



**PENERBIT  
YAYASAN MMI**



# METODE KNN DAN ID3 UNTUK KLASIFIKASI DATA

Disusun Oleh:

Mayang Alvi Perdana Nasution

Ibnu Rasyid Munthe

Deci Irmayani

Angga Putra Juledi



# METODE KNN DAN ID3 UNTUK KLASIFIKASI DATA



## **UU No 28 tahun 2014 tentang Hak Cipta**

### **Fungsi dan sifat hak cipta Pasal 4**

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

### **Pembatasan Pelindungan Pasal 26**

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

- i. penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
- ii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
- iii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
- iv. penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

### **Sanksi Pelanggaran Pasal 113**

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

# METODE KNN DAN ID3 UNTUK KLASIFIKASI DATA

Mayang Alvi Perdana Nasution;

Ibnu Rasyid Munthe;

Deci Irmayani;

Angga Putra Juledi;



**PENERBIT  
YAYASAN MMI**

# METODE KNN DAN ID3 UNTUK KLASIFIKASI DATA

**Mayang Alvi Perdana Nasution;  
Ibnu Rasyid Munthe;  
Deci Irmayani;  
Angga Putra Juledi;**

Desain Cover :  
**Musthafa Haris Munandar**

Sumber :  
<https://munandar.yayasanmmi.com/metode-knn-dan-id3-untuk-klasifikasi-data>

Tata Letak :  
**Deci Irmayani, S.Kom., M.Kom**

Proofreader :  
**Ibnu Rasyid Munthe, S.T., M.Kom**

Ukuran :  
**Jml hal judul: 1, Jml hal isi naskah: 72, Uk: 15x23 cm**

ISBN :  
978-634-04-2085-2

Cetakan Pertama :  
**Agustus 2025**

---

Hak Cipta 2025, Pada Penulis  
Isi diluar tanggung jawab percetakan

---

**Copyright © 2025 by Yayasan MMI**  
All Right Reserved

Hak cipta dilindungi undang-undang  
Dilarang keras menerjemahkan, memfotokopi, atau  
memperbanyak sebagian atau seluruh isi buku ini  
tanpa izin tertulis dari Penerbit.

**PENERBIT YAYASAN MUNANDAR MEMBANGUN INDONESIA**  
Jl. Pasar Banjar Dusun XVI Desa Simpang Empat, Asahan, Sumatera Utara 21271  
Telp/Wa : 082363080181  
<https://munandar.yayasanmmi.com/>  
E-mail: [mmipublisher@yayasanmmi.com](mailto:mmipublisher@yayasanmmi.com)

## **PRAKATA**

Segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga penulisan karya ini yang berjudul “Metode K-Nearest Neighbor (KNN) dan ID3 untuk Klasifikasi Data” dapat diselesaikan dengan baik.

Klasifikasi merupakan salah satu teknik utama dalam bidang data mining dan machine learning, yang memiliki peranan penting dalam membantu pengambilan keputusan berbasis data. Di era digital saat ini, di mana data semakin melimpah dan kompleks, kebutuhan akan metode klasifikasi yang efektif dan efisien menjadi sangat mendesak. Oleh karena itu, dalam karya ini penulis membahas dua algoritma klasifikasi yang populer dan banyak digunakan, yaitu K-Nearest Neighbor (KNN) dan ID3.

Penulisan ini bertujuan untuk memberikan pemahaman lebih mendalam mengenai prinsip kerja kedua algoritma, cara implementasinya, serta contoh penerapannya dalam berbagai domain seperti kesehatan, pendidikan, dan keuangan. Penulis berharap, karya ini dapat menjadi referensi bermanfaat bagi mahasiswa, peneliti, maupun praktisi yang ingin mengembangkan sistem klasifikasi berbasis KNN dan ID3.

Akhir kata, penulis menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan dan masukan selama proses penyusunan karya ini. Kritik dan saran yang membangun sangat diharapkan demi kesempurnaan karya ini di masa mendatang.

