



UNIVERSITAS TEKNOLOGI SUMBAWA

**IMPLEMENTASI CRISP-DM MENGGUNAKAN METODE DECISION
TREE UNTUK PREDIKSI CURAH HUJAN BERPOTENSI BANJIR DI
PROVINSI NTB**

TUGAS KULIAH

Diajukan kepada Universitas Teknologi Sumbawa sebagai salah satu persyaratan
menyelesaikan Mata Kuliah Sains Data

OLEH:

WIDIYA LIS SUSANA

(211001002)

**FAKULTAS REKAYASA SISTEM
PROGRAM STUDI INFORMATIKA
SUMBAWA JANUARI 2022**

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Modul Ajar yang berjudul “Implementasi Crisp-Dm Menggunakan Metode Decision Tree Untuk Prediksi Curah Hujan Berpotensi Banjir di Provinsi NTB”. Laporan ini disusun sebagai syarat Ujian Akhir Semester Mata Kuliah Sains Data.

Dalam proses penyusunan Laporan ini, penulis mendapatkan bantuan dan motivasi dari berbagai pihak baik secara langsung maupun tidak langsung yang selama ini telah banyak membantu. Untuk itu pada kesempatan ini penulis mengucapkan terimakasih kepada :

1. Chairul Hudaya, Ph.D, selaku Rektor Universitas Teknologi Sumbawa.
2. Rodianto, S.Kom, M.Kom, selaku Kepala Program Studi Informatika Universitas Teknologi Sumbawa.
3. Herfandi, M. Kom selaku dosen pengampu Mata Kuliah Analisis Perancangan Sistem.
4. Teman – teman yang selalu memberikan bantuan, dorongan dan semangat kepada penulis untuk menyelesaikan Laporan ini. Semua pihak yang telah membantu dan memberikan dukungan dalam penyusunan Laporan ini yang tidak dapat disebutkan satu persatu.

Karena keterbatasan pengalaman dan wawasan penulis, Laporan ini masih jauh dari sempurna. Masukan saran dan kritik sangat penulis harapkan untuk peningkatan dalam proses penyusunan Laporan ini. Akhir kata penulis mengucapkan terima kasih.

Sumbawa Besar, Januari 2023

Penulis

BAB I

PENDAHULUAN

1.1 Latar Belakang

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 Intelegent*) 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 Ditetentukan [6]. Pada laporan 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 iniyaitu CRISP-DM. Banyak model prosedural dan upaya untk 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 *Confution 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 penambahan 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 laporan 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 laporan 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.

1.2 Tujuan

Untuk mengetahui intensitas hujan lebat dan sangat lebat yang dapat berpotensi banjir di Provinsi NTB menggunakan crisp-dm dan metode decision tree

BAB 11

LANDASAN TEORI

Landasan teori merupakan serangkain teori yang relevan yang dijadikan sebagai landasan dalam penelitian. Sehingga dalam bab ini akan dibahas hal-hal atau teori-teori yang berkaitan dengan permasalahan. Berikut adalah beberapa penelitian terdahulu yang membahas tentang implementasi crisp-dm menggunakan metode decision tree untuk prediksi curah hujan berpotensi banjir di kabupaten sumbawa.

Tabel 2.1 Tabel Penelitian Terkait

No	Nama Peneliti	Topik
1	A. K. Neno, H. Harijanto, and A. Wahid	Hubungan Debit Air dan Tinggi Muka Air di Sungai Lambagu Kecamatan Tawaeli Kota Palu,” <i>War. Rimba</i> , 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),” <i>J. Sains Teknol. Modif. Cuaca</i> , vol. 3, no. 2, pp. 91–97, 2002.
3	B. K. Tjasyono, I. Juaeni, and W. B. Harijono,	Proses Meteorologis Bencana Banjir,” <i>J. Mkg</i> , vol. 8, no. 2, pp. 64–78, 2007.
4	B. P. T.P and R. D. Indah Sari,	Penerapan Data Mining Untuk Prakiraan Cuaca Di Kota Malang Menggunakan Algoritma Iterative Dichotomiser Tree (Id3),” <i>Jouticla</i> , vol. 2, no. 2, pp. 101–108, 2017, doi: 10.30736/jti.v2i2.68.
5	P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno	Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5,” <i>Prism. Pros.Semin. Nas. Mat.</i> , vol. 3, pp. 64–71, 2020.
6	Msy Aulia Hasanah, Sopian Soim, dan Ade Silvia Handayani	Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir.

BAB III

METODOLOGI PENELITIAN

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

3.1 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 definisi masalah pada data mining dan kemudian menentukan rencana serta strategi untuk mencapai tujuan data mining.

3.2 Data Understanding (Pemahaman Data)

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

3.3 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.

3.4 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*).

3.5 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.

3.6 *Deployment* (Penyebaran)

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

BAB IV

HASIL dan PEMBAHASAN

Bab ini membahas mengenai perancangan proses analisis kebutuhan dan proses perancangan implementasi crisp-dm menggunakan metode decision tree untuk prediksi curah hujan berpotensi banjir di kabupaten sumbawa. Analisis kebutuhan terdiri atas enam langkah yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*.

