# Klasifikasi Curah Hujan Berpotensi Banjir Menggunakan Algoritma Decision Tree C4.5 Dan C5.0

Jason Suhali[1], Maria Darlene Kusnadi[2], Fardhila Zahra Dwi Wardhani[3], Michael Abhinaya Bagioyuwono[4]
[1]*Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara,* Tangerang, Indonesia
jason.suhali@student.umn.ac.id
[2]*Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara,* Tangerang, Indonesia
maria.kusnadi@student.umn.ac.id
[3]*Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara,* Tangerang, Indonesia
fardhila.wardhani@student.umn.ac.id
[4]*Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara,* Tangerang, Indonesia
michael.bagioyuwono@student.umn.ac.id

***Abstract*—Curah hujan menjadi salah satu faktor dinamis sebagai penyebab utama banjir, dengan ini dibutuhkan teknologi dan informasi untuk mengolah data mengenai curah hujan. Tujuan dari penelitian ini adalah untuk dapat membuat model yang bisa memprediksi curah hujan yang terjadi dan dapat mengetahui tingkat intensitas curah hujan yang dapat menyebabkan banjir. Pendekatan *data mining* yang diterapkan dalam tahapan penelitian ini adalah *CRISP-DM (Cross-Industry Standard Process Model for Data Mining)*. Penelitian ini dilakukan dengan menggunakan algoritma Decision Tree untuk menemukan *Decision Tree* yang paling tepat yaitu algoritma C4.5 dan C5.0. Pada penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma *decision tree* C4.5 dan C5.0 dapat diterapkan pada penelitian ini. Algoritma C4.5 memiliki akurasi sebesar 90,1% dan algoritma C5.0 memiliki akurasi sebesar 95,2%.**

***Keywords—Curah Hujan, Banjir, Data Mining, CRISP-DM, Decision Tree, Algoritma C4.5 dan C5.0***

## I. PENDAHULUAN

### *A.* Latar Belakang

Bencana alam merupakan fenomena alam yang bersifat merusak tatanan lingkungan disekitarnya. Berdasarkan UU No. 24 Tahun 2007, bencana alam merupakan sebuah rangkaian kejadian yang mengganggu dan mengancam penghidupan dan kehidupan masyarakat sekitar yang disebabkan oleh faktor alam, non alam, atau faktor manusia yang menelan korban jiwa manusia, rusaknya lingkungan, kehilangan harta benda, dan dampak pada psikologis. Salah satu negara yang paling sering dilanda bencana alam adalah Indonesia. Pada tahun 2018, World Risk Report mengatakan bahwa Indonesia menempati urutan ke 36 dengan indeks risiko 10,36 dari 172 negara paling rawan bencana alam. Secara hidro klimatologis, Indonesia memiliki fenomena ENSO (*El-Nino Southern Oscillation*) dan La Nina. Hal tersebut menyebabkan Indonesia menjadi sering dilanda bencana tanah longsor, angin puting beliung, dan banjir [1].

Banjir merupakan fenomena alam yang dapat terjadi kapan saja, terutama pada saat intensitas curah hujan sedang tinggi. Indonesia menjadi salah satu wilayah yang sering dilanda banjir. Salah satu penyebab terjadinya banjir adalah karena perubahan iklim yang semakin ekstrim di Indonesia.Iklim merupakan salah satu faktor utama dari penyebab banjir di indonesia. Iklim sendiri dapat diartikan Sebagai Kondisi rata-rata suhu udara, curah hujan tekanan udara, arah angin, kelembapan dan parameter Lainnya dalam kurun waktu yang lama (Tjasyono, 2004). Indonesia sendiri memiliki dua unsur iklim utama yaitu suhu dan curah hujan, hal inilah yang membuat indonesia memiliki iklim tropis. Iklim Tropis Indonesia sendiri memiliki beberapa fenomena yang dapat menyebabkan lebih banyak curah hujan seperti fenomena *monsoon Tropical Convergence Zone, El nino southern oscillation Madden-julian Oscillation, Tropical Cyclone, Indian Ocean Dipole Mode (IODM)*. Hal ini disebabkan karena perubahan pola curah hujan, peningkatan suhu udara dan kenaikan permukaan laut yang disebabkan oleh iklim [2].

Hujan menjadi bagian dari fenomena alam yang diindikasikan dengan jatuhnya titik air dari atmosfer ke permukaan bumi. Hujan adalah peristiwa presipitasi yang berwujud air (Pettersen, 1958). Sumber utama air di permukaan bumi ini berasal dari hujan [3]. Hujan yang terjadi antara satu daerah dengan daerah lainnya memiliki perbedaan. Hujan memiliki karakteristik yang khas dan karena ini lah perbedaan tersebut dapat terjadi. Hujan memegang peranan penting dalam siklus hidrologi atau siklus perputaran air dan. Ada beberapa faktor yang dapat mempengaruhi terjadinya hujan, antara lain garis lintang, ketinggian tempat, jarak dari laut, posisi di dalam dan ukuran massa tanah daratan, arah angin terhadap sumber air, relief, dan suhu relatif tanah. Umumnya, ketika dua atau lebih lebih faktor-faktor tersebut terjadi secara bersamaan, maka itu akan mempengaruhi variasi dan tipe curah hujan. Curah hujan menjadi salah satu parameter hujan yang dapat diukur, curah hujan akan menyatakan seberapa besar dan tingginya air yang diakibatkan oleh hujan di suatu daerah.

Curah hujan merupakan jumlah air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap, dan tidak mengalir selama periode tertentu yang diukur dengan satuan tinggi *milimeter* (mm) diatas permukaan horizontal [4]. Indonesia sendiri memiliki penguapan dan intensitas curah hujan yang cukup tinggi, hal ini dikarenakan Indonesia adalah negara beriklim tropis yang berada di ekuator sehingga mendapatkan penyinaran matahari yang maksimal. Curah hujan menjadi salah satu faktor dinamis sebagai penyebab utama banjir, dengan ini dibutuhkan teknologi dan informasi untuk mengolah data mengenai curah hujan. Data curah hujan bisa didapatkan dengan pengukuran pada stasiun hujan [5]. Pola penempatan dan penyebaran stasiun pencatatan curah hujan harus tepat dan dapat mewakili lokasi dimana stasiun tersebut berada, hal

ini dikarenakan intensitas, penyebaran, serta kedalaman hujan terjadi secara tidak merata di setiap wilayahnya. Maka dari itu, diperlukan pengolahan data secara teknis untuk mendukung pengambilan keputusan serta pembuatan kebijakan dalam memberikan informasi curah hujan yang sangat berpengaruh terhadap keselamatan masyarakat dan sosial-ekonomi.

