

Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir

Musy Aulia Hasanah^{1*}, Sopian Soim^{2**}, Ade Silvia Handayani^{3*}

* Program Studi D4 Teknik Telekomunikasi, Jurusan Teknik Elektro
Politeknik Negeri Sriwijaya

msy.auliahasanah1234@gmail.com¹, sopiansoim@gmail.com², ade_silvia@polsri.ac.id³

Article Info

Article history:

Received 2021-07-26

Revised 2021-08-31

Accepted 2021-09-01

Keyword:

Confusion Matrix,
CRISP-DM,
Data Mining,
Decision Tree,
Classification Algorithms.

ABSTRACT

Indonesia is part of a tropical climate with high rainfall intensity. High rainfall intensity can potentially cause flooding. To minimize this, accurate weather predictions are needed to be able to anticipate beforehand. This research was conducted with the aim of classifying based on the rain category with the dichotomy of heavy rain and very heavy rain using data mining techniques with the CRISP-DM methodology. The algorithm used in the classification technique is CART (*Classification And Regression Tree*) with *Confusion Matrix* test parameters. Based on the results of the model evaluation, it shows that the CART algorithm has a fairly good performance in classifying with an accuracy value of 89.4%.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Banjir salah satu bencana yang menjadi masalah bagi sebagian masyarakat, terutama yang tinggal didaerah dataran rendah maupun dibantaran sungai. Fenomena banjir ini sangatlah berpengaruh, baik itu dari segi ekonomi, lingkungan maupun keselamatan masyarakat. Faktor penyebab banjir yang paling utama ialah curah hujan. Tingginya intensitas curah hujan dapat mempengaruhi jumlah volume debit air yang mengalir pada saluran sungai yang melebihi kapasitas alirannya [1]. Sehingga aliran sungai tersebut meluap dan menggenangi dataran yang rendah disekitaran dataran banjir [2]. Menurut pendapat [3], banjir diakibatkan oleh ketidaksetaraan antara aliran masuk intensitas hujan yang lebih tinggi dari pada aliran keluar, terutama bila drainase saluran air dan daerah resapan tidak berjalan dengan lancar.

Mengingat curah hujan salah satu faktor dinamis sebagai penyebab utama banjir, dibutukannya teknologi dan informasi untuk mengelola data. Data tersebut akan kita olah menjadi pengetahuan sebagai acuan dalam membaca serta mengetahui pola pendekatan tersembunyi dari kumpulan data, melakukan analisis tentang pengelompokan antara data dan atribut untuk mendukung pengambilan keputusan serta pembuatan kebijakan dalam memberikan informasi curah hujan yang

berpengaruh terhadap segala macam aktifitas seperti keselamatan masyarakat dan sosial-ekonomi. Data ini diolah menjadi sebuah pengetahuan agar dapat bermanfaat bagi banyak orang, dengan mengubah menjadi pengetahuan, manusia dapat melakukan prediksi dan estimasi tentang apa yang akan terjadi kedepan. Maka dari itu, perlu adanya proses yang menggunakan teknik statistik, matematik, kecerdasan buatan (*Artificial Inteligent*) dan *Machine Learning* untuk mengestrak pengetahuan atau menemukan pola dari suatu data yang besar [4].

Data Mining suatu proses mencari pola atau informasi dalam kumpulan data yang terpilih dengan menggunakan teknik atau metode tertentu. Data mining terbagi menjadi 5 bagian menurut peran utamanya yaitu estimasi, prediksi, klasifikasi, clustering, dan asosiasi [5]. Teknik pengolahan data mining yang sering digunakan yaitu klasifikasi. Klasifikasi yaitu proses memetakan data kedalam kelompok atau kelas yang telah Ditetukan [6]. Pada makalah ini, pengklasifikasian data curah hujan dibagi menjadi dua kategori yaitu, hujan lebat dan hujan sangat lebat. Dikarenakan intensitas curah hujan tersebut sangat intens terhadap semua potensi bencana yang terjadi. Proses klasifikasi ini dapat dilakukan berdasarkan kriteria tertentu. Kriteria intensitas curah hujan di verifikasi berdasarkan nilai ambang batas (*threshold*) BMKG (*Badan Meteorologi*