Labuhanbatu, Agustus 2025

Penulis

# DAFTAR ISI

|                                                |     |
|------------------------------------------------|-----|
| PRAKATA.....                                   | vi  |
| DAFTAR ISI .....                               | vii |
| BAB I .....                                    | 1   |
| PENGANTAR MACHINE LEARNING.....                | 1   |
| SEJARAH MACHINE LEARNING .....                 | 3   |
| BAB II.....                                    | 3   |
| KONSEP DASAR MACHINE LEARNING.....             | 9   |
| Pengertian Machine Learning .....              | 9   |
| Knowledge Discovery in Database (KDD).....     | 10  |
| BAB III .....                                  | 11  |
| PENGENALAN METODE K-NN .....                   | 12  |
| Pengertian K-Nearest Neighbor (kNN).....       | 13  |
| Klasifikasi .....                              | 14  |
| Preprocessing Data.....                        | 15  |
| Perhitungan Data .....                         | 15  |
| Uji Performa.....                              | 16  |
| BAB IV .....                                   | 18  |
| PENGENALAN METODE ID3 .....                    | 18  |
| Pengertian Iterative Dichotomiser 3 (ID3)..... | 19  |
| Preprocessing Data.....                        | 23  |

|                                                                                                   |    |
|---------------------------------------------------------------------------------------------------|----|
| Uji Performa.....                                                                                 | 23 |
| BAB V.....                                                                                        | 26 |
| ORANGE SEBAGAI ALAT BANTU PENGAPLIKASIAN<br>.....                                                 | 26 |
| Pengertian Orange.....                                                                            | 27 |
| Implementasi Orange pada Metode kNN dan ID3.....                                                  | 34 |
| Kelebihan dan Kekurangan Orange: Platform Visual untuk Data<br>Science dan Machine Learning ..... | 37 |
| BIOGRAFI PENULIS .....                                                                            | 61 |
| SINOPSIS .....                                                                                    | 64 |

# **BAB I**

## **PENGANTAR MACHINE LEARNING**

Machine Learning (ML) atau pembelajaran mesin merupakan salah satu cabang dari kecerdasan buatan yang memungkinkan komputer mempelajari pola dari data secara mandiri dan membuat prediksi maupun keputusan tanpa perlu diprogram dengan aturan yang eksplisit. Berbeda dari pemrograman konvensional yang sepenuhnya mengandalkan aturan baku dari manusia, ML memanfaatkan data sebagai bahan bakunya untuk membaca pola, mengenali hubungan antar variabel, dan membuat model prediktif yang dapat digunakan untuk berbagai kebutuhan. Berbagai metode dalam ML dapat diaplikasikan sesuai dengan kebutuhan dan pola data yang dianalisis, mulai dari metode klasifikasi yang digunakan untuk mengelompokkan data ke dalam kategori tertentu, metode regresi yang digunakan untuk memprediksi nilai numerik, hingga metode clustering yang digunakan untuk mengelompokkan data berdasarkan pola kesamaan. Keberadaan ML ini telah membawa perubahan signifikan dalam berbagai bidang, memungkinkan teknologi dapat digunakan untuk memecahkan permasalahan yang kompleks dan sulit dijawab dengan metode biasa, termasuk dalam memahami pola perilaku manusia dan pola kebutuhan teknologi di tingkat daerah.

Konsep dasar dari penerapan ML terdiri dari tiga tahap yang saling terkait, yaitu pelatihan (training), pengujian (testing), dan penerapan (deployment). Pada tahap pelatihan, komputer diajarkan pola dari data dengan atribut dan label yang sudah diketahui sehingga dapat digunakan sebagai acuan untuk membuat prediksi dari data yang belum pernah dianalisis sebelumnya. Setelah itu, model yang dihasilkan dari pelatihan tersebut diuji dengan data uji yang belum pernah digunakan untuk

pelatihan guna mengukur tingkat akurasi, serta mengevaluasi sejauh mana model dapat digunakan dengan baik. Model dengan tingkat akurasi tinggi dan nilai evaluasi yang memadai kemudian diterapkan dalam kebutuhan nyata, seperti untuk membuat klasifikasi pola data yang belum diketahui atau untuk membuat prediksi kebutuhan manusia terkait teknologi tertentu. Siklus ini menjadikan ML sebagai metode adaptif yang dapat digunakan berkesinambungan untuk memahami pola data yang terus berubah dari waktu ke waktu.

