

NASKAH PUBLIKASI (*MANUSCRIPTS*)

**PENERAPAN PARTICLE SWARM OPTIMIZATION (PSO) UNTUK
MENINGKATKAN AKURASI ALGORITMA NAIVE BAYES PADA BANJIR
DI KOTA SAMARINDA**

***THE APPLICATION OF PARTICLE SWARM OPTIMIZATION
(PSO) TO IMPROVE THE ACCURACY OF NAIVE BAYES
ALGORITHM IN FLOOD PREDICTION IN SAMARINDA CITY***

Trisha NurHalisha, Faldi, Wawan Joko Pranoto



DISUSUN OLEH :

TRISHA NURHALISHA

1911102441099

**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
SAMARINDA**

2023

Naskah Publikasi (*Manuscripts*)

**Penerapan Particle Swarm Optimization (PSO) untuk
Meningkatkan Akurasi Algoritma Naive Bayes pada Banjir
Di Kota Samarinda**

***The Application of Particle Swarm Optimization (PSO) to
Improve the Accuracy of Naive Bayes Algorithm in Flood
Prediction in Samarinda City***

Trisha NurHalisha, Faldi, Wawan Joko Pranoto



Disusun Oleh :

Trisha Nurhalisha

1911102441099

**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
SAMARINDA
2023**

HALAMAN PENGESAHAN

**PENERAPAN PARTICLE SWARM OPTIMIZATION (PSO)
UNTUK MENINGKATKAN AKURASI ALGORITMA NAIVE
BAYES PADA BANJIR DI KOTA SAMARINDA**

NASKAH PUBLIKASI

DISUSUN OLEH :
TRISHA NURHALISHA
1911102441099

Dosen Pembimbing



Faldi, S.Kom., M.Ti
NIDN. 1121079101

Penguji



Wawan Joko Pranoto, S.Kom., M.Ti
NIDN. 1102057701

Dekan



Prof. Ir. Saqito, MT., Ph.D
NIDN. 0610116204

Ketua Program Studi



Prof. Ir. Saqito, MT., Ph.D
NIDN. 1124098902

Penerapan *Particle Swarm Optimization (PSO)* Untuk Meningkatkan Akurasi Algoritma *Naive Bayes* Pada Banjir Di Kota Samarinda

Trisha NurHalisha 1), Faldi 2), Wawan Joko Pranoto 3)

Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur

Jl. Ir. H. Juanda No. 15 Samarinda, Indonesia

E-Mail : 1911102441099@umkt.ac.id¹⁾; fal146@umkt.ac.id²⁾; wjp337@umkt.ac.id³⁾;

ABSTRAK

Banjir adalah permasalahan yang sering terjadi di berbagai kota, termasuk Kota Samarinda. Penelitian ini bertujuan untuk menggunakan teknik Particle Swarm Optimization (PSO) untuk meningkatkan akurasi algoritma Naive Bayes dalam memprediksi banjir di Kota Samarinda. Nilai akurasi dari hasil pengujian yang diperoleh dengan menggunakan metode Naive Bayes saja sebesar 91,12%. Sementara terjadi peningkatan akurasi setelah dilakukan pengujian teknik optimasi berbasis Particle Swarm Optimization (PSO) dan Algoritma Naive Bayes. Pengujian yang telah dilakukan, dimana nilai akurasi mencapai 94,38%. Hasil akurasi ini lebih tinggi daripada pengujian tanpa tanpa optimasi.

Kata Kunci – Banjir, Akurasi, Naive Bayes, Particle Swarm Optimization (PSO), Optimasi.

1. PENDAHULUAN

Badan Nasional Penanggulangan Bencana (BNPB) menyatakan sepanjang tahun 2022 mencatat adanya 3.522 bencana alam di Indonesia yang terjadi. Kejadian bencana alam banjir sering terjadi tahun 2022, dengan jumlah 1.520 peristiwa (Mustajab, 2023). Meningkatnya bencana banjir dapat menyebabkan berbagai macam kerugian, baik itu kerugian ekonomi atau masalah kesehatan. Berbagai upaya telah dilakukan untuk mengatasi bencana banjir yang datang, akan tetapi hingga saat ini bencana banjir belum bisa diatasi dengan baik.

Ibu Kota Provinsi Kalimantan Timur yakni Kota Samarinda, memiliki 27 aliran sungai dan wilayah dataran rendah dengan sistem drainase air yang kurang baik. Kondisi topografi wilayah Kota Samarinda menyebabkan seringnya terjadi bencana banjir. Curah hujan yang tinggi atau naiknya permukaan air sungai menjadi dua penyebab utama terjadinya banjir di Kota Samarinda. 75 banjir tercatat di Kota Samarinda antara tahun 2019 dan 2022 menurut data BPBD dan BMKG. Algoritma Naive Bayes yang digunakan dalam data mining dapat memberikan angka yang tepat untuk kejadian banjir di Kota Samarinda, oleh karena itu diperlukan analisis bencana banjir.

Untuk mengatasi jumlah data yang besar, mengatasi nilai yang kosong, dan menghadapi beragam atribut dan masalah pada data, digunakan algoritma Naive Bayes (Arifin & Ariesta, 2019). Algoritma Particle Swarm Optimization (PSO) digunakan sebagai teknik optimasi untuk meningkatkan akurasi Algoritma Naive Bayes dalam memprediksi banjir di Kota Samarinda. Penelitian ini menggunakan metode optimasi Particle Swarm Optimization (PSO) untuk meningkatkan akurasi Algoritma Naive Bayes terhadap banjir di Kota Samarinda.