Klimatologi dan Geofisika). Pendekatan data mining yang dapat diterapkan dalam melakukan tahapan penelitian ini yaitu CRISP-DM. Banyak model prosedural dan upaya untuk menstandarisasi proses penambangan data yang telah dilakukan, salah satunya pendekatan CRISP-DM [7].

CRISP-DM (*Cross Industry Standard Process for Data Mining*) suatu standarisasi pemrosesan *data mining* yang telah dikembangkan dimana data yang ada akan melewati setiap fase terstruktur dan terdefinisi dengan jelas dan efisien [4]. Selain menerapkan suatu model dalam proses penambangan data, pemilihan algoritma sangat mempengaruhi terhadap komparasi kinerja metode data mining.

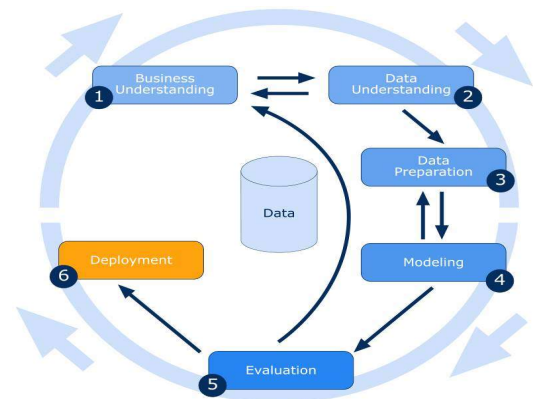
Penelitian mengenai algoritma *data mining* telah banyak digunakan, contoh pada penelitian Aditya [8] melakukan prediksi curah hujan dengan menerapkan metode data mining *Forecasting* dengan menggunakan algoritma *Artificial Neural Network* dengan model sistem *backpropagation*. Model ini mendapatkan nilai terbaik yang menghasilkan nilai MSE sebesar 163.28 dan SMAPE sebesar 67.4%. Penelitian telah dilakukan Prasetya [9] menggunakan metode klasifikasi dengan algoritma *Classification Tree*. Algoritma tersebut digunakan untuk prediksi hujan dengan parameter akurasi *Confusion Matrix*. Pemilihan dan pengujian model menunjukkan bahwa tingkat akurasi algoritma *Classification Tree* yaitu sebesar 74.7% dengan kategori *fair classification* dimana jumlah prediksi benar sebanyak 818 dataset dari total jumlah data uji yaitu 1095 dataset.

Penelitian tentang implementasi data mining untuk menentukan potensi hujan dalam harian juga telah dilakukan, Ahmad Subhan [10] dengan menggunakan algoritma *Naïve Bayes* pada analisis data cuaca untuk menentukan cuaca yang berpeluang hujan atau tidak hujan. Berdasarkan hasil pengujian menunjukkan bahwa klasifikasi dengan menggunakan *naïve bayes* mendapatkan nilai akurasi sebesar 82,5%, *recall* sebesar 82,65% dan *presisi* sebesar 80%. Reza dkk [11] melakukan teknik penambangan data dengan metode *K-NN (K-Nearest Neighbor)* dalam pengklasifikasian indeks cuaca kebakaran berlandaskan kelas *FWI (Fire Weather Index)* penelitian ini menggunakan beberapa parameter cuaca seperti suhu, kelembapan, kecepatan angin dan curah hujan. Dilakukannya pengujian dengan menghitung jarak data training terhadap data testing dengan perolehan keberhasilan sebesar 80,16% dimana nilai $K=5$.