Saat ini, penerapan ML dapat ditemukan di berbagai bidang yang membutuhkan analisis pola dari data yang kompleks, mulai dari teknologi finansial, kesehatan, pertanian, edukasi, komunikasi, hingga pemerintahan. Salah satu contoh penerapan yang signifikan adalah dalam pengklasifikasian pola data terkait pola perilaku manusia, termasuk pola ketertarikan atau kebutuhan masyarakat dalam memanfaatkan teknologi tertentu. Model ML memungkinkan pengambil kebijakan dan tenaga ahli memahami pola dari data yang sulit terlihat dengan pengamatan biasa, membuat prediksi kebutuhan di masa depan, dan bahkan membantu menentukan langkah strategis untuk pengembangan teknologi yang sesuai dengan kebutuhan masyarakat. Dengan kemampuan adaptasi dan skala analisis yang luas, ML dapat digunakan sebagai landasan untuk berbagai pengambilan keputusan yang berdampak positif bagi pembangunan teknologi dan masyarakat, khususnya dalam konteks daerah dengan pola kebutuhan yang belum sepenuhnya terpetakan.

Melalui metode klasifikasi yang digunakan, yaitu K Nearest Neighbor (KNN) dan ID3, pola data dari data yang dikumpulkan dapat dianalisis dengan lebih mendalam guna menemukan hubungan yang signifikan antara atribut tertentu. Dengan memanfaatkan metode ini, para peneliti dapat membuat rekomendasi yang lebih akurat dan relevan bagi kebutuhan

masyarakat terkait teknologi komunikasi, mulai dari kebutuhan edukasi teknologi, perluasan jaringan komunikasi, hingga pola pola penerimaan teknologi dari tingkat individu maupun kelompok masyarakat. Dengan demikian, penerapan ML dalam hal ini tidak hanya memberikan gambaran pola kebutuhan teknologi masyarakat, tetapi juga dapat digunakan sebagai landasan bagi pengambil kebijakan, tenaga ahli, maupun pelaku usaha dalam merancang dan mengimplementasikan teknologi yang berdampak positif bagi pembangunan daerah.

## **SEJARAH MACHINE LEARNIING**

Machine Learning (ML), atau pembelajaran mesin, adalah cabang dari kecerdasan buatan (AI) yang memungkinkan komputer belajar dari data dan membuat keputusan tanpa diprogram secara eksplisit. Perjalanan panjang ML telah melalui berbagai tahap perkembangan, mulai dari pemikiran teoritis hingga penerapan praktis dalam berbagai bidang kehidupan. Berikut adalah uraian sejarahnya secara kronologis:

### **1. Awal Mula Konsep (1940–1950-an)**

Konsep awal dari machine learning muncul sebagai bagian dari eksplorasi terhadap kecerdasan buatan. Tokoh utama dalam masa ini adalah Alan Turing, seorang ilmuwan Inggris yang dikenal dengan "Turing Machine", yaitu model teoritis komputer universal. Pada tahun 1950, Turing memperkenalkan "Turing Test", sebuah uji untuk menentukan apakah mesin dapat menunjukkan kecerdasan setara manusia.

Pada tahun 1952, Arthur Samuel menciptakan salah satu program pembelajaran mesin pertama. Ia merancang sebuah program bermain dam (checkers) yang dapat belajar dari permainannya sendiri. Program ini tidak hanya mengikuti aturan tetap, tetapi

juga meningkatkan strateginya berdasarkan pengalaman. Pada 1959, Samuel secara eksplisit menyebut istilah "machine learning" untuk menggambarkan pendekatan ini, menjadikannya tokoh penting dalam sejarah ML.

## 2. Perkembangan Awal dan Teori Dasar (1960–1970-an)

Pada tahun 1960-an, para peneliti mulai merancang algoritma awal untuk pembelajaran simbolik dan statistik. Sistem seperti ELIZA (1966) karya Joseph Weizenbaum menunjukkan bagaimana komputer bisa “meniru” percakapan manusia, walaupun tanpa pemahaman sebenarnya.

Di sisi lain, para ilmuwan mulai mengembangkan algoritma klasifikasi dan regresi, serta mulai membahas teori pembelajaran, seperti teori PAC (Probably Approximately Correct) dan teori keputusan Bayes. Metode statistik menjadi dasar dalam pendekatan machine learning saat ini.

Namun, keterbatasan komputasi dan data membuat ML masih dalam tahap eksperimental. Banyak eksperimen mengalami kegagalan karena model terlalu sederhana untuk menangani kompleksitas dunia nyata.

