

**TESIS**

**PREDIKSI BENCANA BANJIR BERDASARKAN CURAH HUJAN  
MENGUNAKAN ALGORITMA NAÏVE BAYES**



Disusun oleh:

**Nama : Slamet Triyanto**  
**NIM : 21.55.1088**  
**Konsentrasi : Teknologi Media Digital**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

**TESIS**

**PREDIKSI BENCANA BANJIR BERDASARKAN CURAH HUJAN  
MENGUNAKAN ALGORITMA NAÏVE BAYES**

**PREDICTION OF FLOOD DISASTER BASED ON RAINFALL USING  
NAÏVE BAYES ALGORITHM**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

**Nama : Slamet Triyanto**  
**NIM : 21.55.1088**  
**Konsentrasi : Teknologi Media Digital**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

**HALAMAN PENGESAHAN**

**PREDIKSI BENCANA BANJIR BERDASARKAN CURAH HUJAN  
MENGUNAKAN ALGORITMA NAÏVE BAYES**

**PREDICTION OF FLOOD DISASTER BASED ON RAINFALL USING  
NAÏVE BAYES ALGORITHM**

Dipersiapkan dan Disusun oleh

**Slamet Triyanto**

**21.55.1088**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari Rabu, 05 Oktober 2022

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer.

Yogyakarta, 05 Oktober 2022

**Rektor**

**Prof. Dr. M. Suvanto, M.M.**

**NIK. 190302001**

**HALAMAN PERSETUJUAN**

**PREDIKSI BENCANA BANJIR BERDASARKAN CURAH HUJAN  
MENGUNAKAN ALGORITMA NAÏVE BAYES**

**PREDICTION OF FLOOD DISASTER BASED ON RAINFALL USING  
NAÏVE BAYES ALGORITHM**

Dipersiapkan dan Disusun oleh

**Slamet Triyanto**

**21.55.1088**

Telah Dijikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari Rabu, 05 Oktober 2022

**Pembimbing Utama**

**Anggota Tim Penguji**

**Dr. Andi Sunyoto, M.Kom**  
NIK. 190302052

**Dr. Arief Setyanto, S.Si, M.T.**  
NIK. 190302036

**Pembimbing Pendamping**

**Alva Hendi Muhammad, S.T., M.Eng., Ph.D**  
NIK. 190302493

**M. Rudyanto Arief, M.T**  
NIK. 190302098

**Dr. Andi Sunyoto, M.Kom**  
NIK. 190302052

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, 05 Oktober 2022  
**Direktur Program Pascasarjana**

**Prof.Dr. Kusriani, M.Kom.**  
NIK. 190302106

## HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Slamet Triyanto  
NIM : 21.55.1088  
Konsentrasi : Teknologi Media Digital

Menyatakan bahwa Tesis dengan judul berikut:

**Prediksi Bencana Banjir Berdasarkan Curah Hujan Menggunakan Algoritma Naïve Bayes**

Dosen Pembimbing Utama : Dr. Andi Sanyoto, M.Kom  
Dosen Pembimbing Pendamping : M. Ruddyanto Arief, M.T.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, pemikiran dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat plagiasi atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 05 Oktober 2022

Yang Menyatakan,



Slamet Triyanto

## KATA PENGANTAR

Alhamdulillah atas izin Allah SWT, pengerjaan penelitian dan penulisan Laporan Tesis ini dapat diselesaikan pada waktu yang tepat. Sholawat dan Salam untuk Rosullullah atas safaat beliau pengetahuan berkembang dengan sangat baik, Sehingga pemikiran dan diskusi terbuka menjadi hal yang tidak tabu lagi. Dalam kesempatan penulisan kata pengantar ini, penulis bermaksud mengucapkan terimakasih yang sangat mendalam kepada semua pihak yang telah berperan serta dalam penyelesaian penelitian dan penulisan Tesis ini. Terimakasih yang sangat mendalam khususnya kepada :

1. Bapak Dr. Andi Sunyoto, M.Kom selaku pembimbing 1 dalam pengerjaan Tesis;
2. Bapak M. Rudyanto Arief, M.T selaku Pembing 2 dalam penelitian ini;
3. Rektor Amikom, Direktur Pasca Sarjana dan Admisi Magister Teknik Informatika Amikom Yogyakarta yang sangat luar biasa;
4. Ananda tercinta Oryza Sativa Handayani, Istri, Emak, Mertua dan seluruh Keluarga yang terus mendukung Penulis dalam berbagai aspek;
5. Komandan PUSDALOPS-PB BPBD Kampar Bapak Adi Chandra Lukita,MM dan Rekan-Rekan Satgas terimakasih atas Support yang Luar Biasa;
6. Guru dan Rekan-Rekan Tenaga Pengajar serta Mahasiswa Teknik Informatika Politeknik Kampar atas dukungan Moril maupun Materil;

7. dan semua pihak yang telah mendukung penyelesaian penelitian ini,  
Terimakasih yang sangat mendalam dari Penulis.

Seperti kata pepatah “Tak Ada Gading yang Tak Retak”, begitu pula dalam tulisan ini. Banyak kekeliruan banyak pernyataan maupun kesimpulan yang ternyata tidak tepat. Melalui tulisan ini, Penulis secara terbuka membuka ruang kepada pembaca untuk memberikan kritik maupun saran untuk perbaikan keilmuan dimasa yang akan datang.

Mohon maaf atas segala khilaf dan salah, Akhir kata segala kebaikan datangnya dari Allah SWT, dan segala kekurangan datangnya dari Penulis.

Yogyakarta, 05 Oktober 2022

Penulis

## DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
INTISARI.....	xiii
<i>ABSTRACT</i> .....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. Tinjauan Pustaka.....	7
2.2. Keaslian Penelitian.....	17
2.3. Landasan Teori.....	28
2.3.1 Data Mining.....	28



2.3.2 Prediction/prediksi/peramalan <i>Data Mining</i> .....	33
<b>BAB III METODE PENELITIAN</b> .....	<b>38</b>
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	38
3.2. Metode Pengumpulan Data.....	38
3.3. Metode Analisis Data .....	46
3.4. Alur Penelitian.....	47
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN</b> .....	<b>53</b>
4.1. Prediksi Banjir Menggunakan Seleksi Fitur .....	53
4.2. Komparasi algoritma hasil pengujian .....	55
<b>BAB V PENUTUP</b> .....	<b>66</b>
5.1. Kesimpulan .....	66
5.2. Saran .....	66
<b>DAFTAR PUSTAKA</b> .....	<b>68</b>
<b>LAMPIRAN</b> .....	<b>76</b>

## DAFTAR TABEL

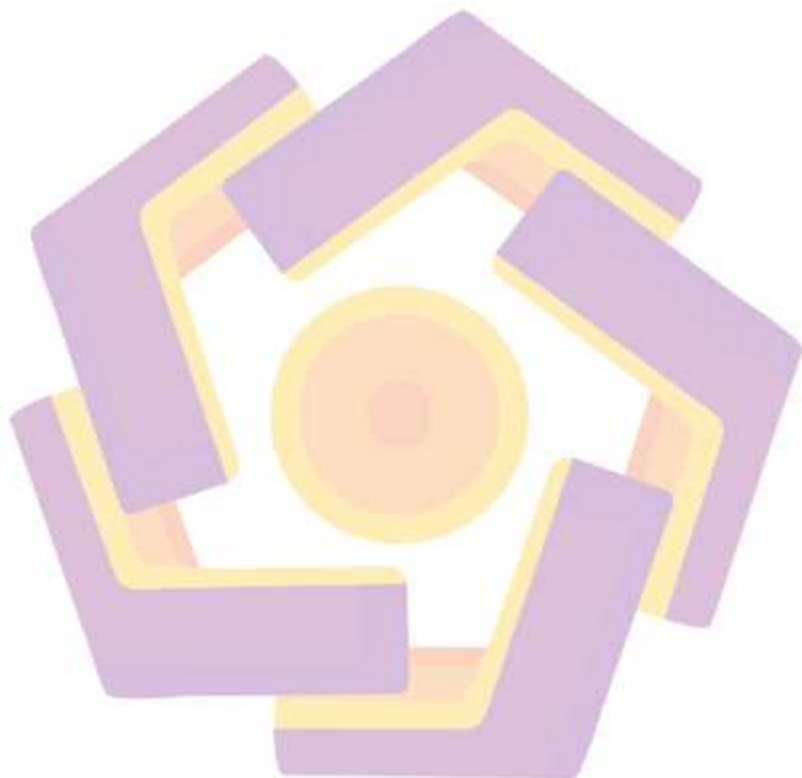
Tabel 2.1 Hasil Classifikasi K-Nearest Neighbor, Support Vector Machiness, Multi Layer Perception, Logistic Regression, dan Naïve Bayes (Gopal et al., 2020).....	9
Tabel 2.2 Hasil penelitian Cancer Disease Prediction using Naïve Bayes, K-Nearest Neighbor and J48 algorithm (Maliha et al., 2019) .....	13
Tabel 2.3 Matriks literatur review dan posisi penelitian Prediksi Bencana Banjir Dengan Algoritma Supervisi .....	17
Tabel 3.1. Menampilkan 5 (lima) data teratas pada <i>data sets</i> .....	50
Tabel 4.1. Performa algoritma dengan seleksi fitur.....	54
Tabel 4.2. Performa algoritma tanpa seleksi fitur .....	54
Tabel 4.3. Perbandingan accuracy, fl score, specificity, Precission Tanpa Seleksi Fitur dan dengan Seleksi Fitur.....	57
Tabel 4.4. Perbandingan Performa Akurasi Algoritma Tanpa Seleksi Fitur ..61	
Tabel 4.5. Perbandingan Performa Akurasi Algoritma Dengan Seleksi Fitur.61	
Tabel 4.6. Perbandingan Performa Precision Algoritma Tanpa Seleksi Fitur 63	
Tabel 4.7. Perbandingan Performa Precision Algoritma Tanpa Seleksi Fitur .....	64
Tabel 4.8. Perbandingan Performa Precision Algoritma Tanpa Seleksi Fitur .....	64
Tabel 4.9. Perbandingan Performa Precision Algoritma Tanpa Seleksi Fitur .....	65

## DAFTAR GAMBAR

Gambar 2.1 <i>Frame Work</i> Penelitian (Domala et al., 2020).....	11
Gambar 2.2 Klasifikasi SMS dengan Naïve Bayes.....	14
Gambar 2.3 Evolusi data(Agarwal, 2014).....	29
Gambar 2.4 Proses data mining (Han & Kamber, 2001).....	30
Gambar 2.5 Tahap-tahap mendapatkan pengetahuan dari database (Agarwal, 2014).....	31
Gambar 2.6 Machine Learning Algorithm Cheat Sheet (Azure, 2021).....	32
Gambar 2.7 DTC dengan <i>test</i> pada variabel X dan Y(Agarwal, 2014).....	36
Gambar 3.1 Halaman dataonline.bkg.go.id.....	40
Gambar 3.2 Halaman data jumlah penduduk.....	40
Gambar 3.3 Halaman luas wilayah.....	41
Gambar 3.4 Data bencana banjir.....	41
Gambar 3.5 Data curah hujan stasiun aceh besar.....	42
Gambar 3.6 Nama kolom 1 (satu) s/d 16.....	43
Gambar 3.7 Nama kolom 1 (satu) s/d 3 (tiga).....	44
Gambar 3.8 Nama kolom 4 (empat) hingga kolom 16.....	45
Gambar 3.9 Kolom POPUL_DENSITY.....	46
Gambar 3.10 Kolom RIVERS.....	46
Gambar 3.11 Kolom FLOODS.....	46
Gambar 3.12 Alur penggunaan algoritma pada penelitian.....	47
Gambar 3.13 Tahap prediksi banjir.....	47
Gambar 3.14 Kerangka kerja penelitian.....	48

Gambar 3.15 Persentase perbandingan Banjir 63,88% Tidak Banjir 32,12%.....51

Gambar 3.16 Persentase perbandingan setelah dilakukan *Random Over Sampling*.....52



## INTISARI

Mendapatkan informasi yang menarik dari kumpulan data / *data sets* disebut dengan *data mining*. Dari proses *data mining* ini dapat diperoleh berbagai informasi penting salah satunya adalah prediksi. Untuk memprediksi banjir sebagai salah satu bencana yang sering terjadi dapat menggunakan algoritma *data mining*. Selain algoritma, hal utama yang diperlukan untuk melakukan proses *data mining* adalah *data sets*.

Untuk membangun *data sets* banjir diperlukan faktor yang cukup banyak namun setidaknya terdapat 3 (tiga) perwakilan faktor yang diharapkan mampu merepresentasikan penyebab banjir sehingga dapat diprediksi secara tepat. Ketiga faktor tersebut adalah curah hujan, kepadatan penduduk dan daerah aliran sungai (DAS). Faktor-faktor ini umumnya terjadi di Indonesia. Sehingga dengan dibentuknya *data sets* yang bersumber dari data di Indonesia harapannya dapat memberikan kontribusi dalam penanganan banjir di Indonesia. Juga dapat menjadi rujukan research tentang kebencanaan dimasa yang akan datang. Selanjutnya Dengan membandingkan berbagai algoritma yang memungkinkan untuk dapat digunakan dalam memprediksi banjir, diharapkan mampu memberikan kontribusi terhadap perkembangan *data mining*

Dari hasil penelitian, akurasi yang dihasilkan oleh algoritma terhadap *data sets* cenderung kecil. Hal ini menjadikan *data sets* perlu untuk dikaji ulang karena hampir semua algoritma yang diuji memiliki akurasi yang kurang optimal. Dari sisi persentase akurasi ketika *data sets* diuji dengan seleksi fitur atau tanpa seleksi fitur cenderung berubah-ubah. Hal ini yang menjadi justifikasi perlunya *data sets* untuk dikaji ulang.

Kata kunci: banjir, *data sets*, *data mining*

## **ABSTRACT**

*Getting interesting information from data sets is called data mining. From this data mining process, various important information can be obtained, one of which is prediction. To predict floods as one of the disasters that often occur, data mining algorithms can be used. In addition to algorithms, the main thing needed to carry out the data mining process is data sets.*

*To build flood data sets, quite a lot of factors are needed but there are at least 3 (three) representative factors that are expected to be able to represent the causes of floods so that they can be predicted accurately. The three factors are rainfall, population density and watershed (DAS). These factors generally occur in Indonesia. So with the establishment of data sets sourced from data in Indonesia, it is hoped that they can contribute to handling floods in Indonesia. It can also be a reference for research on future disasters. Furthermore, by comparing various possible algorithms that can be used in predicting floods, it is hoped that they will be able to contribute to the development of data mining*

*From the research results, the accuracy generated by the algorithm on data sets tends to be small. This makes the data sets need to be reviewed because almost all the algorithms tested have less than optimal accuracy. In terms of the percentage of accuracy when data sets are tested with feature selection or without feature selection, it tends to fluctuate. This justifies the need for data sets to be reviewed.*

*Keywords: floods, data sets, data mining*

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Proses untuk mendapatkan informasi yang menarik dari data yang terhimpun disebut dengan *data mining* (Han & Kamber, 2001; Prihandoko et al., 2017). Informasi yang diperoleh dengan data mining setidaknya dapat dibagi menjadi 6 (enam) kategori yakni Deteksi Anomali, Aturan Asosiasi pembelajaran, *Clustering*, Klasifikasi dan Regresi (Refonaa et al., 2015). Dengan berbagai kemampuan ini, maka *data mining* saat ini banyak digunakan untuk mendapatkan informasi dan pengetahuan yang berguna di berbagai bidang (Kantardzic, 2020).

Bidang *data mining* juga digunakan untuk memprediksi bencana. Analisis probabiliti bencana menjadi penting mengingat bencana merupakan kondisi yang tidak diinginkan yang mengakibatkan kerugian kehidupan manusia dan menyebabkan pengungsian (Arjun Singh & Saxena, 2016). Pada prinsipnya bencana merupakan suatu kejadian yang mengganggu kondisi normal dan menimbulkan penderitaan yang melebihi kemampuan masyarakat yang terimbas (Panafrican Emergency Training Centre, 2002).

Salah satu bencana yang sangat berbahaya di Dunia adalah banjir (Naik et al., 2021). Selain kejadiannya yang sering terjadi dan mengakibatkan kerugian yang cukup besar seperti di Indonesia dalam Laporan Index Risiko Bencana Indoensia (IRBI) tahun 2020 menjabarkan bahwa bencana alam yang paling

banyak terjadi adalah Banjir yang mencapai angka 1.070 Kejadian (Sesa et al., 2020).

Karakteristik banjir di Indonesia telah berubah dari tahun tujuh puluhan yang pada awalnya lebih diakibatkan oleh faktor alam (Nugroho, 2002), berubah menjadi berbagai faktor yang lebih kompleks termasuk sosial (Nugroho, 2002). Meski terdapat berbagai faktor, namun untuk faktor curah hujan sebagai faktor alam, dan kepadatan penduduk sebagai faktor sosial menjadi perwakilan dalam penyebab banjir. Selain itu, menurut (Safitri, 2021) banjir ternyata terjadi pada wilayah yang dilalui sungai. Dengan adanya informasi ini, maka tambahan faktor jumlah sungai pada suatu wilayah menjadi penting sehingga harapannya dapat diketahui prediksi bencana dengan ada atau tidak adanya aliran sungai di wilayah tersebut.

Alasan pemilihan tiap faktor sebagai variabel penyebab banjir dijelaskan oleh salah satunya pernyataan (Nugroho, 2002) yang menyatakan salah satunya faktor alam yang menjadi penyebab banjir adalah jumlah curah hujan (Nugroho, 2002). Faktor curah hujan sebagai penyebab banjir juga diperkuat oleh (Naik et al., 2021). Untuk faktor sosial yang menjadi penyebab banjir lebih disebabkan oleh manusia salah satunya adalah kepadatan penduduk atau populasi (Tingsanchali, 2012). Selain curah hujan dan populasi, banjir juga disebabkan adanya daerah aliran sungai (DAS) (Safitri, 2021).

Setelah ditetapkannya faktor penyebab banjir, tahap selanjutnya mendapatkan data faktor penyebab yang sudah ditetapkan yakni : data curah hujan, data kepadatan penduduk, dan data jumlah sungai. Mengingat faktor



penyebab merupakan karakteristik yang ada di Indonesia, maka fokus penelusuran data juga berfokus pada wilayah Indonesia. Data yang digunakan sebanyak 24 Provinsi sebagai perwakilan wilayah di Indonesia. Yang tersebar di 3 (tiga) Pulau besar yakni Pulau Sumatera, Pulau Jawa dan Pulau Kalimantan selain itu beberapa pulau kecil seperti Kepulauan Riau, Bali dan Kepulauan Nusa Tenggara ditambahkan untuk dapat melihat daerah dengan luasan daratan yang lebih kecil. Pemilihan Provinsi pada Pulau-Pulau tersebut diharapkan merefleksikan seluruh wilayah di Indonesia, mengingat daerah tersebut memiliki jumlah kepadatan penduduk dan aliran sungai yang berbeda-beda.

Dengan ditentukannya Provinsi yang akan menjadi target penelitian, selanjutnya menentukan durasi data. Merujuk pada ketersediaan data dari <http://dataonline.bmkg.go.id> sebagai penyedia data, maka data yang dipilih adalah data dari tahun 2015-2020 pada tiap Provinsi. Hal ini dikarenakan data yang disajikan pada penyedia data tersebut cukup lengkap pada durasi tahun 2015-2020. Pemilihan pengunduhan data curah hujan dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) sendiri, hal ini lebih dikarenakan dikarenakan salah satu wewenang BMKG memberikan informasi tersebut (BMKG, 2021). Sedangkan untuk kepadatan penduduk yang menggunakan rumus luas wilayah dirujuk dari Badan Pusat Statistik (BPS) (B. Indonesia, 2019) dibagi dengan jumlah penduduk yang juga menggunakan sumber data dari BPS (B. Indonesia, 2020). Jumlah Sungai yang mengalir pada suatu Provinsi sebagai salah satu faktor penyebab banjir dirujuk dari data BPS (B. Indonesia, 2014) selain itu juga mengambil data dari halaman <http://p2k.unkris.ac.id/> pada artikel Daftar

sungai di Indonesia (Krisnadwipayana, -) pada artikel ini tidak menyertakan waktu penulisan.

Dengan adanya informasi sumber data penyebab banjir. Maka selanjutnya diperlukan data tentang kejadian banjir pada tiap daerah untuk menentukan label banjir dan tidak banjir. Untuk mendapatkan data ini diambil dari Data Informasi Bencana yang dikeluarkan oleh BNPB (B. Indonesia, 2022a).

Selain itu, pemilihan algoritma yang tepat dalam melakukan prediksi menjadi sangat penting, salah satu algoritma tradisional yang masih digunakan dalam *machine learning* yakni Naïve Bayes (Azure, 2021). Meskipun sebagai algoritma tradisional Naïve Bayes merupakan algoritma yang mampu mengklasifikasikan memberikan output yang tepat (Domingos & Michael Pazzani, 1997). Bukan hanya dianggap sebagai algoritma dalam *machine learning*, namun Naïve Bayes juga sering digunakan dalam menggali informasi melalui proses data mining (Peling et al., 2017).