2. TINJAUAN PUSAKA

A. Banjir

Fenomena alam yang dikenal sebagai banjir dipicu oleh tingginya intensitas hujan atau melebihi kapasitas tampung debit air di suatu daerah (santoso dian, 2019). Secara singkat, banjir dapat diartikan sebagai kondisi di mana air menutupi luas wilayah permukaan bumi. Beberapa jenis banjir antara lain: (Putra et al., 2019).

B. Data Mining

Data Mining adalah metode untuk memperoleh pengetahuan yang relevan dan informasi yang dapat digunakan dari basis data besar menggunakan teknik statistik, machine learning, kecerdasan buatan (Utomo & Mesran, 2020). Data Mining juga dikenal sebagai knowledge discovery in database (KDD) yaitu tindakan mengumpulkan dan menganalisis data historis untuk mencari pola, keteraturan, atau korelasi dalam data dalam jumlah besar. Kata “pattern recognition” jarang digunakan karena merupakan komponen dari proses Data Mining, yang dapat dimanfaatkan untuk memperbaiki pengambilan keputusan di masa mendatang (Sibarani, 2020).

Pembersihan data untuk menghilangkan noise dan data yang tidak konsisten, integrasi data untuk menggabungkan data dari banyak sumber, pemilihan data, dan proses pemilihan adalah beberapa prosedur yang harus diselesaikan dalam data mining sebelum informasi baru diperoleh (Mandar, 2022). Data Mining sendiri dibagi menjadi beberapa kelompok, yang dijelaskan oleh (Naldy & Andri, 2021):

1. Deskripsi (Description): Proses ini mencakup penyajian pola dan kecenderungan yang memberikan penjelasan pada data. Misalnya, pada pemilihan presiden, petugas pengumpul suara tidak menemukan pernyataan atau fakta tentang kurangnya profesionalitas yang sedikit mendukung pemilihan presiden.
2. Pengklusteran (Clustering): Pengelompokan data menjadi kelompok-kelompok berdasarkan kesamaan

mereka. Data yang mirip satu sama lain dan berbeda dari cluster lain dikelompokkan bersama sebagai cluster. Clustering berbeda dari klasifikasi karena tidak memiliki variabel target.

3. Asosiasi (Association): Proses ini digunakan untuk mengidentifikasi kejadian atau proses tertentu dengan atribut yang muncul bersamaan dalam setiap kejadian. Asosiasi mencoba menyusun nilai yang muncul bersamaan pada setiap baris data dan menampilkan hasil keluaran dalam bentuk aturan.

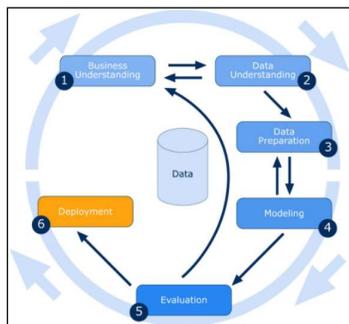
4. Klasifikasi (Classification): Proses ini mencari model atau fungsi yang mendeskripsikan data dan mengkategorikannya ke dalam kelompok tertentu. Klasifikasi melibatkan pemeriksaan karakteristik objek dan mengklasifikasikan objek ke dalam kategori yang telah ditentukan sebelumnya

5. Estimasi (Estimation): Estimasi mirip dengan klasifikasi, tetapi dalam estimasi variabel target menghasilkan angka bukan kategori. Saat memodelkan menggunakan record lengkap dan nilai variabel target diubah menjadi nilai prediksi.

6. Prediksi (Prediction): Mirip dengan klasifikasi, namun data diklasifikasikan menurut pola perilaku atau nilai yang akan terjadi di masa yang akan datang. Contohnya adalah perkiraan perekonomian atau tingkat pengangguran dalam beberapa tahun mendatang.

C. CRISP-DM

CRISP-DM merupakan standarisasi proses dari data mining yang menjadi pendekatan umum untuk pemecahan sebuah masalah (Feblian & Daihani, 2017). Kerangka kerja ini terdiri dari enam tahap dalam siklus hidupnya, dan tahap-tahap tersebut bersifat adaptif, artinya bergantung pada hasil dari tahap sebelumnya. Hubungan ketergantungan yang signifikan antar tahapan ditunjukkan oleh panah pada gambaran siklus CRISP-DM (Hasanah et al., 2021).



Gambar 1. Tahapan CRISP-DM
Sumber : (Hasanah et al., 2021)

1. Business Understanding (Pemahaman Bisnis)
Pada tahap ini, langkah-langkah yang dilakukan meliputi memahami kebutuhan dan tujuan dari sudut pandang bisnis serta menerjemahkan pengetahuan ke dalam definisi masalah dalam data mining. Selain itu, identifikasi rencana dan strategi juga dilakukan untuk mencapai tujuan data mining.

2. Data Understanding (Pemahaman Data)

Tahap ini dimulai dengan pengumpulan data, eksplorasi data untuk menganalisisnya, dan melakukan evaluasi terhadap kualitas data yang ada.

