut, satuan dalam derajat Celcius (°C).
- Suhu rata-rata adalah nilai suhu yang diambil rata-rata ketika dilakukan pengukuran suhu secara perjam dalam satu hari tersebut, satuan dalam derajat Celcius (°C).
- Kelembaban adalah kelembaban udara rata-rata pada saat dilakukan pengukuran, dinyatakan dalam persen (%).
- Curah hujan adalah seperti yang didefinisikan pada bab II penelitian ini dinyatakan dalam satuan milimeter per hari (mm/hari). Dalam penelitian ini, curah hujan digunakan untuk menentukan jenis hujan.
- Lama penyinaran adalah lama matahari menyinari bumi dalam waktu 8 jam tanpa tertutup awan (cerah) di daerah tersebut, dinyatakan dalam satuan jam.
- Kecepatan angin rata-rata adalah kecepatan rata-rata angin yang terukur dalam satu hari tersebut, satuan dalam Knot.
- Arah angin terbanyak adalah modus dari arah angin yang terlihat pada satu hari tersebut, dinyatakan dengan arah utara (N), timur laut (NE), timur (E), tenggara (SE), selatan (S), barat daya (SW), barat (W), atau barat laut (NW).
- Kecepatan angin terbesar adalah kecepatan angin maksimum yang terukur pada satu hari tersebut, satuan dalam Knot.
- Arah angin saat kecepatan maksimum adalah sudut angin saat, terjadi kecepatan maksimum tersebut, dinyatakan dalam derajat (°).

2. Praproses

- Data Cleansing* untuk *data training*

Data training menggunakan data sebanyak 18.976 *dataset* seluruh stasiun pengamatan cuaca di Jawa Tengah yang diambil dari situs <http://dataonline.bmkg.go.id> yang masih mentah. Dari *dataset* tersebut, sudah terlihat atribut yang nanti akan dijadikan sebagai tolok ukur penentuan prediksi curah hujan. Namun dalam tabel tersebut ada beberapa hari yang terkadang tidak ada pengukuran (dinyatakan dalam kode: 8888) maupun tidak ada data yang masuk (dinyatakan dalam kode: 9999).

Tabel 2 *Dataset* Cuaca yang Masih Terdapat Data tidak Valid

Curah Hujan	Lama Penyinaran	Kecepatan Angin Rata-Rata	Arah Angin Terbanyak
46,6	0	9999	S
17,3	0,8	2	SW
8888	0,7	2	N
46	0,9	1	NW

Sehingga yang pertama kali dilakukan adalah membersihkan baris maupun kolom yang mengandung data seperti ini dengan cara menghapusnya. Karena dalam penelitian ini,

seluruh nilai-nilai atribut yang akan dipakai harus terukur dan terdata (lengkap).

Berikutnya, data yang dibutuhkan untuk prediksi ini hanya diambilkan data dari stasiun pengamatan cuaca yang ada di Jawa Tengah bagian utara, yaitu: Stasiun Meteorologi Tegal, Stasiun Klimatologi Semarang, Stasiun Meteorologi Maritim Tanjung Mas, dan Stasiun Meteorologi Ahmad Yani. Sedangkan data sebelumnya pada tabel di bawah ini terdapat Stasiun Meteorologi Cilacap yang terdapat di Jawa Tengah bagian selatan, sehingga harus dihapus dari dalam tabel data.

Tabel 3 Data Stasiun Pengamatan Cuaca di Jawa Tengah untuk *Data Model*

Nama Stasiun	WMO ID	Tanggal	Suhu Min (°C)	Suhu Max (°C)
Stasiun Meteorologi Cilacap	96805	23/06/2008	22,2	29,1
Stasiun Meteorologi Tegal	96797	23/06/2008	23	31,5
Stasiun Klimatologi Semarang	96835	23/06/2008	25,8	32,6
Stasiun Meteorologi Maritim Tanjung Mas	96837	23/06/2008	23,2	33
Stasiun Meteorologi Ahmad Yani	96839	23/06/2008	23,1	32,8
Stasiun Meteorologi Cilacap	96805	24/06/2008	23,2	29,2

Dalam penelitian ini, ada penghapusan data dari 23 Juni 2008 sampai dengan 31 Desember 2008 dan dari tanggal 1 Januari 2018 sampai dengan 22 Juli 2018, disesuaikan dengan kemampuan software tools (Rapidminer) untuk klasifikasi data mining ini, data trainingl berkurang menjadi 9.832 dataset.