## 1.2. Rumusan Masalah

Berdasarkan informasi awal pada latar belakang, dalam penelitian ini berusaha untuk :

- a. Bagaimana mengumpulkan data penyebab banjir di Indonesia yakni data curah hujan, data kepadatan penduduk dan data sungai serta data kejadian banjir sehingga membentuk *data sets*?
- b. Bagaimana *data sets* yang telah terbentuk dapat dianalisis lebih lanjut untuk memprediksi bencana dimasa yang akan datang dengan *data mining*?

- c. Apakah algoritma Naïve Bayes memiliki performa yang baik untuk melakukan prediksi banjir?

### 1.3. Batasan Masalah

Agar penelitian dapat mencapai target yang tepat, maka penelitian memiliki batasan-batasan sebagai berikut :

- a. Penelitian menggunakan data yang di *release* oleh badan resmi yang ditunjuk oleh Pemerintah Republik Indonesia.
- b. Pengukuran performa algoritma menggunakan berbagai *library* yang di sediakan oleh scikitlearn.org.
- c. Perbandingan performa bertujuan untuk mendapatkan algoritma yang mampu untuk menyelesaikan permasalahan sesuai dengan *data sets* yang tersedia.

### 1.4. Tujuan Penelitian

Tujuan Penelitian adalah sebagai berikut :

- a. Mendapatkan *data sets* faktor-faktor dan label (kelas) banjir bersumber dari data Indonesia.
- b. Memprediksi banjir dimasa depan berdasarkan hasil analisis *data sets* dengan *data mining*.
- c. Mengetahui apakah algoritma Naïve Bayes memiliki performa terbaik untuk memprediks banjir dengan *data sets* yang telah terbentuk.

### 1.5. Manfaat Penelitian

Penelitian diharapkan memiliki manfaat sebagai berikut :

- a. Terbentuk *data sets* banjir yang menjadi rujukan dalam bidang *data mining* bencana.
- b. Adanya informasi yang jelas mengenai probability banjir dimasa mendatang sehingga menjadi *early warning* bagi pemangku kebijakan.
- c. Diperoleh performa algoritma yang lebih baik dan tepat dalam memprediksi banjir.



## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Tinjauan Pustaka**

Curah hujan, merupakan salah satu faktor yang menyebabkan terjadinya banjir (Naik et al., 2021; Nugroho, 2002). Dengan memanfaatkan jumlah curah hujan sebagai satu-satunya variabel faktor penyebab banjir, penelitian yang dilakukan oleh (Naik et al., 2021) melakukan prediksi kemungkinan terjadinya banjir dimasa yang akan datang. Namun ternyata faktor curah hujan bukan hanya satu-satunya faktor penyebab banjir. Lebih jauh faktor socio-ekonomi juga ternyata sangat berpengaruh sebagai penyebab banjir (Ganguly. et al., 2019). Untuk faktor socio-ekonomi, (Tingsanchali, 2012) menyebutkan bahwa penyebab banjir diperkotaan sangat erat kaitannya dengan masalah kepadatan penduduk. Faktor lain penyebab banjir menurut (Safitri, 2021) mengatakan bahwa daerah aliran sungai (DAS) menjadi salah satu penyebab terjadinya banjir. Luapan DAS menjadi salah satu penyebab terjadinya banjir (Mosavi et al., 2018).

Dengan diperolehnya beberapa faktor yang menjadi penyebab banjir, selanjutnya daerah-daerah yang menjadi daerah rawan banjir sangat perlu diidentifikasi. Penelitian yang dilakukan oleh Saruni Dwiasnati dan Yudo Devianto (2021) meneliti tentang penentuan daerah rawan bencana banjir (Dwiasnati & Devianto, 2021). Dalam penelitian ini, tidak menginformasikan jenis data yang digunakan untuk dilakukan analisis. Namun menyertakan informasi bahwa proses pengunduhan data yang menggunakan dataminer (Dwiasnati & Devianto, 2021).

Tantangan dari penelitian (Dwiasnati & Devianto, 2021) tidak menyertakan informasi tentang detail data yang digunakan. Padahal jika merujuk (Mosavi et al., 2018) dekomposisi data sangat penting sehingga dapat memberikan informasi tentang kualitatif ketangguhan, akurasi, efektivitas, dan kecepatan algoritma. Meski tidak dapat disangkal bahwa hasil dari penelitian menyatakan bahwa akurasi yang dihasilkan oleh Support Vector Machines (SVM) mencapai 97,62% dengan menggunakan feature selection (Dwiasnati & Devianto, 2021).

Adanya informasi tentang sumber data yang digunakan oleh (Dwiasnati & Devianto, 2021) bahwa data diunduh dari website resmi BNPB, BPS Kabupaten Bandung dan BMKG. Hal ini menjadi informasi penting tentang sumber data dari Pemerintah dapat digunakan dan terpercaya keabsahannya. Penggunaan sumber data dari Pemerintah seperti yang dilakukan oleh (Naik et al., 2021) yang menggunakan sumber data dari pemerintah India mengukuhkan bahwa sumber data dari pemerintah sangat dapat digunakan.

Pemilihan algoritma tidak secara acak tanpa memperhatikan algoritma yang dimungkinkan tepat untuk menyelesaikan prediksi bencana banjir. Pemilihan merujuk dari sumber yang telah melakukan penelitian terdahulu seperti penelitian yang dilakukan oleh (Azlan et al., 2020) berusaha mengkomparasi 3 (tiga) algoritma yakni Naïve Bayes, K-Nearest Neighbour, dan Support Vector Machines (Azlan et al., 2020). Hasil komparasi dengan *data sets* yang diperoleh dari social media, algoritma Naïve Bayes cepat dalam proses perhitungan namun akurasinya rendah, K-Nearest Neighbour memiliki akurasi tinggi dan waktu yang sedang (Azlan et al., 2020). Sedangkan Support Vector Machines memiliki akurasi

tinggi dengan waktu rendah. Untuk akurasi sendiri, Naïve Bayes 59%, K-Nearest Neighbour 79% dan Support Vector Machines 79%(Azlan et al., 2020). Membandingkan beberapa algoritma juga dilakukan oleh (Gopal et al., 2020). Dengan *data sets* yang digunakan adalah mengkoleksi 211 artikel yang release 2020 (Gopal et al., 2020) yang berisi tentang tanah longsor, banjir dan hujan. Algoritma yang digunakan K-Nearest Neighbor, Support Vector Machines, Multi Layer Perception, Logistic Regression, dan Naïve Bayes, adapun hasil yang diperoleh sebagaimana terdapat pada Tabel 2.

Tabel 2.1 Hasil Classifikasi K-Nearest Neighbor, Support Vector Machines, Multi Layer Perception, Logistic Regression, dan Naïve Bayes (Gopal et al., 2020)

Algorithm	best n-gram	accuracy	f1	precision	recall
K-nearest Neighbor	bi-gram	75%	53%	68%	45%
Support Vector Machines	bi-gram	69%	69%	100%	5%
Multi Layer Perceptrons	bi-gram	77%	58%	72%	49%
Logistic Regression	uni-gram	72%	24%	97%	13%
Naïve Bayes	bi-gram	72%	21%	98%	12%

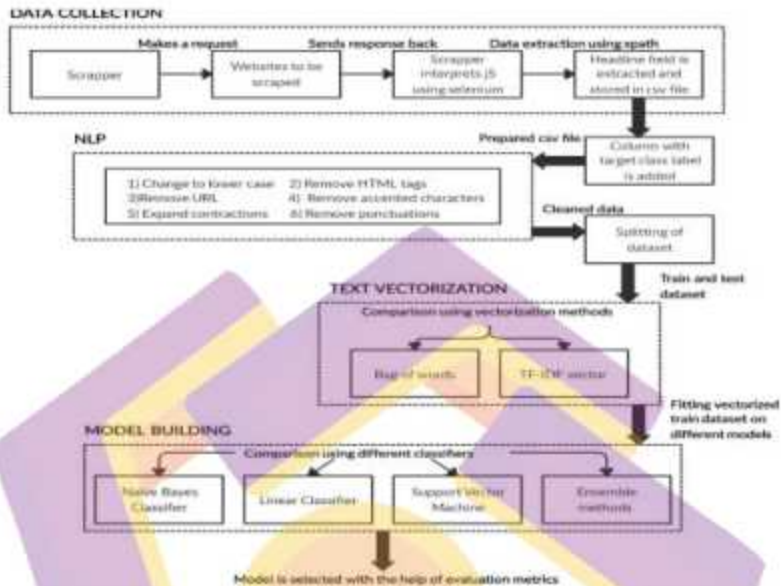
Informasi mengenai performa algoritma baik *accuracy*, *f1 Score*, *precision*, dan *recall* menurut (Ghoneim, 2019) dapat dijadikan acuan performa dengan catatan :

- *Accuracy* dapat menjadi rujukan performa jika *false negative* dan *false positif* memiliki jumlah yang hampir sama (*symmetric*);

- *F1 Score* menjadi rujukan performa jika kebalikan penggunaan *accuracy* yakni *false positif* memiliki jumlah yang berbeda dengan *false negative*.
- *Recall* dapat menjadi rujukan performa jika *false positif* lebih diharapkan terjadi dibandingkan dengan *false positive*.
- *Precision* dapat dijadikan sebagai ukuran performa jika diharapkan terjadi *true positif* dibandingkan dengan *false positif*.
- Pemilihan algoritma dengan *Specificity* tinggi luaran algoritma adalah *false positif*.

Penelitian mengenai bencana dari sisi data mining selanjutnya yang dijadikan tinjauan pustaka yakni penelitian yang dilakukan oleh (Domala et al., 2020). Penelitian ini melakukan scrapping data dari portal berita terpercaya yakni Times of Indina, NDTV India dan Indian Express(Domala et al., 2020). Hasil scrapping data tersebut selanjutnya data ditambah kolom label, baru dilakuan pembersihan data baik itu menghapus tag html, mangganti huruf kecil dan setelah didapatkan data bersih, data di bagi untuk dijadikan sebagai data *test* dan data *train*(Domala et al., 2020). Selanjutnya *data sets* dimasukkan ke dalam model/algoritma. Secara keseluruhan proses yang digunakan terdapat pada gambar berikut :





Gambar 2.1 *Frame Work* Penelitian (Domala et al., 2020)

Untuk algoritma yang digunakan sesuai dengan diagram adalah Multinomial Naïve Bayes, Logistic Regression, Support Vector Machines, Random Forest, dan Xtreme Gradient Boosting Model (Domala et al., 2020). Dari sisi presisi, diperoleh informasi Naïve Bayes sebesar 86%, Logistic Regression 88%, Support Vector Machines 87%, Random Forest 88%, dan Xtreme Gradient Boosting Model 88% (Domala et al., 2020). Unikinya ketika menggunakan *data sets* yang berbeda, ternyata masing-masing algoritma memiliki performa yang berbeda, seperti yang dilakukan oleh (Gopal et al., 2020). Selain itu penelitian yang mengkombinasikan antara metode klustering dan klasifikasi untuk memetakan jumlah sebaran pandemi covid-19 di Indonesia (Windarto et al.,

2020). Adapun data yang digunakan adalah data yang diperoleh dari covid19.go.id pertanggal 7 juli 2020(Windarto et al., 2020). Hasil dari penelitian ini yakni kombinasi metode klustering dan klasifikasi dapat diterapkan pada kasus persebaran pandemi covid-19 di Indonesia dengan memanfaatkan software bantu RapidMiner(Windarto et al., 2020). Selain itu, kombinasi dari hasil pemetaan wilayah berupa klaster diperoleh 9 (Sembilan) provinsi berada di klaster tinggi (C1=zona merah), 3 (tiga) provinsi berada di klaster waspada (C2=zona kuning) dan 22 provinsi berada di klaster rendah (C2=zona hijau)(Windarto et al., 2020).

Untuk melihat performa algoritma data *data mining*, penelitian untuk memprediksi kanker dengan *data sets* yang mencapai 1059 data (Maliha et al., 2019). Dengan jumlah data yang semakin besar ternyata memberikan hasil yang lebih baik. Terbukti dengan data yang merupakan data dikumpulkan dari Dokter dan para Expert(Maliha et al., 2019). Penelitian berusaha menemukan algoritma yang paling efektif untuk memprediksi kemungkinan pasien terkena *cancer*(Maliha et al., 2019). Terdapat 3 (tiga) algoritma supervise yang diterapkan. Merujuk dari (Azure, 2021) algoritma Naïve Bayes, KNN dan J48 merupakan algoritma supervisi. Mengingat pola hasil yang cenderung fluktuatif pada tiap performa yakni *accuracy, error rate, sensitivity, specificity, precision dan F-Score*. Hasil akhir dari penelitian ini diperoleh perbandingan sebagaimana terlihat pada tabel 3 berikut(Maliha et al., 2019).

Tabel 2.2 Hasil penelitian Cancer Disease Prediction using Naïve Bayes, K-Nearest Neighbor and J48 algorithm (Maliha et al., 2019)

<i>Class</i>	<i>Accuracy</i>	<i>Error rate</i>	<i>Sensitivity</i>	<i>Specify</i>	<i>Precision</i>	<i>F-Score</i>
Naïve Bayes	98.2%	2%	92.5%	99.1%	95.3%	93%
KNN	98.8%	1.2%	94.4%	99.2%	94.9%	94.6%
J48	98.5%	2%	94.4%	98.9%	95.1%	94.2%

Penerapan data mining dalam bidang kebencanaan juga telah dilakukan oleh (Tempola et al., 2018). Penelitian ini berusaha memprediksi status gunung berapi (Tempola et al., 2018). Sebagaimana faktor yang menyebabkan bencana seperti banjir yang memiliki banyak faktor (Prihandoko et al., 2017), namun dalam penelitian ini hanya menggunakan data *time series* berupa *history*. Dilatarbelakangi dari banyaknya gunung berapi di Indonesia yakni sebanyak 127 gunung berapi aktif. Dengan jumlah gunung berapi tersebut, penelitian memperoleh data bahwa pada 2017 terdapat 4 (empat) gunung berapi yang berstatus siaga (Tempola et al., 2018).

Dengan keadaan ini, dilakukan penelitian untuk memprediksi status gunung berapi yang ada di Indonesia (Tempola et al., 2018). Adapun untuk melakukan prediksi tersebut, Penelitian menggunakan metode metode Naïve Bayes Classifier (NBC) dengan data yang telah diolah sebelumnya menggunakan metode *K-Fold cross validation* (Tempola et al., 2018). Hasil dari penelitian ini diperoleh informasi bahwa Naïve Bayes bekerja sangat baik meski dengan sedikit

data training. Dari uji coba yang dilakukan diperoleh informasi bahwa akurasi yang dapat dicapai dengan metode ini mencapai 80,44%(Tempola et al., 2018).

Informasi mengenai keadaan bencana dapat diperoleh dari berbagai sumber termasuk salah satunya adalah SMS(Ordoñez et al., 2018). Dengan mengklasifikasikan SMS yang diterima oleh agensi menjadi 5 (lima) klasifikasi (Ordoñez et al., 2018), yakni Spam, Invalid, Alert 1, Alert 2, dan Alert 3(Ordoñez et al., 2018). Dengan memanfaatkan NLP (*Natural Language Processing*) SMS yang dikumpulkan dilakukan dilakukan tahap pembelajaran untuk klasifikasi selanjutnya diperoleh hasil klasifikasi tentang kebenaran informasi SMS yang diterima(Ordoñez et al., 2018).



Gambar 2.2 Klasifikasi SMS dengan Naïve Bayes

Dari hasil klasifikasi ini Naïve Bayes mampu memberi label dengan benar sebesar 89%.

Selain penelitian tentang klasifikasi bencana, penelitian ini juga merujuk kepada penelitian sebelumnya tentang *clustering* bencana. penelitian yang dilakukan oleh (Prihandoko et al., 2017) meneliti kemampuan algoritma K-Means dan K-Medoids dalam melakukan prediksi bencana di Indonesia(Prihandoko et al., 2017). Hal pertama yang dilakukan adalah mengunduh data bencana dari Badan Nasional Penanggulangan Bencana (BNPB) dari tahun 2011-2015 yang berisi informasi tentang provinsi, kota, tanggal kejadian bencana, jumlah

kematian, jumlah kehilangan, jumlah luka-luka, jumlah yang menderita, jumlah pengungsi, jumlah rumah rusak, jumlah kerusakan fasilitas, jumlah kerusakan jalan, dan jumlah kerusakan tanah (Prihandoko et al., 2017). Selanjutnya juga mengambil data dari Badan Meteorologi, Klimatologi dan Geofisika (BMKG) dari tahun 2011-2015 yang berisi informasi tentang stasiun cuaca, tanggal, temperature terendah, temperatur tertinggi, rata-rata temperatur dan curah hujan (Prihandoko et al., 2017).

Sebagai *sample data*, penelitian menggunakan 3 (tiga) provinsi sebagai representasi dari seluruh wilayah di Indonesia (Prihandoko et al., 2017). Adapun provinsi yang dijadikan *sample* adalah provinsi Jawa Barat, provinsi Jawa Tengah, dan provinsi Jawa Timur. Pemilihan ketiga provinsi ini lebih disebabkan karena jumlah penduduk yang banyak yakni sekitar 50% dari total penduduk Indonesia (Prihandoko et al., 2017).

Adapun proses penelitian yang dilakukan pertama adalah memproses data dengan algoritma K-Means. Proses ini meliputi hal-hal sebagai berikut (Prihandoko et al., 2017):

1. Menempatkan titik  $k$  ke ruang yang mewakili objek yang akan di kelompokkan. Titik ini sebagai representasi pusat group.
2. Tetapkan setiap objek dengan group yang paling dekat dengan titik pusat.
3. Setelah semua objek dimasukkan, kalkulasi ulang posisi central  $k$ ;
4. Ulangi langkah 2 (dua) dan 3 (tiga) hingga titik pusat tidak lagi berubah.

Prosedur ini merupakan pemisah objek kedalam grup dari matrik sehingga dapat diperkecil dan dapat dilakukan perhitungan.

Sedangkan untuk algoritma K-Medoids dilakukan dengan cara sebagai berikut (Prihandoko et al., 2017):

1. Pilih secara acak  $k$  dari  $n$  titik data sebagai medoid;
2. Gabungkan setiap titik data untuk mendekati medoid;
3. Untuk setiap medoid  $m$  dan setiap titik data  $o$  digabung dengan  $m$ , pindahkan  $m$  dan  $o$  dan hitung total biaya dari konfigurasi (yaitu ketidaksamaan rata-rata dari semua titik data yang tergabung ke dalam  $m$ ). pilih medoid  $o$  dengan nilai terendah dari konfigurasi.
4. Ulangi langkah 2 (dua) dan 3 (tiga) hingga tidak ada lagi perubahan pekerjaan.

Kesimpulan dari penelitian ini menyebutkan bahwa K-Means lebih tepat dibandingkan K-Medoids dalam menganalisis bencana alam dan data cuaca untuk menemukan pola di dalam tabel (Prihandoko et al., 2017).