3. Data Preparation (Pengolahan Data)

Pada tahap ini, dataset final dibangun dari data mentah. Tindakan yang dilakukan meliputi pembersihan data (Data Cleaning), pemilihan data yang relevan (Data Selection), pengaturan ulang record dan atribut-atribut, serta melakukan transformasi data (Data Transformation) agar dapat digunakan sebagai input pada tahap pemodelan.

4. Modeling (Pemodelan)

Pada tahap ini, machine learning digunakan untuk memilih teknik data mining, alat bantu data mining, serta algoritma data mining yang akan digunakan.

5. Evaluation (Evaluasi)

Tahap ini digunakan untuk menilai kinerja pola yang dihasilkan algoritma untuk memastikan kualitas dan efektivitas. Selain itu, mengevaluasi apakah model telah mencapai tujuan yang dinyatakan dan mengidentifikasi masalah atau bidang penelitian yang memerlukan perbaikan. Keputusan yang relevan dibuat dengan menggunakan temuan data mining.

6. Deployment (Penyebaran)

Tahap ini melibatkan penyusunan laporan berdasarkan model yang telah dihasilkan agar dapat disebarkan atau digunakan secara lebih luas.

D. Naïve Bayes

Naive Bayes Classifier (NBC) adalah salah satu algoritma cukup sederhana, namun sering digunakan dalam metode data mining untuk menerapkan teori Bayes dan melakukan klasifikasi (Tasya, 2021).

Klasifikasi Naive Bayes merupakan metode pengklasifikasian dalam data mining yang berdasarkan teorema Bayes, menggunakan pendekatan statistik independen dan probabilitas (Monowati & Setyadi, 2023). Pengklasifikasi Naive Bayes juga dikenal menghasilkan tingkat akurasi yang tinggi dan beroperasi dengan cepat pada database yang besar. Dalam penerapannya, Naive Bayes menghitung probabilitas setiap atribut pada kelas yang relevan. Persamaan umum dari Naive Bayes adalah sebagai berikut (Wiratama & Pradnya, 2022).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Keterangan :

X : Data pada class yang belum ditemukan,

H : Hipotesis data X menunjukkan suatu class

spesifik ,

P(H|X) : Probabilitas hipotesis H beralaskan kondisi X (posteriori prob.),

P(H) : Probabilitas hipotesis H (prior prob.),

P(X|H) : Probabilitas X berdasarkan kondisi dari hipotesis H,

P(X) : Probabilitas dari X

E. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) adalah teknik yang menarik inspirasi dari perilaku alami kawanan serangga, burung, lebah, dan hewan lainnya. Algoritma Particle Swarm Optimization (PSO) meniru bagian dari perilaku alami organisme ini, termasuk rutinitas dalam aktivitas sehari-hari dan bagaimana seseorang memengaruhi orang lain dalam suatu populasi. "Partikel" adalah anggota populasi, seperti burung dalam kawanan burung. Setiap orang atau partikel memiliki kecerdasannya masing-masing dan juga dipengaruhi oleh cara kelompok masyarakat lain bertindak. Terlepas dari lokasi asalnya, yang mungkin jauh dari kelompoknya, jika salah satu partikel menemukan rute yang lebih cepat atau lebih efisien ke sumber makanannya, partikel lain akan mengikuti rute tersebut (Rizki & Nurlaili, 2021).

Algoritma Particle Swarm Optimization (PSO) memiliki langkah-langkah sebagai berikut: (Sa'diyah et al., 2020):

Langkah 1: Inisialisasi jumlah partikel secara acak, dan memberikan nilai kecepatan awal partikel = 0.

Langkah 2: Menghitung nilai fitness dari setiap partikel menggunakan persamaan 2 berikut.

$$f = \frac{\sum_{i=1}^{Nc} \left(\frac{\sum_{j=1}^{Nx} d(x_j, c_i)}{Nx} \right)}{Nc} \quad (2)$$

Keterangan :

Nc = Jumlah cluster

Nx = Jumlah data pada cluster ke $-i$

$d(x_j, c_i)$ = Jarak data dengan cluster

Langkah 3: Memperbarui nilai Pbest dengan menggunakan persamaan 3 berikut ini.

$$Pbest = \begin{cases} Pbest_i, & \text{jika } F(x_i) \geq F(Pbest_i) \\ x_i, & \text{jika } F(x_i) < F(Pbest_i) \end{cases} \quad (3)$$

Keterangan :

$F(x_i)$ = Nilai fitness pada partikel ke $-i$

$F(Pbest_i)$ = Nilai fitness partikel baik ke $-i$

Langkah 4 : Perbarui nilai Gbest menggunakan persamaan 4 berikut.

$$Gbest_i = \text{MIN}(F(Pbest_i), \dots, F(Pbest_n)) \quad (4)$$

Langkah 5 : Perbarui nilai kecepatan menggunakan persamaan 5 berikut.

$$v_{ij}(t+1) = w \cdot v_{ij}(t) + c_1 r_1 [P_{ij}(t) - x_{ij}(t)] + c_2 r_2 [G_j(t) - x_{ij}(t)] \quad (5)$$