Data yang akan diolah ini merupakan data yang besar, luas, dan kompleks mengenai detail-detail karakteristik curah hujan terutama. Data ini perlu dianalisis dengan cepat untuk menghasilkan suatu informasi yang penting agar menghasilkan sebuah keputusan dan pengetahuan yang dapat diprediksi dan estimasi mengenai apa yang terjadi kedepannya. Oleh karena itu, dibutuhkan teknik yang dapat memecahkan permasalahan ini, teknik statistik, matematik, kecerdasan buatan (*Artificial Intelligence*) dan *Machine Learning* untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar, dan teknik tersebut adalah teknik data mining. Data mining menjadi teknik yang telah dikenal pada akhir 80an dan umum digunakan oleh komunitas riset oleh ahli statistik, analisis data, dan komunitas Manajemen Information System. Secara tradisional, teknik data mining dapat dikategorikan sebagai berikut, (1) *pattern extraction/ identification*, (2) *data clustering*, (3) *classification/ categorisation* [6]. Salah satu dari teknik pengolahan data mining yang paling sering digunakan adalah klasifikasi. Klasifikasi menjadi salah satu teknik yang sangat umum digunakan, teknik ini telah diterapkan diberbagai bidang ilmu. dalam melakukan model klasifikasi, data training dibutuhkan untuk membangun model klasifikasi untuk memprediksi label kelas untuk membentuk sampel baru. Pada masalah ini, pengklasifikasian data curah hujan dibagi menjadi dua kategori, yaitu hujan lebat dan hujan sangat lebat. Proses klasifikasi pada data ini dilakukan berdasarkan kriteria tertentu, seperti intensitas curah hujan. Pendekatan *data mining* yang diterapkan dalam tahapan penelitian ini adalah *CRISP-DM (Cross-Industry Standard Process Model for Data Mining)*.

Penelitian ini dilakukan dengan menggunakan algoritma *Decision Tree.* Decision Tree menjadi salah satu teknik klasifikasi yang mudah diimplementasikan terutama pada data curah hujan ini Decision Tree memiliki visualisasi data yang lebih mudah dilihat dan dipahami karena visualisasinya dinilai lebih sederhana. Selain itu, dengan menggunakan Algoritma ini proses pembagiannya juga lebih jelas karena Algoritma ini menggambarkan dengan sangat jelas karakteristik-karakteristik data yang menghasilkan sebuah keputusan, sehingga pada kasus banjir lebih mudah mengetahui gejalanya [7].

*B.* Tujuan

Tujuan dari penelitian ini adalah untuk dapat membuat model yang bisa memprediksi curah hujan yang terjadi dan dapat mengetahui tingkat intensitas curah hujan yang dapat menyebabkan banjir.

## II. DASAR TEORI

*A. Banjir*

Banjir merupakan salah satu bencana alam yang cukup sering terjadi pada daerah Indonesia dan bersifat merusak lingkungan disekitar masyarakat. Menurut KBBI (Kamus Besar Bahasa Indonesia), banjir adalah berair banyak dan deras, kadang-kadang meluap. Menurut BNPB (Badan Nasional Penanggulangan Bencana), banjir merupakan peristiwa atau kejadian alami dimana sebidang tanah atau area yang biasanya merupakan lahan kering, tiba-tiba terendam air karena volume air meningkat

*B. Curah Hujan*

Menurut KBBI (Kamus Besar Bahasa Indonesia), curah hujan adalah banyaknya hujan yang tercurah atau turun di suatu daerah ataupun tempat dalam jangka waktu tertentu. Ada pengertian dari penelitian lain, yang menyatakan bahwa curah hujan merupakan intensitas ketinggian air hujan yang jatuh pada permukaan yang datar, tidak menyerap, dan tidak mengalir pada suatu tempat. Pengukuran satuan curah hujan menggunakan milimeter agar mampu mendapatkan hasil dan akurasi yang mendekati kesempurnaan.

*C. Data Science*

Menurut pengertian para ahli, *data science* merupakan sebuah ilmu pengetahuan mengenai sebuah keahlian pada bidang tertentu yang dipadukan dengan ilmu statistik, pemrograman, dan statistik. Tujuan dari implementasi *data science* adalah untuk menghasilkan sebuah informasi yang dibutuhkan untuk tujuan tertentu dari data yang dianalisis tersebut. Menurut Urban Institute, data science adalah sebuah keahlian yang memerlukan ilmu pada bidang komputer, pemrograman, teknologi, dan statistik untuk dapat mengolah data yang digunakan dan bisa mendapatkan hasil dari data tersebut [8].

*D. Data Mining*

Data mining merupakan sebuah metode yang digunakan untuk mengolah data tertentu agar dapat menemukan pola pada data tersebut. Menurut David Hand dari MIT, data mining merupakan proses analisa pada data yang berukuran besar untuk dapat menemukan hubungan dan pola yang sebelumnya tidak pernah diketahui untuk dipahami dan disimpulkan sehingga hasil tersebut dapat berguna bagi pemilik data tersebut.

*E. Stratified Sampling*

Menurut Taro Yamane, *stratified sampling* merupakan metode yang digunakan untuk proses pengambilan sampel dengan cara membagi suatu populasi menjadi populasi yang lebih kecil dan mengkategorikan menjadi strata tertentu, serta memilih sampel tersebut secara acak dan menggabungkannya untuk menetapkan parameter populasi yang dianalisis [9].

*F. Algoritma Klasifikasi*

Algoritma klasifikasi adalah suatu algoritma yang digunakan pada analisis data mining untuk mengelompokan data menjadi suatu kategori tertentu dengan memahami data sebelumnya. Berikut beberapa contoh algoritma klasifikasi, yaitu Hierarchical clustering, K-Means *Clustering*, *Decision Tree, Random Forest*, dan algoritma lainnya

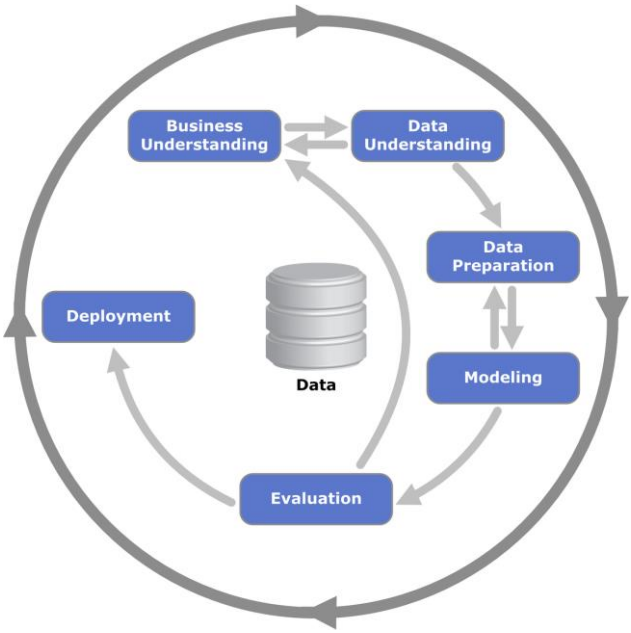
*G. Decision Tree*

*Decision Tree* adalah algoritma yang berbentuk seperti pohon keputusan yang terdiri dari struktur diagram

alur dan setiap node mewakili tes pada atribut, cabang mewakili hasil tes, dan label kelas diwakili oleh setiap *node* terminal. Algoritma ini diperkenalkan pada tahun 1986 oleh Quinlan Ross. Algoritma pohon keputusan menggunakan model prediktif yang memetakan pengamatan tentang suatu item untuk kesimpulan tentang nilai target *item*. Bahkan, algoritma decision tree memiliki akurasi yang sama atau lebih baik jika dibandingkan dengan algoritma lain.