Banyak algoritma yang diterapkan dalam penelitian sebelumnya namun dalam makalah ini akan mengimplementasikan salah satu algoritma klasifikasi yang sering digunakan untuk mencari suatu pola tertentu. *Decision Tree* salah satu teknik yang mudah diinterpretasikan dan divisualisasikan. Ada beberapa macam algoritma *Decision Tree* yaitu: CART, C4.5, ID3 dan ada banyak algoritma lainnya. Maka dari itu pada makalah ini akan menggunakan algoritma CART (*Classification And Regression Tree*) dengan harapan algoritma CART ini mampu menghasilkan suatu pola pengetahuan dengan mengklasifikasi curah hujan dengan intensitas tinggi sehingga pola yang dihasilkan tersebut memberikan informasi tentang parameter-parameter apa saja yang mempengaruhinya serta untuk mengetahui kinerja dari algoritma ini.

II. METODE PENELITIAN

Pada penelitian ini, digunakannya metodologi data mining CRISP-DM sebagai pemecah masalah yang umum untuk bisnis dan penelitian. Metodologi ini terdiri dari enam tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Proses metodologi ini terdiri dari 6 tahapan yang dapat dijelaskan sebagai berikut.



Gambar 1. Metodologi CRISP-DM

1) *Business Understanding* (Pemahaman Bisnis)

Beberapa hal yang dilakukan pada tahap ini seperti memahami kebutuhan serta tujuan dari sudut pandang bisnis selanjutnya mengartikan pengetahuan ke dalam bentuk pendefinisian masalah pada data mining dan kemudian menentukan rencana serta strategi untuk mencapai tujuan data mining.

2) *Data Understanding* (Pemahaman Data)

Tahapan ini diawali dengan mengumpulkan data, mendeskripsikan data, serta mengevaluasi kualitas data.

3) *Data Preparation* (Persiapan Data)

Dalam tahapan ini yaitu membangun dataset akhir dari berupa data mentah. Ada beberapa hal yang akan dilakukan mencakup melakukan pembersihan data (*Data Cleaning*), melakukan pemilihan data (*Data Selection*), record dan atribut-atribut, dan juga melakukan transformasi terhadap data (*Data Transformation*) untuk dijadikan masukan dalam tahap pemodelan.

4) *Modelling* (Pemodelan)

Pada tahapan ini secara langsung melibatkan *Machine Learning* untuk penentuan teknik *data mining*, alat bantu *data mining* serta algoritma data mining. Penelitian ini menggunakan model klasifikasi *Decision Tree* dengan algoritma yang dipilih CART (*Classification And Regression Tree*). Cara algoritma CART untuk membentuk pohon keputusan dengan kriteria pemecah *Gain Information* [12]:

- Menentukan Akar/Root

- Menghitung *Gain information* masing-masing calon cabang
- Memilih atribut dengan *Gain information* yang mempunyai nilai terbesar sebagai splitting atribut yang akan dipilih sebagai cabang.
- Ulangi langkah 2,3 hingga terdapat *leaf node*

Gain information mengukur nilai impurity dari suatu partisi, A, dengan perhitungan;

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2$$

Dimana:

- Gini(D) = nilai impurity dari partisi D
- m = jumlah indeks
- p_i = peluang sebuah *tuple* D pada indeks ke i

Average Gini Impurity dapat dihitung dengan.

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2).$$

Dimana:

- D = *tuple* D
- D₁ = partisi pertama *tuple* D
- D₂ = partisi kedua *tuple* D
- Gini_A(D) = impurity dari partisi D pada atribut A
- Gini(D₁) = impurity dari partisi pertama *tuple* D
- Gini(D₂) = impurity dari partisi kedua *tuple* D
- Penurunan tingkat impurity yang diperoleh terhadap atribut A, bisa dihitung dengan :

$$\Delta Gini(A) = Gini(D) - Gini_A(D)$$

Dimana:

- ΔGini(A) = tingkat impurity

5) *Evaluation* (Pengujian)

