

## PENERAPAN DATA MINING UNTUK PRAKIRAAN CUACA DI KOTA MALANG MENGGUNAKAN ALGORITMA *ITERATIVE DICHOTOMISER TREE* (ID3)

**Broto Poernomo T.P.<sup>1)</sup>, Rina Dewi Indah Sari<sup>2)</sup>**

Dosen STMIK ASIA Malang

e-mail: papung@gmail.com, rndinsr@gmail.com

### **ABSTRACT**

*In the preparation of weather forecasts information there are several obstacles, involving many sources of data and weather forecasts relying on the ability of the forerunner, so that the interpretations produced may differ between forecasters because of their own experience. Differences in interpretation can confuse users and potentially reduce the quality of information submitted. Based on this problem the author intends to study the forecast model with data mining using ID3 algorithm to obtain the appropriate model so as to facilitate the process of weather analysis and forecast. In building the application, data obtained from BMKG website is addressed dataonline.bmkg.go.id and the data taken is datacuaca on January 1, 2012 through November 30, 2015 for a total of 2414 data and 9 attributes. After the selection of attributes (only weather related attributes) and removing damaged data (incomplete data and outliers), the data is reduced to 1790 and attributes reduced to 6 pieces. In the testing process performed with 179 data (10% of the data set) random, it is known that there are 112 databases corresponding to the actual weather data. So it can be concluded that the accuracy of 73.74%.*

**Kata Kunci:** *Data Mining, Weather Forecasting, Introduction, Iterative Dichotomiser Tree (ID3)*

### **PENDAHULUAN**

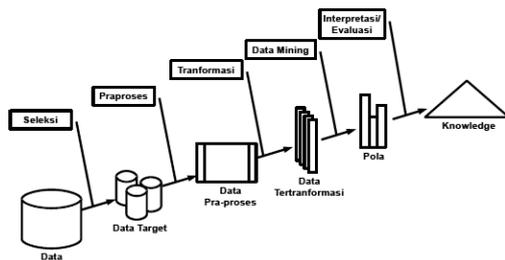
Cuaca merupakan keadaan udara pada suatu waktu yang singkat dan tempat yang sempit. Kondisi cuaca di suatu tempat dapat ditentukan oleh beberapa faktor, seperti suhu udara, kelembaban udara, intensitas cahaya matahari, arah angin, kecepatan angin, dan sebagainya. Sehingga dengan melihat faktor-faktor ini dapat diprediksi cuaca yang akan terjadi pada keesokan harinya. Petani dan nelayan merupakan bidang pekerjaan yang erat kaitannya dengan cuaca. Prediksi cuaca yang tepat sangat dibutuhkan oleh bidang tersebut dalam melakukan berbagai kegiatan. Tentu tidak hanya petani dan nelayan yang bisa memanfaatkan informasi prakiraan cuaca, masih banyak bidang pekerjaan lain yang berkaitan, seperti pariwisata, pelayaran, perkebunan, kehutanan, pembangunan gedung, penataan wilayah, dan kesehatan. Sebagian besar stasiun meteorologi melakukan pengamatan sinoptik dan sebagian di antaranya memberikan pelayanan analisa dan prakiraan cuaca. 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. Dengan pesatnya teknologi yang berkembang saat ini, prakiraan cuaca dapat dilakukan dengan membangun sebuah sistem penunjang keputusan berbasis komputer yang mengimplementasikan algoritma klasifikasi data mining. Setiap harinya Badan Meteorologi dan Geofisika (BMKG) melakukan pengukuran terhadap faktor-faktor cuaca. Ketersediaan data yang melimpah tentang faktor-faktor cuaca

tersebutakan mendukung penggalian informasi untuk prakiraan cuaca menggunakan data mining. Data mining mampu memberikan suatu prakiraan kondisi cuaca yang akan terjadi.