## III. METODOLOGI

Framework metodologi yang digunakan dalam penelitian ini adalah *Cross Industry Standard Process for Data Mining (CRISP-DM)*. *CRISP-DM* adalah langkah-langkah dari siklus proses penemuan pengetahuan yang berguna dari sebuah dataset untuk menyelesaikan suatu masalah. *CRISP-DM* merupakan model yang sering digunakan dalam *data mining* karena selalu berhasil memberikan suatu bentuk hasil, meski masalah tidak selalu terselesaikan. Karena *CRISP-DM* berbentuk sebuah siklus, proses pengolahan data dapat berjalan dengan lebih fleksibel dan mudah untuk dikoreksi, sehingga lebih tahan terhadap kesalahan dan mudah untuk digunakan. Siklus *CRISP-DM* dapat digambarkan menggunakan diagram berikut.




Gambar 1. Framework Metodologi CRISP-DM.

### A. Business Understanding

*Business Understanding* adalah tahap pertama yang dilakukan untuk mendapatkan pemahaman mengenai tujuan dan persyaratan penelitian dari perspektif bisnis secara keseluruhan. [10]. Lalu, menerjemahkan tujuan dari penelitian serta menentukan batasan dalam perumusan data mining, sehingga selanjutnya dapat mempersiapkan strategi awal untuk mencapai tujuan tersebut.

### B. Data Understanding

*Data Understanding* merupakan tahap kedua yang dilakukan untuk mendapatkan pemahaman mengenai data yang akan dipakai dan tipe data yang dipakai dalam penelitian. *Data understanding* ini peneliti akan melakukan kualifikasi data menjadi data *categorical* dan *numerical* data.

Data categorical merupakan tipe data yang dapat disimpan kedalam group atau kategori. Dimana *categorical* data dapat diklasifikasikan menjadi *categories*. Sedangkan numerical data merupakan tipe data yang berbentuk angka dan bersifat *numeric*, kedua tipe data ini akan sangat dibutuhkan dalam melakukan penelitian [11].

Pada penelitian ini para peneliti menggunakan python *programming language* dalam melakukan pemrosesan data. Python merupakan bahasa pemrograman yang mudah dipelajari dan sangat berguna dalam pengolahan data. Python sendiri memiliki struktur data yang tinggi dan pendekatan yang sederhana di dalam data modelling namun efektif [12].

### C. Data Preparation

*Data Preparation* adalah tahap ketiga yang dilakukan untuk mempersiapkan dataset dan atribut akhir akhir yang akan digunakan pada langkah selanjutnya. Pada tahap ini digunakan *stratified sampling*, *data cleansing*, *data exploration*, dan melihat korelasi antar fitur.

*Stratified sampling* adalah pengambilan sampel data secara acak dengan proporsi tertentu dari sekumpulan subkelompok yang telah dibagi sesuai dengan karakteristik dari anggota dan memastikan bahwa subjek yang dipilih akan mewaliki setiap populasi yang diinginkan [13].

*Data cleansing* merupakan proses membersihkan data untuk menghilangkan *noise* dan data yang tidak relevan [14]. Umumnya, data yang diperoleh dari manapun memiliki struktur yang tidak sempurna, seperti data yang hilang, data yang tidak valid, dan lainnya. Data seperti itu akan dibuang untuk meningkatkan performa dari teknik *data mining.*

*Data Exploration* adalah teknik untuk memahami pola awal dan karakteristik kumpulan data seperti struktur data, distribusi nilai, hubungan timbal balik data, kuantitas, dan ukuran data dengan melibatkan penggunaan *visualization tool* [15]. *Data Exploration* kemungkinan penelitian mendapatkan pengetahuan yang lebih luas mengenai data mentah.

Korelasi adalah salah satu teknik statistik yang digunakan untuk melihat hubungan antara dua variabel. Korelasi terbagi menjadi dua, yaitu korelasi *bivariat* dan korelasi parsial. Korelasi bivariat digunakan untuk uji korelasi antara dua variabel, sedangkan korelasi parsial digunakan untuk menghitung koefisien korelasi antara dua variabel. Keeratan hubungan yang dimiliki antara variabel satu dengan lainnya disebut dengan Koefisien korelasi.

### D. Modeling

Salah satu algoritma klasifikasi s*upervised machine learning* yang paling banyak digunakan dalam penelitian adalah *decision tree. decision tree* merupakan sebuah algoritma yang memiliki bentuk seperti pohon dimana decision tree akan memiliki pohon utama yang disebut *root node*, cabang yang disebut *branches*, dan ujung yang dinamakan *leaf nodes*. *Decision tree* biasanya digunakan untuk mengkategorikan atau membuat prediksi berdasarkan bagaimana serangkaian pertanyaan sebelumnya dijawab.

Decision Tree memiliki visualisasi data yang lebih mudah dilihat dan dipahami karena visualisasinya dinilai lebih sederhana. Selain itu, dengan menggunakan Algoritma

ini proses pembagiannya juga lebih jelas karena Algoritma ini menggambarkan dengan sangat jelas karakteristik-karakteristik data yang menghasilkan sebuah keputusan, sehingga pada kasus banjir lebih mudah mengetahui gejalanya.

*Decision tree* merupakan algoritma yang dianggap tepat untuk penelitian ini karena alasan diatas. *Decision tree* memiliki banyak jenis, sehingga peneliti melakukan perbandingan antara 2 algoritma Decision Tree untuk menemukan *Decision Tree* yang paling tepat yaitu algoritma C4.5 dan C5.0.