Setelah proses *cleansing* selesai dataset yang siap ditransformasikan.

b. *Data Cleansing* untuk *data testing*

Data testing diambilkan data dari seluruh stasiun pengamatan cuaca di Jawa Tengah situs <http://dataonline.bmkg.go.id> dari bulan Mei 2018 sampai Juli 2018 dengan jumlah 356 dataset yang masih berupa data mentah.

Tabel 4 *Dataset* Cuaca untuk *Testing* yang Masih Terdapat Data tidak Valid

Tanggal	Suhu Min (°C)	Suhu Max (°C)	Suhu Rata-Rata (°C)
24/06/2018	20,8	26,2	9999
24/06/2018	24,7	32,6	27,6
24/06/2018	24,8	33,4	28,3
24/06/2018	25,8	29,1	26,3

Sehingga yang pertama kali dilakukan adalah membersihkan baris maupun kolom yang mengandung data seperti ini dengan cara menghapusnya. Karena dalam penelitian ini, seluruh nilai-nilai atribut yang akan dipakai harus terukur dan terdata (lengkap).

Berikutnya, data yang dibutuhkan untuk prediksi ini hanya diambilkan data dari stasiun pengamatan cuaca yang ada di Jawa Tengah bagian utara, yaitu: Stasiun Meteorologi Tegal, Stasiun Klimatologi Semarang, Stasiun Meteorologi Maritim Tanjung Mas, dan Stasiun Meteorologi Ahmad Yani. Sedangkan data sebelumnya pada tabel di bawah ini terdapat Stasiun Meteorologi Cilacap yang terdapat di Jawa Tengah bagian selatan, sehingga harus dihapus dari dalam tabel data.

Tabel 5 Data Stasiun Pengamatan Cuaca di Jawa Tengah untuk *Data Testing*

Nama Stasiun	WMO ID	Tanggal	Suhu Min (°C)	Suhu Max (°C)
Stasiun Meteorologi Tegal	96797	24/06/2018	24,7	32,6
Stasiun Meteorologi Ahmad Yani	96839	24/06/2018	24,8	33,4
Stasiun Meteorologi Cilacap	96805	24/06/2018	25,8	29,1
Stasiun Meteorologi Maritim Tanjung Mas	96837	24/06/2018	25,2	33,5
Stasiun Klimatologi Semarang	96835	24/06/2018	24,8	33,8

Setelah proses *cleansing* selesai, data testing berkurang menjadi 137 dataset yang siap ditransformasikan.

3. Integrasi

a. Integrasi Data untuk *Data Training*

Dalam penelitian ini, objek utama dalam menentukan kategori hujan adalah jumlah intensitas curah hujan, yang dalam tabel dinyatakan dalam bentuk numerik.

Tabel 6 Tabel Curah Hujan Sebelum Diintegrasikan

Suhu Min	Suhu Max	Suhu Rata-Rata	Kelembaban	Curah Hujan
24,3	31,8	27,7	76	37,3
25,2	33,2	28,8	66	0
25,3	33,1	28,2	72	0

Dalam hal ini, data curah hujan akan diintegrasikan menjadi kategori hujan dengan skala yang mengacu pada Tabel 7.

Tabel 7 Kriteria Curah Hujan BMKG

No	Kategori	Intensitas Curah Hujan	
		Setiap Jam (mm/jam)	Setiap Hari (mm/hari)
1.	Tidak hujan (berawan)	-	-
2.	Hujan ringan	0.1 – 4.9	0.1 – 9.9
3.	Hujan sedang	5.0 - 9.9	20.0 – 4.9
4.	Hujan sangat lebat	10.0 - 20	50.0 – 100
5	Hujan sangat lebat	> 20.0	> 100.0

Jenis data menjadi polinomial. Proses integrasi dapat dilihat pada Tabel 8.