## 2.2. Keaslian Penelitian

Tabel 2.3 Matriks literatur review dan posisi penelitian  
Prediksi Bencana Banjir Dengan Algoritma Supervisi

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1.	Flood Prediction using Logistic Regression for Kerala State	Saies naik, Anang Verma, Srushti Ashok Patil, Prof. Anil Hingmire, International Journal of Engineering Research & Technology (IJERT), 2021, pp. 36-38,	<ul style="list-style-type: none"> <li>a. Mendapatkan informasi probabilistic banjir menggunakan Logistic Regression;</li> <li>b. Memperoleh akurasi penggunaan algoritma Logistic Regression</li> </ul>	<ul style="list-style-type: none"> <li>a. Logistic Regression mampu melakukan prediksi banjir dengan akurasi 86,08%</li> </ul>	<ul style="list-style-type: none"> <li>a. Belum/tidak ditemukan perbandingan dengan algoritma lainnya sebagai pembandingan</li> </ul>	Penelitian hanya menjadikan curah hujan sebagai variabel x tanpa melihat berbagai faktor lain sebagai penyebab banjir
2.	Analyzing Algorithms to Detect Disaster Events using Social Media	F. A. Azlan, A. Ahmad, S. Yussof and A. A. Ghapar, 8th International Conference on Information Technology and Multimedia (ICIMU), 2020, pp. 384-389, doi: 10.1109/ICIMU49871.2020.9243599.	<ul style="list-style-type: none"> <li>a. melakukan perbandingan terhadap 3 (tiga) metode klasifikasi yakni KNN, NB dan SVM;</li> <li>b. Data set yang digunakan adalah pesan dari tweter;</li> </ul>	<ul style="list-style-type: none"> <li>a. Algoritma Naive Bayes cepat dalam memprediksi namun kurang akurat,</li> <li>b. K-Nearest Neighbour memiliki akurasi tinggi dan memberikan</li> </ul>	<ul style="list-style-type: none"> <li>a. tidak ada kejelasan informasi tentang jenis bencana yang diparsing dari twiter. Informasi hanya dijadikan sample yakni</li> </ul>	Karena tidak adanya informasi detail mengenai <i>data set</i> yang digunakan, maka dalam penelitian ini akan memberikan informasi detail mengenai <i>data sets</i> dan variabel-variabel yang akan dijadikan sebagai

Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				<p>prediksi yang baik,</p> <p>c. Support vector Machine memiliki akurasi tinggi namun lambat dalam memprediksi</p>	<p>b. gempa bumi dan kejadian tidak ditemukan apakah proses perhitungan akurasi ataupun perhitungan waktu eksekusi menggunakan system (program) yang ditulis ulang atau menggunakan system yang telah tersedia</p>	<p>rujukan dalam proses <i>data mining</i>.</p>
3.	Tsunami Potential Identification based on Seismic	A. Novianty, C. Machbub, S. Widiyantoro, I. Meilano and H. Irawan, IEEE 7th Conference on Systems, Process and Control (ICSPC), 2019, pp. 155-160, doi:	a. mengkaji kemungkinan pemanfaatan pendekatan machine learning dalam	a. hasil penelitian diketahui bahwa machine learning memiliki	a. tidak adanya perbandingan antar algoritma sebagai	Ada perbandingan dengan algoritma yang digunakan sebelumnya



Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	Features using KNN Algorithm	10.1109/ICSPC47137.2019.9068095.	<p>a. mengidentifikasi potensi tsunami dari suatu peristiwa gempa bumi sesuai dengan parameter seismic;</p> <p>b. analisis pola karakteristik gempa yang terkait dengan tsunami dan kinerja k-NN</p>	prospek yang baik untuk menangani masalah klasifikasi yang berkaitan dengan bencana alam, khususnya gempa bumi dan tsunami.	pembandingan akurasi	
4.	Automated Identification of Disaster News for Crisis Management using Machine Learning and Natural Language Processing	J. Domala et al., International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 503-508, doi: 10.1109/ICESC48915.2020.9156031.	<p>a. Mendapatkan solusi terbaik untuk menganalisis kebenaran web berita tentang kondisi bencana menggunakan beberapa metode klasifikasi yakni : naïve Bayes, Logistic Regression, SVM, Xtreme Gradient Boosting Dan Random Forest</p>	<p>a. Logistic Regression lebih baik dalam melakukan klasifikasi dibandingkan dengan algoritma lain;</p> <p>b. Akurasi Logistic Regression mencapai 89%</p>	a. Tidak ditemukan informasi detail mengenai tool yang digunakan	Akan menggunakan library dari scikit-learn.org dan disandingkan dengan rapid minner sebagai pembandingan
5.	An Analysis	Prihandoko, dkk (IEE-2017)	a. Menemukan korelasi	a. Ditemukan	a. Penelitian	Dalam penelitian

Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	of Natural Disaster Data by Using K-Means and K-Medoids Algorithm of Data Mining Techniques		<p>antara kondisi cuaca dan saat kejadian bencana alam;</p> <p>b. Membandingkan metode yang paling tepat dalam pengelompokan dan prediksi pada 2 (dua) metode yakni k-means dan k-medoids</p>	<p>informasi bahwa jumlah korban jiwa meningkat seiring dengan semakin tingginya intensitas dan kerusakan lahan serta peningkatan jumlah pengungsi merupakan objek yang paling terpengaruh terhadap tingginya intensitas hujan.</p> <p>b. Hasil penelitian menyimpulkan bahwa metode k-means lebih</p>	<p>menyasar pada daerah tingkat Provinsi. Sehingga belum dapat menjadi rujukan bagi Kabupaten/Kota;</p>	<p>yang akan dilakukan akan lebih spesifik terhadap Kabupaten Kota sehingga diharapkan dapat menjadi rujukan ilmiah bagi <i>stake holder</i> Kabupaten Kota yang menjadi objek penelitian</p>

Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				tepat dibandingkan k-medoids dalam menganalisa data bencana alam dan cuaca.		
6.	SMS Classification Method for Disaster Response Using Naïve Bayes Algorithm	A. Ordonez, R. E. Paje and R. Naz, International Symposium on Computer, Consumer and Control (IS3C), 2018, pp. 233-236, doi: 10.1109/IS3C.2018.00066.	<p>a. Penelitian ini adalah melakukan klasifikasi terhadap SMS yang diterima pada saat terjadi bencana sehingga dapat memberikan respon yang tepat;</p> <p>b. naïve bayes dianggap sebagai metode yang paling efektif dalam penanganan permasalahan ini</p>	<p>a. Naïve Bayes Formula untuk mengklasifikasi pesan SMS menurut lima klasifikasi yang berbeda berdasarkan kumpulan informasi praklasifikasi yang digunakan sebagai pengklasifikasi yang dipelajari;</p> <p>b. Akurasi Naïve Bayes sebesar</p>	c. Tidak ada perbandingan antara Naive Bayes dengan algoritma lainnya	Perbandingan antara penelitian sebelumnya juga terdapat perbandingan antara tools yang digunakan

Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
7.	Prediction Future Disaster Using Convex Hull & K-Mean, An Approach	Abhinab Basak, dkk (IEE 2019)	<p>a. Mendapatkan informasi probabilitas bencana berdasarkan informasi bencana sebelumnya dan pengaruhnya terhadap kehidupan sosial masyarakat</p> <p>b. Memperoleh akurasi jenis dan probabilitas bencana dengan menggunakan metode Convex Hull &amp; K-Means</p>	<p>89%</p> <p>a. Diperoleh informasi akurasi kemungkinan bencana sebesar 76% dengan rumus <math>Accuracy = \frac{\text{no of matched records}}{\text{total no of records}}</math></p>	<p>a. Perolehan akurasi probabilitas hanya menggunakan n 1 (satu) metode yakni K-Means, dan belum diuji lebih lanjut dengan metode lainnya</p>	Menggunakan 2 (dua) metode sehingga akan diperoleh informasi yang diharapkan lebih akurat
8.	Naïve Bayes Classifier For Prediction Of Volcanic Status In Indonesia	Firman Tempola dkk (IEEE 2018)	<p>a. Memperoleh akurasi informasi probabilitas bencana dengan menggunakan metode naïve bayes</p>	<p>a. Diperoleh informasi bahwa naïve bayes dengan 23 data training diperoleh akurasi sebesar 80,44% dan ketika diuji dengan 44 data</p>	<p>a. Tidak ditemukan informasi sumber data yang digunakan</p>	Menggunakan sumber data dari Badan Penanggulangan Bencana Daerah Kabupaten Kampar

Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				terjadi penurunan yakni menjadi 78,26%		
9.	Analysis and prediction of natural disaster using spatial data mining technique	J. Refonaa, M. Lakshmi and V. Vivek International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], 2015, pp. 1-6, doi: 10.1109/ICCPCT.2015.7159379.	a. implementasi mapreduce untuk mengolah data spasial	implementasi berhasil dilakukan dalam pengembangan dan penerapan teknik komputasi baru untuk menganalisis basis data spasial yang besar. permasalahan yang muncul dalam pengembangan yakni penggunaan <i>data sets</i> spasial seperti data cuaca, letusan gunung berapi, gempa bumi, pola kolokasi, asosiasi spasial pertambangan. Oleh karena itu,	a. penelitian merupakan penelitian terapan. Namun tidak ditemukan kejelasan informasi dan tidak ditemukan abstraksi sebagai penggambaran umum tentang tujuan dan kesimpulan dari paper	Penerapan/implementasi Tidak menjadi prioritas dalam penelitian

Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				pekerjaan lebih lanjut akan diproses dengan kerangka CMR yang akan diimplementasikan di infrastruktur awan untuk mengidentifikasi wilayah pesisir yang terkena badai.		
10.	Hybrid Data Model for Prediction of Disaster using Data Mining Approaches	Singh, A., & Saxena, A. (2016). international journal of engineering trends and technology, 41, 384-392.	<p>a. melakukan ekstrak data dari web data source dengan menggunakan google news api kemudian melakukan analisis terhadap data tersebut.</p> <p>b. Untuk webdata source yang digunakan <a href="https://www.emdat.be/">https://www.emdat.be/</a></p>	a. proses ekstrak data berhasil dilakukan dengan bantuan google news api. Kemudian data selanjutnya langkah pre-processed data menggunakan Bays classifier untuk mengklasifikik	a. Tidak ditemukan informasi tentang kata yang digunakan untuk ekstrak data text yang berhasil di unduh.	Akan menggunakan <i>data sets</i> yang dibuka secara <i>public</i>

Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				asikan data berupa text. Dan terahir menggunakan algoritma K-Means untuk melakukan klustering terhadap data dan menggunakan Hiden Markov Model (HMM).		
11.	Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir	Dwiasnati, S., & Devianto, Y. (2021). Prosiding SISFOTEK, 5(1), 202-207.	<p>a. Tujuan penelitian adalah untuk mengetahui daerah rawan bencana menggunakan teknik data mining,</p> <p>b. Algoritma data mining menggunakan SVM (Support Vector Machine)</p>	a. Setelah melakukan penelitian dengan model algoritma SVM tanpa feature selection dan dibandingkan dengan model SVM dengan feature selection dapat kita simpulkan	a. penggunaan tool rapid miner dalam proses pengolahan data menjadikan kebenaran proses dan apa yang terjadi terhadap data set	Akan menggunakan 2 (dua) tools berbeda sebagai perbandingan yakni library dari scikit-learn.org serta RapidMiner

Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				<p>bahwa penggunaan feature selection Weight by Correlation dapat meningkatkan nilai dari Accuracy dan AUC. Peningkatan yang didapatkan sangat signifikan yang sebelumnya model SVM tanpa feature selection hanya menghasilkan 66,49% dan nilai AUC 0,716 setelah ditambahkan penggunaan</p>	<p>kurang bisa dipahami. Sehingga kebenaran informasi perlu untuk diuji untuk mengetahui kebenaran.</p>	



Tabel 2.3 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				feature selection menjadi 81.18% untuk accuracy dan nilai AUC 0.943.		
12.	Analyzing Algorithms to Detect Disaster Events using Social Media	F. A. Azlan, A. Ahmad, S. Yussof and A. A. Ghapar, 8th International Conference on Information Technology and Multimedia (ICIMU), 2020, pp. 384-389	<ul style="list-style-type: none"> <li>a. Mendeteksi informasi menggunakan data dari sosial media</li> <li>b. Algoritma yang digunakan Naïve Bayes, KNN, SVM</li> </ul>	<ul style="list-style-type: none"> <li>a. Algoritma Naive Bayes cepat dalam melakukan prediksi namun kurang akurat;</li> <li>b. KNN akurasi tinggi dengan kecepatan prediksi sedang;</li> <li>c. SVM akurasi tinggi dengan kecepatan prediksi lambat</li> </ul>	a. Penelitian tidak menyertakan <i>data sets</i> sehingga tidak dapat dilakukan penelitian lebih lanjut	Akan menyertakan <i>data sets</i> dan bersifat public

### 2.3. Landasan Teori

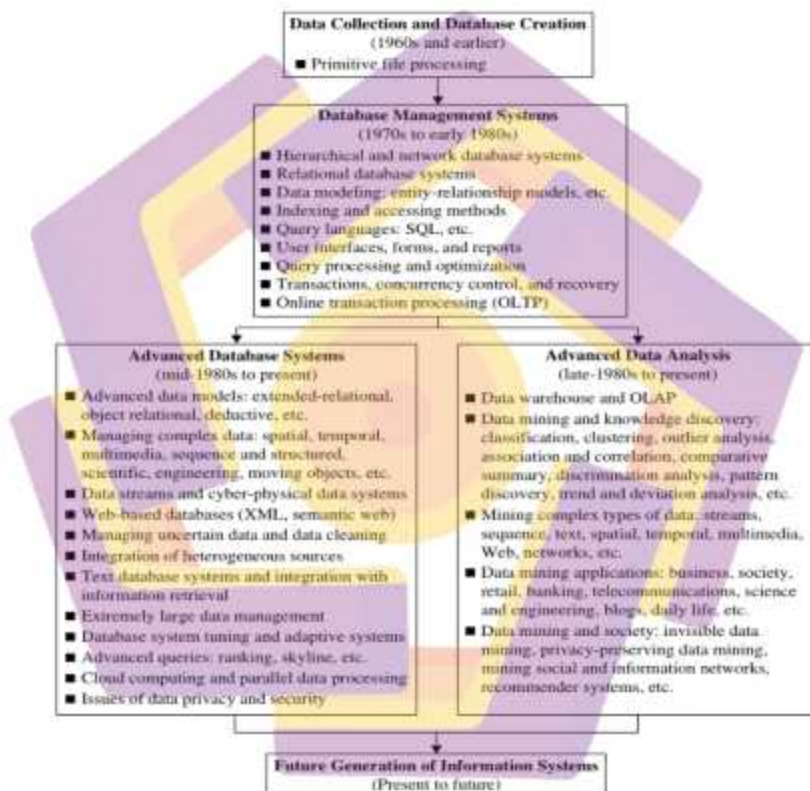
Perkembangan *data mining* diawali dari *database management system* (DBMS) (Han & Kamber, 2001). Setelah DBMS berjalan dengan baik, selanjutnya teknologi berkembang dengan meningkatkan kemampuan dari DBMS selanjutnya yakni *data warehousing*, dan terakhir saat ini adalah *data mining* yang digunakan untuk meningkatkan analisis data dan database pada web (Agarwal, 2014). Sejalan dengan itu, perkembangan data baik yang tersimpan dalam komputer, jaringan maupun Pemerintah, Lembaga ilmiah dan bisnis mampu mengoleksi data yang cukup banyak (Kantardzic, 2020).

Data yang dikoleksi oleh Badan Pemerintah sesuai dengan kewenangannya merujuk dari undang-undang keterbukaan informasi publik sedianya dapat diunduh dan dimanfaatkan oleh Masyarakat sesuai peruntukannya (R. Indonesia, 2008). Salah satu lembaga non Kementerian yang ditunjuk oleh Pemerintah untuk menyediakan data baik untuk Pemerintah maupun Masyarakat adalah BPS (B. Indonesia, 2022b). Selain itu, sesuai dengan kewenangan BMKG menyediakan data berkaitan dengan Meteorologi, Klimatologi dan Geofisika (BMKG, 2021) termasuk di dalamnya adalah data curah hujan. Lembaga Pemerintah lainnya sesuai dengan kewenangannya untuk menyediakan data kebencanaan adalah BNPB (Republik Indonesia, 2008).

#### 2.3.1 Data Mining

*data mining* pada prinsipnya adalah cara untuk mendapatkan informasi (*knowledge*) dari data (Agarwal, 2014). Sedangkan Evolusi data sendiri dimulai semenjak awal 1960 dimulai dengan pemrosesan file secara primitif baru

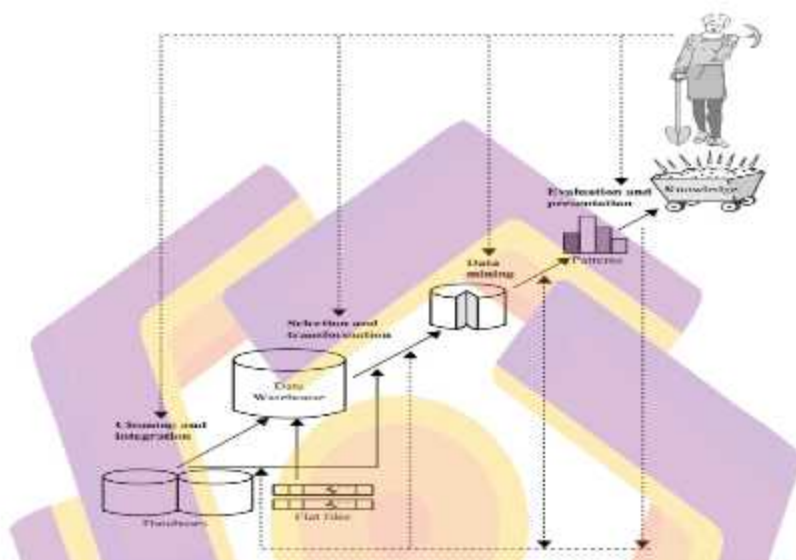
kemudian masuk ketahap *Database Management System*, *Advanced Database System*, *Advanced Data Analysis* bbaru terakhir adalah *data mining* (Han & Kamber, 2001). Secara keseluruhan tahapan evolusi data terdapat pada gambar berikut :



Gambar 2.3 Evolusi data(Agarwal, 2014)

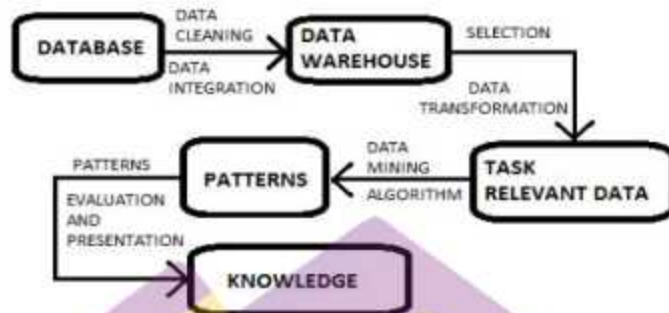
Pada gambar evolusi data, data mining dimulai pada akhir 1980-an hingga saat ini. Pada prinsipnya proses *mining* bertujuan untuk mendapatkan informasi dari gudang data (Han & Kamber, 2001). Proses *data mining* untuk mendapat

pengetahuan dari data atau disebut dengan *knowledge discovery from data* (KDD) terlihat pada gambar berikut :



Gambar 2.4 Proses data mining (Han & Kamber, 2001)

Selain itu, Shivam Agrawal (2014) lebih lanjut menjelaskan tentang proses mendapatkan informasi dari database yang penyebutan proses ini dengan istilah KDD, memperlihatkan tahap demi tahap proses mendapatkan pengetahuan dari database sebagai berikut :



Gambar 2.5 Tahap-tahap mendapatkan pengetahuan dari database (Agarwal, 2014)

Sehingga, untuk mendapatkan informasi menurut han & kember (2001) terdapat 7 (tujuh) langkah yang harus dilakukan yakni :

- (1) *Data Cleaning* (berfungsi untuk membuang *noise* dan data yang tidak konsisten);
- (2) *Data Integration* (berfungsi untuk menggabung data dari berbagai sumber);
- (3) *Data Selection* (yakni proses menseleksi data untuk mendapatkan data yang relevan);
- (4) *Data transformation* (yakni proses untuk mentransformasi/mengubah data kedalam bentuk yang sesuai dengan informasi yang sedang digali);
- (5) *Data mining* (merupakan tahap untuk memperoleh informasi dengan metode tertentu);
- (6) *Pattern evaluation* (langkah ini adalah melakukan identifikasi kebenaran informasi yang diperoleh dari hasil data mining);



Meski demikian, tidak menutup kemungkinan untuk menggunakan algoritma dengan berbagai kebutuhan yang tidak sama dengan *cheatshet*. Hal ini dikarenakan *data mining* merupakan proses berulang tanpa ada gagasan yang telah ditentukan sebelumnya tentang hal yang akan ditemukan (Kantardzic, 2020).

### 2.3.2 Prediction/prediksi/peramalan *Data Mining*

Dalam metode *clustering*, pengelompokan objek hanya dilakukan dengan melihat persamaannya (Agarwal, 2014). Berbeda dengan *prediction* yang justru mencoba menghubungkan suatu *cluster* dengan *cluster* lainnya untuk mendapat informasi yang berguna lainnya (Agarwal, 2014). Salah satu algoritma yang digunakan untuk melakukan prediksi atau kemungkinan adalah Naïve Bayes (Domingos & Michael Pazzani, 1997). Naïve Bayes merupakan metode utama untuk menentukan probabilitas (Albon, 2018). Penggunaan metode ini sangat populer dalam *mesin learning* lebih dikarenakan hal-hal sebagai berikut (Albon, 2018):

1. Pendekatan intuitif;
2. Mampu bekerja dengan data yang kecil;
3. Hasil yang diperoleh baik.

dengan formula sebagai berikut (Albon, 2018):

$$(1) \quad P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Dengan mengikuti prinsip ini, kemungkinan informasi baru  $P(A|B)$  dapat diperoleh.