*Decision tree* C4,5 merupakan sebuah algoritma yang dibuat oleh oleh Ross Quinlan untuk mengatasi permasalahan yang ada di dalam *decision tree* ID3. C4.5 algorithm merupakan algoritma yang dibuat untuk melakukan improvisasi dai akurasi dalam *decision tree*, mengurangi *training time*, dan kemampuan untuk menentukan *training set* yang paling baik. selain itu C4.5 algorithm juga mengimplementasikan *IGR* dimana implementasi ini membuat algoritma decision tree C4.5 tidak terlalu bias terhadap atribut yang memiliki banyak cabang dan membuat *gain ratio* lebih besar dan merata serta membuat semua values lebih kecil jika dimasukan ke dalam satu branch [16].

$$Entropy(S) = \sum_{i=1}^{..} - pi * log_2\ pi$$

Keterangan :
S : Himpunan kasus
A : fitur
N : Jumlah Partisi S
Pi : Proporsi dari Si terhadap S

Algoritma Decision Tree C5.0 merupakan penyempurnaan kembali dari algoritma C4.5 yang oleh Ross pad tahun 1987. di dalam algoritma c5.0 ini dilakukan pemilihan proses atribut menggunakan ratio. dimana gain ratio yang paling besar akan dipilih sebagai parent node. Perbedaan dari c4.5 dengan c5.0 adalah c5.0 akan membuat algoritma yang lebih ringkas dibandingkan c4.5 [17].

$$\text{Info}_A(D) = \sum_{J=I}^{Y} \frac{|Dj|}{D} x Info\ (Dj)$$

$$\text{GAIN (A)} = \text{Info(D)} - \text{Info(Dj)}$$

## E. *Evaluation*

*Evaluation* adalah tahapan akhir yang perlu dilakukan untuk melihat tingkat keberhasilan model untuk bisnis tersebut. Tahapan evaluasi dipecah menjadi tiga, yaitu mengevaluasi hasil, proses tinjauan, dan menentukan langkah selanjutnya. Mengevaluasi hasil dengan melihat sejauh mana model telah memenuhi tujuan bisnis dan menguji model. Proses tinjauan adalah melakukan peninjauan yang lebih menyeluruh terhadap keterlibatan data mining. Menentukan langkah selanjutnya dengan menilai bagaimana jika peneliti melanjutkan penelitian.

Dalam pengujian model, biasanya digunakan teknik *Confusion Matrix*. *Confusion Matrix* adalah proses yang yang akan menilai dan memprediksi validitas model klasifikasi. Penggunaan *Confusion Matrix* memungkinkan peneliti memperoleh lebih banyak informasi mengenai perbedaan antara data yang dengan baik diklasifikasikasikan dan yang salah diklasifikasikan. *Confusion Matrix* pada penelitian ini digunakan untuk melihat nilai akurasi, presisi, dan recall [18]. Nilai tersebut diperoleh dengan perhitungan berikut.

$$\text{Accuracy} = \frac{\text{Jumlah klasifikasi benar}}{\text{Total sampel testing yang diuji}} \text{ x100\%}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}} \text{ x100\%}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}} \text{ x100\%}$$

## F. *Deployment*

*Deployment* merupakan hasil akhir dimana output data mining akan dipakai dan digunakan untuk kebutuhan user maupun perusahan. Tahap ini mencakup beberapa tugas seperti merencanakan penerapan, pemantauan dan pemeliharaan, melaporkan hari akhir dan meninjau hasil akhir dari penerapan tersebut.

# IV. HASIL DAN PEMBAHASAN

Setelah menentukan metodologi untuk melaksanakan penelitian, maka proses analisa data pun dimulai.

## A. *Business Understanding*

Penerapan hasil dari penggunaan data mining yang digunakan dalam penelitian ini dikarenakan peneliti ingin mengetahui bagaimana pola curah hujan yang turun dan intensitasnya dalam membuat menimbulkan potensi bencana alam banjir, serta memiliki parameter yang mempengaruhi terjadinya curah hujan yang tinggi.

## B. *Data Understanding*

Data yang digunakan merupakan data curah hujan yang diambil dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Klimatologi Palembang dari tahun 2011 hingga tahun 2020. Dataset yang didapatkan memiliki 3.645 data, dengan 7 atribut yang dapat digambarkan dengan gambar berikut.

| Atribut | Tipe | Keterangan |
|---|---|---|
| Tanggal | Numerical | Tahun data diambil |
| 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 |
| Kelembapan Rata-rata (RH_avg) | Numerical | Kelembapan rata -rata dalam satu hari |
| Lama Penyinaran Matahari (ss) | Numerical | Lama penyinaran matahari dalam satu hari |
| Curah Hujan (RR) | Nominal | Curah hujan harian |

Gambar 2. Tabel 7 Atribut Dataset

Atribut pertama adalah Tanggal, yaitu data tipe numerik yang menggambarkan tahun kondisi curah hujan direkam. Atribut kedua adalah suhu minimum yang diberi simbol "Tn", yaitu data numerik yang menggambarkan suhu minimum dalam hari data tersebut diambil. Atribut ketiga adalah suhu maksimum yang diberi simbol "Tx", yaitu data numerik yang menggambarkan suhu maksimum dalam hari data tersebut diambil. Atribut keempat adalah suhu rata-rata (Tavg), yaitu data numerik yang menggambarkan suhu rata-rata dalam hari data tersebut diambil. Atribut kelima adalah Kelembaban rata-rata yang diberi simbol "RH_avg", yaitu data numerik yang menggambarkan kelembaban rata-rata dalam hari data tersebut diambil. Atribut keenam adalah lama penyinaran matahari yang diberi simbol "ss", yaitu data numerik yang menggambarkan lama penyinaran matahari dalam hari data tersebut diambil. Atribut terakhir adalah curah hujan, variabel target dari penelitian ini, yang diberi simbol "RR" , yaitu data numerik yang menggambarkan curah hujan hari data tersebut diambil. Data curah hujan ini disimpan dalam file excel yang berjudul "LaporanIklimHarian2011-2020.xlsx".

```
df = pd.read_excel("LaporanIklimHarian2011-2020.xlsx")
df.info()
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3654 entries, 0 to 3653
Data columns (total 7 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Tanggal  3654 non-null   object
 1   Tn       3519 non-null   object
 2   Tx       3562 non-null   object
 3   Tavg     3583 non-null   object
 4   RH_avg   3544 non-null   object
 5   RR       2986 non-null   object
 6   ss       3532 non-null   object
dtypes: object(7)
memory usage: 200.0+ KB
```

| | Tanggal | Tn | Tx | Tavg | RH_avg | RR | ss |
|---|---|---|---|---|---|---|---|
| 0 | 01-01-2011 | 24 | 32.4 | 27.5 | 83 | 0 | 3.2 |
| 1 | 02-01-2011 | 25 | 33.5 | 27.7 | 81 | 0 | 3.9 |
| 2 | 03-01-2011 | 23 | 30.3 | 26.6 | 84 | 0.5 | 0 |
| 3 | 04-01-2011 | 23 | 32.4 | 27 | 82 | 3.3 | 5.5 |
| 4 | 05-01-2011 | 22 | 29.9 | 24.7 | 93 | 26.2 | 0.3 |

Gambar 3. Head Data Laporan Iklim Harian 2011-2020

## C. *Data Preparation*

Dalam tahap data preparation, langkah pertama yang dilakukan adalah *data cleansing*. Terlihat dari fungsi info() yang telah digunakan sebelumnya, terdapat jumlah data *non-null* yang tidak imbang antara setiap atribut, artinya terdapat data-data dengan nilai kosong pada atribut-atribut dataset. Jumlah data yang hilang pada setiap atribut dicari menggunakan fungsi is.na().sum() pada setiap atribut.

```
#memeriksa apakah ada data null
dataHilang = [df['Tanggal'].isna().sum(),
              df['Tn'].isna().sum(),
              df['Tx'].isna().sum(),
              df['Tavg'].isna().sum(),
              df['RH_avg'].isna().sum(),
              df['RR'].isna().sum(),
              df['ss'].isna().sum()]
dataHilang

#Tanggal tidak memiliki data hilang,
#Tn memiliki 132 data hilang,
#Tx memiliki 92 data hilang,
#Tavg memiliki 71 data hilang,
#RH_avg memiliki 110 data hilang,
#RR memiliki 668 data hilang,
#ss memiliki 122 data hilang
```

```
[0, 135, 92, 71, 110, 668, 122]
```