**1. KNOWLEDGE DISCOVERY AND DATA MINING(KDD)**

KDD mengacu pada keseluruhan proses menemukan pengetahuan yang bermanfaat dari data. Ini melibatkan evaluasi dan kemungkinan interpretasi pola untuk membuat keputusan tentang apa yang memenuhi syarat sebagai pengetahuan.



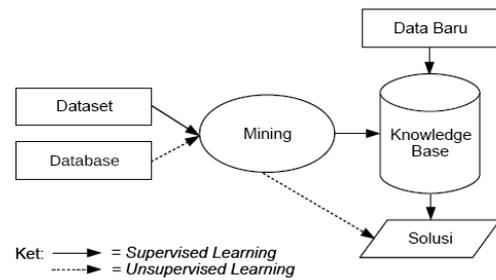
Gambar 1. Tahapan dalam KDD

Itu juga mencakup pilihan skema encoding, preprocessing, pengambilan sampel, dan proyeksi data sebelum langkah data mining.

Secara garis besar, Knowledge Discovery and Data Mining (KDD) adalah proses yang dibantu oleh komputer untuk menggali dan menganalisis sejumlah besar himpunan data dan mengekstrak informasi dan pengetahuan yang berguna.

**2. DATA MINING**

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Data mining adalah teknik yang merupakan gabungan metode metode analisis data secara statistik dengan algoritma-algoritma untuk memproses data berukuran besar. Data mining merupakan proses menemukan informasi atau pola yang penting dalam basis data berukuran besar. (Susanto, 2010)



Gambar 2. Gambaran sistem kerja data mining

**3. Tugas Data Mining**

Tugas data mining secara prediktif adalah untuk mengambil kesimpulan terhadap data terakhir untuk membuat prediksi (Fadli, 2010).

*a. Association Analysis*

Association analysis adalah penemuan association rules yang menunjukkan nilai kondisi suatu attribute yang terjadi bersamaan secara terus-menerus dalam memberikan dataset. Association analysis secara luas dipakai untuk market basket atau analisa data transaksi.

*b. Klasifikasi dan Prediksi*

Klasifikasi dan prediksi mungkin perlu diproses oleh analisis yang relevan, yang berusaha untuk mengidentifikasi atribut-atribut yang tidak ditambahkan pada proses klasifikasi dan prediksi. Atribut-atribut ini kemudian dapat dikeluarkan.

*c. Cluster Analysis*

Tidak seperti klasifikasi dan prediksi, yang menganalisis objek data dengan kelas yang terlabeli, clustering menganalisis objek data tanpa mencari keterangan pada label kelas yang diketahui. Pada umumnya, label kelas tidak ditampilkan di dalam latihan data simply, karena mereka tidak tahu bagaimana memulainya. Clustering dapat digunakan untuk menghasilkan label-label.

**4. Dataset**

Dataset merupakan sekumpulan data yang saling berhubungan dan dikumpulkan dari sebuah sumber tunggal. Istilah ini memiliki beberapa penerapan, mulai dari informasi yang dikompilasi dari hasil survei (surveyresults) sampai dengan sejumlah hasil penelitian ilmiah (scientific research results). Pada dunia komputer dan internet

dataset merupakan sekumpulan angka atau byte, seringkali ditampilkan pada suatu tabel dengan kolom yang mengkategorikan data tersebut ke dalam subset data.

Dataset dapat digunakan untuk memberikan pemahaman mengenai tema atau konsep tertentu. Mereka menyimpan data yang dibutuhkan oleh aplikasi atau sistem operasi agar dapat berfungsi dengan baik. Pada umumnya, data yang terdapat di dalamnya mencakup macro libraries, source program dan sistem parameter atau sistem variabel. Dataset ini dapat disusun sedemikian rupa sehingga dapat dengan mudah dimengerti tanpa perlu referensi pada area penyimpanan tertentu.