Selain itu, naïve bayes juga terus dikembangkan. Salah satu algoritma dalam Naïve Bayes adalah Gaussian Naïve Bayes (<https://scikit-learn.org/>, n.d.). Dalam algoritma Gaussian Naïve Bayes, formulai yang digunakan adalah sebagai berikut (<https://scikit-learn.org/>, n.d.):

$$(2) \quad P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \quad (2)$$

Sebagai pembandingan, untuk mengetahui performa Naïve Bayes, maka digunakan 4 (empat) algoritma pembandingan. Ke empat algoritma ini dipilih dikarenakan memiliki ciri khas yang berbeda. Algoritma yang akan dijadikan sebagai komparai yakni :

1. K-Nearest Neighbour (KNN)
2. Logistic Regression (LR)
3. Decision Tree Classification (DTC)
4. Random Forest (RMF).

Salah satu algoritma yang biasa digunakan dalam melakukan klasifikasi dengan *data sets* yang besar adalah k-nearest neighbor (KNN) *classification* (Kantardzic, 2020). KNN biasa disebut sebagai pembelajaran berbasis memori (Kantardzic, 2020). Secara matematis KNN dari scikit-learn.org menjabarkan formula KNN dalam menentukan kelas sebagai berikut :

$$(3) \quad P_{ij} = \frac{\exp(-\|Lx_i - Lx_j\|^2)}{\sum_{k \neq i} \exp(-\|Lx_i - Lx_k\|^2)}, p_{ii} = 0$$



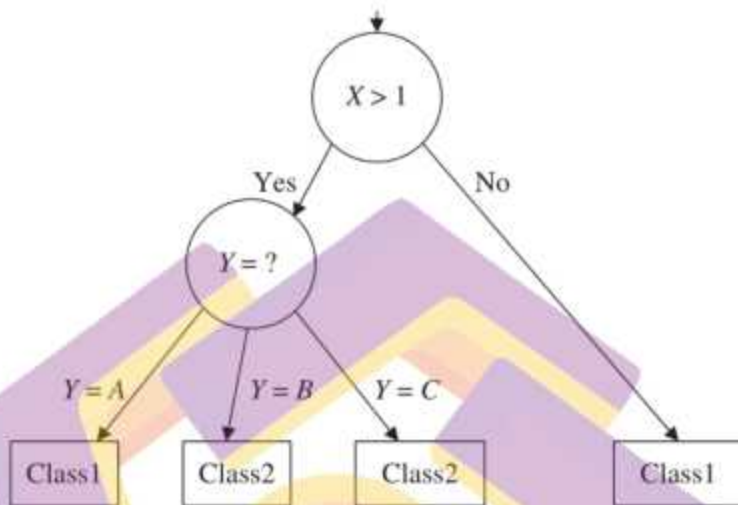
Untuk memodelkan fungsi nilai kontinu salah satu yang digunakan adalah model Regresi linier. Umumnya model regresi mewakili landasan teoretis yakni pendekatan regresi linier yang diterapkan pada model variabel respons kategoris. Salah satu jenis model linier adalah *logistic regression* (LR) (Agarwal, 2014).

Dalam proses memprediksi LR akan menilai variabel dependen, metode regresi logistik mencoba memperkirakan probabilitas variabel dependen yang memiliki nilai tertentu (Agarwal, 2014). Secara matematis, formula dalam LR dapat dilihat sebagai berikut :

$$(4) \quad \log\left(\frac{p_j}{(1-p_j)}\right) = \alpha + \beta_1 \cdot X_{1j} + \beta_2 \cdot X_{2j} + \beta_3 \cdot X_{3j} + \dots + \beta_n \cdot X_{nj}$$

Formula LR (Agarwal, 2014)

Algoritma DTC merupakan salah satu metode logika yang paling banyak digunakan (Agarwal, 2014). Metode ini sangat efisien untuk mengklasifikasi data dengan menghasilkan pohon keputusan (Agarwal, 2014). Metode DTC merupakan metode pembelajaran yang diawasi (supervised) yang membangun pohon keputusan (*decision tree*) dari satu set sampel input-output. Hal ini menjadikan DTC merupakan metode nonparametrik yang efisien untuk klasifikasi dan regresi (Agarwal, 2014). DTC merupakan model hierarki untuk pembelajaran terawasi dengan cara identifikasi lokal dan dilakukan pengurutan pemisahan rekursif melalui simpul keputusan dengan fungsi uji (Agarwal, 2014).



Gambar 2.7 DTC dengan test pada variabel X dan Y(Agarwal, 2014)

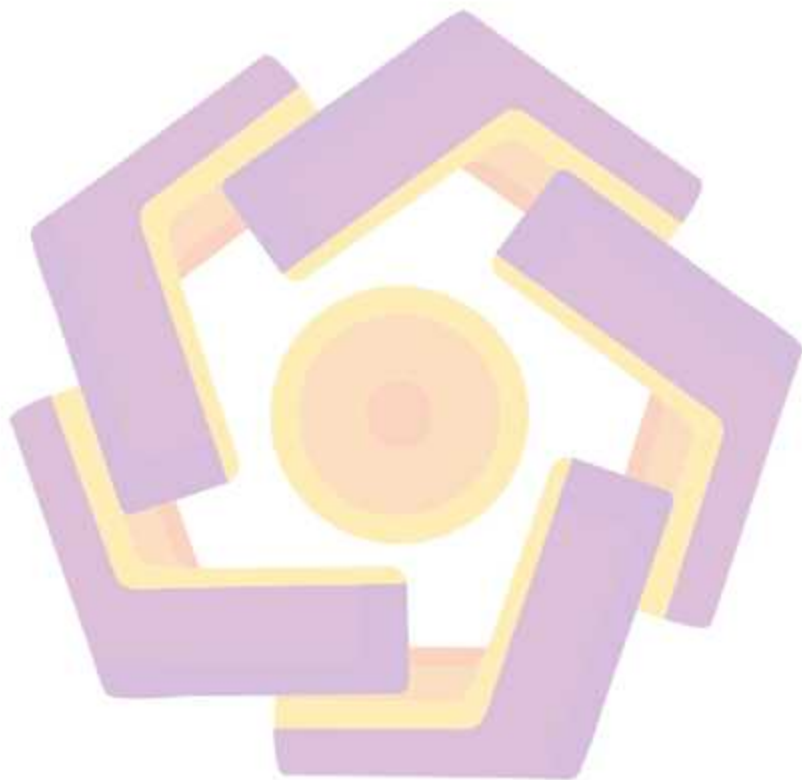
Dalam DTC menggunakan kriteria yang disebut dengan gain untuk memilih atribut yang akan diuji (Agarwal, 2014). Merujuk dari konsep ini, maka formula DTC adalah sebagai berikut :

$$(5) \quad \text{Info}(T) = - \sum_{i=1}^k \left( \left( \frac{\text{freq}(C_i)}{|T|} \right) \cdot \log_2 \left( \frac{\text{freq}(C_i)}{|T|} \right) \right)$$

Formula DTC(Agarwal, 2014)

Salah satu algoritma yang cocok untuk menganalisis struktur data kompleks dengan *data sets* kecil hingga sedang yang biasanya berisi kurang dari 10.000 baris namun tetap memungkinkan ketika *data sets* lebih dari 1 juta baris adalah RF (Agarwal, 2014). RF sendiri menurut (defeloper, 2022) merupakan meta estimator yang mampu mengklasifikasi *decision tree* dari *data sets* dan

mampu untuk meningkatkan akurasi prediksi dan kontrol over-fitting (defeloper, 2022).



## BAB III

### METODE PENELITIAN

#### 3.1. Jenis, Sifat, dan Pendekatan Penelitian

Dalam menentukan performa algoritma, penelitian menggunakan *data sets* yang terukur dengan justifikasi yang jelas terhadap penggunaan setiap variabel. Dengan keadaan ini, maka penelitian dapat digolongkan sebagai penelitian kuantitatif. Penelitian ini bukan merupakan penelitian baru namun menggali performa algoritma Naïve Bayes dalam menangani datasets yang telah terbentuk. Sebagai perbandingan, maka Naïve Bayes akan dibandingkan berbagai algoritma klasifikasi lainnya dengan melihat kemampuan dalam menghitung *accuracy*, *f1 score*, *specify*, dan *precision*.

Dikarenakan merupakan penelitian kuantitatif, angka dan informasi lainnya merupakan angka yang memiliki sumber yang valid dan bukan merupakan asumsi. Sumber data yang digunakan merupakan data yang berasal dari agensi Pemerintah yang memiliki wewenang untuk mengeluarkan data. Pendekatan dengan menggunakan data yang dikeluarkan oleh agensi resmi menjadikan data dan informasi dalam penelitian lebih mudah untuk ditelusuri kebenarannya.

#### 3.2. Metode Pengumpulan Data

Data dan informasi yang akan dikumpulkan sesuai dengan rencana penelitian terdiri atas data curah hujan (mm), kepadatan penduduk yang berisi jumlah jiwa dalam suatu provinsi dibagi dengan luas wilayah provinsi

(jiwa/km<sup>2</sup>), jumlah sungai (buah), dan status banjir dan tidak banjir. Data ini akan diunduh dari berbagai sumber yakni :

1. Data curah hujan diunduh dari <https://dataonline.bmkg.go.id>;
2. Data kepadatan penduduk merupakan hasil dari formula sebagai berikut :

$$(6) \text{ Kepadatan Penduduk} = \frac{\text{Jumlah penduduk (jiwa)}}{\text{Luas Wilayah (KM}^2\text{)}}$$

- Jumlah penduduk diunduh dari halaman :

<https://www.bps.go.id/indikator/12/1886/1/jumlah-penduduk-hasil-proyeksi-menurut-provinsi-dan-jenis-kelamin.html>

- Luas wilayah diunduh dari halaman :

[https://www.bps.go.id/indikator/indikator/view\\_data\\_pub/0000/api\\_public/UFpWMmJZOVZIZTJnc1pXaHhDV1hPQT09/da\\_01/1](https://www.bps.go.id/indikator/indikator/view_data_pub/0000/api_public/UFpWMmJZOVZIZTJnc1pXaHhDV1hPQT09/da_01/1)

3. Data jumlah sungai dimabil dari laman <http://p2k.unkris.ac.id/id3/1-3065-2962/Daftar-Sungai-Di-Indonesia-26973-stie-walisongo-p2k-unkris.html>
4. Sedangkan data untuk menentukan suatu provinsi terjadi banjir atau tidak merupakan hasil dari formula

$$(7) \text{ Rata - Rata Banjir} = \frac{\text{jumlah kejadian dalam Tahun}}{12 \text{ bulan}}$$

Data jumlah kejadian bencana di ambil dari <https://dibi.bnpb.go.id/xdibi>  
Masing-masing halaman pengunduhan data terlihat pada tangkapan layar sebagai berikut :



Gambar 3.1 Halaman dataonline.bkg.go.id



Gambar 3.2 Halaman data jumlah penduduk

Perencanaan Pembangunan

Luas Daerah dan Jumlah Pulau Menurut Provinsi, 2021

Provinsi/Provinsi	Luas Wilayah (km <sup>2</sup> )	Jumlah Wilayah (Pulau)	Perubahan Tahunan Luas Wilayah	Jumlah Pulau
Bangka	57.065,00			500
Bangka-Bangka	57.065,00			500
Bengkulu	42.912,20			229
Bengkulu-Bengkulu	42.912,20			229
Buru	87.021,00			104

Gambar 3.3 Halaman luas wilayah

No	Kode	Nama	Tanggal	Kategori	Status
1	43334-01	Kab. Muar	08/01/2018	03/01/18	0-0-0
2	43334-02	Kab. Muar	08/01/2018	03/01/18	0-0-0
3	43334-03	Kab. Muar	08/01/2018	03/01/18	0-0-0
4	43334-04	Kab. Muar	08/01/2018	03/01/18	0-0-0
5	43334-05	Kab. Muar	08/01/2018	03/01/18	0-0-0
6	43334-06	Kab. Muar	08/01/2018	03/01/18	0-0-0
7	43334-07	Kab. Muar	08/01/2018	03/01/18	0-0-0
8	43334-08	Kab. Muar	08/01/2018	03/01/18	0-0-0
9	43334-09	Kab. Muar	08/01/2018	03/01/18	0-0-0
10	43334-10	Kab. Muar	08/01/2018	03/01/18	0-0-0

Gambar 3.4 Data bencana banjir

Jenis dan dan cara pengolahannya pada setiap data berbeda. Seperti halnya data curah hujan yang diunduh dari BMKG merupakan data dalam file berformat *xlsx*. Data curah hujan merupakan data *time series* harian yang dapat diunduh perbulan dengan bentuk sebagai berikut :



ID WMO : 96013  
 Nama Stasiun : Stasiun Hidrologi Aceh Besar  
 Ulang : 5.49.000  
 Rujuk : 95.29600  
 Elevasi : 52

Tanggal	Tm	Tu	Tang	Rujuk	RR
01-01-2015					1,5
02-01-2015					8888
03-01-2015					0
04-01-2015					0
05-01-2015					1,5
06-01-2015					0
07-01-2015					
08-01-2015					0
09-01-2015					8888
10-01-2015					11,5
11-01-2015					
12-01-2015					1
13-01-2015					0
14-01-2015					9,0
15-01-2015					4
16-01-2015					5,5
17-01-2015					1,5
18-01-2015					8888
19-01-2015					0
20-01-2015					0
21-01-2015					
22-01-2015					
23-01-2015					0
24-01-2015					0
25-01-2015					0
26-01-2015					
27-01-2015					8888
28-01-2015					0
29-01-2015					
30-01-2015					0
31-01-2015					0

Keterangan:  
 8888: data tidak terukur  
 8888: Tidak Ada Data (tidak dilakukan pengukuran)  
 Tm: Temperatur maksimum (°C)  
 Tu: Temperatur minimum (°C)  
 Tang: Temperatur rata-rata (°C)  
 Rujuk: kelembapan rata-rata (%)  
 RR: Curah hujan (mm)

Gambar 3.5 Data curah hujan stasiun aceh besar

Hasil pada kolom curah hujan (RR) selanjutnya dilakukan *summary*. Untuk melakukan penjumlahan sesuai dengan keterangan baris yang bernilai 8888 merupakan baris yang tidak terukur dan baris ini dianggap bernilai 0. Sedangkan untuk baris yang tidak terisi (*blank space*) juga dianggap bernilai 0 atau tidak



diperhitungkan. Sehingga formula yang digunakan dalam melakukan penjumlahan adalah sebagai berikut :

(8) =SUMIFS(\$F\$10:\$F\$540;\$F\$10:\$F\$540;"<>8888")

Hasil dari *summary* ini nantinya akan di tempatkan pada kolom dan baris sesuai dengan Provinsi data diunduh dan tahun serta bulan. Sehingga penyusunan *data set* berbentuk sebagai berikut :

No	PROV	YEARS	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
1	ALIR	2023	80	35	47	143,4	76,8	73,3	181,9	170,4	203,5	51,8	223,3	188,3				
2	ACDI	2020	222,2	140	98,3	62,4	237	112,9	112,2	111,7	170,9	446,4	260,7	209,1				
3	ATJH	2021	310	89	173,3	48	201	112,9	112,2	111,7	170,9	446,4	260,7	209,1				
4	ACDI	2020	141	46,2	36	227	214,3	88,3	120	120,6	142,8	261,2	227,5	276,8				
5	ALIR	2024	144,5	89	140	118,9	113,9	9	39	47,9	1,4	90	391,2	121				
6	ACDI	2020	116	34	34	15,8	182,8	421,4	141,7	117,2	1,8	28,6	72,8	220,8	174,1			
7	DIKAWITANA UTARA	2021	187,6	84,8	219,8	171	118,2	188,2	189,7	189,7	189,6	627,9	488,4	184,8				
8	DIKAWITANA UTARA	2020	130,4	302,4	360,8	35,7	171,7	113,4	181,2	201,8	280,8	327,2	197	154,8				
9	DIKAWITANA UTARA	2021	284,3	116,6	111,9	188,8	181,1	119,9	181,2	201,8	280,8	327,2	197	154,8				
10	DIKAWITANA UTARA	2020	224,3	281	38,8	381,8	113,8	284,3	307	116,7	171,5	191,4	243,8	277,1				
11	DIKAWITANA UTARA	2020	336,1	288,7	188,4	193,4	428,4	178,4	146,1	191,9	76,6	30,4	86,1	191,4				
12	DIKAWITANA UTARA	2020	80,9	80	101,7	137	137	137	137	137,1	401,7	401,7	136,6	136,1				
13	DIKAWITANA BARAT	2021	272,6	284,1	271	195,9	146,4	284,4	146,1	191,9	76,6	30,4	86,1	191,4				
14	DIKAWITANA BARAT	2020	220,1	301,1	470,4	412,8	361,1	411,2	281,3	191,9	401,9	377,4	340,1	491,1				
15	DIKAWITANA BARAT	2021	280,1	179	211,8	242,1	444,4	111,1	279,4	189,7	143,7	280,4	19	404,4				
16	DIKAWITANA BARAT	2020	131,4	118,1	117	117,7	471,9	271	113,1	140,1	301,1	191,1	191	470,4				
17	DIKAWITANA BARAT	2020	272,6	284,1	271	195,9	146,4	284,4	146,1	191,9	76,6	30,4	86,1	191,4				
18	DIKAWITANA BARAT	2021	481,9	301,4	171,1	211,9	401,1	191,1	181,4	181,9	401,1	440	481,1	281,1				
19	DIKAWITANA BARAT	2021	181,9	111	111	111,1	141,1	91	111	181,1	161	111	181,4	281,1				
20	DIKAWITANA BARAT	2020	281,9	191,8	171,8	181,8	181,1	181,1	181,1	181,1	41,1	111	181,4	111	171,8			

Gambar 3.6 Nama kolom 1 (satu) s/d 16

Kolom pertama pada *data sets* merupakan kolom Nomor yang diberi *header* (No) selanjutnya kolom kedua merupakan kolom Nama Provinsi yang diberi *header* (PROV) dan pada kolom ke tiga adalah kolom Tahun yang diberi *header* Years. Pada kolom ketiga ini berisi angka tahun 2015,2016,2017,2018,2019, dan 2020 yang dimaksudkan sebagai penanda bahwa baris tersebut merupakan data-data yang diambil pada tahun sesuai baris.

No	PROV	YEARS
1	ACEH	2015
2	ACEH	2016
3	ACEH	2017
4	ACEH	2018
5	ACEH	2019
6	ACEH	2020
7	SUMATERA UTARA	2015
8	SUMATERA UTARA	2016
9	SUMATERA UTARA	2017
10	SUMATERA UTARA	2018
11	SUMATERA UTARA	2019
12	SUMATERA UTARA	2020
13	SUMATERA BARAT	2015
14	SUMATERA BARAT	2016
15	SUMATERA BARAT	2017
16	SUMATERA BARAT	2018
17	SUMATERA BARAT	2019
18	SUMATERA BARAT	2020

Gambar 3.7 Nama kolom 1 (satu) s/d 3 (tiga)

Kolom selanjutnya pada *data sets* dari kolom ke 4 (empat) hingga kolom ke 16 merupakan kolom yang berisi data hasil *summary* mengikuti formula (8). Kolom ke 4 (empat) merupakan kolom curah hujan Januari dan diberi *header* RAIN\_JAN. Secara berurutan RAIN\_FEB, RAIN\_MARC, RAIN\_APR, RAIN\_MEI, RAIN\_JUN, RAIN\_JUL, RAIN\_AUG, RAIN\_SEPT, RAIN\_OKT, RAIN\_NOV, RAIN\_DEC. Maksud dari masing-masing header adalah hujan Februari hingga hujan Desember.