Gambar 4. Pengecekan Data Null

Menurut perhitungan, dari atribut tanggal tidak ada data yang hilang, dari atribut suhu minimal terdapat 132 data yang hilang, dari atribut suhu maksimal terdapat 92 data yang hilang, dari atribut suhu rata-rata terdapat 71 data yang hilang, dari atribut kelembaban rata-rata terdapat 110 data yang hilang, dari atribut lama penyinaran matahari terdapat 122 data yang hilang, dan terakhir, dari atribut curah hujan terdapat 668 data yang hilang.

Data dengan nilai yang hilang harus ditangani, namun sebelumnya dilakukan transformasi data untuk memperbaiki format data terlebih dahulu. Atribut-atribut pada dataset memiliki tipe data numerik, namun masih tercatat sebagai *object* pada Python. Oleh karena itu, digunakan fungsi to_numeric() untuk memperbaiki tipe data yang salah.

```
#Mengubah format data
df['Tanggal'] = pd.to_numeric(df['Tanggal'],errors = 'coerce')
df['Tn'] = pd.to_numeric(df['Tn'],errors = 'coerce')
df['Tx'] = pd.to_numeric(df['Tx'],errors = 'coerce')
df['Tavg'] = pd.to_numeric(df['Tavg'],errors = 'coerce')
df['RH_avg'] = pd.to_numeric(df['RH_avg'],errors = 'coerce')
df['RR'] = pd.to_numeric(df['RR'],errors = 'coerce')
df['ss'] = pd.to_numeric(df['ss'],errors = 'coerce')
```

Gambar 5. Perubahan Format Data

Data dengan nilai yang hilang tersebut ditangani dengan dua cara, yaitu dengan menghilangkan baris data dengan nilai hilang, dan dengan melakukan imputasi data. Pertama-tama, data hilang dari atribut curah hujan dihapus karena curah hujan merupakan variabel target dari penelitian. Data-data dengan nilai yang tidak hilang diambil menggunakan fungsi notna().

Membuang baris yang memiliki curah hujan dengan nilai yang tidak diketahui

```
df = df[df['RR'].notna()]
```

Gambar 6. Penghapusan Variabel Curah Hujan

Untuk atribut-atribut lain dengan data yang hilang, data-data tersebut diimputasi menggunakan nilai rata-rata dari atribut tersebut dengan menggunakan fungsi *mean*() dan *fillna*().

```
#mengisi data null dengan rata-rata
#Tn
rata_Tn = df['Tn'].mean()
df['Tn'] = df['Tn'].fillna(rata_Tn)

#Tx
rata_Tx = df['Tx'].mean()
df['Tx'] = df['Tx'].fillna(rata_Tx)

#Tavg
rata_Tavg = df['Tavg'].mean()
df['Tavg'] = df['Tavg'].fillna(rata_Tavg)

#RH_avg
rata_RH_avg = df['RH_avg'].mean()
df['RH_avg'] = df['RH_avg'].fillna(rata_RH_avg)

#ss
rata_ss = df['ss'].mean()
df['ss'] = df['ss'].fillna(rata_Tn)
```

Gambar 7. Pengisian Data Null Dengan Nilai Rata-rata

Setelah melakukan *cleansing*, data kembali dicek untuk memastikan bahwa sudah tidak ada lagi data dengan nilai hilang dengan fungsi info(). Setelah *cleansing*, ternyata jumlah data yang dapat dipakai tersisa 2.623 data.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2623 entries, 0 to 3653
Data columns (total 7 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Tanggal  2623 non-null   int64
 1   Tn       2623 non-null   float64
 2   Tx       2623 non-null   float64
 3   Tavg     2623 non-null   float64
 4   RH_avg   2623 non-null   float64
 5   RR       2623 non-null   float64
 6   ss       2623 non-null   float64
dtypes: float64(6), int64(1)
memory usage: 163.9 KB
```

Gambar 8. Pengecekan Kembali Data Null

Setelah melakukan *cleansing data*, kembali dilakukan perbaikan format pada variabel curah hujan. Variabel curah hujan diubah menjadi variabel kategorikal dengan mengelompokkan intensitas curah hujan berdasarkan pembagian kategori hujan milik BMKG, yaitu sebagai berikut. Kategori 'Tidak Hujan' memiliki intensitas curah hujan sebesar 0 hingga 0.5. Kategori 'Hujan Ringan' memiliki intensitas curah hujan sebesar 0.5 hingga 20. Kategori 'Hujan Sedang' memiliki intensitas curah hujan sebesar 20 hingga 50. Kategori 'Hujan Lebat' memiliki intensitas curah hujan sebesar 50 hingga 100. Terakhir, kategori 'Hujan Sangat Lebat' memiliki intensitas curah hujan lebih dari 100.

```
bins = [-0.5, 0.5, 19.9, 49.9, 100, 173]
group_names = ['Tidak Hujan','Hujan Ringan','Hujan Sedang', 'Hujan Lebat', 'Hujan Sangat Lebat']
df['RR'] = pd.cut(df['RR'], bins, labels=group_names)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2623 entries, 0 to 3653
Data columns (total 7 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Tanggal  2623 non-null   int64
 1   Tn       2623 non-null   float64
 2   Tx       2623 non-null   float64
 3   Tavg     2623 non-null   float64
 4   RH_avg   2623 non-null   float64
 5   RR       2623 non-null   category
 6   ss       2623 non-null   float64
dtypes: category(1), float64(5), int64(1)
memory usage: 146.2 KB
```

Gambar 9. Perbaikan Format Pada Variabel Curah Hujan

Tujuan penelitian ini adalah untuk membuat model yang dapat memprediksi potensi banjir dengan curah hujan, oleh karena itu variabel curah hujan dikategorikan kembali menjadi curah hujan yang berpotensi banjir dan curah hujan yang tidak berpotensi banjir. Curah hujan yang tidak berpotensi banjir merupakan hujan dengan kategori 'Tidak Hujan', 'Hujan Ringan', dan 'Hujan Sedang', sedangkan curah hujan yang berpotensi banjir merupakan hujan dengan kategori 'Hujan Lebat' dan 'Hujan Sangat Lebat'.

| | Tanggal | Tn | Tx | Tavg | RH_avg | RR | ss |
|---|---|---|---|---|---|---|---|
| 0 | 2011 | 24.0 | 32.4 | 27.5 | 83.0 | Tidak potensi banjir | 3.2 |
| 1 | 2011 | 25.0 | 33.5 | 27.7 | 81.0 | Tidak potensi banjir | 3.9 |
| 2 | 2011 | 23.0 | 30.3 | 26.6 | 84.0 | Tidak potensi banjir | 0.0 |
| 3 | 2011 | 23.0 | 32.4 | 27.0 | 82.0 | Tidak potensi banjir | 5.5 |
| 4 | 2011 | 22.0 | 29.9 | 24.7 | 93.0 | Tidak potensi banjir | 0.3 |

Gambar 10. Pengkategorian Kembali variabel Curah Hujan

Agar kategori ini dapat diproses oleh algoritma Decision Tree, maka kategori 'Potensi banjir' digambarkan dengan angka 1, dan 'Tidak potensi banjir' digambarkan dengan angka 0.