Tabel 8 Tabel Curah Hujan Saat Diintegrasikan

Kecepatan Angin	Arah Saat Kecepatan Angin Maksimum	Curah Hujan	Jenis Hujan
5	70	0	TIDAK HUJAN / BERAWAN
8	90	0	TIDAK HUJAN / BERAWAN
5	50	37,3	HUJAN SEDANG

Dari Tabel 8 dapat dilihat bahwa nilai numerik dari curah hujan sudah diintegrasikan ke dalam kategori hujan. Selanjutnya, untuk lebih memudahkan dalam proses penambangan data, maka atribut curah hujan setelah melalui proses integrasi, dihapus. Lebih jelasnya dapat dilihat pada Tabel 9.

Tabel 9 Tabel Curah Hujan Setelah Diintegrasikan

Arah Angin Terbanyak	Kecepatan Angin Terbesar	Arah Angin Saat Kecepatan Maksimal	Kategori Hujan
SE	6	120	TIDAK HUJAN / BERAWAN
E	5	70	TIDAK HUJAN / BERAWAN
SE	8	90	TIDAK HUJAN / BERAWAN
S	5	50	HUJAN SEDANG

b. Integrasi Data untuk *Data Testing*

Untuk data testing juga dilakukan hal yang sama seperti transformasi data untuk data training.

4. Agregasi

Setelah dilakukan integrasi data dari data numerik menjadi data polinomial kategori hujan, berikutnya data polinomial ini agregasikan menjadi binomial yang hanya terdiri dari dua kelas, yaitu “tidak hujan/berawan” dan “hujan”.

Kategori “hujan ringan”, “hujan sedang”, “hujan lebat”, dan “hujan sangat lebat” diagregasikan ke

dalam “hujan”. Sedangkan kategori “tidak hujan/berawan” masih tetap.

Nantinya klasifikasi hanya akan memiliki dua label, yaitu “tidak hujan/berawan” dan “hujan”.

Tabel 10 Tabel Agregasi

Tidak hujan/berawan	Tidak hujan/berawan
Hujan ringan	Hujan
Hujan sedang	
Hujan lebat	
Hujan sangat lebat	

Adapun hasil agregasi dapat dilihat pada Tabel 11.

Tabel 11 Tabel Hasil Agregasi

Arah	Kec	Sudut Arah	Kategori
N	5	360	HUJAN
NW	5	350	HUJAN
NW	6	315	TIDAK HUJAN / BERAWAN
NW	7	320	TIDAK HUJAN / BERAWAN
NW	8	200	TIDAK HUJAN / BERAWAN

Dengan melihat data Tabel 11, maka data telah siap untuk ditambang untuk memperoleh informasi dari kumpulan data-data tersebut. Dalam hal ini, penambangan data menggunakan algoritma C4.5.

5. Penerapan Data Mining

a. Menggunakan Metode Klasifikasi dan Prediksi

Metode untuk melakukan penambangan data ini adalah klasifikasi, karena nanti di akhir dari penambangan data ini diharapkan akan menghasilkan informasi yang sudah terklasifikasi untuk kemudian digunakan sebagai acuan prediksi data berikutnya.

b. Menggunakan Algoritma C4.5

Dalam penelitian ini menggunakan aplikasi Framework Rapidminer 8.2 untuk mempermudah mengklasifikasikan menggunakan algoritma C4.5.

Penerapan Algoritma C4.5

1. Tahap Awal

Untuk menggunakan Rapidminer, terlebih dahulu dikelompokkan nilai dari setiap variabel sesuai tipe datanya dan status keanggotaannya.

- Tanggal termasuk dalam tipe data date_time dan digunakan sebagai id.
- Suhu minimum termasuk dalam tipe data real dan digunakan sebagai atribut.
- Suhu maksimum termasuk dalam tipe data real dan digunakan sebagai atribut.
- Suhu rata-rata termasuk dalam tipe data real dan digunakan sebagai atribut.
- Kelembaban rata-rata termasuk dalam tipe data integer dan digunakan sebagai atribut.
- Lama penyinaran termasuk dalam tipe data real dan digunakan sebagai atribut.

- g. Kecepatan angin termasuk dalam tipe data integer dan digunakan sebagai atribut.
- h. Arah angin terbanyak termasuk dalam tipe data polynominal dan digunakan sebagai atribut.
- i. Kecepatan angin termasuk dalam tipe data integer dan digunakan sebagai atribut.
- j. Arah angin saat kecepatan maksimum termasuk dalam tipe data integer dan digunakan sebagai atribut.
- k. Kategori hujan termasuk dalam tipe data polynominal dan digunakan sebagai label.