## 3. Munculnya Neural Networks dan Kekecewaan AI (1980-an)

Pada 1980-an, terjadi lonjakan minat terhadap jaringan saraf tiruan (artificial neural networks/ANN). Metode ini terinspirasi oleh struktur dan cara kerja otak manusia. Salah satu terobosan besar terjadi pada 1986, saat Rumelhart, Hinton, dan Williams mempopulerkan algoritma backpropagation, yang memungkinkan jaringan saraf belajar dari kesalahan dengan cara mengatur bobot secara iteratif.

Namun, masih ada keterbatasan serius: jaringan saraf saat itu hanya efektif untuk masalah kecil dan tidak dapat menangani data kompleks secara efisien. Akibatnya, pada akhir 1980-an dan awal 1990-an, terjadi masa yang disebut sebagai “AI Winter”, yaitu penurunan drastis minat dan pendanaan dalam penelitian AI dan ML akibat ekspektasi yang tidak terpenuhi.

#### 4. Kebangkitan Melalui Metode Statistik (1990-an)

Pada 1990-an, pendekatan statistik mulai mendominasi ML. Para peneliti mengembangkan berbagai metode baru yang lebih terukur dan kuat secara matematis, seperti:

- a. Support Vector Machines (SVM)
- b. Decision Trees
- c. Naïve Bayes
- d. Ensemble Methods seperti Random Forest dan Boosting

Era ini juga ditandai dengan penggunaan data training yang lebih sistematis dan evaluasi model menggunakan cross-validation serta metrik kinerja seperti akurasi, presisi, dan recall.

ML mulai digunakan dalam aplikasi dunia nyata seperti pengenalan tulisan tangan (handwriting recognition), diagnosis medis, dan sistem rekomendasi awal (rekomendasi film atau buku).

#### 5. Revolusi Big Data dan Awal Era Modern (2000–2010)

Memasuki awal 2000-an, ledakan data akibat internet, media sosial, dan sensor digital memicu revolusi big data. Bersamaan dengan peningkatan kapasitas penyimpanan dan pemrosesan (cloud computing, GPU), machine learning memasuki tahap kematangan baru.

Algoritma pembelajaran berbasis unsupervised learning dan semi-supervised learning mulai banyak digunakan, seperti:

- a. K-means clustering
- b. Principal Component Analysis (PCA)
- c. Hidden Markov Models (HMM) untuk pemrosesan bahasa alami (NLP)

Perusahaan teknologi besar seperti Google, Amazon, dan Facebook mulai menerapkan ML secara masif untuk meningkatkan pengalaman pengguna melalui personalisasi, prediksi perilaku, dan pengolahan bahasa alami.

## 6. Deep Learning dan Perkembangan Pesat (2010–sekarang)

Periode ini menjadi tonggak besar dalam sejarah ML. Deep Learning, sebagai bagian dari neural network yang terdiri dari banyak lapisan tersembunyi (deep neural networks), menjadi sangat populer karena kemampuannya dalam mengenali pola kompleks secara otomatis.

Tiga faktor utama yang mendukung revolusi deep learning:

1. Data: Tersedia dalam jumlah besar (big data)
2. Komputasi: Adanya GPU yang mendukung komputasi paralel
3. Algoritma: Inovasi model seperti CNN, RNN, LSTM, dan Transformer

Beberapa momen penting:

2012: Tim dari University of Toronto (Geoffrey Hinton dkk.) memenangkan kompetisi ImageNet dengan model deep convolutional neural network bernama AlexNet, mengalahkan pendekatan lain dengan margin besar.

2014–2018: Munculnya berbagai arsitektur seperti VGGNet, ResNet, dan GAN (Generative Adversarial Network) mendorong

aplikasi ML di bidang pengenalan gambar, suara, dan pembuatan konten.

2017: Transformer diperkenalkan oleh Vaswani et al. dalam paper “Attention is All You Need”, menjadi dasar bagi model NLP modern seperti BERT dan GPT.

2020–2023: Model bahasa besar (LLM) seperti GPT-3, ChatGPT, dan Bard menjadi populer, membuka jalan bagi aplikasi berbasis AI dalam komunikasi, penulisan, dan pemrosesan informasi.