```
df['RR'] = df['RR'].replace(['Tidak potensi banjir','Potensi banjir'],[0,1])
df.head()
```

| | Tanggal | Tn | Tx | Tavg | RH_avg | RR | ss |
|---|---|---|---|---|---|---|---|
| 0 | 2011 | 24.0 | 32.4 | 27.5 | 83.0 | 0 | 3.2 |
| 1 | 2011 | 25.0 | 33.5 | 27.7 | 81.0 | 0 | 3.9 |
| 2 | 2011 | 23.0 | 30.3 | 26.6 | 84.0 | 0 | 0.0 |
| 3 | 2011 | 23.0 | 32.4 | 27.0 | 82.0 | 0 | 5.5 |
| 4 | 2011 | 22.0 | 29.9 | 24.7 | 93.0 | 0 | 0.3 |

Gambar 11. Mengubah Value Variabel Curah Hujan

Setelah melakukan *cleansing data*, dilakukan juga eksplorasi data untuk melihat dan mengerti kondisi data yang akan digunakan untuk analisis. Pertama, dilakukan eksplorasi deskriptif dengan fungsi *describe*().

```
df.describe()
```

| | Tanggal | Tn | Tx | Tavg | RH_avg | RR | ss |
|---|---|---|---|---|---|---|---|
| count | 2623.000000 | 2623.000000 | 2623.000000 | 2623.000000 | 2623.000000 | 2623.000000 | 2623.000000 |
| mean | 2015.205490 | 24.182133 | 32.779229 | 27.430223 | 84.494899 | 0.046893 | 4.597891 |
| std | 3.007126 | 0.900366 | 1.528170 | 1.009749 | 5.623474 | 0.211450 | 3.661053 |
| min | 2011.000000 | 20.000000 | 24.800000 | 23.300000 | 58.000000 | 0.000000 | 0.000000 |
| 25% | 2012.000000 | 24.000000 | 31.900000 | 26.800000 | 82.000000 | 0.000000 | 2.400000 |
| 50% | 2015.000000 | 24.000000 | 33.000000 | 27.500000 | 85.000000 | 0.000000 | 4.400000 |
| 75% | 2018.000000 | 25.000000 | 33.800000 | 28.100000 | 88.000000 | 0.000000 | 6.300000 |
| max | 2020.000000 | 27.200000 | 38.800000 | 33.300000 | 98.000000 | 1.000000 | 24.229376 |

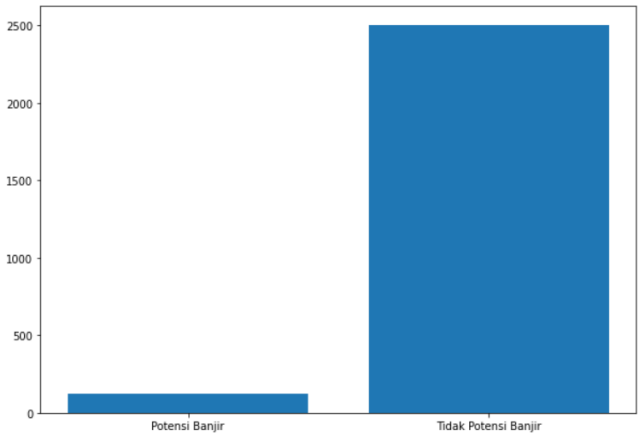
Gambar 12. Eksplorasi Deskriptif Dengan Fungsi Describe()

Melalui eksplorasi deskriptif, dapat terlihat persebaran data setiap variabel. Eksplorasi deskriptif memberikan informasi mengenai jumlah data (*count*), rata-rata (*mean*), standar deviasi (*std*), nilai minimum (*min*), nilai maksimum (*max*), dan kuartil pertama (25%), kedua (50%), dan ketiga (75%). Setelah melihat eksplorasi deskriptifnya, dibuat *heatmap* diagram korelasi yang menunjukkan korelasi antara setiap atribut.


Gambar 13. Visualisasi Heatmap Diagram Korelasi

Dari diagram korelasi, terlihat bahwa ada empat fitur yang memiliki korelasi yang lebih kuat, artinya lebih berpengaruh terhadap variabel target penelitian, yaitu curah hujan. Atribut pertama adalah suhu minimal dengan korelasi -0,19. Atribut kedua adalah suhu maksimal dengan korelasi 0,098, suhu rata-rata dengan korelasi 0,13, dan kelembaban rata-rata dengan korelasi 0,11. Atribut tanggal dan lama penyinaran matahari memiliki korelasi yang terlalu kecil dengan nilai -0,02 dan -0,03. Oleh karena itu, atribut tanggal dan lama penyinaran matahari tidak digunakan dalam analisa.

Setelah melihat korelasi data, distribusi variabel target juga dilihat menggunakan diagram batang.



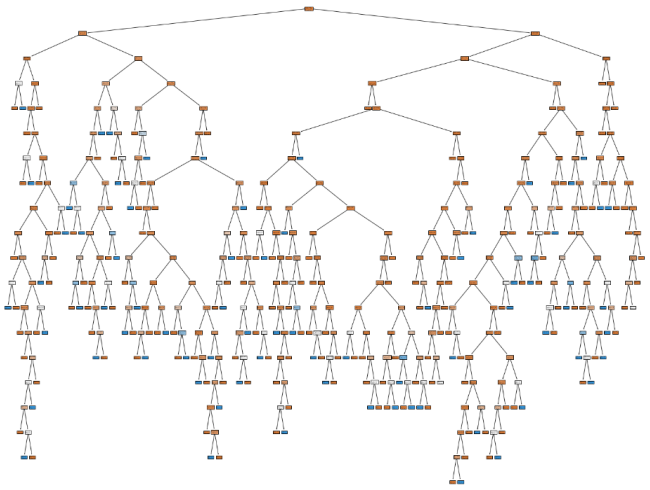

Gambar 14. Barplot Perbandingan Banjir Dan Tidak Banjir

Melalui barplot terlihat bahwa persebaran curah hujan dengan potensi banjir tidak seimbang dengan curah hujan yang tidak berpotensi banjir. Jumlah data dengan kondisi tidak berpotensi banjir jauh lebih tinggi dibandingkan dengan kondisi potensi banjir. Melihat ketidakseimbangan ini, sampling yang digunakan untuk membuat algoritma klasifikasi membutuhkan teknik stratified sampling.

*D.* Modeling

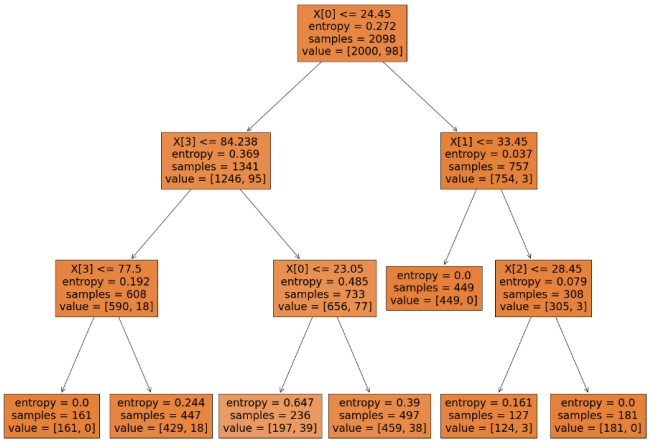
Proses modeling pertama dilakukan dengan melakukan *split data* menjadi data *training* dan *testing* menggunakan teknik *stratified sampling*. Untuk penelitian ini, digunakan data *training* sebesar 80%, dan data *testing* sebesar 20%. Data *training* memiliki 2.098 data, dan data testing memiliki 525 data. Setelah melakukan *splitting* data, maka dapat mulai dibangun model algoritma klasifikasi *Decision Tree* C4.5 dan C5.0.

Pertama, dibangun *Decision Tree* C4.5 yang memiliki kedalaman sebanyak 19 tingkat karena tidak dilakukan *pruning*. *Decision Tree* ini dibuat dengan menggunakan fungsi dari *library sklearn*.



Gambar 15. Visualisasi Hasil Decision Tree C4.5

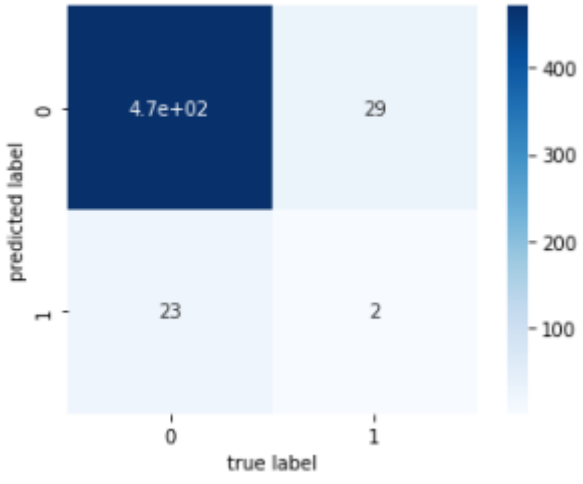
Kedua, dibangun *Decision Tree* C5.0 yang juga menggunakan *library sklearn*. Pada *Decision Tree* ini dilakukan pruning dengan menentukan kedalaman maksimal sebesar 3 tingkat. *Decision Tree* ini memiliki 6 daun terminal. Contoh keputusan yang menghasilkan daun terminal ini adalah, apabila suhu minimal lebih kecil atau sama dengan 24,45 derajat *celcius*, dan kelembaban rata-rata lebih kecil atau sama dengan 77,5% , maka hujan tidak berpotensi banjir, namun apabila kelembaban di bawah atau sama dengan 84,238% tapi di atas 77,5%, dan suhu minimal di bawah atau sama dengan 23,05 derajat *celcius*, maka terdapat kemungkinan hujan memiliki potensi banjir.




Gambar 16. Visualisasi Hasil Decision Tree C5.0

*E.* Evaluation

Setelah membangun kedua model *Decision Tree*, dilakukan evaluasi menggunakan *Confusion Matrix*. Model *Decision Tree* diaplikasikan pada dataset testing, kemudian hasilnya direkam dan dievaluasi dalam bentuk *Confusion Matrix.*



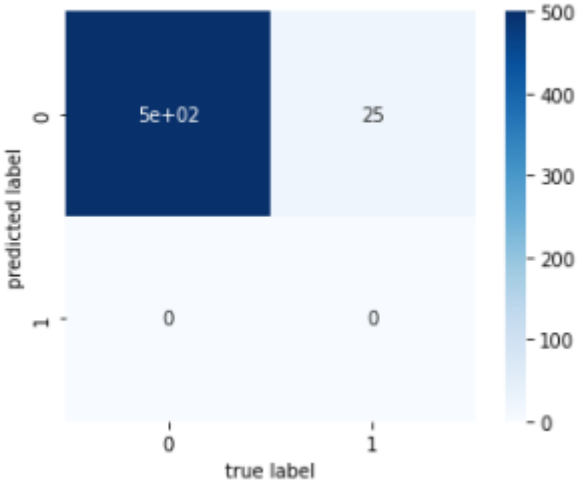

Gambar 17. Visualisasi Hasil Confusion Matrix C4.5