Keterangan :

w = bobot inertia

$v_{ij}(t)$ = kecepatan inertia ke $-i$

c_1 dan c_2 = nilai konstanta akselerasi

r_1 dan r_2 = nilai random dari 0 sampai 1

$x_{ij}(t)$ = partikel ke $-i$ dengan atribut ke $-j$ pada iterasi ke t

$P_{ij}(t)$ = partikel baik ke $-i$ dengan atribut ke $-j$

$G_j(t)$ = atribut ke $-j$ pada partikel terbaik

Partikel yang bergerak dalam Particle Swarm Optimization (PSO) bisa keluar dari batas ruang pencarian akibat rentang kecepatan yang terlalu besar, yang mengakibatkan penurunan efisiensi dan kecepatan konvergensi algoritma. Untuk mengatasi masalah ini, beberapa batasan diterapkan pada nilai komponen kecepatan. Setelah menghitung kecepatan menggunakan Persamaan 6, semua komponen dari setiap dimensi akan dibatasi dalam rentang $[-V_{max}, V_{max}]$ untuk mengurangi probabilitas partikel keluar dari batas ruang pencarian.

Langkah 6: Nilai posisi diperbarui dengan rumus yang dapat dilihat pada Persamaan 6 berikut ini:

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (6)$$

Keterangan:

$v_{ij}(t+1)$ = nilai kecepatan pada iterasi ke $-t+1$

$x_{ij}(t)$ = partikel ke $-i$ dengan atribut ke $-j$ pada iterasi ke $-t$

Langkah 7: Ulangi langkah 2, 3, dan 4 hingga mencapai jumlah iterasi maksimum.

F. K-Fold Cross Validation

Cross-validation adalah teknik dalam data mining yang bertujuan untuk mencapai akurasi maksimal, data dibagi menjadi dua subset yaitu data training dan data testing. Salah satu jenis uji cross-validation yang umum digunakan adalah K-Fold cross-validation, yang bertujuan untuk mengevaluasi efisiensi proses algoritma dengan membagi data sampel menjadi beberapa bagian secara acak dan menggabungkan data dengan nilai K dari K-fold. Dengan menggunakan metode K-Fold Cross Validation, kumpulan data dibagi secara acak menjadi beberapa bagian. Data partisi tersebut kemudian diproses dalam beberapa percobaan di mana setiap percobaan menggunakan satu bagian sebagai data testing (Ke-K) dan bagian lain sebagai data training (Arisandi et al., 2022).

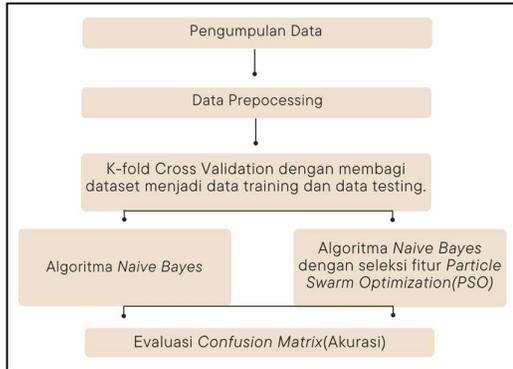
G. Confusion Matrix

Confusion Matrix adalah metode dalam mengukur efektivitas atau tingkat akurasi dari proses klasifikasi. Terdiri dari empat istilah dari proses klasifikasi yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) (trivusi, 2022).

H. RapidMiner

RapidMiner adalah perangkat lunak open source yang menyediakan berbagai metodologi deskriptif dan preskriptif, untuk membantu pengguna membuat keputusan sebaik mungkin. Lebih dari 500 operator data mining, termasuk operator input, output, persiapan data, dan visualisasi, tersedia di RapidMiner. Perangkat lunak RapidMiner ditulis menggunakan bahasa pemrograman Java, sehingga dapat berfungsi pada berbagai sistem operasi (Manullang et al., 2021).

3. METODE PENELITIAN



Gambar 2. Metode Penelitian

Pengumpulan Data

Untuk mengumpulkan data akurat yang dapat digunakan dalam memecahkan masalah, pada tahap pengumpulan data. Peneliti mendapatkan informasi data banjir dengan melakukan observasi secara langsung pada unit BPBD dan BMKG kota Samarinda. Berdasarkan ketersediaan data yang di peroleh yaitu tahun 2019 hingga periode tahun 2022 dengan jumlah data 1461 yang terdiri dari 20 atribut. Adapun atribut yang diperoleh seperti tabel berikut : Tabel 1 Atribut Data BMKG dan BPBD Kota Samarinda

Atribut	Keterangan
Tn	Temperatur minimum (°C)
Tx	Temperatur maksimum (°C)
Tavg	Temperatur rata-rata (°C)
RH_avg	Kelembapan rata-rata (%)
RR	Curah hujan (mm)
ss	Lamanya penyinaran matahari (jam)
ff_x	Kecepatan angin maksimum (m/s)
ddd_x	Arah angin saat kecepatan maksimum (°)
ff_avg	Kecepatan angin rata-rata (m/s)
ddd_car	Arah angin terbanyak (°)
Tanggal	Waktu Kejadian
Terjadi Banjir	Ya/ Tidak
Jam Kejadian	Jam terjadinya bencana
Jenis Bencana	Peristiwa bencana alam yang terjadi
Lokasi Wilayah	Tempat terjadinya banjir
Luas Area M2	Luas area yang terdampak
Objek Terkena Bencana	Fasilitas yang terdampak bencana
Korban	Jumlah korban yang terdampak bencana
Kerugian	Nominal kerugian yang dialami
Keterangan	Detail kejadian bencana

Setelah pengumpulan data, langkah selanjutnya adalah Pemahaman Bisnis (Business Understanding) dan Pemahaman Data (Data Understanding).