Tabel 1: Contoh Dataset

Outlook	Temperature	Humidity	Play
Sunny	34	Normal	Yes
Rainy	19	High	No
Overcast	22	High	No
Overcast	34	Normal	Yes

**5. ITERATIVE DICHOTOMISER 3 (ID3)**

Algoritma Iterative Dichotomiser 3 adalah salah satu yang paling populer karena mudah diinterpretasikan oleh manusia. Decision Tree digunakan untuk pengenalan pola dan termasuk dalam pengenalan pola secara statistik. Pengembang metode klasifikasi algoritma Iterative Dichotomiser 3 adalah Ross Quinlan pada akhir dekade 70-an, di mana dalam upaya mewujudkan suatu sistem pakar yang mampu belajar dari kumpulan contoh. Quinlan memperbaiki algoritma Iterative Dichotomiser 3 menjadi C4.5 pada tahun 1993. Jauh hari sebelumnya, Hunt dari disiplin Psychology menggunakan metode pencarian penuh pada decision tree untuk memodelkan belajar konsep pada manusia pada tahun 60-an. Algoritma pada metode ini menggunakan konsep dari entropy informasi. Secara ringkas cara kerja algoritma ID3 dapat digambarkan sebagai berikut:

- a. Pemilihan atribut dengan menggunakan Information Gain.
- b. Pilih atribut yang mempunyai nilai Information Gain terbesar.
- c. Buat simpul yang berisi atribut tersebut.

- d. Proses perhitungan Information Gain akan terus dilakukan hingga semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikutkan lagi dalam perhitungan nilai Information Gain.

Pemilihan atribut pada ID3 dilakukan dengan properti statistik, yang disebut dengan Information Gain. Information Gain mengukur seberapa baik suatu atribut memisahkan training example ke dalam kelas target. Atribut dengan informasi tertinggi akan dipilih. Dengan tujuan untuk mendefinisikan gain, pertama-tama digunakanlah ide dari teori informasi dengan entropy. Entropy mengukur jumlah dari informasi yang ada.

Kemudian untuk menghitung nilai Information Gain, yang digunakan untuk mengukur efektifitas suatu atribut pengklasifikasian data adalah:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{S} Entropy(S_v)$$

Keterangan dari rumus tersebut adalah:

- A : Atribut
- v : Menyatakan suatu nilai yang mungkin untuk atribut A
- Values(A) : Himpunan nilai-nilai yang mungkin untuk atribut A
- |S<sub>v</sub>| : Jumlah sampel untuk nilai v
- S : Jumlah seluruh sampel data
- Entropy : Entropy untuk sampel-sampel yang memiliki nilai v

**6. Struktur dari Iterative Dichotomiser 3**

Iterative Dichotomiser 3 dibentuk dari 3 tipe simpul, yaitu simpulroot, simpul perantara, dan simpul leaf, yang masing-masing mempunyai pengertian berikut:

- a. Simpul root adalah titik awal dari suatu decision tree.

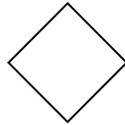


Gambar 3. SimbolRoot

- b. Setiap simpul perantara berhubungan dengan suatu pertanyaan atau pengujian atau Branch

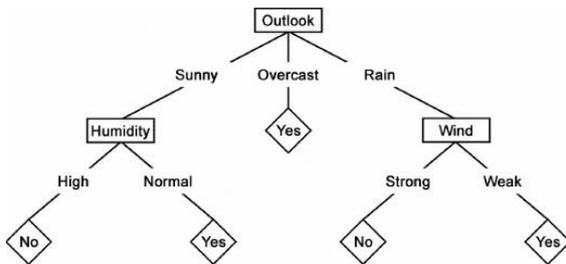
→  
Gambar 4. Simbol Branch

c. Simpul leaf memuat suatu keputusan akhir atau kelas target untuk suatu pohon keputusan.