RAIN	JAN RAIN	FEB RAIN	MAR RAIN	APR RAIN	MAY RAIN	JUN RAIN	JUL RAIN	AUG RAIN	SEP RAIN	OKT RAIN	NOV RAIN	DEC
88	35	47	243,4	74,6	75,3	186,6	175,4	203,5	92,8	222,5	116,5	
153,5	183	96,3	67,1	237	92,9	21,2	393,3	178,9	445,4	265,7	199,1	
512	89	173,5	48	221	70,5	72,5	187,5	437	64,5	206	238	
181	45,5	91	107	216,5	88,5	121	108,5	143,3	263,3	287,9	178,8	
114,5	96	192	115,5	22,5	8	26	47,5	2,5	10	105,2	127	
7,6	36	19,8	183,6	433,6	149,7	137,2	3,4	29,8	75,8	230,8	174,2	
190,4	64,4	233,6	173	230,1	100,2	180,2	180,7	176,4	327,9	448,6	134,5	
190,4	303,4	160,6	99,7	371,7	123,4	180,1	281,8	580,8	327,2	197	154,8	
284,5	139,6	317,4	186,5	282,1	158,9	205,2	262,9	524	234	179,8	251,8	
224,3	39,1	88,6	300,6	225,9	200,5	609	136,7	375,9	393,4	243,8	257,9	
208,2	204,7	130,6	151,4	424,8	374,4	199,7	130,9	388,1	455,6	455,6	319,2	
86,9	80	103,9	329	837	613	300,3	229,1	465,7	465,7	336,6	396,5	
272,6	204,5	273	185,9	185,9	236,4	166,1	319,9	75,6	80,4	483,3	391,4	
229,3	305,5	470,4	412,8	501,8	415,2	280,3	631,9	458,9	697,4	340,2	489,5	
369,5	179	227,6	262,5	644,4	131,3	278,4	589,7	543,7	280,4	719	404,4	
150,4	228,3	229	319,7	475,9	233	123,1	140,1	385,5	621,4	551	470,6	
272,8	96,2	297,4	291,8	147,1	427,8	240	126	58,1	184,7	130,1	398,4	
443,9	203,4	377,2	227,9	456,2	199,7	324,4	188,5	416,7	443	644,5	245,7	

Gambar 3.8 Nama kolom 4 (empat) hingga kolom 16

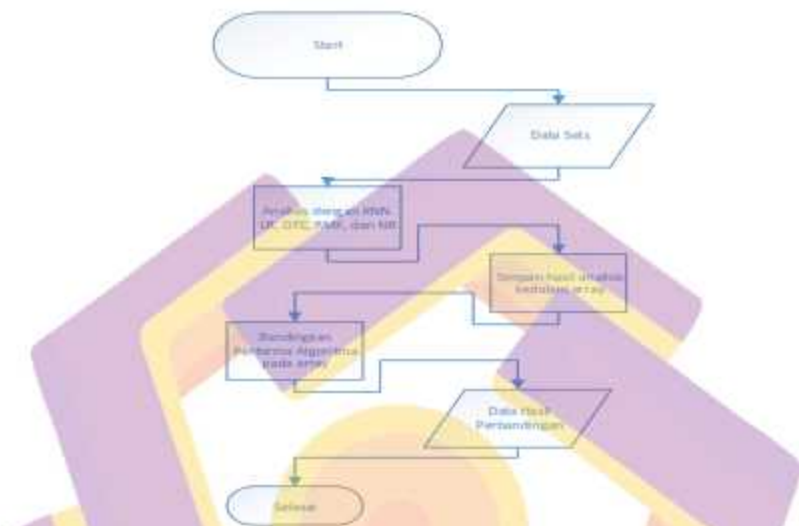
Pada kolom 17 merupakan kolom yang berisi jumlah kepadatan penduduk. Kolom ini diberi label POPUL\_DENSITY, dengan isi kolom merujuk pada formula (6).

Untuk kolom 18 merupakan kolom jumlah sungai yang terdapat pada tiap Provinsi. Kolom ini diberi label RIVERS. Data jumlah sungai pada tiap Provinsi merupakan sungai besar yang terdapat di Provinsi tersebut.

Kolom terakhir yang merupakan kolom kelas pada *data sets* berisi data string berupa Banjir dan Tidak Banjir. *Header* pada kolom ini adalah FLOODS. Untuk memberikan label merujuk dari formula (7).



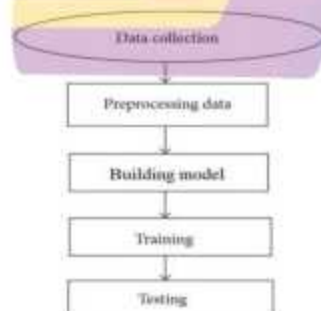
(Pedregosa et al., 2011). Library ini menggunakan Python sebagai bahasa pemrograman.



Gambar 3.12 Alur penggunaan algoritma pada penelitian

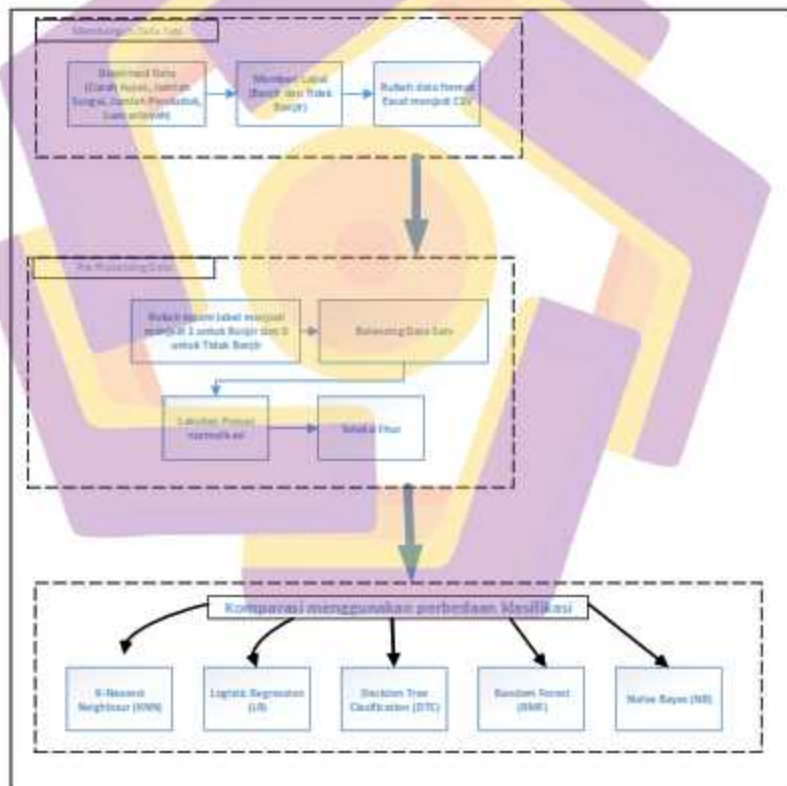
### 3.4. Alur Penelitian

Merujuk pada (Naik et al., 2021) tahapan dalam melakukan prediksi banjir setidaknya terdapat 5 (lima) tahap sebagai berikut :



Gambar 3.13 Tahap prediksi banjir

Adapun detail tahapan dalam penelitian terlihat pada gambar 19. Detail tahap penelitian. Dimulai dengan membangun *data sets* melakukan perubahan string menjadi *number* pada kolom kelas. Selanjutnya dilakukan balancing data untuk mendapatkan keseimbangan antar kelas, tahap selanjutnya yakni normalisasi data set yang berada pada variabel X. Kemudian dilakukan seleksi fitur untuk mendapatkan fitur terbaik dalam *data set*. Setelah tahap preprocessing data selesai, baru dilakukan pengujian menggunakan berbagai algoritma.



Gambar 3.14 Kerangka kerja penelitian

Pada tahap awal penelitian dalam proses pengumpulan data untuk membangun *data sets*, pengunduhan dilakukan secara manual untuk hampir semua jenis data baik itu curah hujan, jumlah penduduk, jumlah sungai maupun penentuan label Banjir dan Tidak Banjir. Hal ini dilakukan dikarenakan untuk request menggunakan *Application Programming Interface* (API) hampir disemua sumber data belum disediakan secara bebas.

*Data sets* yang sebelumnya berformat *xlsx*, disimpan dalam bentuk *csv*. Agar *data sets* benar-benar dipisahkan menggunakan koma (,) maka terlebih dahulu data yang tersimpan dipastikan tidak menggunakan pemisah koma namun menggunakan titik untuk data float. Sedangkan untuk data Ribuan tidak menggunakan pemisah. Setelah semua data telah sesuai dengan kebutuhan, selanjutnya *data sets* secara *programmatic* siap di analisis menggunakan bahasa *python*.

Tahap pertama *data sets* dalam bentuk *csv* di *import* ke dalam program dan merubah pemisah titik koma (;) menjadi koma (,).

```
import pandas as pd
import numpy as np
from imblearn.pipeline import Pipeline
import matplotlib.pyplot as plt
from sklearn.datasets import make_classification
from sklearn.model_selection import cross_validate
from sklearn.model_selection import
RepeatedStratifiedKFold
from sklearn.ensemble import AdaBoostClassifier
from imblearn.combine import SMOTEENN
from imblearn.under_sampling import
EditedNearestNeighbours
#import dataset dan mengganti ; menjadi ,
data=pd.read_csv('data_id.csv',
```

Setelah dilakukan tahap import untuk memastikan bahwa dokumen telah ter *import* dengan tepat dapat mengecek 5 (lima) data teratas *data sets*.

Tabel 3.1. Menampilkan 5 (lima) data teratas pada *data sets*

	PROV	YEARS	RAIN_JAN	RAIN_FEB	RAIN_MAR	RAIN_APR	RAIN_MEI	RAIN_JUN	RAIN_JUL	RAIN_AUG	RAIN_SEPT	RAIN_OKT	RAIN_NOV	RAIN_DEK
0	ACSH	2015	98.0	35.0	47.0	231.4	11.0	75.5	188.0	175.4	201.5	92.0	222.0	195
1	ACSH	2016	153.5	183.0	95.2	97.1	227.8	50.5	21.2	390.3	176.9	449.4	256.7	189
2	ACSH	2017	132.0	89.0	173.5	48.9	221.9	52.9	72.5	187.0	437.9	94.3	206.6	248
3	ACSH	2018	181.0	45.5	41.0	187.9	219.1	88.5	121.9	100.0	143.3	201.3	257.8	105
4	ACSH	2019	114.0	96.0	182.0	155.5	22.0	0.0	38.9	47.0	2.5	18.0	181.2	127

Setelah tahap import dilakukan perubahan pada kolom FLOODS dari sebelumnya Banjir dan Tidak Banjir menjadi 1 dan 0.

```
data['FLOODS'].replace(['Banjir', 'Tidak
Banjir'], [1,0], inplace=True)
```

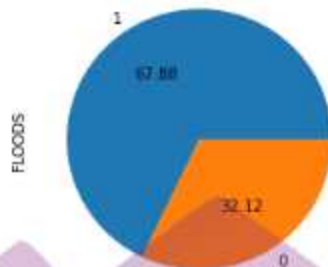
Perubahan data pada kolom FLOODS menjadikan kolom ini sebagai label pada *datasets*. Sehingga kolom FLOODS merupakan variabel Y sedang untuk kolom lainnya dijadikan sebagai variabel X. namun sebelum masuk ke dalam variabel X kolom PROV yang berisi string dihapus dari data dan data yang berulang-ulang yakni kolom YEARS juga dihapus dari data.

Tahapan selanjutnya *data sets* dilakukan *balancing*. Sebelum dilakukan *balancing* terlebih dahulu dilihat besarnya masing-masing kelas. Adapun jumlah data pada masing-masing kelas yakni 93 data kelas 1 (satu) dan 44 data kelas 0.

```
data['FLOODS'].value_counts()
```

```
1    93
0    44
```





Gambar 3.15 Persentase perbandingan Banjir 63,88% Tidak Banjir 32,12%

Proses balancing menggunakan metode *Random Over Sampling*. Pemilihan penggunaan metode ini dikarenakan perbandingan antara kelas yang terpaut cukup jauh. Sehingga untuk menghindari penghapusan data yang besar pada *data sets* maka *Random Over Sampling* dipilih menjadi metode untuk melakukan balancing.

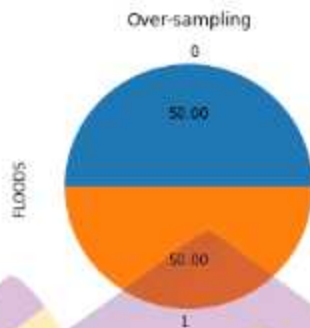
```

from imblearn.over_sampling import RandomOverSampler

#ros = RandomOverSampler(sampling_strategy=1) # float
ros = RandomOverSampler(sampling_strategy="not majority")
# String
X_res, y_res = ros.fit_resample(X, Y)

ax = y_res.value_counts().plot.pie(autopct='%0.2f')
_ = ax.set_title("Over-sampling")

```



Gambar 3.16 Persentase perbandingan setelah dilakukan *Random Over Sampling*

Jumlah Data setelah dilakukan *Random Over Sampling* adalah 93 untuk kelas 1 dan 93 untuk kelas 0.

```
y_res.value_counts()
0    93
1    93
Name: FLOODS, dtype: int64
```

Selanjutnya, *data sets* dibagi menjadi *data train* dan *data test*. Rasio yang digunakan adalah 80 : 20.

```
#Let's divide the dataset into 2 sets: train and test in r
atio (4:1)
from sklearn import model_selection, neighbors
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X_res, y_re
s, test_size=0.2, random_state=25)
print(f"No. of training examples: {x_train.shape[0]}")
print(f"No. of testing examples: {x_test.shape[0]}")

No. of training examples: 148
No. of testing examples: 38
```

Setelah *data sets* dibagi maka *data sets* siap untuk dilakukan analisis dengan algoritma yang telah dipilih.

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1. Prediksi Banjir Menggunakan Seleksi Fitur

Sesuai dengan kerangka kerja penelitian, pemilihan fitur diperlukan untuk mendapatkan fitur yang paling berpengaruh terhadap kelas. Seleksi fitur menggunakan *library* f-classif dan mendapatkan *score* fitur sebagai berikut :

```
from sklearn.feature_selection import SelectKBest, f_classif
sif
#Urutan 15 variabel yang paling berpengaruh hingga kurang
berpengaruh
selector = SelectKBest(f_classif, k = 14)
#New dataframe with the selected features for later use i
n the classifier. fit() method works best, if you want onl
y the feature names and their corresponding scores
X_new = selector.fit_transform(data_x, data_y)
names = data_x.columns.values[selector.get_support()]
scores = selector.scores_[selector.get_support()]
names_scores = list(zip(names, scores))
ns_df = pd.DataFrame(data = names_scores, columns=['Feat_
names', 'F_Scores'])
#Sort the dataframe for better visualization
ns_df_sorted = ns_df.sort_values(['F_Scores', 'Feat_names
'], ascending = [False, True])
print(ns_df_sorted)
```

	Feat_names	F_Scores
13	RIVERS	19.619971
2	RAIN_MARC	6.369618
4	RAIN_MEI	4.747333
5	RAIN_JUN	4.007605
9	RAIN_OKT	1.810807
1	RAIN_FEB	1.574202
3	RAIN_APR	1.541844
11	RAIN_DEC	1.399031
0	RAIN_JAN	0.731930
10	RAIN_NOV	0.281070
6	RAIN_JUL	0.257442
12	POPUL_DENSITY	0.220445
7	RAIN_AUG	0.171315
8	RAIN_SEPT	0.027190

Dari hasil seleksi fitur, diperoleh performa algoritma dengan perbandingan pada tabel *confusion matrix* berikut :

Tabel 4.1. Performa algoritma dengan seleksi fitur

No	Algoritma	Pengukuran Performa			
		Accuracy (%)	F1 Score (%)	Specificity (%)	Precision (%)
1.	k-Nearest Neighbour (KNN)	65	87	83	86
2.	Logistic Regression	67	76	77	76
3.	Decision Tree Classification (DTC)	49	80	94	83
4.	Random Forest (RMF)	65	82	77	81
5.	Naïve Bayes (NB)	60	74	88	78

Performa pada tiap algoritma memiliki *score* yang cenderung rendah dari sisi *accuracy*. Semua algoritma yang diuji terhadap *data sets* tidak ada yang mencapai 70%.

Pengujian terhadap dataset bukan hanya menggunakan seleksi fitur namun juga dengan menguji tanpa seleksi fitur. Hal ini bertujuan untuk melihat letak *gap* jika *data sets* secara utuh digunakan untuk memprediksi banjir. Hasil dari pengujian *data sets* dengan performa algoritma terlihat pada tabel berikut :

Tabel 4.2. Performa algoritma tanpa seleksi fitur

No	Algoritma	Pengukuran Performa			
		Accuracy (%)	F1 Score (%)	Specificity (%)	Precision (%)
1.	k-Nearest Neighbour (KNN)	57	82	85	84
2.	Logistic Regression	60	72	71	73
3.	Decision Tree Classification (DTC)	63	91	90	91

Tabel 4.2 (Lanjutan)

No	Algoritma	Pengukuran Performa			
		Accuracy (%)	F1 Score (%)	Specificity (%)	Precision (%)
4.	Random Forest (RMF)	60	85	85	86
5.	Naïve Bayes (NB)	68	68	71	70

Setelah diperoleh hasil pengujian *data sets* dengan performa 5 (lima) algoritma tanpa melakukan seleksi fitur, akurasi yang diperoleh tidak memiliki perbedaan yang signifikan. masing-masing algoritma menunjukkan angka akurasi yang tidak lebih dari 70%.

Dalam hal *data sets* hanya dihilangkan kolom PROV dan YEARS atau tanpa menggunakan seleksi fitur, akurasi Naïve Bayes mencapai 68%. Namun akurasi ini berubah setelah dilakukan fitur seleksi dan variabel X

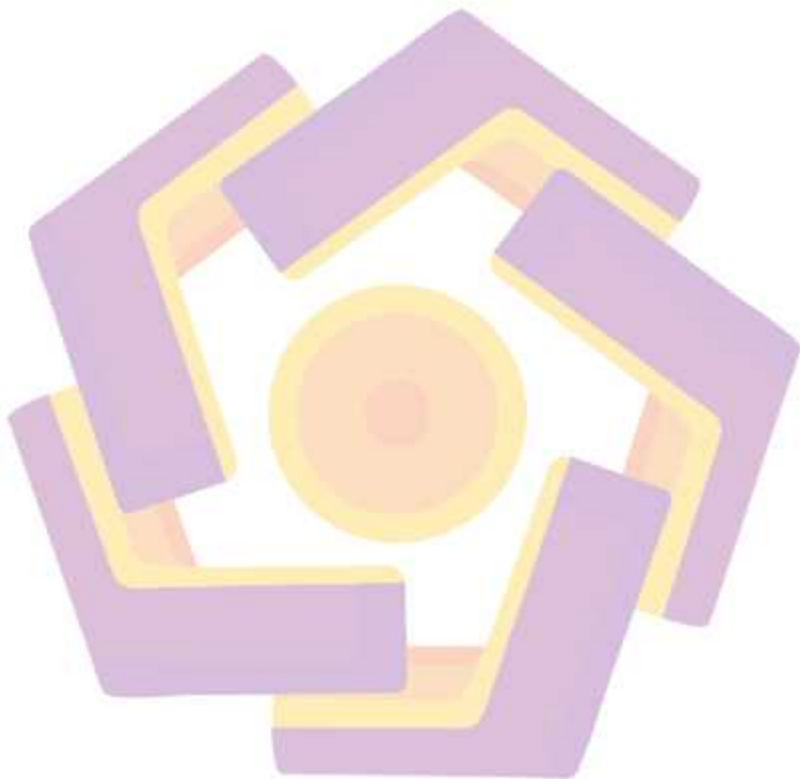
```
#ganti data x
data_x=data_x.drop(columns=['RAIN_SEPT', 'RAIN_AUG', 'POPUL
_DENSITY', 'RAIN_JUL', 'RAIN_NOV', 'RAIN_JAN'])
```

Akurasi berubah dan menjadikan Logistic Regression lebih unggul yakni dengan akurasi 67%.

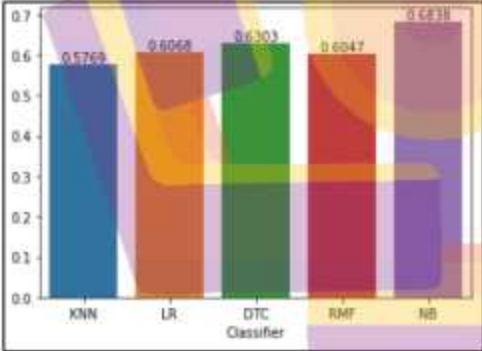
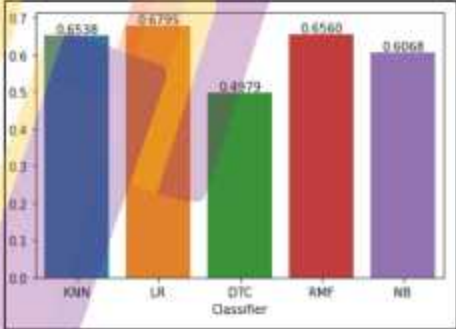
#### 4.2. Komparasi algoritma hasil pengujian

Komparasi pada tiap pengujian perlu untuk dilakukan. Hal ini bertujuan untuk melihat permasalahan yang ada dan mengakibatkan kemampuan algoritma yang digunakan untuk menguji *data sets* menjadi kurang/tidak optimal. Dengan demikian harapannya dapat diperoleh informasi permasalahan yang sebenarnya.

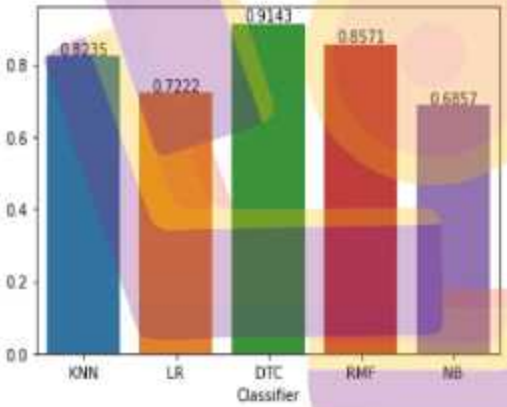
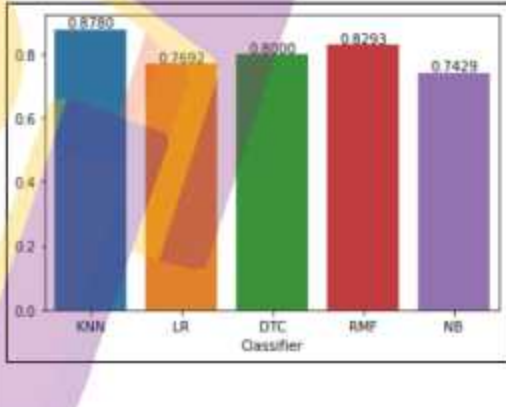
Untuk itu berikut adalah tabel hasil pengujian dan matrik pada masing-masing algoritma dan performa yang dihasilkan.