## 7. Aplikasi Kontemporer Machine Learning

Saat ini, machine learning telah merambah ke berbagai bidang kehidupan:

- a. Kesehatan: Diagnosa otomatis, prediksi penyakit, analisis citra medis
- b. Transportasi: Kendaraan otonom, optimisasi rute
- c. Keuangan: Deteksi penipuan, prediksi pasar saham
- d. Pertanian: Prediksi panen, pemantauan hama melalui citra satelit
- e. Industri Kreatif: Pembuatan musik dan seni menggunakan AI generatif
- f. Edukasi: Pembelajaran adaptif dan sistem rekomendasi konten

## 8. Tantangan dan Masa Depan Machine Learning

Meski banyak kemajuan telah dicapai, machine learning masih menghadapi beberapa tantangan besar:

- a. Etika dan privasi: Penggunaan data pribadi dan bias algoritma

- b. Transparansi: Model yang sangat kompleks sulit dijelaskan (black box)
- c. Ketergantungan data besar: Model modern butuh jutaan data pelatihan
- d. Energi dan keberlanjutan: Pelatihan model besar memakan energi sangat tinggi

## **BAB II**

### **KONSEP DASAR MACHINE LEARNING**

#### **Pengertian Machine Learning**

Machine Learning (ML) atau pembelajaran mesin merupakan cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan sistem komputer mempelajari pola dari data dan membuat prediksi maupun keputusan berdasarkan pola yang ditemukan, tanpa memerlukan pemrograman aturan yang eksplisit. Berbeda dari metode konvensional yang sepenuhnya mengandalkan kode atau logika dari manusia, ML memberi komputer kemampuan untuk beradaptasi dengan pola yang belum pernah terlihat sebelumnya. Hal ini menjadikan ML sebagai teknologi yang sangat relevan dan strategis bagi berbagai kebutuhan analisis data yang terus tumbuh dari waktu ke waktu.

Saat ini, penerapan ML dapat ditemukan hampir di berbagai bidang, mulai dari teknologi finansial, kesehatan, komunikasi, pertanian, pemerintahan, hingga edukasi dan bisnis. Berbagai metode dan algoritma digunakan sesuai dengan kebutuhan dan pola dari data yang dianalisis. Ada tiga metode utama dalam ML, yaitu supervised learning, yang memanfaatkan data dengan label untuk membuat model prediktif, unsupervised learning, yang digunakan untuk menemukan pola atau struktur dari data yang belum diberi label, serta reinforcement learning, yang memungkinkan sistem belajar dari pengalaman dan mendapatkan umpan balik dari lingkungannya.

Dalam konteks penerapan teknologi, ML berfungsi sebagai pondasi bagi banyak sistem cerdas yang digunakan untuk membuat prediksi atau pengambilan keputusan, mulai dari rekomendasi produk di situs e-commerce hingga diagnosa awal

suatu penyakit. Model ML dapat digunakan untuk mengenali pola dari data yang jumlah dan kompleksitasnya terus berkembang, memungkinkan pengambil keputusan membuat strategi berdasarkan bukti dan pola yang terlihat dari data, bukan asumsi semata. Keberadaan ML juga memberikan nilai tambah dalam pengolahan data, di mana pola-pola yang sulit terlihat dapat diekstrak dan digunakan untuk kebutuhan pengambilan keputusan yang lebih efisien.

Selain itu, perkembangan ML juga memberikan peluang bagi daerah-daerah dengan kebutuhan teknologi spesifik, termasuk daerah pedesaan atau daerah dengan tingkat adopsi teknologi yang belum merata. Dengan penerapan metode ML yang tepat, pola kebutuhan dan tingkat penerimaan teknologi dapat diidentifikasi dan dianalisis dengan lebih mudah, memungkinkan para pengambil kebijakan dan tenaga ahli membuat langkah strategis yang lebih relevan dengan kebutuhan daerah tersebut. Dengan berbagai metode dan algoritma yang tersedia, ML menjelma sebagai teknologi vital yang memungkinkan data digunakan tidak hanya sebagai angka atau catatan biasa, tetapi juga sebagai landasan untuk membuat langkah perubahan yang signifikan.