Tahap ini dilakukan dengan melihat tingkat performa dari pola yang dihasilkan oleh algoritma. Parameter yang digunakan untuk evaluasi komparasi algoritma adalah *Confusion Matrix* dengan aturan nilai akurasi, presisi dan *recall*. Nilai tersebut dapat diperoleh melalui perhitungan:

$$Accuracy = \frac{\text{Jumlah klasifikasi benar}}{\text{Total sampel testing yang diuji}} \times 100\%$$

$$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \times 100\%$$

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \times 100\%$$

6) *Deployment* (Penyebaran)

Tahapan ini dilakukan dengan pembuatan laporan dan artikel jurnal menggunakan model yang dihasilkan.

III. HASIL DAN PEMBAHASAN

A. *Business Understanding*

Penerapan *data mining* pada penelitian ini berhubungan langsung dengan data Curah hujan untuk menggali pengetahuan tentang suatu pola terhadap intensitas curah hujan yang berpotensi terhadap bencana banjir. Serta untuk melihat parameter-parameter apa saja yang mempengaruhinya terhadap tingginya intensitas curah hujan tersebut

B. *Data Understanding*

Penelitian ini menggunakan data yang berasal dari BMKG Stasiun Klimatologi Kelas 1 Palembang. Data tersebut merupakan data iklim dari periode 2011-2020. Parameter yang digunakan seperti suhu (T), kelembaban (RH), lama penyinaran matahari (ss) serta curah hujan. Verifikasi terhadap klasifikasi curah hujan yang berpotensi banjir dikategorikan hujan lebat dan hujan sangat lebat berdasarkan nilai ambang (*threshold*) BMKG.

TABEL I
KATEGORI HUJAN BMKG

No	Kategori Hujan	Intensitas Curah Hujan (mm/hari)
1	Tidak Hujan/Berawan	0
2	Hujan Ringan	0.5 – 19.9
3	Hujan Sedang	20 – 49.9
4	Hujan Lebat	50 - 100
5	Hujan Sangat Lebat	>100

TABEL II
DATASET BMKG

Tanggal	Tn	Tx	Tavg	RH	ss	RR
01-01-2011	24	32.4	27.5	83	3.2	0
02-01-2011	25	33.5	27.7	81	3.9	0
03-01-2011	23	30.3	26.6	84	0	0.5
04-01-2011	23	32.4	27	82	5.5	3.3
05-01-2011	22	29.9	24.7	93	0.3	26.2
06-01-2011	23	31.8	26	85	4	7
07-01-2011	23	31.5	25.9	88	1.5	4
08-01-2011	24	32.5	27.2	85	7	7.8
09-01-2011	23	29.4	25.6	92	1.5	0
10-01-2011	23	28.8	24.7	90	1.4	8.5
20-03-2011	22	33.4	27.2	86	6.8	129.9
13-01-2012	23	32.2	26.8	82	6	60.5
					
24-12-2020	22.2	30.4	26	91	5.4	45.2
25-12-2020	22	33.2	27.7	86	1.4	11.4
26-12-2020	22.8	33.2	26	87	5.4	9.9
27-12-2020	23.4	31.8	26.7	89	5.5	0.6
28-12-2020	23	29.2	24.8	94	3.1	60
29-12-2020	23	31.6	26.6	84	1.4	11.4
30-12-2020	23.4	32.4	27.3	83	6.1	11.4
31-12-2020	22.6	28.3	24.1	93	4.5	17.6

C. *Data Preparation*

Pada Tahap *Data Cleaning*, dataset yang terdiri dari 3.653 data hanya terdapat 123 *record* yang dikotomi dengan intensitas hujan lebat dan hujan sangat lebat. Sehingga data

yang tersisa harus dibersihkan. Data meteorologi yang digunakan memiliki 7 atribut dan terdapat beberapa *Missing Value* seperti data kosong, data tidak terukur, dan data yang tidak dilakukannya pengukuran. Hal tersebut diatasi dengan menghitung nilai pengganti (*Imputation*) seperti mengganti dengan nilai rata-rata