Tabel 4.3. Perbandingan accuracy, f1 score, specificity, Precision Tanpa Seleksi Fitur dan dengan Seleksi Fitur

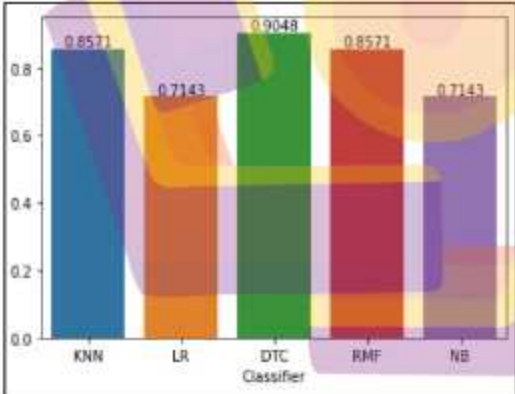
No	Pengukuran Performance	Tanpa Seleksi Fitur	Seleksi Fitur																								
1.	Accuracy (Akurasi)	<table border="1"> <thead> <tr> <th>Name</th> <th>Score</th> </tr> </thead> <tbody> <tr> <td>0 KNN</td> <td>0.576923</td> </tr> <tr> <td>1 LR</td> <td>0.606838</td> </tr> <tr> <td>2 DTC</td> <td>0.630342</td> </tr> <tr> <td>3 RMF</td> <td>0.604701</td> </tr> <tr> <td>4 NB</td> <td>0.683761</td> </tr> </tbody> </table> 	Name	Score	0 KNN	0.576923	1 LR	0.606838	2 DTC	0.630342	3 RMF	0.604701	4 NB	0.683761	<table border="1"> <thead> <tr> <th>Name</th> <th>Score</th> </tr> </thead> <tbody> <tr> <td>0 KNN</td> <td>0.653846</td> </tr> <tr> <td>1 LR</td> <td>0.679487</td> </tr> <tr> <td>2 DTC</td> <td>0.497863</td> </tr> <tr> <td>3 RMF</td> <td>0.655983</td> </tr> <tr> <td>4 NB</td> <td>0.606838</td> </tr> </tbody> </table> 	Name	Score	0 KNN	0.653846	1 LR	0.679487	2 DTC	0.497863	3 RMF	0.655983	4 NB	0.606838
Name	Score																										
0 KNN	0.576923																										
1 LR	0.606838																										
2 DTC	0.630342																										
3 RMF	0.604701																										
4 NB	0.683761																										
Name	Score																										
0 KNN	0.653846																										
1 LR	0.679487																										
2 DTC	0.497863																										
3 RMF	0.655983																										
4 NB	0.606838																										

Tabel 4.3 (Lanjutan)

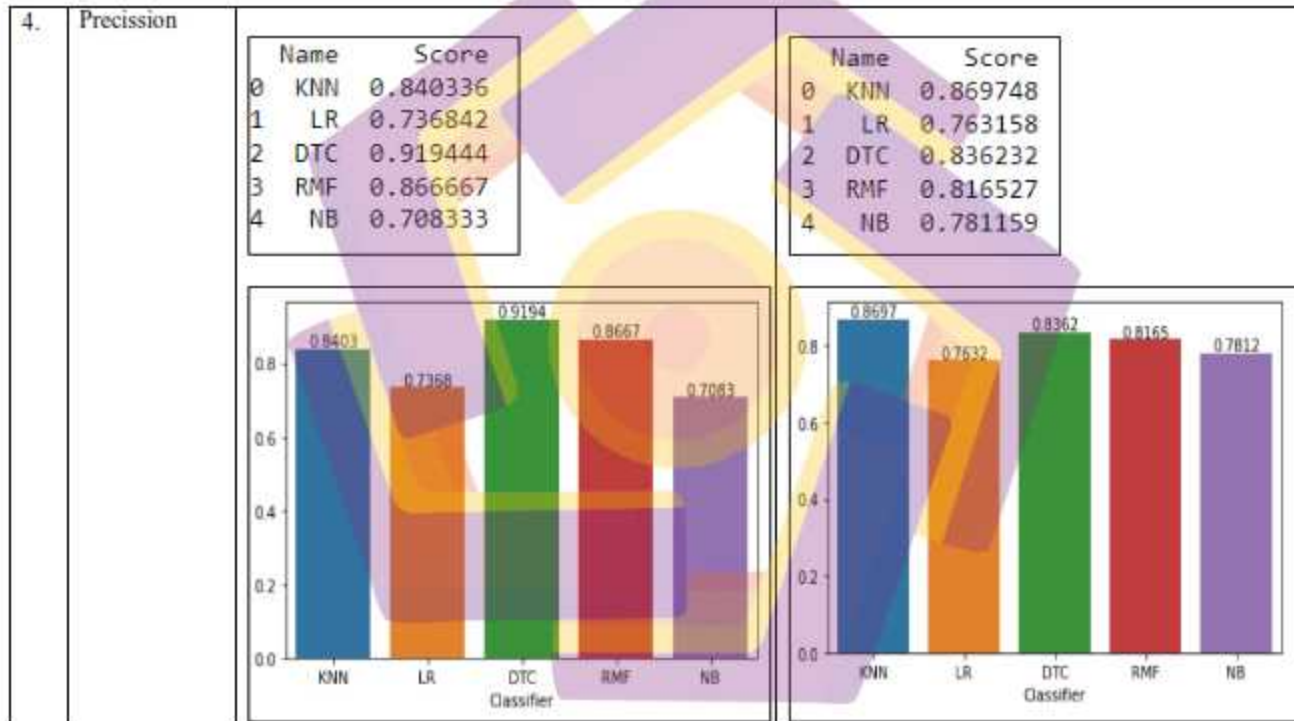
2.	F1 Score	<table border="1" data-bbox="403 294 705 529"> <thead> <tr> <th></th> <th>Name</th> <th>Score</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>KNN</td> <td>0.823529</td> </tr> <tr> <td>1</td> <td>LR</td> <td>0.722222</td> </tr> <tr> <td>2</td> <td>DTC</td> <td>0.914286</td> </tr> <tr> <td>3</td> <td>RMF</td> <td>0.857143</td> </tr> <tr> <td>4</td> <td>NB</td> <td>0.685714</td> </tr> </tbody> </table> 		Name	Score	0	KNN	0.823529	1	LR	0.722222	2	DTC	0.914286	3	RMF	0.857143	4	NB	0.685714	<table border="1" data-bbox="940 294 1243 529"> <thead> <tr> <th></th> <th>Name</th> <th>Score</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>KNN</td> <td>0.878049</td> </tr> <tr> <td>1</td> <td>LR</td> <td>0.769231</td> </tr> <tr> <td>2</td> <td>DTC</td> <td>0.800000</td> </tr> <tr> <td>3</td> <td>RMF</td> <td>0.829268</td> </tr> <tr> <td>4</td> <td>NB</td> <td>0.742857</td> </tr> </tbody> </table> 		Name	Score	0	KNN	0.878049	1	LR	0.769231	2	DTC	0.800000	3	RMF	0.829268	4	NB	0.742857
	Name	Score																																					
0	KNN	0.823529																																					
1	LR	0.722222																																					
2	DTC	0.914286																																					
3	RMF	0.857143																																					
4	NB	0.685714																																					
	Name	Score																																					
0	KNN	0.878049																																					
1	LR	0.769231																																					
2	DTC	0.800000																																					
3	RMF	0.829268																																					
4	NB	0.742857																																					



Tabel 4.3 (Lanjutan)

3.	Specificity	<table border="1"> <thead> <tr> <th></th> <th>Name</th> <th>Score</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>KNN</td> <td>0.857143</td> </tr> <tr> <td>1</td> <td>LR</td> <td>0.714286</td> </tr> <tr> <td>2</td> <td>DTC</td> <td>0.904762</td> </tr> <tr> <td>3</td> <td>RMF</td> <td>0.857143</td> </tr> <tr> <td>4</td> <td>NB</td> <td>0.714286</td> </tr> </tbody> </table> 		Name	Score	0	KNN	0.857143	1	LR	0.714286	2	DTC	0.904762	3	RMF	0.857143	4	NB	0.714286	<table border="1"> <thead> <tr> <th></th> <th>Name</th> <th>Score</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>KNN</td> <td>0.833333</td> </tr> <tr> <td>1</td> <td>LR</td> <td>0.777778</td> </tr> <tr> <td>2</td> <td>DTC</td> <td>0.944444</td> </tr> <tr> <td>3</td> <td>RMF</td> <td>0.777778</td> </tr> <tr> <td>4</td> <td>NB</td> <td>0.888889</td> </tr> </tbody> </table> 		Name	Score	0	KNN	0.833333	1	LR	0.777778	2	DTC	0.944444	3	RMF	0.777778	4	NB	0.888889
	Name	Score																																					
0	KNN	0.857143																																					
1	LR	0.714286																																					
2	DTC	0.904762																																					
3	RMF	0.857143																																					
4	NB	0.714286																																					
	Name	Score																																					
0	KNN	0.833333																																					
1	LR	0.777778																																					
2	DTC	0.944444																																					
3	RMF	0.777778																																					
4	NB	0.888889																																					

Tabel 4.3 (Lanjutan)



Dari Tabel 4.3 performa algoritma Naïve Bayes di komparasi dengan berbagai algoritma lainnya. Dalam tahap komparasi, dilakukan dengan 2 (dua) tahap yakni tanpa seleksi fitur dan dengan seleksi fitur. Pada tahap tanpa seleksi fitur, masing-masing algoritma diterapkan untuk melakukan prediksi dengan menggunakan datasets yang tersedia. Perbandingan sesuai dengan informasi yang diperoleh pada masing-masing algoritma adalah sebagai berikut :

Tabel 4.4. Perbandingan Performa Akurasi Algoritma Tanpa Seleksi Fitur

No	Algoritma yang digunakan	Performa Akurasi
1.	Naïve Bayes	68%
2.	Random Forest	60%
3.	Decition Tree Classification	63%
4.	Logistic Regression	60%
5.	K Nearest Neighbour	57%

Hasil yang berbeda diperoleh setelah dilakukan seleksi fitur, Naïve Bayes sebagai algoritma yang diharapkan memberikan keluaran yang lebih baik setelah dilakukan seleksi fitur dengan Chi Square, justru mengalami penurunan performa akurasi dalam melakukan prediksi.

Tabel 4.5. Perbandingan Performa Akurasi Algoritma Dengan Seleksi Fitur

No	Algoritma yang digunakan	Performa Akurasi
1.	Naïve Bayes	65%
2.	Random Forest	67%
3.	Decition Tree Classification	49%

Tabel 4.5 (Lanjutan)

No	Algoritma yang digunakan	Performa Akurasi
4.	Logistic Regression	65%
5.	K Nearest Neigboard	60%

Dari perbandingan akurasi yang diharapkan dapat menjadi patokan dalam menentukan prediksi kemungkinan banjir dimasa yang akan datang memberikan keluaran yang kurang optimal. Penyebab dari permasalahan ini setidaknya dikarenakan hal-hal sebagai berikut :

1. Jumlah datasets yang kecil mengakibatkan data train setelah di over sampling hanya 148 data sedangkan untuk testing menggunakan 38 data;
2. Data kurang variatif, mengakibatkan algoritma machine learning yang secara teoritis mampu menagai prediksi dengan datasets kecil seperti Random Forest (Agarwal, 2014) tidak dapat memberikan keluaran yang tepat;
3. True positif (banjir) memiliki jumlah data yang lebih banyak yakni sebanyak 98 data hal ini dikarenakan dilakukan rerata pada kejadian banjir tiap bulannya padahal jika merujuk dari kejadian harian jumlah kejadian banjir seharusnya lebih sedikit.

Meskipun demikian, berdasarkan target akurasi yang diharapkan mampu diterapkan dalam memprediksi bencana banjir, maka Naïve Bayes sebagai algoritma yang diunggulkan dalam penelitian memberikan keluaran yang lebih

stabil. Hal ini terbukti dengan penurunan yang tidak signifikan ketika prediksi dilakukan tanpa seleksi fitur maupun setelah dilakukan seleksi fitur.

Meskipun akurasi menjadi prioritas dalam memperhitungkan kemungkinan benar terjadi banjir atau tidak, namun kesalahan dalam menentukan kondisi kebenaran kemungkinan terjadi banjir (*True Positive*) dibandingkan dengan seluruh data benar terjadi banjir hasil prediksi baik benar *True Positive* dan *False Positive* untuk mendapatkan *Precision*.

$$\text{Precision} = \text{True Positive} / (\text{True Positive} + \text{False Positive})$$

*Precision* menjadi sangat penting mengingat jika *false positif* terjadi dikhawatirkan akan menimbulkan polemik dimasyarakat.

Dari hasil penggunaan algoritma pada *machine learning*, maka diperoleh kemampuan *Decision Tree Classification* (DTC) memiliki kemampuan dalam *precision* lebih baik dan stabil tanpa menggunakan seleksi fitur maupun setelah menggunakan seleksi fitur.

Tabel 4.6. Perbandingan Performa Precision Algoritma Tanpa Seleksi Fitur

No	Algoritma yang digunakan	Performa Precision
1.	Naïve Bayes	70%
2.	Random Forest	86%
3.	Decition Tree Classification	91%
4.	Logistic Regression	73%
5.	K Nearest Neighbour	84%

Tabel 4.7. Perbandingan Performa Precision Algoritma Tanpa Seleksi Fitur

No	Algoritma yang digunakan	Performa Precision
1.	Naïve Bayes	78%
2.	Random Forest	81%
3.	Decition Tree Classification	83%
4.	Logistic Regression	76%
5.	K Nearest Neighbour	86%

Mengingat kesalahan dalam memberikan informasi kebenaran terjadi banjir (*False Positive*) berbanding lurus dengan resiko kesalahan ketika memberikan informasi tidak terjadi banjir (*False Negative*) maka pengukuran performa algoritma juga dilakukan dengan mengukur f1 Score. Adapun formula dalam menentukan f1 Score yang digunakan adalah sebagai berikut :

$$F1\text{-score} = 2 * ((Precision * Recall) / (Precision + Recall))$$

Sehingga keluaran dari f1 Score sebelum dilakukan seleksi fitur yang dihasilkan pada masing-masing algoritma terlihat pada tabel berikut :

Tabel 4.8. Perbandingan Performa Precision Algoritma Tanpa Seleksi Fitur

No	Algoritma yang digunakan	Performa Precision
1.	Naïve Bayes	68%
2.	Random Forest	85%
3.	Decition Tree Classification	91%
4.	Logistic Regression	72%
5.	K Nearest Neighbour	82%

Setelah dilakukan pengujian tanpa seleksi fitur selanjutnya dilakukan pengujian dengan seleksi fitur untuk mendapatkan informasi performa masing-masing algoritma.

Tabel 4.9. Perbandingan Performa Precision Algoritma Tanpa Seleksi Fitur

No	Algoritma yang digunakan	Performa Precision
1.	Naïve Bayes	74%
2.	Random Forest	82%
3.	Decition Tree Classification	80%
4.	Logistic Regression	76%
5.	K Nearest Neighbour	87%

## **BAB V**

### **PENUTUP**

#### **5.1. Kesimpulan**

Dari hasil penelitian yang telah dilakukan, beberapa kesimpulan dapat ditarik merujuk dari tujuan sebagai berikut :

1. Data sets faktor penyebab banjir yakni curah hujan, kepadatan penduduk, dan daerah aliran sungai (DAS) berhasil dibuat dengan sumber data dari berbagai sumber yang ada di Indonesia;
2. Dari penelitian, perhitungan probability banjir belum dapat sepenuhnya dilakukan dengan menggunakan *data sets* yang telah terbentuk;
3. Diperolehnya Naïve Bayes dan Logistic Regression dengan akurasi tertinggi namun belum dapat dijadikan sebagai acuan, dikarenakan tingkat akurasinya rendah.

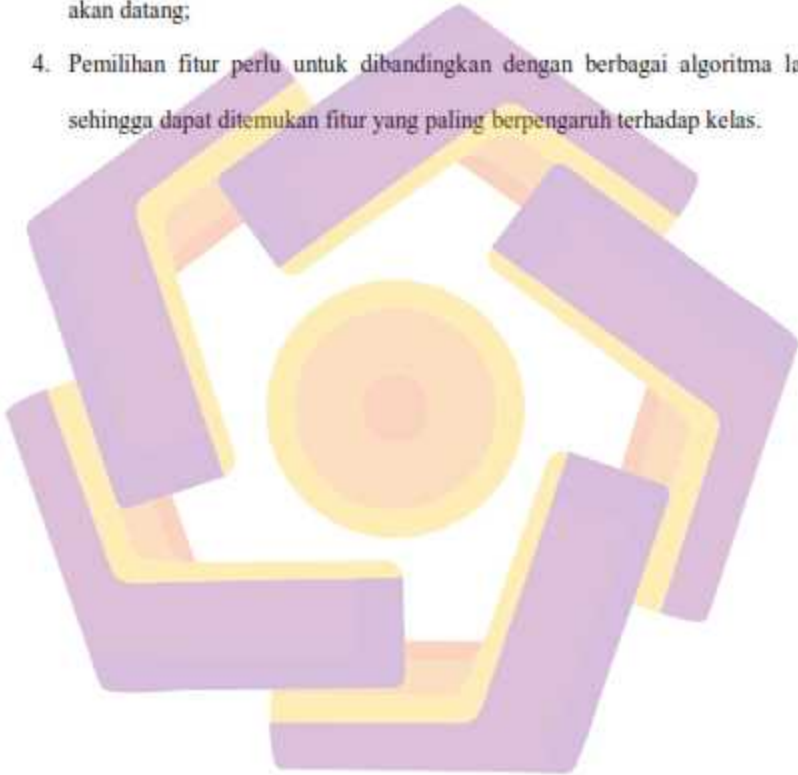
#### **5.2. Saran**

Kesulitan untuk membangun *data sets*, kemudian menemukan algoritma yang tepat untuk menangani *data set* dan berbagai tantangan lainnya, menjadikan penelitian ini masih memiliki banyak kekurangan. Sehingga untuk perbaikan dimasa yang akan datang beberapa saran berikut diharapkan mampu untuk memberikan kontribusi perbaikan pada penelitian selanjutnya :

1. Faktor penyebab banjir perlu untuk ditambah merujuk dari informasi yang diperoleh para ahli hidrologi maupun kebencanaan;



2. Pola *data sets* perlu untuk diperbaiki dengan mengguna format data harian disesuaikan dengan laporan bencana pada data informasi bencana Indonesia.
3. Percobaan pada berbagai algoritma lain yang diharapkan mampu untuk memberikan informasi yang lebih baik dalam memprediksi banjir dimasa yang akan datang;
4. Pemilihan fitur perlu untuk dibandingkan dengan berbagai algoritma lain, sehingga dapat ditemukan fitur yang paling berpengaruh terhadap kelas.



## DAFTAR PUSTAKA

### PUSTAKA BUKU

- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceedings—2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>
- Albon, C. (2018). *Python Machine Learning Cookbook* (first). O'Reilly Media, Inc.
- Arjun Singh, & Saxena, A. (2016). A Hybrid Data Model for Prediction of Disaster using Data Mining Approaches. *International Journal of Engineering Trends and Technology (IJETT)*, 384–392.
- Azlan, F. A., Ahmad, A., Yussof, S., & Ghapar, A. A. (2020). Analyzing Algorithms to Detect Disaster Events using Social Media. *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*, 384–389. <https://doi.org/10.1109/ICIMU49871.2020.9243599>
- Azure, M. (2021, November 4). *Machine Learning Algorithm Cheat Sheet for Azure Machine Learning designer*. <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet>
- BMKG, I. (2021). *Tugas dan Fungsi*. <https://www.bmkg.go.id/profil/?p=tugas-fungsi>
- defeloper, scikit-learn. (2022). *Sklearn.ensemble.RandomForestClassifier*. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn-ensemble-randomforestclassifier>
- Domala, J., Dogra, M., Masrani, V., Fernandes, D., D'souza, K., Fernandes, D., & Carvalho, T. (2020). Automated Identification of Disaster New for Crisis Management using Machine Learning and Natural Language Processing. *Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020)*, 503–508.
- Domingos, P., & Michael Pazzani. (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Kluwer Academic Publisher*, 29(Machine Learning), 103–130.
- Dwiasnati, D., & Devianto, Y. (2021). Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana

Banjir. *Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (SISFOTEK)*, 5, 202–207.