Gambar 5. Simbol Leaf

Setelah melewati berbagai proses, maka akan didapat sebuah tree Iterative Dichotomiser 3, berikut adalah contoh dari sebuah tree:



Gambar 6. Contoh Decesion tree

**7. Penghitungan Impurity**

Dalam dataset yang memiliki atribut dan atribut kelas, dapat dihitung keseragaman (atau ketidakseragaman) dari tabel berdasarkan kelasnya. Dapat disimpulkan bahwa tabel tersebut seragam atau homogen jika hanya terdapat satu kelas saja (tunggal). Jika tabel memiliki beberapa kelas, dapat disimpulkan bahwa tabel tersebut tidak seragam, heterogen, atau impure. Ada beberapa cara untuk menghitung derajat ketidakseragaman/impurity. Cara yang paling banyak diketahui yaitu:

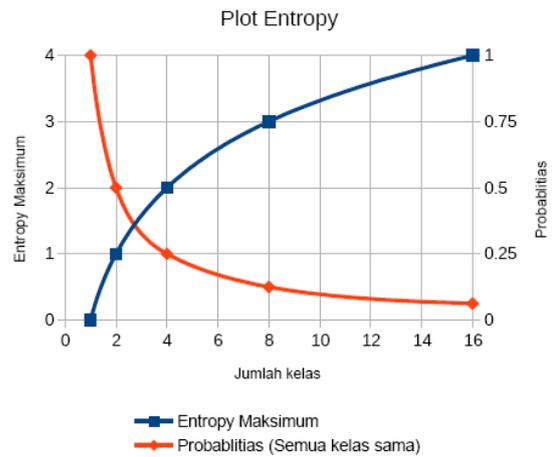
$$Entropy (S) = \sum_{i=1}^c -p_i \log_2 p_i$$

Keterangan dari rumus tersebut adalah:

c : Jumlah nilai yang ada pada atribut target (jumlah kelas klasifikasi).

pi : Jumlah sampel untuk kelas i

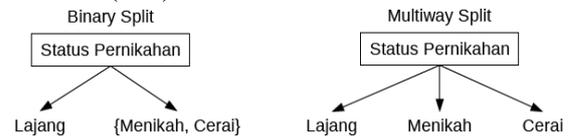
I : Jumlah sampel



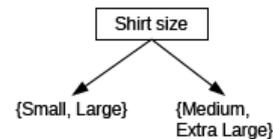
Gambar 7. Plot Entropy

**8. Splitting Way**

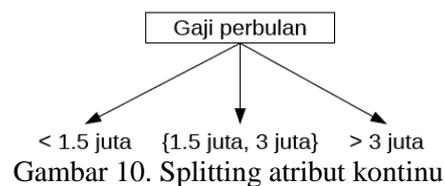
Secara umum splitting memiliki arti pemisahan atau pemecahan. Dalam ID3 splitting merupakan sebuah proses pembagian atau pemecahan atribut suatu node untuk menentukan node berikutnya dalam proses mencari kelas baru. Dari banyaknya nilai, splitting dinyatakan dalam dua cara, yaitu binary split dan multiway split. Binary split digunakan pada atribut yang hanya memiliki dua keluaran (leaf), sedangkan untuk multiway split digunakan pada atribut yang memiliki lebih dari dua keluaran (leaf).



Gambar 8. Splitting atribut nomina



Gambar 9. Splitting atribut ordinal



Gambar 10. Splitting atribut kontinu

**9. METEOROLOGI**

Meteorologi berasal dari kata Yunani, yaitu meteoros, yang artinya benda yang ada di dalam udara dan logos artinya ilmu atau

kajian. Jadi meteorologi didefinisikan sebagai ilmu yang mempelajari proses fisis dan gejala cuaca yang terjadi di dalam atmosfer terutama pada lapisan bawah, yaitu troposfer. Kajian meteorologi diperlukan dalam pembangunan irigasi, objek wisata, tempat peristirahatan, perkebunan, perikanan, lapangan terbang, pelayaran, proyek industri, dan sebagainya.