TABEL III
PROSES DATA CLEANING

Tanggal	Tn	Tx	Tavg	RH	ss	INTENSITAS
29-01-2011	23	30	25.2	93	3	LEBAT
10-02-2011	24	34.4	28	86	5.7	LEBAT
27-02-2011	24	29.8	25.5	94	0.9	LEBAT
20-03-2011	22	33.4	27.2	86	6.8	S. LEBAT
23-04-2011	23	32.2	26.9	88	2.5	S. LEBAT
27-05-2011	23	33.4	27.1	87	4.6	LEBAT
23-10-2011	24	32.8	26.9	88	4.7	LEBAT
12-12-2011	23	30.3	26	88	1	LEBAT
13-01-2012	23	32.2	26.8	82	6	LEBAT
15-02-2012	23	31.8	26.2	89	4.8	LEBAT
31-03-2012	23	32.5	27.6	82	4.7	LEBAT
11-04-2012	24	31	26.2	88	2.8	LEBAT
					
05-03-2020	23.2	31.7	27	88	4.1	LEBAT
01-04-2020	24.6	33	27.5	88	4.2	LEBAT
11-04-2020	24	31.6	27.1	90	2.8	LEBAT
09-05-2020	23.4	32.6	28.1	85	1.1	LEBAT
24-10-2020	22.2	31.8	26.4	89	4.4	LEBAT
25-11-2020	23.2	33.8	26	91	6.6	LEBAT
28-11-2020	23.8	33	27.5	82	6.3	LEBAT
28-12-2020	23	29.2	24.8	94	3.1	LEBAT

Pada Tahap *Data Selection* dilakukan analisis dan memilih setiap variabel data yang digunakan antara lain, suhu min harian, suhu max harian, suhu rata-rata harian, kelembaban rata-rata harian, dan lama penyinaran matahari dalam satu hari, serta variabel curah hujan harian yang akan dijadikan label/class/target dalam proses klasifikasi data mining. Data tersebut terdapat 123 dataset dengan 6 atribut, terdiri dari 5 atribut numerical (kontinyu) dan 1 atribut nominal. Hasil Pengelolaan data pada tahap ini terdapat pada Tabel II. Kemudian dataset tersebut disimpan ke dalam dataset baru menggunakan *Microsoft Excel*.

TABEL IV
ATRIBUT DATASET BMKG

Atribut	Tipe	Keterangan
<i>Year</i>	Numerical	<i>Year considered</i>
<i>Month</i>	Numerical	<i>Month considered</i>
Suhu Minimum (Tn)	Numerical	Suhu minimum dalam satu hari
Suhu Maksimum (Tx)	Numerical	Suhu maksimum dalam satu hari
Suhu rata-rata (Tavg)	Numerical	Suhu rata-rata dalam satu hari
Kelembaban rata-rata (RH avg)	Numerical	Kelembaban rata-rata dalam satu hari
Lama Penyinaran Matahari (ss)	Numerical	Lama penyinaran matahari dalam satu hari
Curah Hujan	Nominal/label	Jumlah curah hujan harian

Pada Tahap *Data Transformation* kumpulan data yang akan diubah ke bentuk data yang sesuai dalam proses data mining. Data tersebut dilakukan perubahan dari format xls

menjadi format csv, Dataset intensitas hujan lebat diberi label angka 0 (nol) sedangkan intensitas hujan sangat lebat diberi label angka 1 (satu). Setelah itu data tersebut disimpan, dan diimport ke tools data mining menggunakan python.