Data PreProcessing

Proses Data Preprocessing melibatkan tahapan-tahapan berikut:

- **Data Selection:** Setelah proses pengumpulan data, langkah ini melibatkan seleksi atribut yang akan digunakan dari data BPBD Kota Samarinda dan BMKG Kota Samarinda.
- **Data Integration:** Pada tahap ini, data dari BMKG Kota Samarinda dan BPBD Kota Samarinda digabungkan menjadi satu dataset.
- **Data Transformation:** Tahap ini dilakukan untuk mengubah isi data atau tipe data sebelum digunakan dalam pemodelan.
- **Data Cleaning:** Tahap ini dilakukan untuk membersihkan data dari missing value.

Pembagian Data

Dataset dibagi menjadi dua kelompok, yaitu data training dan data testing. Data training digunakan untuk melatih model atau algoritma, sedangkan data testing digunakan untuk menguji model dengan memberikan input data testing dan membandingkan hasil prediksi model dengan nilai target yang sebenarnya. Dalam penelitian ini, terdapat parameter $cv = 10$ diimplementasikan menggunakan perangkat lunak RapidMiner, dan diterapkan teknik validasi K-Fold Cross Validation.

Pemodelan (Modeling)

Pada tahapan ini, dilakukan klasifikasi data banjir kota Samarinda menerapkan Algoritma Naive Bayes serta penambahan teknik optimasi dari Particle Swarm Optimization (PSO). Tahap awal dari Algoritma Naive Bayes melibatkan perhitungan probabilitas kelas dan atribut. Kemudian, probabilitas gabungan dari kelas dan atribut dihitung dan digunakan dalam proses klasifikasi. Algoritma Naive Bayes memiliki kemampuan untuk mengklasifikasikan data baru, dan performanya dievaluasi dengan membandingkan hasil klasifikasi dengan label kelas yang sebenarnya. Berbagai metrik evaluasi seperti akurasi, presisi, dan recall dapat digunakan untuk mengukur kinerja algoritma ini.

Selanjutnya, diterapkan teknik optimasi dari Particle Swarm Optimization (PSO). Tahapan PSO dimulai dengan inialisasi, yaitu menentukan ukuran populasi partikel, inialisasi posisi dan kecepatan awal setiap partikel secara acak, serta inialisasi posisi partikel terbaik (pbest) dan posisi global terbaik (gbest). Setelah itu, tahap evaluasi dilakukan dengan menghitung nilai fungsi tujuan (fitness) untuk setiap partikel berdasarkan posisinya saat ini. Selanjutnya, dilakukan pembaruan Pbest dan Gbest, serta dilakukan pembaruan kecepatan dan posisi partikel. Proses evaluasi konvergensi dan iterasi berlanjut hingga menghasilkan output dengan posisi partikel terbaik (gbest) yang memiliki nilai fitness optimal.

Evaluasi (Evaluation)

Pada tahap Evaluasi (Evaluation), dilakukan validasi dan pengukuran keakuratan hasil yang diperoleh menggunakan metode CRISP-DM dan

Confusion Matrix untuk mengukur tingkat akurasi algoritma tersebut.

4. HASIL DAN PEMBAHASAN

Pengumpulan Data

Penelitian ini menggunakan data yang diperoleh dari data BPBD dan data BMKG Kota Samarinda. Berdasarkan ketersediaan data yang di peroleh yaitu tahun 2019 hingga periode tahun 2022 dengan jumlah data 1461 yang terdiri dari 20 atribut. Data dari BMKG Kota Samarinda diperoleh informasi yang akan dijadikan atribut pada penelitian ini diantaranya seperti Temperatur minimum, Temperatur maksimum, Temperatur rata-rata, Kelembapan rata-rata, Curah hujan, Lamanya penyinaran matahari, Kecepatan angin maksimum, Arah angin saat kecepatan maksimum, Kecepatan angin rata-rata, dan Arah angin terbanyak. Sementara itu, data dari BPBD Kota Samarinda berisi label atau kelas untuk penentuan kejadian banjir.

Data Preparation

• Data Selection

Tahapan ini dilakukan setelah pengumpulan data, di mana data yang sudah dikumpulkan akan diseleksi untuk menentukan atribut yang akan digunakan. Terdapat dua tahapan pada proses seleksi. Pada tahap pertama, atribut yang tidak diperlukan dari data BPBD Kota Samarinda, seperti jam kejadian, jenis bencana, lokasi wilayah, luas area, jumlah objek terkena bencana, korban, kerugian, dan keterangan, dihapus. Pada tahap kedua, atribut tanggal dari data BMKG Kota Samarinda dihapus karena informasi tanggal tersebut hanya merupakan rekapan waktu harian, sementara pada penelitian ini atribut tanggal yang diperlukan sebagai informasi waktu terjadinya banjir berasal dari data BPBD.