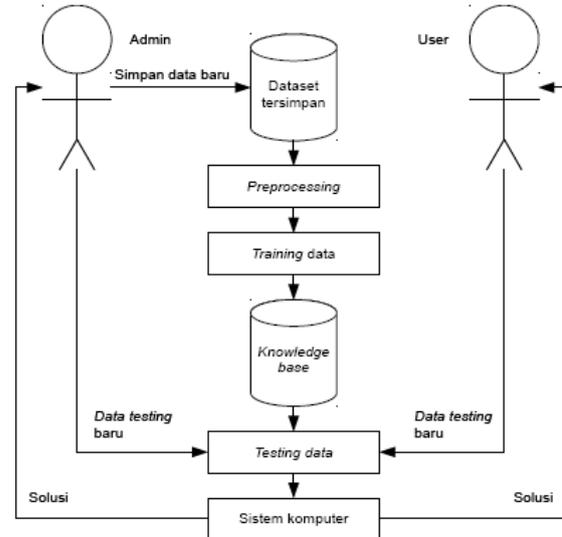
**10. Unsur Cuaca dan Iklim**

Suhu udara, kelembaban udara, curah hujan, tekanan udara, angin, durasi sinar matahari, dan beberapa unsur iklim yang kurang penting lain merupakan faktor yang mempengaruhi unsur iklim dan cuaca. Matahari adalah kendali iklim yang sangat penting dan sumber energi di bumi yang menimbulkan gerak udara dan arus laut.

**METODE PENELITIAN**

Sistem ini merupakan sistem yang dikembangkan untuk membuat analisis terhadap data cuaca yang jumlahnya cukup banyak hingga ditemukan hasil akhir untuk memperkirakan cuaca yang terjadi. Data cuaca digunakan sebagai dataset. Dataset ini harus melalui tahap preprocessing untuk mendapatkan dataset yang siap dihitung. Dataset yang telah dipreprocessing akan detrain menggunakan algoritma Iterative Dichotomizer Three (ID3) untuk menemukan rule yang nantinya akan digunakan sebagai dasar prakiraan. Dalam sistem ini digunakan algoritma ID3 (Iterative Dichotomizer 3) sebagai algoritma dasar. Record tabel yang beragam berupa simbolik dan numerik merupakan alasan dasar penggunaan algoritma ini. Sistem akan berhenti bekerja jika sudah mampu menemukan target cuaca.

Berikut ini akan digambarkan mengenai gambaran diagram blok dari sistem yang dibuat



Gambar 11. Blok Diagram

Dataset yang digunakan sebagai sampel adalah dataset yang didapat dari situs BMKG (dataonline.bmkg.go.id) pada bulan Januari 2012 hingga bulan November 2015. Record dataset berjumlah 2415 dan memiliki delapan atribut yaitu Nama Stasiun, Tanggal, Suhu Minimal, Suhu Maksimal, Suhu Rata Rata, Kelembaban, SunShine dan intensitas Hujan.

PreProcessing yang dilakukan adalah dengan menghapus atribut Nama Stasiun dan Tanggal, penghapusan missing Value dan merubah nilai intensitas hujan ke dalam bentuk linguistik.

Tabel 2. Kategori Intensitas Hujan

Kategori Intensitas Hujan	Mm/Jam
Tidak Hujan	< 1
Hujan Ringan	1 – 5
Hujan Sedang	5 – 10
Hujan Lebat	– 20
Hujan Sangat Lebat	>20

Berikut ini adalah sampel Dataset setelah preprocessing.

Tabel 3. Sampel Dataset

ShMn	ShMx	ShRt	Klbn	Sunshine	Intsts
20,7	26,8	22,8	87	1,5	Ringan
20,	28,5	23,4	86	3,5	Sedang
5	21	23,7	79	5	Tdk Hujan
20,8	28	24	85	2	Ringan
19	27,2	24,1	84	3,4	Sedang

**Contoh Perhitungan**

Penghitungan nilai gain bertujuan untuk menentukan root node pada sebuah tree yang akan dibuat. Nilai root node didapat dari nilai gain tertinggi. Nilai gain didapat setelah menghitung impurity tiap value pada atribut. Dalam penghitungan ini, rumus yang digunakan dalam mencari nilai impurity menggunakan rumus entropy. Nilai gain terdiri dari atribut SUHU MIN (SMI), SUHU MAKS (SMA), SUHU RATARATA (SR), KELEMBABAN (K), dan LAMA PENYINARAN (LP).