TABELV
DATASET TRANSFORMATION

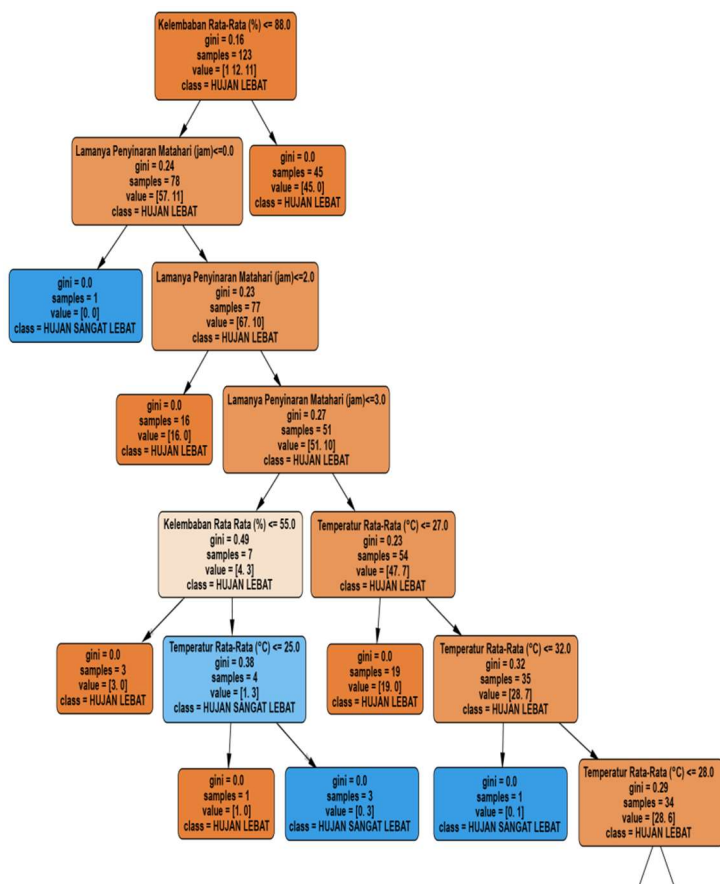
No	Tgl.	Tn (°C)	Tx (°C)	Tavg (°C)	RH_avg (%)	Lama Penyinaran Matahari (jam)	Intensitas Hujan
1	#####	23	30	25.2	93	3	0
2	#####	24	34.4	28	86	5.7	0
3	#####	24	29.8	25.5	94	0.9	0
4	#####	22	33.4	27.2	86	6.8	1
5	#####	23	32.2	26.9	88	2.5	1
6	#####	23	33.4	27.1	87	4.6	0
7	#####	24	32.8	26.9	88	4.7	0
8	#####	23	32.8	26	88	1	0
9	#####	23	32.2	26.8	82	6	0
10	#####	23	31.8	26.2	89	4.8	0
11	#####	23	32.5	27.6	82	4.7	0
12	#####	24	31	26.2	88	2.8	0
13	#####	24	31.5	26.4	88	3	0
14	#####	22	32.8	26.4	87	5.4	0
15	#####	23	34.1	27.8	85	6.3	0
16	#####	24	34.3	24.9	96	1	0
17	#####	24	33.5	27.7	80	3.3	0
18	#####	24	33.4	26.7	88	2.7	1
19	#####	23	32.5	26.6	90	3.7	0
20	#####	24	33	27.9	84	3.9	1
21	#####	24	32.7	27.5	88	4.2	0
22	#####	23	31.3	26.7	86	1.3	0
23	#####	23	33	27.1	83	6.1	0
24	#####	23	28.6	24.7	94	0.4	0

D. Modelling

Pada tahap ini, proses dimana model dengan teknik klasifikasi yang dipilih menghasilkan suatu pola informasi yang dapat memudahkan pihak yang berkepentingan. Pola klasifikasi yang dihasilkan teknik data mining ini digunakan untuk memprediksi intensitas curah hujan dengan predictor yang mempengaruhinya. Adapun tools atau alat bantu dalam proses pengolahan data mining yaitu *Jetbrains Pycharm Community Edition 2021.1.0.0* dengan dasar Python 3.7.2. Tahap ini menampilkan serta memberikan informasi terhadap kinerja algoritma dalam metode klasifikasi.