### **Knowledge Discovery in Database (KDD)**

Knowledge Discovery in Database (KDD) merupakan proses sistematis yang digunakan untuk menemukan pola, pengetahuan, atau informasi berharga dari kumpulan data yang jumlah dan kompleksitasnya sangat besar. KDD bukanlah proses tunggal, tetapi terdiri dari serangkaian tahap yang saling terkait, mulai dari pemilihan data, pembersihan data, transformasi data, penerapan metode data mining, hingga interpretasi dan evaluasi pola yang ditemukan. Tahapan ini memungkinkan para peneliti dan praktisi memahami pola dari data yang belum pernah terlihat sebelumnya, menjadikan data sebagai sumber pengetahuan yang

dapat digunakan untuk kebutuhan analisis dan pengambilan keputusan.

Tahap pertama dari KDD ialah pemilihan data, yaitu menentukan data apa saja yang relevan dengan kebutuhan penelitian atau kebutuhan bisnis. Setelah itu, data yang terkumpul masuk ke tahap pembersihan, di mana data yang tidak lengkap, duplikasi, atau tidak relevan dihapus guna memastikan kualitas data yang digunakan. Selanjutnya, data yang sudah bersih dan relevan ini diubah atau ditransformasi ke dalam bentuk yang sesuai untuk kebutuhan analisis lebih lanjut, agar dapat digunakan secara efisien oleh metode data mining yang digunakan.

Pada tahap penerapan metode data mining, berbagai algoritma digunakan untuk mengeksplorasi pola, tren, atau pengetahuan dari data yang telah disiapkan. Beberapa metode yang biasa digunakan dalam tahap ini termasuk klasifikasi, clustering, regresi, dan asosiasi. Melalui penerapan metode ini, pola atau hubungan antar atribut dapat ditemukan, memungkinkan para peneliti atau pengambil keputusan mendapatkan gambaran yang lebih jelas mengenai pola dari data yang dianalisis.

Tahap terakhir dari proses KDD ialah evaluasi pola dan interpretasi pola yang ditemukan. Pola atau pengetahuan yang dihasilkan dari metode data mining perlu dievaluasi untuk memastikan nilai, relevansi, dan kegunaannya bagi kebutuhan yang ditentukan. Dengan KDD, data yang awalnya hanya berupa angka atau catatan biasa dapat diubah menjadi pengetahuan yang dapat digunakan untuk membuat strategi bisnis, menentukan kebutuhan teknologi, atau bahkan membuat kebijakan publik yang lebih berdasar dan terukur. Dengan kata lain, KDD memungkinkan pemanfaatan data yang lebih maksimal untuk menjawab kebutuhan dan tantangan yang terus berubah dari waktu ke waktu.

### **BAB III**

## **PENGENALAN METODE K-NN**

K-Nearest Neighbor (KNN) merupakan salah satu metode klasifikasi dalam pembelajaran mesin yang bekerja berdasarkan pola kedekatan atau kesamaan nilai atribut antar data. Proses kerja KNN diawali dengan menghitung nilai jarak antar data, sehingga pola pola dari data yang belum diketahui dapat diklasifikasikan berdasarkan pola dari data yang sudah diketahui (data training). Dengan penerapan metode ini, pola pola kebutuhan dan preferensi masyarakat terkait teknologi komunikasi dapat diidentifikasi dengan tingkat akurasi yang tinggi, memungkinkan peneliti membuat gambaran yang jelas mengenai pola kebutuhan teknologi komunikasi di daerah tersebut.

Metode KNN dapat digunakan untuk memetakan pola pola kebutuhan teknologi komunikasi dari berbagai atribut yang digunakan, mulai dari desain antarmuka, kualitas handphone, kualitas jaringan, performa mesin, hingga harga jual. Dengan memanfaatkan nilai nilai atribut ini, pola pola kebutuhan dapat diklasifikasi berdasarkan nilai nilai atribut dari data yang telah diketahui. Kelebihan dari metode KNN ialah kemampuannya untuk mengklasifikasi pola pola dari data yang belum pernah terlihat sebelumnya berdasarkan pola pola data yang sudah ada, membuat metode ini sangat sesuai untuk kebutuhan penelitian terkait pola kebutuhan teknologi komunikasi masyarakat.

Pada penerapan metode KNN, data yang digunakan terdiri dari data training dan data testing dengan atribut atribut terkait kebutuhan dan pola penerimaan teknologi komunikasi dari masyarakat. Model KNN digunakan untuk mengklasifikasi data testing dengan menghitung nilai kedekatan dari data training, kemudian menentukan pola pola kebutuhan dari data yang belum diketahui. Hasil dari penerapan metode ini memungkinkan

peneliti untuk memahami pola pola kebutuhan dan penerimaan teknologi komunikasi dari berbagai lapisan masyarakat, mulai dari pola pola kebutuhan teknologi komunikasi dari tingkat usia, tingkat pendidikan, hingga pola pola kebutuhan teknologi komunikasi dari berbagai tingkat daya beli.