Berikut ini adalah contoh pencarian root node pada penghitungan atribut Suhu Minimum.

Pada dataset terdapat total 1790 record berupa 1110 Tidak Hujan (TH), 225 Ringan (R), 140 Sedang (S), 139 Lebat (L), dan 176 Sangat Lebat (SL). Sehingga proses penghitungan Entropy(S) adalah:

$$\begin{aligned} \text{Entropy}(S) &= - (1110/1790)\log_2(1110/1790) \\ &\quad - (225/1790)\log_2(225/1790) \\ &\quad - (140/1790)\log_2(140/1790) - \\ &\quad (139/1790)\log_2(139/1790) \\ &\quad - (176/1790)\log_2(176/1790) \\ &= \underline{1.70645} \end{aligned}$$

Setelah entropy sudah didapat, selanjutnya mencari nilai gain tertinggi tiap atribut. Pada contoh ini dihitung hanya pada atribut Suhu Minimal. Berikut adalah langkah-langkah perhitungan gain untuk suhu minimal.

a. Pengurutan Data dan pencarian transisi output

20.1 203 20.4 20.5  
TH | S

b. Perhitungan gain untuk tiap titik transisi yang ditemukan yaitu pada titik :

$$\begin{aligned} < 20.45 [354,10,6,5,4] \text{ dan} \\ > 20.45 [756,215,134,134,172] \\ e[354,10,6,5,4] &= - (354/379).\log_2(354/379) \\ &\quad - (10/379).\log_2(10/379) \\ &\quad - (6/379).\log_2(6/379) - \\ &\quad (5/379).\log_2(5/379) \\ &\quad - (4/379).\log_2(4/379) \\ &= \underline{0.47668} \end{aligned}$$

$$\begin{aligned} e[756,215,134,134,172] \\ = - (756/1411).\log_2(756/1411) - \\ (215/1411).\log_2(215/1411) \\ 172] - (134/1411).\log_2(134/1411) - \end{aligned}$$

$$\begin{aligned} (134/1411).\log_2(134/1411) \\ - (172/1411).\log_2(172/1411) \\ = \underline{1.91116} \end{aligned}$$

$$\begin{aligned} \text{Gain } 20.45 &= \text{Entropy}(S) - \\ (379/1790)e[354,10,6,5,4] \\ - (1411/1790) \\ e[756,215,134,134,172] \\ &= 1.70645 - (379/1790)0.48 - \\ (1411/1790)1.91 \\ &= \underline{0.09916} \end{aligned}$$

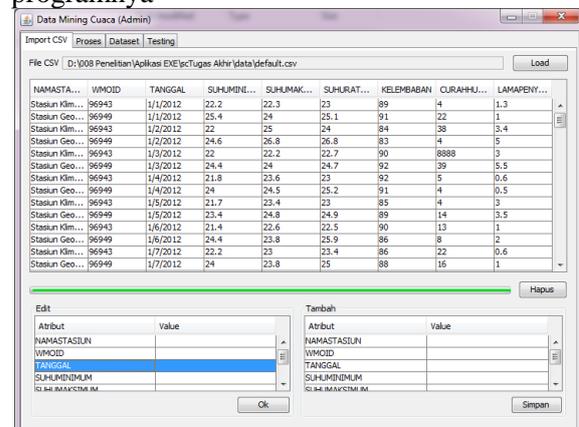
Nilai 0.09916 adalah nilai gain pada titik 20.5. Nilai gain ini masih harus dibandingkan dengan nilai dari semua titik pada atribut SUHU MINIMAL. Nilai gain terbesar dari pilihan gain pada titik-titik split yang ada yang akan mewakili nilai gain atribut SUHU MINIMAL.