Kelembaban Rata-rata (%) = range1 [-∞ - 78.500]: TIDAK HUJAN/BERAWAN {HUJAN=1480, TIDAK HUJAN/BERAWAN=3884}

Kelembaban Rata-rata (%) = range2 [78.500 - ∞]: HUJAN {HUJAN=3197, TIDAK HUJAN/BERAWAN=1271}

Tabel 12 Tabel Nilai Atribut setelah Didiskritisasi dengan Frekuensi

No	Kategori Hujan	Suhu Rata-rata (°C)	Kelembaban Rata-rata (%)	Lama Penyinaran (jam)
1	Hujan	range1[-∞ - 28.050]	range2[78.500 - ∞]	range1[-∞ - 6.550]
2	Hujan	range1[-∞ - 28.050]	range2[78.500 - ∞]	range1[-∞ - 6.550]
3	Hujan	range1[-∞ - 28.050]	range1[-∞ - 78.500]	range1[-∞ - 6.550]
4	Hujan	range1[-∞ - 28.050]	range2[78.500 - ∞]	range2[6.550 - ∞]
5	Hujan	range1[-∞ - 28.050]	range2[78.500 - ∞]	range1[-∞ - 6.550]
6	Hujan	range1[-∞ - 28.050]	range2[78.500 - ∞]	range1[-∞ - 6.550]
7	Hujan	range1[-∞ - 28.050]	range2[78.500 - ∞]	range1[-∞ - 6.550]
8	Tidak Hujan/Berawan	range2[28.050 - ∞]	range1[-∞ - 78.500]	range2[6.550 - ∞]
9	Hujan	range1[-∞ - 28.050]	range2[78.500 - ∞]	range1[-∞ - 6.550]
10	Hujan	range1[-∞ - 28.050]	range2[78.500 - ∞]	range1[-∞ - 6.550]

Hasil Klasifikasi Menggunakan C4.5

1. Klasifikasi C4.5 Langsung



Gambar 2. Pohon Keputusan C4.5 yang Terbentuk

Menghasilkan *rule* sebagai berikut:

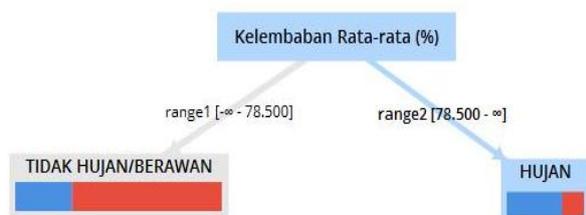
```

Kelembaban Rata-rata (%) > 73.500
| Suhu Minimum (°C) > 27.650: TIDAK HUJAN/BERAWAN {HUJAN=0, TIDAK HUJAN/BERAWAN=9}
| Suhu Minimum (°C) ≤ 27.650
| | Suhu Rata-rata (°C) > 30.350: TIDAK HUJAN/BERAWAN {HUJAN=0, TIDAK HUJAN/BERAWAN=3}
| | Suhu Rata-rata (°C) ≤ 30.350: HUJAN {HUJAN=4169, TIDAK HUJAN/BERAWAN=2580}
Kelembaban Rata-rata (%) ≤ 73.500
| Suhu Maksimum (°C) > 27.600: TIDAK HUJAN/BERAWAN {HUJAN=507, TIDAK HUJAN/BERAWAN=2563}
| Suhu Maksimum (°C) ≤ 27.600: HUJAN {HUJAN=1, TIDAK HUJAN/BERAWAN=0}
    
```

2. Klasifikasi C4.5 Menggunakan Diskritisasi dengan Frekuensi

Diskretisasi membagi *range* dari setiap atribut menjadi beberapa interval yang dapat dilihat pada Tabel 12.

Setelah dataset tersebut didiskritisasi, maka pohon keputusan yang terbentuk menjadi lebih sedikit.