- Ganguly, K. K., Nahar, N., & Hossain, B. M. (2019). A machine learning-based prediction and analysis of flood affected households: A case study of floods in Bangladesh. *International Journal of Disaster Risk Reduction*, 34, 283–294. <https://doi.org/10.1016/j.ijdr.2018.12.002>
- Ghoneim, S. (2019, February 4). *Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on?* <https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f1124>
- Gopal, L. S., Prabha, R., Pullarkatt, D., & Ramesh, M. V. (2020). Machine Learning based Classification of Online News Data for Disaster management. *2020 IEEE Global Humanitarian Technology Conference (GHTC)*.
- Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.
- <https://scikit-learn.org/>. (n.d.). *Naive Bayes*. Retrieved November 20, 2021, from [https://scikit-learn.org/stable/modules/naive\\_bayes.html](https://scikit-learn.org/stable/modules/naive_bayes.html)
- Indonesia, B. (2014). *Rata-Rata Harian Aliran Sungai, Tinggi Aliran, dan Volume Air di Beberapa Sungai yang Daerah Pengalirannya Lebih dari 1.000 km2, 2000-2011*. <https://www.bps.go.id/statictable/2014/09/05/1371/rata-rata-harian-aliran-sungai-tinggi-aliran-dan-volume-air-di-beberapa-sungai-yang-daerah-pengalirannya-lebih-dari-1-000-km2-2000-2011-.html>
- Indonesia, B. (2019). *Luas Daerah dan Jumlah Pulau Menurut Provinsi, 2019*. [https://www.bps.go.id/indikator/indikator/view\\_data\\_pub/0600/api\\_pub/U\\_FpWMmJZOVZIZTInc1pXaHhDV1hPQT09/da\\_01/1](https://www.bps.go.id/indikator/indikator/view_data_pub/0600/api_pub/U_FpWMmJZOVZIZTInc1pXaHhDV1hPQT09/da_01/1)
- Indonesia, B. (2020). *Jumlah Penduduk Hasil Proyeksi Menurut Provinsi dan Jenis Kelamin (Ribu Jiwa), 2018-2020*. <https://www.bps.go.id/indikator/12/1886/1/jumlah-penduduk-hasil-proyeksi-menurut-provinsi-dan-jenis-kelamin.html>
- Indonesia, B. (2022a). *Data Informasi Bencana Indonesia*. <https://dibi.bnpb.go.id/xdibi>
- Indonesia, B. (2022b). *Tentang Profil BPS*. <https://www.bps.go.id/menu/1/informasi-umum.html>
- Indonesia, R. (2008). *UNDANG UNDANG REPUBLIK INDONESIA NOMOR 14 TAHUN 2008 TENTANG KETERBUKAAN INFORMASI PUBLIK*. Republik Indonesia.

<https://www.kemenkeu.go.id/sites/default/files/uu%2014%20tahun%202008.pdf>

- Kantardzic, M. (2020). *DATA MINING Concepts, Models, Methods, and Algorithms* (third). John Wiley & Sons, Incy.
- Krisnadwipayana, U. (-). *Daftar sungai di Indonesia*. [http://p2k.unkris.ac.id/id3/1-3065-2962/Daftar-Sungai-Di-Indonesia\\_26973\\_stie-walisongo\\_p2k-unkris.html](http://p2k.unkris.ac.id/id3/1-3065-2962/Daftar-Sungai-Di-Indonesia_26973_stie-walisongo_p2k-unkris.html)
- Maliha, S. K., Ema, R. R., Ghosh, S. K., Ahmed, H., Mollick, Md. R. J., & Islam, T. (2019). Cancer Disease Prediction Using Naive Bayes, K-Nearest Neighbor and J48 algorithm. *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–7. <https://doi.org/10.1109/ICCCNT45670.2019.8944686>
- Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, *10*(11), 1536. <https://doi.org/10.3390/w10111536>
- Naik, S., Verma, A., Ashok Patil, S., & Hingmire, A. (2021). Flood Prediction using Logistic Regression for Kerala State. *International Journal of Engineering Research & Technology*, 36–38.
- Nugroho, S. P. (2002). EVALUASI DAN ANALISIS CURAH HUJAN SEBAGAI FAKTOR PENYEBAB BENCANA BANJIR JAKARTA. *Urnal Sains & Teknologi Modifikasi Cuaca*, *3*, 91–97.
- Ordoñez, A. J., Paje, R. E. J., & Naz, R. N. (2018). SMS Classification Method for Disaster Response using Naïve Bayes Algorithm. *2018 International Symposium on Computer, Consumer and Control (IS3C)*, 233–236. <https://doi.org/10.1109/IS3C.2018.00066>
- Panafrican Emergency Training Centre. (2002). *DISASTERS & EMERGENCIES DEFINITIONS*. <https://apps.who.int/disasters/repo/7656.pdf>
- Pedregosa, f, Varoquaux, G., Gramfort, A., Michel, V., Thirison, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Caurnapeu, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning In Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
- Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., & Janardana, I. (2017). Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm. *International Journal of Engineering and Emerging Technology*, *2*(1).

- Prihandoko, Bertalya, & Ramadhan, M. I. (2017). An Analysis of Natural Disaster Data by Using K-Means and K-Medoids Algorithm of Data Mining Techniques. *2017 15th International Conference on Quality in Research (QiR) : International Symposium on Electrical and Computer Engineering*, 221–225. <https://doi.org/10.1109/QIR.2017.8168485>
- Refonaa, j, Lakshmi, M., & Vivek, v. (2015). Analysis and Prediction Of natural Disaster Using Spatial Data Mining Technique. *2015 International Conference on Circuit, Power and Computing Technologies [ICCPCT]*, 1–6. <https://doi.org/10.1109/ICCPCT.2015.7159379>
- Republik Indonesia, B. (2008). *KEPALA BADAN NASIONAL PENANGGULANGAN BENCANA NOMOR 3 TAHUN 2008 TENTANG PEDOMAN PEMBENTUKAN BADAN PENANGGULANGAN BENCANA DAERAH*. <https://bnpb.go.id/uploads/24/peraturan-kepala/2008/perka-3-tahun-2008.pdf>
- Safitri, D. (2021). Karakteristik Aliran dan Debit Banjir pada Beberapa Sungai di Indonesia: Kajian Literatur. *Journal of Infrastructural in Civil Engineering (JICE)*, 02(02), 1–9.
- Sesa, W., shalih, O., Syauqi, W.Adi, A., Z.Shabrina, F., Rizqi, A., Widiastomo, Y., S.Putra, A., Karimah, R., Eveline, F., Alfian, A., Hafiz, A., BAGaskoro, Y., Nomita Dewi, A., & Rahmawati, I. (2020). *Index Risiko Bencana Indonesia Tahun 2020*. Badan Nasional Penanggulangan Bencana. <https://inarisk.bnpb.go.id/pdf/BUKU%20IRBI%202020%20KP.pdf>
- Tempola, F., Muhammad, M., & Khairan, A. (2018). Naive Bayes Classifier For Prediction Of Volcanic Status In Indonesia. *Proc. of 2018 5th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE)*.
- Tingsanchali, T. (2012). Urban flood disaster management. *Elsevier Ltd*, 32, 25–37.
- Windarto, A. P., Indriani, U., Raharjo, M. R., & Dewi, L. S. (2020). Bagian 1: Kombinasi Metode Klustering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia). *Jurnal Media Informatika Budidarma*, 4(3), 855. <https://doi.org/10.30865/mib.v4i3.2312>

#### PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceedings—2013 International Conference on Machine Intelligence*

*Research and Advancement, ICMIRA 2013.*  
<https://doi.org/10.1109/ICMIRA.2013.45>

- Arjun Singh, & Saxena, A. (2016). A Hybrid Data Model for Prediction of Disaster using Data Mining Approaches. *International Journal of Engineering Trends and Technology (IJETT)*, 384–392.
- Azlan, F. A., Ahmad, A., Yussof, S., & Gharar, A. A. (2020). Analyzing Algorithms to Detect Disaster Events using Social Media. *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*, 384–389. <https://doi.org/10.1109/ICIMU49871.2020.9243599>
- Domala, J., Dogra, M., Masrani, V., Fernandes, D., D'souza, K., Fernandes, D., & Carvalho, T. (2020). Automated Identification of Disaster New for Crisis Management using Machine Learning and Natural Language Processing. *Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020)*, 503–508.
- Dwiasnati, D., & Devianto, Y. (2021). Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir. *Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (SISFOTEK)*, 5, 202–207.
- Ganguly, K. K., Nahar, N., & Hossain, B. M. (2019). A machine learning-based prediction and analysis of flood affected households: A case study of floods in Bangladesh. *International Journal of Disaster Risk Reduction*, 34, 283–294. <https://doi.org/10.1016/j.ijdrr.2018.12.002>
- Gopal, L. S., Prabha, R., Pullarkatt, D., & Ramesh, M. V. (2020). Machine Learning based Classification of Online News Data for Disaster anagement. *2020 IEEE Global Humanitarian Technology Conference (GHTC)*.
- Maliha, S. K., Ema, R. R., Ghosh, S. K., Ahmed, H., Mollick, Md. R. J., & Islam, T. (2019). Cancer Disease Prediction Using Naive Bayes, K-Nearest Neighbor and J48 algorithm. *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–7. <https://doi.org/10.1109/ICCCNT45670.2019.8944686>
- Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, 10(11), 1536. <https://doi.org/10.3390/w10111536>
- Naik, S., Verma, A., Ashok Patil, S., & Hingmire, A. (2021). Flood Prediction using Logistic Regression for Kerala State. *International Journal of Engineering Research & Technology*, 36–38.

- Nugroho, S. P. (2002). Evaluasi Dan Analisis Curah Hujan Sebagai Faktor Penyebab Bencana Banjir Jakarta. *Urnal Sains & Teknologi Modifikasi Cuaca*, 3, 91–97.
- Ordoñez, A. J., Paje, R. E. J., & Naz, R. N. (2018). SMS Classification Method for Disaster Response using Naïve Bayes Algorithm. *2018 International Symposium on Computer, Consumer and Control (IS3C)*, 233–236. <https://doi.org/10.1109/IS3C.2018.00066>
- Pedregosa, f, Varoquaux, G., Gramfort, A., Michel, V., Thirison, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Caurnapeu, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning In Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Prihandoko, Bertalya, & Ramadhan, M. I. (2017). An Analysis of Natural Disaster Data by Using K-Means and K- Medoids Algorithm of Data Mining Techniques. *2017 15th International Conference on Quality in Research (QIR) : International Symposium on Electrical and Computer Engineering*, 221–225. <https://doi.org/10.1109/QIR.2017.8168485>
- Refonaa, j, Lakshmi, M., & Vivek, v. (2015). Analysis and Prediction Of natural Disaster Using Spatial Data Mining Technique. *2015 International Conference on Circuit, Power and Computing Technologies [ICCPCT]*, 1–6. <https://doi.org/10.1109/ICCPCT.2015.7159379>
- Safitri, D. (2021). Karakteristik Aliran dan Debit Banjir pada Beberapa Sungai di Indonesia: Kajian Literatur. *Journal of Infrastructural in Civil Engineering (JICE)*, 02(02), 1–9.
- Tempola, F., Muhammad, M., & Khairan, A. (2018). Naive Bayes Classifier For Prediction Of Volcanic Status In Indonesia. *Proc. of 2018 5th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE)*.
- Tingsanchali, T. (2012). Urban flood disaster management. *Elsevier Ltd*, 32, 25–37.
- Windarto, A. P., Indriani, U., Raharjo, M. R., & Dewi, L. S. (2020). Bagian 1: Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia). *Jurnal Media Informatika Budidarma*, 4(3), 855. <https://doi.org/10.30865/mib.v4i3.2312>

#### **PUSTAKA LAPORAN PENELITIAN**

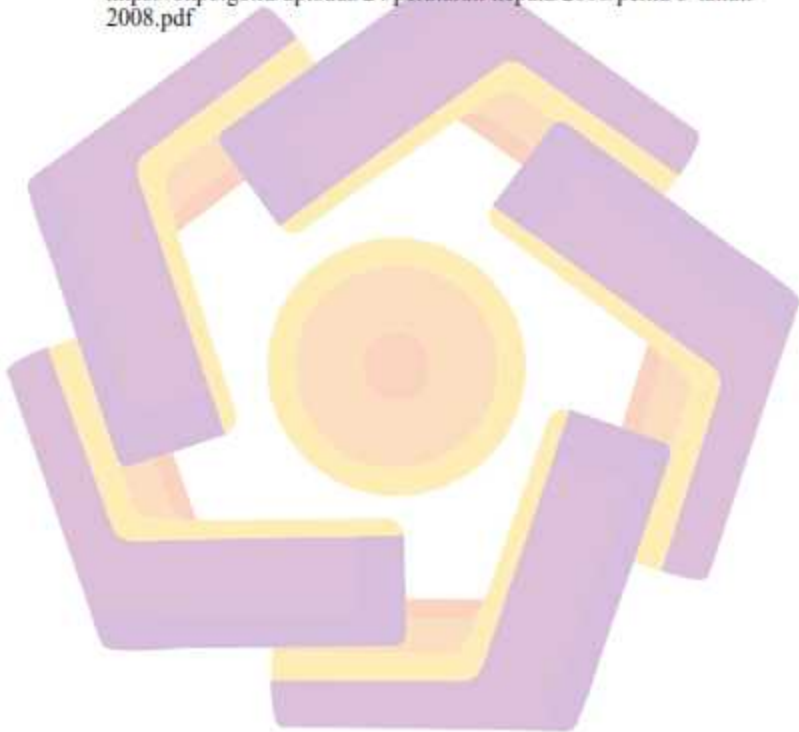
- Panafrican Emergency Training Centre. (2002). *DISASTERS & EMERGENCIES DEFINITIONS*. <https://apps.who.int/disasters/repo/7656.pdf>

## PUSTAKA ELEKTRONIK

- Azure, M. (2021, November 4). *Machine Learning Algorithm Cheat Sheet for Azure Machine Learning designer*. <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet>
- BMKG, I. (2021). *Tugas dan Fungsi*. <https://www.bmkg.go.id/profil/?p=tugas-fungsi>
- defeloper, scikit-learn. (2022). *Sklearn.ensemble.RandomForestClassifier*. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn-ensemble-randomforestclassifier>
- Ghoneim, S. (2019, February 4). *Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on?* <https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124>
- <https://scikit-learn.org/>. (n.d.). *Naive Bayes*. Retrieved November 20, 2021, from [https://scikit-learn.org/stable/modules/naive\\_bayes.html](https://scikit-learn.org/stable/modules/naive_bayes.html)
- Indonesia, B. (2014). *Rata-Rata Harian Aliran Sungai, Tinggi Aliran, dan Volume Air di Beberapa Sungai yang Daerah Pengalirannya Lebih dari 1.000 km2, 2000-2011*. <https://www.bps.go.id/statictable/2014/09/05/1371/rata-rata-harian-aliran-sungai-tinggi-aliran-dan-volume-air-di-beberapa-sungai-yang-daerah-pengalirannya-lebih-dari-1-000-km2-2000-2011-.html>
- Indonesia, B. (2019). *Luas Daerah dan Jumlah Pulau Menurut Provinsi, 2019*. [https://www.bps.go.id/indikator/indikator/view\\_data/pub/0000/api\\_pub/UfPwMmJZOVZIZTJnc1pXaHhDV1hPQT09/da\\_01/1](https://www.bps.go.id/indikator/indikator/view_data/pub/0000/api_pub/UfPwMmJZOVZIZTJnc1pXaHhDV1hPQT09/da_01/1)
- Indonesia, B. (2020). *Jumlah Penduduk Hasil Proyeksi Menurut Provinsi dan Jenis Kelamin (Ribu Jiwa), 2018-2020*. <https://www.bps.go.id/indikator/12/1886/1/jumlah-penduduk-hasil-proyeksi-menurut-provinsi-dan-jenis-kelamin.html>
- Indonesia, B. (2022a). *Data Informasi Bencana Indonesia*. <https://dibi.bnpb.go.id/xdibi>
- Indonesia, B. (2022b). *Tentang Profil BPS*. <https://www.bps.go.id/menu/1/informasi-umum.html>
- Krisnadwipayana, U. (-). *Daftar sungai di Indonesia*. [http://p2k.unkris.ac.id/id3/1-3065-2962/Daftar-Sungai-Di-Indonesia\\_26973\\_stie-walisongo\\_p2k-unkris.html](http://p2k.unkris.ac.id/id3/1-3065-2962/Daftar-Sungai-Di-Indonesia_26973_stie-walisongo_p2k-unkris.html)

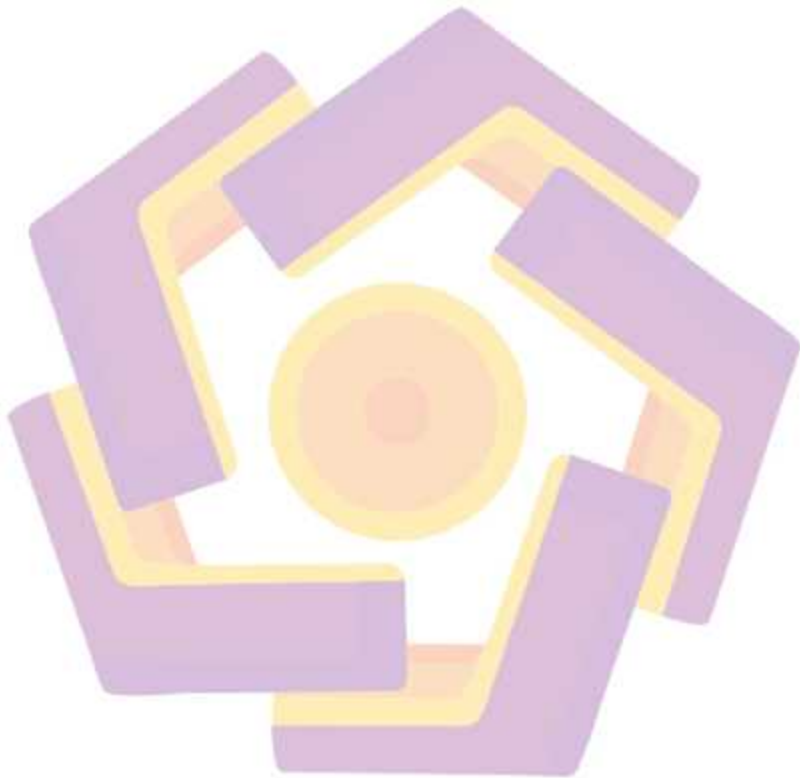


- Indonesia, R. (2008). *UNDANG UNDANG REPUBLIK INDONESIA NOMOR 14 TAHUN 2008 TENTANG KETERBUKAAN INFORMASI PUBLIK*. Republik Indonesia. <https://www.kemenkeu.go.id/sites/default/files/uu%2014%20tahun%202008.pdf>
- Republik Indonesia, B. (2008). *Kepala Badan Nasional Penanggulangan Bencana Nomor 3 Tahun 2008 Tentang Pedoman Pembentukan Badan Penanggulangan Bencana Daerah*. <https://bnpb.go.id/uploads/24/peraturan-kepala/2008/perka-3-tahun-2008.pdf>