Perhitungan ini dilakukan pada semua atribut yang ada. Dan nilai gain terbesar dari seluruh atribut akan menjadi root node (Node pada level 1).

Untuk pencarian node pada level di bawah root node menggunakan cara yang sama. Dengan catatan nilai impurity menggunakan value dari node parent-nya atau node pada level di atasnya.

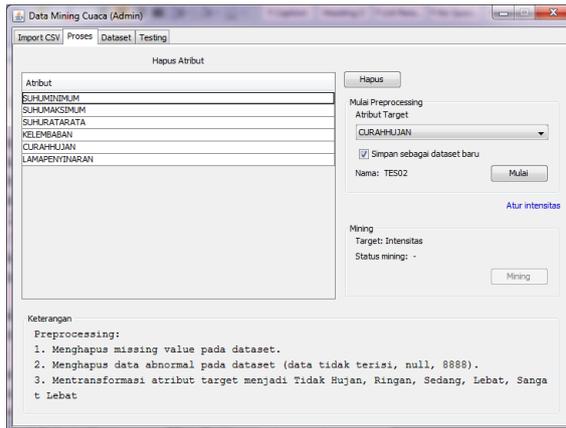
**HASIL DAN PEMBAHASAN**

Setelah dilakukan proses preprocessing maka selanjutnya dilakukan proses mining. Program yang dibangun menggunakan bahasa pemrograman Java dengan JDK (Java Development Kit) Versi 8 dengan editor NetBean 8. Berikut ini adalah hasil programnya



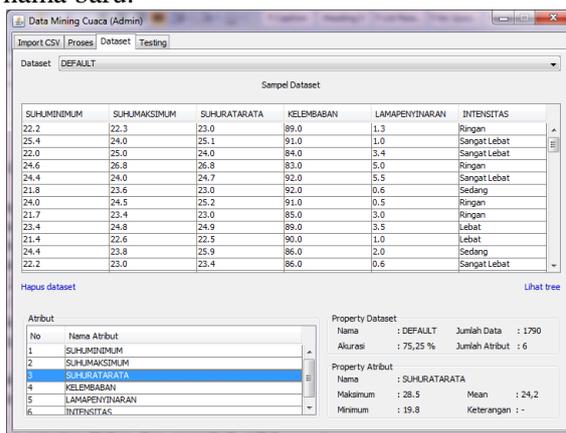
Gambar 12. Form Import data dari CSV

Form ini digunakan untuk mengambil dataset dari format CSV.



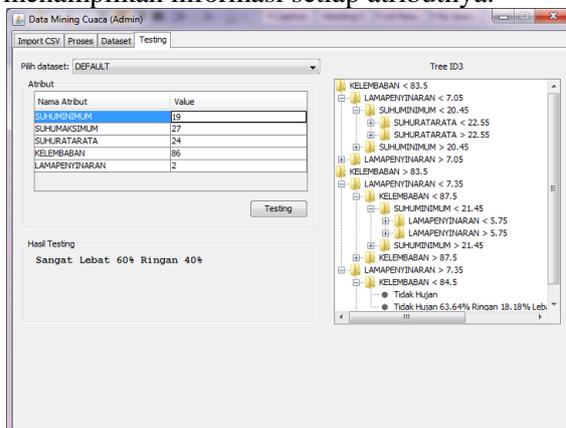
Gambar 13. Form Preprocessing dan Mining

Form ini digunakan untuk melakukan ppreprocessing yang meliputi penghapusan missing value, penghapusan atribut, perubahan value curah hujan ke bentuk linguistik dan penyimpanan dataset dengan nama baru.



Gambar 14. Form Dataset

Form digunakan untuk melihat hasil preprocessing dataset serta untuk menampilkan informasi setiap atributnya.



Gambar 15. Form Dataset

Form di atas digunakan untuk testing data baru dan untuk menampilkan decision tree yang dihasilkan.