• Data Integration

Tahapan ini bertujuan untuk menggabungkan data dari BMKG Kota Samarinda dan BPBD Kota Samarinda. Data yang digabungkan meliputi Tanggal kejadian banjir, Temperatur minimum, Temperatur maksimum, Temperatur rata-rata, Kelembapan rata-rata, Curah hujan, Lamanya penyinaran matahari, Kecepatan angin maksimum, Arah angin saat kecepatan maksimum, Kecepatan angin rata-rata, dan Arah angin terbanyak.

• Data Transformation

Tahapan ini dilakukan untuk mengubah isi atau tipe data sebelum data dimasukkan ke dalam pemodelan. Atribut yang ditransformasi adalah Tanggal, yang merepresentasikan waktu terjadinya bencana banjir. Atribut Tanggal diinisialisasi sebagai indikator terjadinya banjir secara keseluruhan. Atribut Tanggal akan digunakan sebagai kelas target dalam pemodelan. Kelas target yang memiliki dua kriteria, "Ya" dan "Tidak," diubah menjadi data numerikal, yaitu "Ya" menjadi 1 dan "Tidak" menjadi 0, agar dapat dihitung menggunakan RapidMiner.

• Data Cleaning

Tahapan ini dilakukan untuk membersihkan data dari nilai yang kosong atau tidak lengkap. Dari total 1461

data banjir yang diperoleh, terdapat 447 data yang tidak lengkap, termasuk data yang kosong, tidak terukur, atau tidak memiliki data pengukuran. Setelah melalui proses data cleaning, dataset mencakup 1014 data yang lengkap. Sebelum dilakukan data cleaning, terdapat 75 kejadian banjir, tetapi setelah proses ini hanya terdapat 57 kejadian banjir dalam dataset.

Pemodelan (Modeling) dan Evaluation

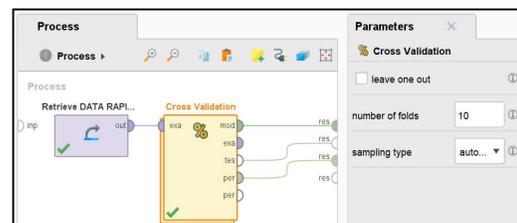
• Pemodelan Naive Bayes dengan RapidMiner

Pada tahap pemodelan menggunakan RapidMiner, algoritma Naive Bayes akan digunakan untuk mengukur tingkat akurasi. Pengujian pemodelan dilakukan dengan menggunakan metode K-Fold Cross Validation dengan nilai K=10. Tahap persiapan data training dimulai dengan mengimpor data menggunakan "Retrieving Data". Tampilan proses pemodelan pada RapidMiner dapat dilihat pada gambar 3.



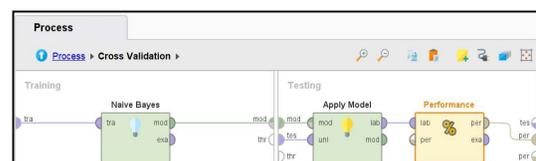
Gambar 3. Tahap Modelling Rapid Miner

Setelah persiapan data training selesai, langkah berikutnya adalah membagi dataset menggunakan "Cross Validation" dengan K=10. Proses ini dapat dilihat pada gambar 4.



Gambar 4. Pemilihan Validation

Selanjutnya, kita memilih model algoritma data mining yang akan digunakan, yaitu model Naive Bayes. Pengujian dilakukan menggunakan "Apply Model" dalam RapidMiner. Evaluasi kinerja model dilakukan dengan menggunakan opsi "Performance". Proses ini dapat dilihat pada gambar 5.



Gambar 5. Pilih Model Algoritma dan Performance

Hasil dari pemodelan menggunakan Naive Bayes akan menghasilkan sebuah matriks Confusion Matrix yang terdiri dari True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). Data yang telah diproses dimasukkan ke dalam

Confusion Matrix sehingga diperoleh hasil seperti yang terlihat pada Tabel 2 di bawah ini:

Tabel 2 Confusion Matrix Algoritma Naive Bayes

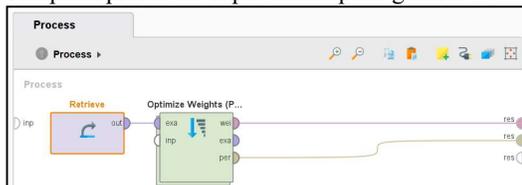
		Actual	
		1 (True)	0 (False)
Predicted	Positive	914	47
	Negative	43	10

Dari Tabel 2, diperoleh hasil akurasi sebesar 91.12%. Perhitungan manual untuk akurasi sebagai berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{914+10}{914+10+47+43} \times 100\% \\
 &= \frac{924}{1014} \times 100\% \\
 &= 91,1242 \text{ atau } 91,12\%
 \end{aligned}$$

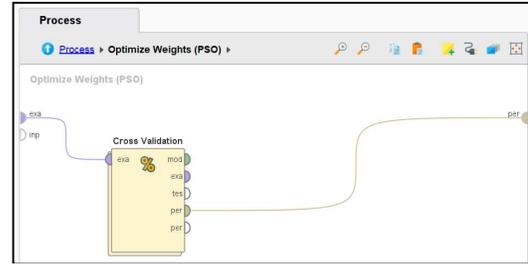
- Pemodelan Naive Bayes dengan PSO pada RapidMiner

Pemodelan Particle Swarm Optimization (PSO) digunakan untuk memberi bobot pada data dan meningkatkan hasil perhitungan. Nilai minimum dan maksimum dipertimbangkan karena bobot ini ditentukan secara acak. Setiap partikel dalam populasi akan memiliki bobot sendiri untuk atribut-atribut dalam dataset, dan algoritma Naive Bayes akan diaplikasikan untuk menghitung tingkat akurasi model yang terbentuk. Proses pemodelan dengan PSO dimulai dengan mengimpor data menggunakan "Retrieving Data" dan memilih Selection - Optimization - Optimize Weight (PSO) untuk mengoptimalkan bobot atribut dengan menerapkan algoritma Particle Swarm Optimization (PSO). Tampilan proses ini dapat dilihat pada gambar 6.