## LAMPIRAN

Data sets faktor banjir dan klasifikasi



No	PROV	YEARS	RAIN_JAN	RAIN_FEB	RAIN_MAR	RAIN_APR	RAIN_MAY	RAIN_JUN	RAIN_JUL	RAIN_AUG	RAIN_SEP	RAIN_OKT	RAIN_NOV	RAIN_DEC	POPUL_DE	RIVERS	FLOODS
1	ACEH	2015	88	35	47	243,4	74,6	75,5	186,6	175,4	203,5	92,8	222,5	116,5	87	10	Banjir
2	ACEH	2016	153,5	183	96,3	67,1	237	92,9	21,2	393,3	178,9	445,4	265,7	199,1	88	10	Banjir
3	ACEH	2017	532	89	173,5	48	221	70,5	72,5	187,5	437	64,5	206	218	89	10	Banjir
4	ACEH	2018	181	45,5	91	107	216,5	88,5	121	108,5	143,3	263,3	287,9	176,8	90	10	Banjir
5	ACEH	2019	114,5	96	192	115,5	22,5	8	26	47,5	2,5	10	105,2	127	92	10	Banjir
6	ACEH	2020	7,6	56	19,8	183,6	433,6	149,7	137,2	3,4	29,8	75,8	230,8	374,2	93	10	Banjir
7	SUMATER	2015	190,4	64,4	233,6	173	230,1	100,2	160,2	180,7	176,4	327,9	448,6	134,5	191	14	Banjir
8	SUMATER	2016	190,4	305,4	160,6	99,7	371,7	123,4	180,1	281,8	580,8	327,2	197	154,8	194	14	Banjir
9	SUMATER	2017	284,5	139,6	317,4	186,5	282,1	158,9	205,2	262,9	524	234	179,8	251,8	196	14	Banjir
10	SUMATER	2018	224,3	39,1	88,6	300,6	225,9	200,5	609	136,7	375,9	393,4	243,8	257,3	198	14	Banjir
11	SUMATER	2019	208,2	204,7	130,6	151,4	424,8	374,4	199,7	130,9	388,1	455,6	455,6	319,2	201	14	Banjir
12	SUMATER	2020	86,9	80	103,9	329	837	615	300,3	229,1	465,7	465,7	336,6	396,5	203	14	Banjir
13	SUMATER	2015	272,6	204,5	273	185,9	185,9	298,4	166,1	319,9	75,6	80,4	485,3	391,4	124	25	Banjir
14	SUMATER	2016	229,3	305,5	470,4	412,8	561,3	415,2	280,3	631,9	458,9	697,4	340,2	499,5	125	25	Banjir
15	SUMATER	2017	369,5	179	227,6	262,5	644,4	151,3	278,4	589,7	543,7	280,4	719	404,4	127	25	Banjir
16	SUMATER	2018	150,4	228,3	229	319,7	475,9	233	123,1	140,1	385,5	621,4	551	470,6	129	25	Banjir
17	SUMATER	2019	272,8	96,2	297,4	251,8	147,1	427,8	240	126	58,1	184,7	130,1	398,4	130	25	Banjir
18	SUMATER	2020	443,9	203,4	377,2	227,9	459,2	199,2	324,4	188,5	416,7	443	644,5	245,7	132	25	Banjir
19	RIAU	2015	163,9	11,1	329	133,8	140,3	98	13,5	300,1	60	57,2	389,4	281,7	73	10	Tidak Ban
20	RIAU	2016	293,5	192,8	177,9	309,8	243,3	105,1	272,7	43,7	274	147,4	514	177,9	74	10	Tidak Ban
21	RIAU	2017	336,2	302,5	424,1	386,2	375,4	323	119	148,3	28	201,4	412,3	237,8	76	10	Banjir
22	RIAU	2018	169,6	118,5	243,2	161,1	217,4	291,3	125,3	107,7	111,1	292,4	335,5	525,6	77	10	Banjir
23	RIAU	2019	185,8	105,2	113,4	283	162,1	275,4	74,3	46	53,5	203,5	313,3	169,4	79	10	Banjir
24	RIAU	2020	121,8	30,2	96,7	351,6	246	196,8	109,2	200,3	113,1	195,3	359,3	104,8	80	10	Banjir
25	JAMBI	2015	143	122,5	169,7	266,8	126,2	34,6	73,1	37,4	28	36	294,4	325,1	68	11	Tidak Ban
26	JAMBI	2016	106,1	190,4	80,2	160,7	76	62,8	75,5	198,6	109,4	119,5	207,3	104,4	69	11	Banjir
27	JAMBI	2017	128,7	181,1	194	311,1	168	146,4	54,6	68,1	205,5	228,2	336,8	236	70	11	Banjir
28	JAMBI	2018	137	216,6	285,6	221	286,6	86,3	31,3	62	219	127,1	388,9	272	70	11	Tidak Ban
29	JAMBI	2019	173,1	291,2	142,9	343,7	148,6	140,8	20,8	2,8	48	176,2	115,2	147,5	71	11	Banjir
30	JAMBI	2020	150,1	89,3	222,6	213,2	218,7	95,1	190,8	95,9	142,5	346,7	290,3	321,7	72	11	Banjir
31	SUMATER	2015	233,3	173,2	36	20,8	62	2,8	58,6	24,4	1	2,1	173,8	192,1	88	12	Banjir
32	SUMATER	2016	217,5	261	321,7	227,3	200,2	111,8	106,3	214,5	229,9	234	256,6	204,6	89	12	Banjir
33	SUMATER	2017	267,6	318,3	293,8	244,9	288,5	127,8	73,8	109	89,9	255,2	232,5	270,1	90	12	Banjir
34	SUMATER	2018	200,5	199,1	451,4	125,1	172,8	142,7	29,6	66,2	140,9	169,3	326,4	289,7	92	12	Banjir

35	SUMATER	2019	260,1	263,9	312	359,6	72,1	138	119,1	7,1	46,2	53,9	164,4	450,3	93	12	Banjir
36	SUMATER	2020	206,3	195,9	334,5	352,7	239,4	275,4	131,2	52,9	56,3	280,4	384,5	240,7	94	12	Banjir
37	BENGKULI	2015	379	322,1	138,5	332,8	55	164	44,1	157,4	79,6	7,1	256,1	558,8	94	22	Banjir
38	BENGKULI	2016	262,5	224,8	637,6	245,9	280,9	128,6	115,8	423,4	208,8	365,5	554,8	280,1	95	22	Tidak Banjir
39	BENGKULI	2017	378,5	471,7	310,5	326,3	235,8	212,1	91,2	318,2	479,5	342,3	314,5	401,2	97	22	Tidak Banjir
40	BENGKULI	2018	220,3	198	268,5	185,4	152,9	206,1	37,1	321,3	155,9	283,3	530,3	489,3	98	22	Tidak Banjir
41	BENGKULI	2019	313,9	120,7	345,9	340	81	115,9	198	7,8	58	42,2	61,3	140,3	99	22	Banjir
42	BENGKULI	2020	406,6	245,3	429,4	582,9	221,1	226,2	90,9	136,4	270,9	461,7	607,2	555	100	22	Banjir
43	LAMPUNG	2015	418,5	279,3	260,3	0	11,5	14,5	22,8	108,9	0	6	45,4	132,1	235	10	Banjir
44	LAMPUNG	2016	215,8	122,1	257,7	322,8	106,8	76,9	57,4	79,2	155,5	77,5	286,1	135	237	10	Tidak Banjir
45	LAMPUNG	2017	117,6	244,2	105,2	158,6	206,8	45,1	10,1	38,6	77,5	28,2	119,7	179,7	240	10	Tidak Banjir
46	LAMPUNG	2018	85,4	181	178	226,2	71,9	42,6	87,9	144,3	11,7	21,9	26,7	180,7	242	10	Banjir
47	LAMPUNG	2019	72,8	69,3	218	163,2	10,1	46	78,3	6,8	0	38,4	47,6	241,9	244	10	Tidak Banjir
48	LAMPUNG	2020	791,7	125,4	410,1	202,3	274,4	255,6	154,3	224,8	33,3	102,5	110,6	266,3	247	10	Banjir
49	KEP. BANGKALAN	2015	171,6	70,3	145,9	83,7	77,2	71,3	19,9	11	65,8	31,9	109,1	230,1	84	0	Tidak Banjir
50	KEP. BANGKALAN	2016	235,2	602	407,2	261,1	252,7	170,9	85,7	158,7	413,5	273,3	90,7	155,4	85	0	Tidak Banjir
51	KEP. BANGKALAN	2017	404,7	227,9	220	251,9	231,3	47,7	184,6	139	71,4	211,1	258,5	358,4	86	0	Tidak Banjir
52	KEP. BANGKALAN	2018	66,1	175,3	284,1	260,5	261,4	158,3	31,9	64,2	173,2	88,1	206,4	318,6	87	0	Banjir
53	KEP. BANGKALAN	2019	161	473,7	270,1	221,4	214,1	101,7	65	15,4	20,3	52,5	41,3	365	88	0	Banjir
54	KEP. BANGKALAN	2020	180,3	137,1	175,7	292,3	413,1	259,3	188,5	49,9	36,1	353	161,2	318,5	89	0	Tidak Banjir
55	DKI JAKARTA	2015	472,6	939,5	207,1	82,9	16,6	10,1	0	5,2	5,1	0	79,5	273,2	15330	13	Tidak Banjir
56	DKI JAKARTA	2016	163,8	516,5	350,1	204	156,3	202,1	259,3	227,2	237,4	136,8	58,1	199,9	15460	13	Banjir
57	DKI JAKARTA	2017	214,2	520,8	138,7	156,5	135	138,5	119,9	0,8	165,8	112,4	195,4	254,1	15585	13	Banjir
58	DKI JAKARTA	2018	215,4	431,2	202	159,1	16,5	12,5	14,5	32,8	64,8	159,6	140,9	52,3	15705	13	Banjir
59	DKI JAKARTA	2019	382,2	270,1	327,3	194,6	47,8	23,1	207,5	178,4	0	1	50,1	263,8	15819	13	Tidak Banjir
60	DKI JAKARTA	2020	692	1043,2	220,7	182,8	50,4	21,1	12,1	101	3,9	208,3	87,3	134,7	15928	13	Banjir
61	JAWA BARAT	2015	167,3	179,7	264,5	231	208,1	50,4	0,3	6,9	43,2	34,5	419,4	307,4	1323	27	Banjir
62	JAWA BARAT	2016	194,3	194,3	376,2	523	317,8	139,3	182,3	128,7	286,2	362,3	442,5	62,1	1339	27	Banjir
63	JAWA BARAT	2017	196,3	196,3	396,5	210,8	222,3	68,4	7,9	45,7	90,8	345,3	442	129,9	1355	27	Banjir
64	JAWA BARAT	2018	239,3	239,3	292	297,5	123,9	33,4	0,3	38,9	40,8	124,8	25	322,9	1370	27	Banjir
65	JAWA BARAT	2019	269,3	269,3	223,3	298,9	243	26,5	13,4	0,2	55	84,2	270,9	315,5	1386	27	Banjir

66	JAWA BAF	2020	336,6	336,6	290,8	271,4	292,3	30,3	63,7	41,6	35,9	327,3	207,3	261,8	1401	27	Banjir
67	JAWA TEN	2016	109,9	109,9	177	32,7	108,5	140,8	5,7	2,3	177,3	99	41	253,7	1035	22	Banjir
68	JAWA TEN	2017	172,5	172,5	139	87,2	11,4	3	0	1,3	4	11	57,7	79	1041	22	Banjir
69	JAWA TEN	2018	171,8	171,8	107,8	88,2	29,4	24,4	47,3	33,7	55,9	21,5	128,3	91,9	1047	22	Banjir
70	JAWA TEN	2019	128,5	128,5	104,5	72,5	11,3	0,5	6,5	3,5	3,5	4,5	42,3	197	1053	22	Banjir
71	JAWA TEN	2020	136,9	136,9	458,4	331,2	234,7	41,5	32,4	12,3	4,8	448,1	269,7	623,5	1059	22	Banjir
72	DI YOGYAI	2015	328,98	328,98	213,6	168,38	94,54	65,56	73,4	65	103,1	127,4	216,4	301,6	1171	17	Tidak Banjir
73	DI YOGYAI	2016	320,2	320,2	409,9	184,7	140,4	296,5	105,7	94,5	240	327,2	508,2	267,1	1187	17	Tidak Banjir
74	DI YOGYAI	2017	347,7	347,7	403,4	243,4	45,7	9,2	12,7	22,5	62,8	60,3	689	370,3	1203	17	Tidak Banjir
75	DI YOGYAI	2018	338	338	190,9	107,5	9,9	17,4	44,9	1,1	20,6	22,2	268,4	171	1219	17	Tidak Banjir
76	DI YOGYAI	2019	240,6	240,6	618,4	99,3	0,8	1	200,1	193,4	185,5	113,3	22,2	214,5	1235	17	Tidak Banjir
77	DI YOGYAI	2020	398,4	398,4	516,8	207	275,9	3,7	3,3	13,7	6,6	114	182,8	341,6	1251	17	Tidak Banjir
78	JAWA TIM	2015	339,1	339,1	282,6	168,6	96,1	0,2	88,3	187,7	166,1	141,4	167,9	137,6	811	22	Banjir
79	JAWA TIM	2016	427	427	154,3	181,8	287,5	77	76,7	55,5	82,5	170,4	190,1	384,1	817	22	Banjir
80	JAWA TIM	2017	244,8	244,8	221,7	157,7	185,7	119,5	24,6	109,9	9,5	57	319,1	276,7	822	22	Banjir
81	JAWA TIM	2018	222,5	222,5	205,8	44	4,1	25,5	24,5	45,2	31,7	33,8	147,8	214,1	827	22	Banjir
82	JAWA TIM	2019	316,5	316,5	183,5	290,2	18,2	164	1,7	61,3	75,7	46,2	16,8	72,8	831	22	Banjir
83	JAWA TIM	2020	481,1	481,1	384,8	436	96,2	29,6	14,7	23,1	22,5	68,2	103,9	532,9	836	22	Banjir
84	BANTEN	2015	241,4	241,4	193,1	130,2	39,1	83,4	4,7	11,7	0,2	28,5	96,2	154,2	1239	4	Banjir
85	BANTEN	2016	271,9	271,9	217,8	88	137,2	93	134,9	86,6	158,9	158,3	138	164	1258	4	Banjir
86	BANTEN	2017	351,1	351,1	112,9	93,5	142,8	100,8	106,9	18,1	47,5	112,3	153,5	240,1	1278	4	Banjir
87	BANTEN	2018	178,2	178,2	222,2	153,7	69,6	196,6	3	38,8	29,9	87,6	116,5	167,9	1297	4	Banjir
88	BANTEN	2019	179,9	179,9	209,9	150,7	93,8	10	14,4	1	8,5	11,2	74,8	179	1316	4	Tidak Banjir
89	BANTEN	2020	298,3	298,3	287,7	202	185,6	79,9	30,9	28,8	9,4	62,1	131,1	257,4	1335	4	Banjir
90	BALI	2015	245,9	245,9	272,2	33,2	53	119,5	0,5	5,8	0,7	0	13	157,7	718	15	Tidak Banjir
91	BALI	2016	448,1	448,1	9,7	31,1	21,4	117,2	191,4	39	235,9	127,4	322,9	398	727	15	Tidak Banjir
92	BALI	2017	374,4	374,4	120,7	165,4	85,2	91,1	18,9	16,8	7,4	21,1	447,7	606	736	15	Tidak Banjir
93	BALI	2018	203,5	203,5	95,6	25,5	3,9	46,4	34,8	88,8	6,5	20	302,9	61,8	746	15	Banjir
94	BALI	2019	129,6	129,6	339,6	38,9	9,4	1,1	1,8	9,6	9,7	0	17,8	142,2	755	15	Tidak Banjir
95	BALI	2020	386,6	386,6	216,8	56	90,9	104	17,7	2,5	7,6	197	41,9	346,9	764	15	Tidak Banjir
96	NUSA TEN	2015	128,3	128,3	157,3	117	2,2	92,2	70,5	1,5	54,7	0,3	38,5	197,6	261	19	Tidak Banjir
97	NUSA TEN	2016	156,6	156,6	157,2	124,9	51,3	112,5	2,6	11,7	38,8	85,3	161,8	200,5	265	19	Banjir
98	NUSA TEN	2017	154	154	229,3	121,9	29,8	62,3	0,4	0,5	0,2	99,7	199,6	175,8	269	19	Banjir
99	NUSA TEN	2018	189,8	189,8	136,5	35	120,4	97,3	84,2	2,2	14,6	7,5	132,5	247,9	273	19	Banjir
100	NUSA TEN	2019	128,4	128,4	125,4	117,7	0,6	81,2	66,5	49,4	2,1	49,4	52,6	201,8	277	19	Banjir

101	NUSA TEN	2020	327,7	327,7	208,4	34	35	0,2	4	13,1	5,8	95,2	33,1	224,2	281	19	Banjir
102	NUSA TEN	2015	107,5	107,5	292,9	60,7	13	122,2	62,3	66,8	84,8	72,3	16,0	200,0	105	13	Tidak Banjir
103	NUSA TEN	2016	106,7	106,7	150,9	121,4	84,4	3,2	17,4	35	26,7	8,2	26,1	229,2	107	13	Tidak Banjir
104	NUSA TEN	2017	384,1	384,1	120,7	18,3	0,0	48,5	3,1	0,4	17,3	65,8	130,1	181,0	108	13	Tidak Banjir
105	NUSA TEN	2018	190	190	91,4	55,2	112,2	2	56,5	4,1	0,5	0,3	177,7	308,2	110	13	Banjir
106	NUSA TEN	2019	132,4	132,4	199,0	29,8	18,0	1,0	16,7	0	8,1	12,8	16,0	180,8	112	13	Tidak Banjir
107	NUSA TEN	2020	226,9	226,9	244,8	90,4	24,2	0,3	38,3	20,9	19,8	35,5	39,1	305,7	113	13	Tidak Banjir
108	KALIMAN	2015	128,8	128,8	151,2	121,9	278,6	258,1	201,5	41,1	5,7	112	166,8	256,0	33	20	Tidak Banjir
109	KALIMAN	2016	292,8	292,8	235,2	130,3	725,0	310,9	308,5	68,0	194	240,1	284,9	282,8	33	20	Banjir
110	KALIMAN	2017	220	220	100,1	101,5	267,4	240,4	321,8	510,5	128,0	108,9	247,2	196,8	33	20	Banjir
111	KALIMAN	2018	103,5	103,5	234,1	252,9	549,8	400,4	50,5	78	319,2	515,8	428,8	449,3	34	20	Banjir
112	KALIMAN	2019	247,7	247,7	107	299,2	246,7	515,8	227,9	73	35,2	582,3	387,5	584,1	34	20	Banjir
113	KALIMAN	2020	421,6	421,6	212,5	237	282,3	278,2	408,0	164,9	281,5	213,5	585	179,7	35	20	Banjir
114	KALIMAN	2015	467	467	427	280,0	322,3	135	31,9	8,4	38,4	32,9	416,2	262,7	16	8	Tidak Banjir
115	KALIMAN	2016	438,2	438,2	248,5	442,8	292	307,4	148,1	188,5	281	293,4	261,2	183,1	17	8	Tidak Banjir
116	KALIMAN	2017	161,4	161,4	475,7	236,1	475,7	322,3	134,4	137,7	60,7	237,3	409,2	403	17	8	Banjir
117	KALIMAN	2018	207,2	207,2	342,7	421,9	133,4	118,0	148,3	73,3	17,3	155,8	265,2	324,3	17	8	Banjir
118	KALIMAN	2019	481	481	395,5	274,9	69,7	35	7,3	58,5	55,1	179,7	133,1	361,3	17	8	Banjir
119	KALIMAN	2020	263,9	263,9	339,3	382,8	307,5	131,2	69,9	119,2	60,9	238	180,5	159,9	17	8	Banjir
120	KALIMAN	2015	322,5	322,5	196,7	0	64,3	104,4	23,7	38,8	55,0	7,4	112,7	327,2	103	25	Tidak Banjir
121	KALIMAN	2016	295,1	295,1	258,5	477,7	208,2	210,5	92,5	72,7	140	112,2	270,4	140	105	25	Banjir
122	KALIMAN	2017	389,1	389,1	235,0	221,7	210,8	196,2	203,1	102,0	90,0	138,2	299,4	364,0	106	25	Banjir
123	KALIMAN	2018	313,4	313,4	369,0	174,4	75	111,7	77,2	78,3	100,7	100,8	227,4	433,1	107	25	Banjir
124	KALIMAN	2019	350,4	350,4	282,1	322,1	54,3	146,1	38,4	12,5	99	59,3	121,2	269,2	109	25	Banjir
125	KALIMAN	2020	334,3	334,3	302,4	266,4	138	218	62,4	57,2	162,9	190,4	282,1	554,1	110	25	Banjir
126	KALIMAN	2015	146,0	146,0	198,8	380,3	229,2	258,3	154,5	57,0	156,8	70,5	70,4	198,4	27	11	Banjir
127	KALIMAN	2016	102,8	102,8	117,5	382,7	243,7	157,8	170,8	101,1	266,8	184,9	292,4	355,0	27	11	Banjir
128	KALIMAN	2017	140,8	140,8	88,1	341,0	310,8	315,3	104	237,4	100,5	151,0	219,8	222,3	27	11	Banjir
129	KALIMAN	2018	97,8	97,8	155,5	182,1	508,4	198,4	125,2	50,7	127,4	152,9	126,9	58,0	28	11	Banjir
130	KALIMAN	2019	20,1	20,1	198,0	142,7	198,7	204,0	52,7	63,4	47,5	197,4	131,5	401,7	28	11	Banjir
131	KALIMAN	2020	91,8	91,8	135,5	176,5	224,7	172,0	148	228,7	154,0	170	185,3	117,9	28	11	Banjir
132	KALIMAN	2015	283,7	283,7	309,9	266,4	232,0	224,8	100,1	188,0	352,1	202,0	372,5	334	9	4	Tidak Banjir
133	KALIMAN	2016	198,0	198,0	266,2	235,9	271	438,8	324,5	417	405,9	818	277	410,5	9	4	Tidak Banjir
134	KALIMAN	2017	218,3	218,3	459,3	399,3	395,3	161,5	294,2	291,4	260,3	292,5	321,8	514,5	9	4	Tidak Banjir
135	KALIMAN	2018	383,7	383,7	315,1	251,2	204,2	138,0	517	126,3	267,0	379,9	343,9	313,1	9	4	Tidak Banjir
136	KALIMAN	2019	68	68	417,4	457,2	431,9	275,9	317,5	140	253,4	414,1	441,9	572,5	9	4	Tidak Banjir
137	KALIMAN	2020	117,2	117,2	381,3	218	435,2	249,8	347,5	288,4	301,7	333,3	358,2	440,2	9	4	Tidak Banjir