Algoritma CART (*Classification and Regression Tree*) suatu model klasifikasi menggunakan struktur pohon yang paling populer dikarenakan mudah untuk diterapkan. Algoritma ini bekerja dengan mengubah himpunan data menjadi suatu pohon keputusan dengan memberikan sebuah aturan-aturan pada keputusan sehingga proses pengambilan keputusan dengan berbagai atribut yang kompleks menjadi lebih mudah dan efisien. Susunan proses pembentukan pohon

keputusan terdiri dari *root node*, *internal node*, dan *leaf node*. Berikut hasil pengolahan data algoritma CART.



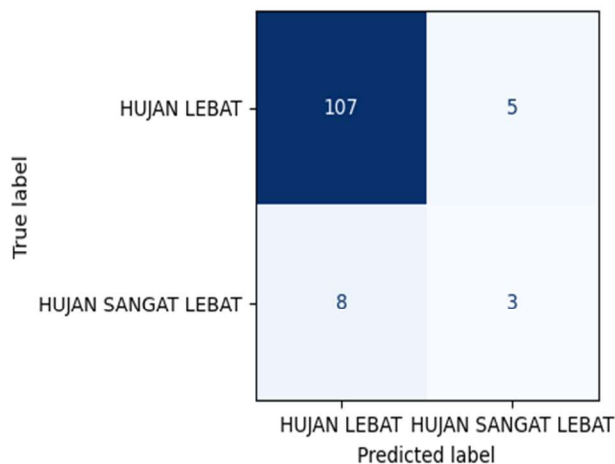
Gambar 2. Classification Tree dalam prediksi curah hujan.

Hasil ini merupakan visualisasi pengolahan data pada algoritma CART menghasilkan 17 nodes dan 18 leaf dengan tingkat kedalaman 13 level. Pada root node memiliki splitting kriteria pada atribut kelembaban rata-rata $\leq 88\%$ dengan kata lain disini memanfaatkan features kelembaban sebagai pemecah kondisi sebelum proses splitting. Berikutnya terdapat beberapa informasi seperti perolehan parameter gini impurity yaitu sebesar 0.16 dengan sampel (jumlah data) 112 data untuk kelas 0 (hujan lebat) dan 11 data untuk kelas 1 (hujan sangat lebat)

Setelah dikenakan proses splitting, dari kriteria yang ada pada root node menghasilkan 2 ruas yaitu 1 internal node dan 1 leaf node. Pada leaf node (ruas kanan) menghasilkan sebuah keputusan bila kelembaban rata-rata $> 88\%$ maka intensitas curah hujan berpeluang dengan kriteria hujan lebat. Kemudian pada internal node (ruas kanan) ini perlu dilakukannya proses splitting lebih lanjut pada atribut lama penyinaran matahari dikarenakan masih terdapat nilai gini impuritynya sebesar 0.24 dengan jumlah data untuk kelas indeks ke-0 (hujan lebat) sejumlah 67 data dan untuk kelas indeks ke-1 (hujan sangat lebat) sebanyak 11 data. Tahap ini terus berlanjut, jika pada akhirnya proses splitting menghasilkan 2 leaf node dengan kata lain leaf ini bisa diambil sebuah keputusan.

E. Evaluation

Setelah pola klasifikasi didapatkan pada algoritma CART selanjutnya dilakukan tahap evaluasi komparasi algoritma dengan parameter yang digunakan ialah *Confusion Matrix* yang pada dasarnya untuk memberikan informasi perbandingan hasil yang telah dilakukan oleh model dengan hasil klasifikasi sebenarnya dengan melihat nilai akurasi, presisi dan recall. Berikut hasil evaluasi pada model yang telah dibuat dengan algoritma ini.