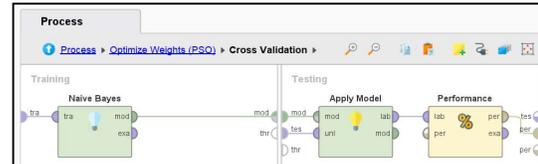
Gambar 6. Pemilihan Optimize Selection

Langkah selanjutnya adalah membagi dataset menggunakan teknik "Cross Validation" dengan nilai K=10. Tampilan pada RapidMiner untuk proses ini dapat dilihat pada gambar 7.



Gambar 7. Pemilihan Validation

Setelah itu, langkah berikutnya adalah memilih model algoritma data mining yang akan digunakan. Dalam penelitian ini, model yang dipilih adalah Naive Bayes. Untuk menguji model tersebut, digunakan operator "Apply Model" pada RapidMiner. Selanjutnya, untuk mengevaluasi kinerjanya, menggunakan opsi "Performance". Tampilan pada RapidMiner untuk proses ini dapat dilihat pada gambar 8.



Gambar 8. Proses Pemilihan Model Algoritma dan Evaluasi Kinerja

Hasil dari pemodelan akan menghasilkan matriks Confusion Matrix yang terdiri dari True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). Data yang telah diproses dimasukkan ke dalam Confusion Matrix sehingga diperoleh hasil pada Tabel 3 di bawah ini:

Tabel 3 Confusion Matrix Algoritma Naive Bayes dengan PSO pada RapidMiner

		Actual	
		1 (True)	0 (False)
Predicted	Positive	957	57
	Negative	0	0

Dari Tabel 3, diperoleh hasil dari Confusion Matrix yang menghasilkan akurasi sebesar 94.38%. Perhitungan manual untuk akurasi adalah sebagai berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+T}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{957+0}{957+0+57+0} \times 100\% \\
 &= \frac{957}{1014} \times 100\% \\
 &= 94,3786 \text{ atau } 94,38\%
 \end{aligned}$$

5. KESIMPULAN

Berdasarkan penelitian yang dilakukan, diperoleh kesimpulan sebagai berikut:

1. Hasil implementasi algoritma Naïve Bayes dalam klasifikasi data banjir Kota Samarinda mencapai akurasi 91.12%. Untuk meningkatkan hasil nilai akurasi dari algoritma tersebut, dilakukan penambahan optimasi dengan penggunaan Algoritma Naïve Bayes berbasis PSO. Hasil pengujian menunjukkan bahwa nilai akurasi mencapai 94.38%, yang lebih tinggi dibandingkan tanpa optimasi.
2. Atribut data yang mempengaruhi terjadinya banjir adalah arah angin saat kecepatan maksimum dan kecepatan angin rata-rata.