Pada uji coba yang akan dilakukan, digunakan tabel yang berisidata yang didapat secara acak dari dataset karena mengingat data cuaca cenderung menghasilkan konklusi yang sama pada bulan atau musim tertentu seperti pada musim kemarau yang cenderung tidak hujan dan musim hujan yang cenderung hujan dengan intensitas sedang hingga sangat lebat. Pengujian dilakukan dengan mengambil 10% data dari dataset secara acak. Sehingga data uji yang digunakan adalah:  $(1790 / 10 = 179)$ .

Setelah melakukan pengujian dengan memasukkan 179 data pengujian pada program, diketahui terdapat 132 data yang sesuai dengan output. Maka dapat disimpulkan bahwa akurasi adalah:

$$\text{Akurasi} = (\text{total benar} / \text{jumlah data}) * 100\% = (132 / 179) * 100\% = 73.74\%$$

### KESIMPULAN

Berdasarkan uraian yang telah dipaparkan sebelumnya, dapat disimpulkan bahwa: Hasil training data menggunakan algoritma Iterative Dichotomiser 3 dapat membangun tree dengan lima level dan 30 rule. Pada proses testing yang dilakukan dengan 179 data (10% daridataset) yang diperoleh secara acak, diketahui bahwa terdapat 132 data yang sesuai dengan data cuaca sebenarnya. Maka dapat disimpulkan bahwa akurasi sebesar 73,74%. Atribut Kelembaban memiliki pengaruh paling dominan dalam menentukan prediksi cuaca. Kelas tidak hujan cenderung muncul lebih banyak daripada kelas lain seperti Hujan Ringan, Hujan Sedang, Hujan Lebat, dan Hujan Sangat Lebat.

Dan untuk penelitian lanjutan kami sarankan untuk melakukan beberapa hal berikut ini: Untuk menjaga dan memelihara keakuratan data, maka perlu dilakukan proses penambahan data dan dilakukan proses mining ulang secara berkala dengan men-training banyak data baru. Dengan bertambahnya data training, maka perlu dilakukan proses pruning pada tree, sehingga

ukuran tree tidak terlalu besar, overfitting, dan dapat menambah keakuratan. Aplikasi sistem dapat dikembangkan dengan menambah atribut baru agar hasil yang diperoleh lebih akurat, seperti kecepatan dan arah angin, tekanan udara, penguapan, dan parameter meteorologi lainnya.

#### REFERENSI

- [1] Bhukya, D. P. dan Ramachandram, S. Decision Tree Induction: An Approach for Data Classification Using AVL-Tree, International Journal of Computer and Electrical Engineering, Vol. 2, No. 4, August, 2010 1793-8163
- [2] Fayyad, Usama. Advances in Knowledge Discovery and Data Mining. California: AAAI Press, 1996
- [3] Kusumadewi, Sri. Artificial Intelligence (Teknik dan Aplikasinya). Yogyakarta: Graha Ilmu, 2003
- [4] Larose, D. T., Discovering Knowledge in Data: An Introduction to Data Mining. Hoboken: Wiley Interscience, John Wiley & Sons, Inc, 2005
- [5] Liang, Y. Daniel. Introduction to Java Programming, Comprehensive (Eight Edition). New Jersey: Prentice Hall, 2010
- [6] Luthfi, E. T., dan Kusriani, Algoritma Data Mining. Yogyakarta: Andi Publisher, 2009
- [7] Nettleton, David. Commercial Data Mining: Processing, Analysis and Modeling for Predictive Analytics Projects. Pennsylvania: Elsevier Science, 2014
- [8] Rajaraman, A, Leskovec, J., dan Ullman, J. D., Mining of Massive Datasets. California: Stanford University, 2014
- [9] Slocum, M. Decision Making Using ID3 Algorithm, Rivier Academic Journal, Vol. 8, No. 2, 2012
- [10] Susanto, S, dan Suryadi, D., Pengantar Data Mining "Menggali Pengetahuan dari Bongkahan Data". Yogyakarta: Andi Publisher, 2010
- [11] Tjasyono, Bayong. Klimatologi. Bandung: Penerbit ITB, 2004