Gambar 2. Pohon Keputusan C4.5 yang Terbentuk setelah Dataset Didiskritisasi

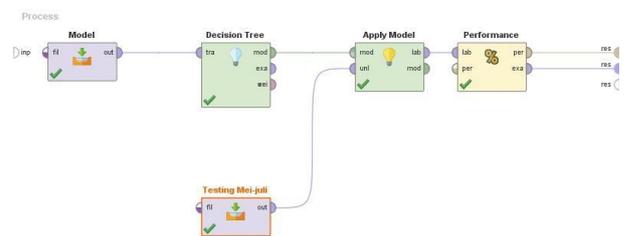
Setelah proses ini, *rule* yang diperoleh menjadi lebih singkat.

Dari hasil klasifikasi C4.5, didapatkan *rule* seperti di atas, yang akan digunakan sebagai *data model* dalam menentukan akurasi dari prediksi prediksi untuk *data testing*.

Proses dan Hasil Validasi

Validasi digunakan untuk mengukur sejauh mana tingkat akurasi antara *data model* dengan *data testing*.

1. Proses dan validasi algoritma C4.5



Gambar 3. Bagan Proses Validasi untuk Algoritma C4.5

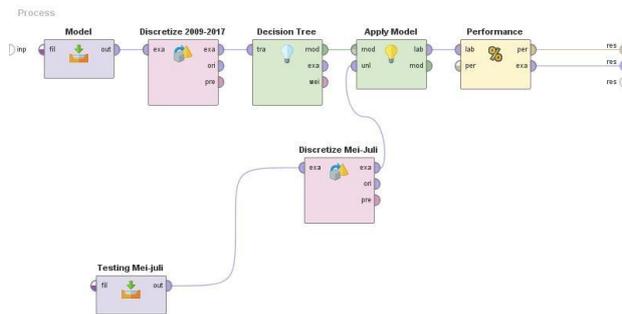
Dalam proses validasi ini, data model sejumlah 9832 dataset akan digunakan untuk memvalidasi 356 dataset dari data testing.

Tabel 13 Tabel Hasil Validasi Algoritma C4.5

Accuracy: 72.99%			
	True Hujan	True Tidak Hujan	class precision
Prediksi Hujan	14	32	30.43%
Prediksi Tidak Hujan	5	86	94.51%
class recall	73.68%	72.88%	

Dari proses di atas, didapatkan hasil validasi dengan nilai akurasi sebesar 72,99%, nilai precision sebesar 94,51%, dan nilai recall sebesar 72,88%.

2. Proses dan validasi algoritma C4.5 dengan Discretize by Frequency



Gambar 4. Bagan Proses Validasi untuk Algoritma C4.5

Dalam proses validasi ini, data model sejumlah 9832 dataset akan digunakan untuk memvalidasi 356 dataset dari *data testing* yang sebelumnya didiskritisasi dengan frekuensi dengan dua range.

1. Range untuk suhu rata-rata
range 1 $[-\infty - 28,050]$
range 2 $[28,050 - \infty]$
2. Range untuk Kelembaban Rata-rata
range 1 $[-\infty - 78,500]$
range 2 $[78,500 - \infty]$
3. Range untuk Lama Penyinaran
range 1 $[-\infty - 6,550]$
range 2 $[6,550 - \infty]$
4. Range untuk Kec. Angin Rata-Rata
range 1 $[-\infty - 2,500]$
range 2 $[2,500 - \infty]$

Tabel 14 Tabel Hasil Validasi Algoritma C4.5

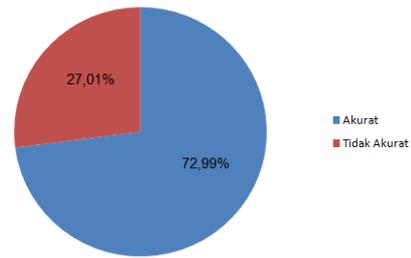
Accuracy: 86.13%			
	True Hujan	True Tidak Hujan	class precision
Prediksi Hujan	0	0	0.00%
Prediksi Tidak Hujan	19	118	86.13%
class recall	0.00%	100.00%	

Berdasarkan proses tersebut, didapatkan hasil validasi dengan nilai akurasi sebesar 86,13%, nilai precision sebesar 86,13%, dan nilai recall sebesar 100%.