6. DAFTAR PUSTAKA

- Amrin, A., Pahlevi, O., & Satriadi, I. (2021). Optimasi Algoritma C4.5 dan Naïve Bayes Berbasis Particle Swarm Optimization Untuk Diagnosa Penyakit Peradangan Hati. *INSANtek*, 2(1), 10–14. <https://doi.org/10.31294/instk.v2i1.399>
- Arifin, T. (2020). Optimasi Decision Tree menggunakan Particle Swarm Optimization untuk klasifikasi sel Pap Smear. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(3), 572–579. <https://doi.org/10.35957/jatisi.v7i3.361>
- Arifin, T., & Ariesta, D. (2019). Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization. *Jurnal Tekno Insentif*, 13(1), 26–30. <https://doi.org/10.36787/jti.v13i1.97>
- Arisandi, R. R. R., Warsito, B., & Hakim, A. R. (2022). Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation. *Jurnal Gaussian*, 11(1), 130–139. <https://doi.org/10.14710/j.gauss.v11i1.33991>
- Feblian, D., & Daihani, D. U. (2017). Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X. *Jurnal Teknik Industri*, 6(1). <https://doi.org/10.25105/jti.v6i1.1526>
- Guntur, G., Putra, S., Swastika, W., Lucky, P., & Irawan, T. (2022). Perbandingan Particle Swarm Optimization dengan Genetic Algorithm dalam Feature Selection untuk Analisis Sentimen pada Permendikbudristek PPKS-. 8(3), 412–421.
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103–108. <https://doi.org/10.30871/jaic.v5i2.3200>
- Mandar, A. A. (2022). *Peqguruang: Conference Series*. 4.
- Manullang, N., Sembiring, R. W., Gunawan, I., Parlina, I., Informatika, T., & Informatika, T. (2021). Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma. 2(2), 1–5.
- Monowati, I. T., & Setyadi, R. (2023). Penerapan Algoritma Naïve Bayes Dalam Memprediksi Pengusulan Penghapusan Peralatan dan Mesin Kantor. *Journal of Information System ...*, 4(2), 483–491. <https://doi.org/10.47065/josh.v4i2.2674>
- Mustajab, R. (2023). BNPB: Indonesia Alami 3.522 Bencana Alam pada 2022. *DataIndonesia.Id*. <https://dataindonesia.id/varia/detail/bnpb-indonesia-alami-3522-bencana-alam-pada-2022>
- Naldy, E. T., & Andri, A. (2021). Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN. *Jurnal Nasional Ilmu Komputer*, 2(2), 89–101. <https://doi.org/10.47747/jurnalnik.v2i2.525>
- Putra, D. B., Suprayogi, A., & Sudarsono, B. (2019). Analisis Kerawanan Banjir pada Kawasan Terbangun Berdasarkan Klasifikasi Indeks EBBI (Enhanced Built-Up and Bareness Index) Menggunakan SIG (Studi Kasus di Kabupaten Demak). *Jurnal Geodesi Undip*, 8(1), 93–102.
- Rachman, R., & Handayani, R. N. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *Jurnal Informatika*, 8(2), 111–122. <https://doi.org/10.31294/ji.v8i2.10494>
- Rizki, A. M., & Nurlaili, A. L. (2021). Algoritme Particle Swarm Optimization (PSO) untuk Optimasi Perencanaan Produksi Agregat Multi-Site pada Industri Tekstil Rumahan. *Journal of Computer, Electronic, and Telecommunication*, 1(2), 1–9. <https://doi.org/10.52435/complete.v1i2.73>
- Sa'diyah, N., Supianto, A. A., & Dewi, C. (2020). Implementasi Algoritme Fuzzy C-Means dengan Particle Swarm Optimization (FCMPSO) untuk Pengelompokan Proses Berpikir Siswa dalam Proses Belajar. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK)*, 4(6), 1625–1632.
- santoso dian. (2019). *Jurnal Penyebab Banjir*. *Jurnal Geografi*, 16(Penanggulangan Bencana Banjir Berdasarkan Tingkat Kerentanan dengan Metode Ecodrainage Pada Ekosistem Karst di Dukuh Tungu, Desa Girimulyo, Kecamatan Panggang, Kabupaten Gunungkidul, DIY), 1–9. <https://doi.org/10.15294/jg.v16i1.17136>
- Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), 1. <https://doi.org/10.24076/citec.2019v6i1.178>
- Sibarani, A. J. P. (2020). Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat. *JATISI (Jurnal Teknik Informatika Dan Sistem*

- Informasi), 7(2), 262–276.
<https://doi.org/10.35957/jatisi.v7i2.195>
- Sugianto, N. A., Cholissodin, I., & Widodo, A. W. (2018). Klasifikasi Keminatan Menggunakan Algoritme Extreme Learning Machine dan Particle Swarm Optimization untuk Seleksi Fitur (Studi Kasus: Program Studi Teknik Informatika FISugianto, N. A., Cholissodin, I., & Widodo, A. W. (2018). Klasifikasi Keminatan Mengg. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya, 2(5), 1856–1865.
- Tasya, E. (2021). Sistem Informasi Monitoring Bencana Alam Dari Data Media Sosial Menggunakan Metode K-Nearest Neighbor Natural Disaster Monitoring Information System From Social Media Data Using K-Nearest Neighbor Method. 8(5), 6684–6693.
- trivusi. (2022). Metriks Evaluasi Sistem Menggunakan Confusion Matrix. Trivusi.Web.Id.
<https://www.trivusi.web.id/2022/04/evaluasi-sistem-dengan-confusion-matrix.html>
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. Jurnal Media Informatika Budidarma, 4(2), 437.
<https://doi.org/10.30865/mib.v4i2.2080>
- Wiratama, M. A., & Pradnya, W. M. (2022). Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes. Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI), 11(1), 1.
<https://doi.org/10.23887/janapati.v11i1.45sS28>



UMKT
UNIVERSITAS MUHAMMADIYAH
Kalimantan Timur

Kampus 1 : Jl. Ir. H. Juanda, No.15, Samarinda
Kampus 2 : Jl. Pelita, Pesona Mahakam, Samarinda
Telp. 0541-748511 Fax.0541-766832



SURAT KETERANGAN ARTIKEL PUBLIKASI

Assalamu'alaikum Warahmatullahi wabarakatuh

Saya yang bertanda tangan dibawah ini:

Nama	: Faldi, S.Kom, M.TI
NIDN	: 1121079101
Nama	: Trisha NurHalisha
NIM	: 1911102441099
Fakultas	: Sains Dan Teknologi
Progam Studi	: S1 Teknik Informatika

Manyatakan bahwa artikel ilmiah yang berjudul "Penerapan Particle Swarm Optimization (PSO) Untuk Meningkatkan Akurasi Algoritma Naive Bayes Pada Banjir Di Kota Samarinda" telah di submit pada jurnal ilmiah ilmu komputer pada tahun 2023. <https://e-journals.unmul.ac.id/index.php/IIM/index>

Demikian surat keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Warahmatullahi wabarakatuh

Samarinda, Rabu 02 Agustus 2023
Dosen Pembimbing

Faldi, S.Kom, M.TI
NIDN. 1121079101