4.1 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.

4.2 Data Understanding

Penelitian ini menggunakan data yang berasal dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) Sumbawa. 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 4.1 Kategori Hujan BMKG

No	Kategori Hujan	Intensitas Curah Hujan(mm/hari)
1	Hujan sangat ringan	< 5 mm
2	Hujan Ringan	5 – 20
3	Hujan Sedang	20 – 50
4	Hujan Lebat	50 - 100
5	Hujan Sangat Lebat	>100

Tabel 4.2 Gangguan Tropis yang Terjadi Selama Bulan Januari 2020

No	Tanggal	Nama	Posisi	Kec. Angin Max.		Tekanan Terendah (milibar)
				mph	Km/h	
1	23 Desember 2019 – 2 Januari 2020	Tropical CycloneCat. 1 SARAI	Samudra Pasifik Bagian Selatan	70	110	972
2	27 Desember 2019 – 1 Januari 2020	Tropical CycloneCat. 1 CALVINIA	Samudra Hindia bagian Barat Daya	75	120	973
3	4 Januari – 11 Januari	Tropical Storm BLAKE	Perairan Australia bagian Barat Laut	45	75	986
4	4 Januari – 17 Januari	Tropical CycloneCat. 1 CLAUDIA	Perairan Australia bagian Barat Laut	85	140	969
5	11 Januari – 20 Januari	Tropical CycloneCat. 1 TINO	Samudra Pasifik Bagian Selatan	75	120	970
6	22 Januari – 26 Januari	Tropical Storm DIANE	Samudra Hindia bagian Barat Daya	60	95	980
7	23 Januari – 26 Januari	Tropical Storm ESAMI	Samudra Hindia bagian Barat Daya	45	75	993

4.3 Tabel Terbit dan Tenggelam Matahari Kota Mataram Bulan Februari 2020

MATARAM								
Location: E116°06'00.9", S 8°34'30.2", 15M								
(Longitude referred to Greenwich meridian)								
Time Zone: 8h 00m east of Greenwich								
Tanggal				Fajar	Terbit	Transit	Terbenam	Senja
				h m	h m	h m	h m	h m
				(WITA)	(WITA)	(WITA)	(WITA)	(WITA)
2020	Feb	1	(Sat)	05.27	06.15	12.29	18.43	19.31
2020	Feb	2	(Sun)	05.27	06.15	12.29	18.43	19.31
2020	Feb	3	(Mon)	05.28	06.15	12.29	18.43	19.31
2020	Feb	4	(Tue)	05.28	06.16	12.29	18.43	19.30
2020	Feb	5	(Wed)	05.29	06.16	12.30	18.43	19.30
2020	Feb	6	(Thu)	05.29	06.16	12.30	18.43	19.30
2020	Feb	7	(Fri)	05.29	06.17	12.30	18.43	19.30
2020	Feb	8	(Sat)	05.30	06.17	12.30	18.43	19.30
2020	Feb	9	(Sun)	05.30	06.17	12.30	18.42	19.29
2020	Feb	10	(Mon)	05.30	06.17	12.30	18.42	19.29
2020	Feb	11	(Tue)	05.31	06.18	12.30	18.42	19.29
2020	Feb	12	(Wed)	05.31	06.18	12.30	18.42	19.29
2020	Feb	13	(Thu)	05.31	06.18	12.30	18.42	19.28
2020	Feb	14	(Fri)	05.31	06.18	12.30	18.41	19.28
2020	Feb	15	(Sat)	05.32	06.18	12.30	18.41	19.28
2020	Feb	16	(Sun)	05.32	06.19	12.30	18.41	19.27
2020	Feb	17	(Mon)	05.32	06.19	12.30	18.40	19.27
2020	Feb	18	(Tue)	05.32	06.19	12.30	18.40	19.27
2020	Feb	19	(Wed)	05.33	06.19	12.29	18.40	19.26
2020	Feb	20	(Thu)	05.33	06.19	12.29	18.40	19.26
2020	Feb	21	(Fri)	05.33	06.19	12.29	18.39	19.25
2020	Feb	22	(Sat)	05.33	06.19	12.29	18.39	19.25
2020	Feb	23	(Sun)	05.33	06.19	12.29	18.39	19.25
2020	Feb	24	(Mon)	05.34	06.20	12.29	18.38	19.24
2020	Feb	25	(Tue)	05.34	06.20	12.29	18.38	19.24
2020	Feb	26	(Wed)	05.34	06.20	12.29	18.37	19.23
2020	Feb	27	(Thu)	05.34	06.20	12.28	18.37	19.23
2020	Feb	28	(Fri)	05.34	06.20	12.28	18.37	19.22
2020	Feb	29	(Sat)	05.34	06.20	12.28	18.36	19.22

BAB 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 REFRENSI

- [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.