Evaluasi

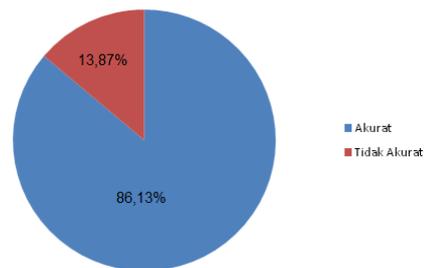
Dari pengukuran di atas, maka didapatkan hasil bahwa tingkat akurasi algoritma C4.5 untuk memprediksi kategori hujan di daerah pantura jalur Tegal-Pekalongan-Semarang adalah sebesar 72,99%.

Dari kedua diagram (Gambar 5 dan Gambar 6), evaluasi menunjukkan bahwa hasil akurasi algoritma C4.5 mengalami peningkatan sebesar 13,14% setelah data mengalami diskritisasi dengan frekuensi.



Gambar 5. Hasil Evaluasi untuk Algoritma C4.5

Sedangkan untuk penerapan diskritisasi dengan frekuensi dari pengukuran di atas, maka didapatkan hasil bahwa tingkat akurasi algoritma C4.5 untuk memprediksi kategori hujan di daerah pantura jalur Tegal-Pekalongan-Semarang adalah sebesar 86,13%.



Gambar 6. Hasil Evaluasi untuk Algoritma C4.5 dengan Discretize by Frequency

IV. PENUTUP

Kesimpulan

Dari penelitian ini dapat diambil kesimpulan bahwa tingkat akurasi algoritma C4.5 untuk memprediksi kategori hujan di daerah pantura jalur Tegal-Pekalongan-Semarang adalah sebesar 72,99% sebelum data didiskritisasi dengan frekuensi dan meningkat sebesar 13,14% menjadi 86,13% setelah data didiskritisasi dengan frekuensi. Diskritisasi dengan frekuensi terbukti dapat digunakan untuk meningkatkan performansi algoritma C4.5 dalam proses klasifikasi.

Saran

Dalam penelitian ini hanya digunakan satu algoritma untuk mengolah data, jadi masih dapat dimungkinkan akurasi ditingkatkan menggunakan algoritma optimisasi. Masih dimungkinkan untuk menambahkan optimasi lain dalam hal meningkatkan keakuratan pengukuran

V. DAFTAR PUSTAKA

Bijleveld, F. and Churchill, T. (2009) ‘The influence of weather conditions on road safety. An assessment of the effect of precipitation and temperature.’, *SWOV Institute for Road Safety Research, Leidschendam*, 04(3), p. 2.

Cipta, S. P. (2014) ‘Penerapan Algoritma Evolving Neural Network Untuk Prediksi Curah Hujan’,

- Jtiulm*, Volume 1(1), pp. 1–8.
- Novandya, Adhika., Oktria, I. (2017) ‘Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi’, *Jurnal Ilmiah Teknik Informatika (FORMAT)*, 6(2), pp. 98–106.
- Novisnky, F., Stevy, H. and Bobanto, M. D. (2017) ‘Pemodelan Sistem Prediksi Intensitas Curah Hujan di Kota Manado Dengan Menggunakan Kontrol Logika Fuzzy’, *JURNAL MIPA UNSRAT ONLINE*, 6(2), pp. 19–23.
- Ommeren, J. van, Koetse, M. J. and Sabir, P. R. M. (2010) ‘Impact of Weather on Daily Travel Demand’, *Amsterdam: VU University, Department of Spatial Economics*, p. 25. doi: 10.13339/j.cnki.sglc.2017.05.002.
- Ratnasari, D. (2017) ‘PENILAIAN PRESTASI KERJA KARYAWAN PADA SALES OFFICE PT GARUDA INDONESIA (PERSERO) TBK. Palembang, Indonesia’, *Politeknik Negeri Sriwijaya*.
- Subhan, A. and Fanani, A. Z. (2017) ‘PENERAPAN DATA MINING UNTUK MENENTUKAN POTENSI HUJAN HARIAN DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES’, *FIK UDINUS*. doi: 10.1021/jf901375e.
- T.P, B. P. and Indah Sari, R. D. (2018) ‘Penerapan Data Mining Untuk Prakiraan Cuaca Di Kota Malang Menggunakan Algoritma Iterative Dichotomiser Tree (Id3)’, *Jouticla*, 2(2), pp. 101–108. doi: 10.30736/jti.v2i2.68.