```
print('Precision: %.3f' % precision_score(y_test, y_pred))
print('Recall: %.3f' % recall_score(y_test, y_pred))
print('Accuracy: %.3f' % accuracy_score(y_test, y_pred))
```

```
Precision: 0.065
Recall: 0.080
Accuracy: 0.901
```

Gambar 18. Hasil Confusion Matrix C4.5

Evaluasi yang dilakukan pada *Confusion Matrix* untuk *Decision Tree* C4.5 memperlihatkan hasil *true positive* sebanyak 2 data, *true negative* sebanyak 470 data, *false positive* sebanyak 23 data, dan *false negative* sebanyak 29 data. Dari data-data tersebut, dapat dihitung hasil akurasi model algoritma adalah sebesar 90,1% , hasil presisi sebesar 6,5%, dan hasil *recall* adalah sebesar 8%.




Gambar 19. Visualisasi Hasil Confusion Matrix C5.0

```
print('Precision: %.3f' % precision_score(y_test, y_pred_p))
print('Recall: %.3f' % recall_score(y_test, y_pred_p))
print('Accuracy: %.3f' % accuracy_score(y_test, y_pred_p))
```

```
Precision: 0.000
Recall: 0.000
Accuracy: 0.952
```

Gambar 20. Hasil Confusion Matrix C5.0

Evaluasi yang dilakukan pada *Confusion Matrix* untuk *Decision Tree* C5.0 memperlihatkan hasil *true positive* sebanyak 0 data, *true negative* sebanyak 500 data, *false positive* sebanyak 0 data, dan *false negative* sebanyak 25 data. Dari data-data tersebut, dapat dihitung hasil akurasi model algoritma adalah sebesar 95,2% , hasil presisi sebesar 0%, dan hasil recall adalah sebesar 0% karena model tidak memiliki *true positive*.