Gambar 3. Confusion Matrix pada algoritma CART

Berdasarkan pada gambar 3, hasil nilai *Accuracy* = 0.89, *Precision* = 0.37 dan *recall* = 0.27. Sehingga didapatkan tingkat akurasi algoritma CART sebesar 89,4% dengan jumlah prediksi benar 110 data dari total data testing 123 dataset.

$$Accuracy = \frac{110}{123} \times 100\% = 89,4\%$$

F. Deployment

Setelah tahap evaluasi dimana menilai secara detail hasil dari sebuah model maka dilakukan pengimplementasian dari keseluruhan model yang telah dibangun. Selain itu juga dilakukan penyesuaian terhadap model sehingga dapat menghasilkan suatu hasil yang sesuai dengan target awal tahap CRISP-DM ini.

V. KESIMPULAN

Pada penelitian ini, telah dilakukan klasifikasi data curah hujan menggunakan metode decision tree algoritma CART (*Classification And Regression Tree*) dengan teknik data mining CRISP-DM. Dataset terdiri 6 atribut yaitu suhu rata-rata, suhu min, suhu max, kelembaban, lama penyinaran matahari dan curah hujan. Dari 3.653 dataset curah hujan hanya 123 record yang dikotomi curah hujan yang berintensitas hujan lebat dan hujan sangat lebat namun dari hasil pengujian algoritma ini memiliki kinerja yang cukup baik

dengan akurasi sebesar 89,4% dengan Evaluasi dan validasi menggunakan parameter uji *Confusion Matrix*, dari perolehan akurasi tersebut didapatkan bahwa jumlah prediksi benar adalah 110 data dari jumlah total data uji yaitu 123 data.

DAFTAR PUSTAKA

- [1] A. K. Neno, H. Harijanto, and A. Wahid., "Hubungan Debit Air dan Tinggi Muka Air di Sungai Lambagu Kecamatan Tawaeli Kota Palu," *War. Rimba*, vol. 4, no. 2, pp. 1–8, 2016.
- [2] S. P. Nugroho, "Evaluasi dan analisis curah hujan sebagai faktor penyebab bencana banjir jakarta (in Bahasa)," *J. Sains Teknol. Modif. Cuaca*, vol. 3, no. 2, pp. 91–97, 2002.
- [3] B. K. Tjasyono, I. Juaeni, and W. B. Harijono, "Proses Meteorologis Bencana Banjir," *J. Mkg*, vol. 8, no. 2, pp. 64–78, 2007.
- [4] A. Вульфин and A. Фрид, "Нейросетевая модель анализа технологических временных рядов в рамках методологии Data Mining," *Информационно-Управляющие Системы*, no. 5, 2011.
- [5] B. P. T.P and R. D. Indah Sari, "Penerapan Data Mining Untuk Prakiraan Cuaca Di Kota Malang Menggunakan Algoritma Iterative Dichotomiser Tree (Id3)," *Jouticla*, vol. 2, no. 2, pp. 101–108, 2017, doi: 10.30736/jti.v2i2.68.
- [6] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [7] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model," *Procedia CIRP*, vol. 79, pp. 403–408, 2019, doi: 10.1016/j.procir.2019.02.106.
- [8] D. S. Informasi, *Artificial Neural Malang Ann Method Implementation To Predict Rainfall in Case of Dengue Fever Anticipation in Malang District*. 2018.
- [9] R. Prasetya, "Penerapan Teknik Data Mining Dengan Algoritma," vol. 2, no. 2, 2020.
- [10] Sugiyono, "Dokumen Karya Ilmiah | Skripsi | Prodi Teknik Informatika - S1 | FIK | UDINUS | 2016," *Fik*, vol. 1, no. 1, pp. 1–2, 2016.
- [11] J. Coding and S. K. Untan, "Kata Kunci: Kebakaran Hutan, Data Mining, K-Nearest Neighbor (KNN), Fire Weather Index(FWI). 1.," vol. 06, no. 2, 2018.
- [12] J. Wijaya, "Implementasi algoritma pohon keputusan cart untuk menentukan klasifikasi data evaluasi mobil skripsi," 2019.