```
print('Accuracy Decision Tree dengan Pruning: %.3f' % accuracy_score(y_test, y_pred))
print('Accuracy Decision Tree tanpa Pruning: %.3f' % accuracy_score(y_test, y_pred_p))
```

```
Accuracy Decision Tree dengan Pruning: 0.901
Accuracy Decision Tree tanpa Pruning: 0.952
```

Gambar 21. Akurasi Decision Tree Dengan Pruning Dan Tidak

Perbandingan *Decision Tree* C4.5 dengan D*ecision Tree* C5.0 menunjukkan bahwa akurasi pada *Decision Tree* C5.0 dengan nilai 95,2% lebih tinggi daripada akurasi pada Decision Tree C4.5 dengan nilai 90,1%. Hal ini menunjukkan bahwa *Decision Tree* C5.0 dapat melakukan prediksi dengan lebih tepat dibandingkan dengan *Decision Tree* C4.5. Akan tetapi, *Decision Tree* C4.5 menunjukkan kemampuan yang lebih tinggi untuk mendeteksi hasil yang positif, karena memiliki nilai presisi dan recall yang lebih tinggi dengan nilai 6,5% dan 8% dibandingkan Decision Tree C5.0 yang memiliki nilai presisi dan recall 0%.

*F.* Deployment

Setelah membuat model dan melakukan pengujian dan evaluasi pada model tersebut, Model *Decision Tree* telah dapat digunakan dalam memprediksi curah hujan yang berpotensi banjir menggunakan variabel-variabel suhu minimal, suhu maksimal, suhu rata-rata, dan kelembaban rata-rata.

## V. KESIMPULAN

Pada penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma *decision tree* C4.5 dan C5.0 dapat diterapkan pada penelitian ini. Algoritma C4.5 memiliki akurasi sebesar 90,1% dan algoritma C5.0 memiliki akurasi sebesar 95,2%. Terlihat bahwa akurasi *decision tree* C5.0 lebih tinggi dari algoritma C4.5. Namun, algoritma decision tree C4.5 lebih cocok diterapkan pada penelitian ini dikarenakan algoritma C5.0 tidak memiliki nilai presisi dan recall, sehingga kurang dapat menentukan tingkat intensitas curah hujan yang dapat menyebabkan banjir.

## DAFTAR PUSTAKA


[1] A. S. Putri, "Apa Itu Banjir? Definisi, Penyebab dan Dampak," Kompas.com, 03 January 2020. [Online]. Available: https://www.kompas.com/skola/read/2020/01/03/060000269/apa-itu-banjir-definisi-penyebab-dan-dampak. [Accessed 07 06 2022].

[2] A. F. L. Ben Arther Molle, "Analisis Anomali Pola Curah Hujan Bulanan Tahun 2019 Terhadap Normal Curah Hujan (30 Tahun) Di Kota Manado Dan Sekitarnya," *Jurnal Meteorologi Klimatologi dan Geofisika*, vol. 7, no. 1, p. 2, 2020.

[3] rimbakita, "Curah Hujan – Pengertian, Jenis, Alat Ukur & Metode Perhitungan," rimbakita.com, 2020. [Online]. Available: https://rimbakita.com/curah-hujan/. [Accessed 07 06 2022].

[4] D. Mulyono, "Analisis Karakteristik Curah Hujan Di Wilayah Kabupaten Garut Selatan," *Jurnal Konstruksi*, vol. 13, no. 1, p. 2, 2014.

[5] F. D. Ezza Qodriatullah Ajr, "Menentukan Stasiun Hujan Dan Curah Hujan Dengan Metode Polygon Thiessen Daerah Kabupaten Lebak," *ejournal lppm Unbaja*, vol. 2, no. 2, p. 140, 2019.

[6] R. S. K. K. A.Dinesh Kumar, "Review on Prediction Algorithms in Educational Data Mining," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 8, pp. 2-3, 2018.

[7] "Analisa Klasifikasi menggunakan Algoritma Decision Tree pada Data Log Firewall," *ejournal stmikgici*, vol. 9, no. 3, p. 259, 2021.

[8] DqLab, "Data Science Adalah: Yuk Kenali Lebih Jauh Tentang Data Science!," dqlab.id, 26 Oktober 2020. [Online]. Available: https://dqlab.id/yuk-kenalan-dengan-data-science. [Accessed 07 06 2022].

[9] Y. S. P. H. Siti Faiqotul Ulya, "Analisis Prediksi Quick Count Dengan Metode Stratified Random Sampling Dan Estimasi Confidence Interval Menggunakan Metode Maksimum Likelihood," *UNNES Journal of Mathematics*, vol. 7, no. 1, p. 109, 2018.

[10] c. a. M. D. F. M. Veronika Plotnikova, "Adaptations of data mining methodologies: a systematic literature review," *PubMed Central*, vol. 267, no. 6, p. 1, 2020.

[11] N. J. G. Priya Ranganathan, "An Introduction to Statistics – Data Types, Distributions and Summarizing Data," *PubMed Central*, p. 1, 2019.

[12] J. Lakshmi, "Machine learning techniques using python for data analysis in performance evaluation," *Acharya Institute of Management and Sciences*, vol. 17, no. 1/2, p. 5, 2018.

[13] R. Arnab, "Stratified Sampling," *sciencedirect*, p. 1, 2017.

[14] A. M. B. Firdaus Akhmad Muttaqin, "Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak "Dodo Kids Browser," *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, p. 2, 2016.

[15] B. D. Vijay Kotu, "Data Exploration," *sciencedirect*, p. 1, 2019.

[16] M. Y. S. Z. W. Ibomoiye Domor, "Procedia Manufacturing," *sciencedirect*, vol. 35, p. 699, 2019.

[17] Y. R. Putri, "Prediksi Pola Kecelakaan Kerja Pada Perusahaan Non Ekstraktif Menggunakan Algoritma Decision Tree: C4.5 Dan C5.0," *Jurnal Sains Dan Seni Pomits*, vol. 1, no. 1, p. 19, 2013.

[18] . Narkhede, "Understanding Confusion Matrix," owardsdatascience.com, 9 May 2018. [Online]. Available: ttps://towardsdatascience.com/understanding-confusion-matrix-9ad42dcfd62. [Accessed 07 06 2022].


## ROLE MEMBER

*1)* Jason Suhali - 00000045379
- a) Latar belakang
- b) Tujuan
- c) Metodologi
- d) Hasil Penelitian
- e) Kesimpulan
- f) Design

*2)* Maria Darlene Kusnadi - 00000045996
- a) Latar belakang
- b) Tujuan
- c) Metodologi
- d) Hasil Penelitian
- e) Kesimpulan
- f) Design

*3)* Fardhila Zahra Dwi Wardhani - 00000044817
- a) Latar belakang
- b) Tujuan
- c) Metodologi
- d) Hasil Penelitian
- e) Kesimpulan
- f) Design

*4)* Michael Abhinaya Bagioyuwono - 00000044426
- a) Latar belakang
- b) Tujuan
- c) Metodologi
- d) Hasil Penelitian
- e) Kesimpulan
- f) Design