Hasil dari penerapan metode KNN memberikan gambaran pola pola kebutuhan teknologi komunikasi yang dapat digunakan untuk membuat berbagai langkah strategis bagi pihak terkait. Pola pola kebutuhan ini dapat digunakan sebagai landasan bagi pemerintahan daerah maupun pihak pihak terkait untuk membuat langkah edukasi teknologi komunikasi yang dapat menjembatani kebutuhan teknologi masyarakat. Dengan pola pola kebutuhan yang lebih jelas, teknologi komunikasi dapat dikembangkan dan diterapkan sesuai dengan kebutuhan dan kemampuan masyarakat, guna memperluas jangkauan teknologi komunikasi dan meningkatkan tingkat adopsinya dari waktu ke waktu.

### **Pengertian K-Nearest Neighbor (kNN)**

K-Nearest Neighbor (KNN) merupakan salah satu algoritma klasifikasi yang termasuk dalam kategori pembelajaran terawasi (supervised learning). KNN bekerja dengan mengukur nilai kedekatan atau kesamaan nilai atribut dari data yang belum diketahui dengan nilai nilai atribut dari data yang telah diketahui sebelumnya (data training). Model ini tidak memerlukan pelatihan atau pembentukan pola yang kompleks, tetapi bekerja dengan menyimpan data training dan menghitung nilai kesamaan dari data yang belum diketahui dengan data yang ada guna menentukan pola pola klasifikasi dari data tersebut.

Kelebihan dari metode KNN ialah kesederhanaannya dalam penerapan dan kemampuannya untuk bekerja dengan berbagai jenis data dan pola pola kebutuhan. Dengan metode ini, pola pola kebutuhan teknologi komunikasi dapat diidentifikasi dan

digunakan sebagai landasan bagi berbagai pihak terkait guna membuat langkah langkah strategis untuk pemerataan teknologi komunikasi di daerah tersebut.

## **Klasifikasi**

Klasifikasi merupakan proses dalam metode data mining yang digunakan untuk mengelompokkan data berdasarkan pola pola atribut tertentu, dengan tujuan membuat pola pola dari data yang belum diketahui berdasarkan pola pola dari data yang telah diketahui.

Proses klasifikasi ini diawali dengan pengumpulan data dari berbagai lapisan masyarakat, diikuti dengan tahap pembersihan data dan transformasi data guna memastikan pola pola kebutuhan teknologi komunikasi dapat digunakan dengan baik. Data yang sudah dibersihkan kemudian digunakan sebagai data training untuk membuat pola pola klasifikasi dengan metode KNN, sehingga pola pola kebutuhan teknologi komunikasi dari data yang belum diketahui dapat diklasifikasi dengan nilai nilai atribut dari data yang telah diketahui. Pola pola klasifikasi ini dapat digunakan sebagai landasan bagi pihak terkait untuk membuat langkah langkah strategis guna pemerataan teknologi komunikasi dari tingkat daerah hingga tingkat lapisan masyarakat.

Hasil dari klasifikasi pola pola kebutuhan teknologi komunikasi dapat digunakan sebagai landasan bagi berbagai pihak terkait guna membuat langkah langkah edukasi teknologi komunikasi dan membuat pola pola teknologi komunikasi yang sesuai dengan kebutuhan dan pola pola penerimaan teknologi dari masyarakat. Dengan pola pola klasifikasi ini, teknologi komunikasi dapat digunakan dengan lebih efisien, relevan, dan berdampak positif bagi tingkat adopsi teknologi komunikasi dari waktu ke waktu. Klasifikasi pola pola kebutuhan teknologi komunikasi juga dapat digunakan sebagai landasan bagi

penelitian selanjutnya guna membuat pola pola kebutuhan teknologi komunikasi yang lebih sesuai dengan kebutuhan daerah daerah lainnya.

### **Preprocessing Data**

Tahap preprocessing merupakan tahapan yang dilakukan untuk membersihkan data. Hal ini dilakukan agar data yang nantinya akan diolah dan dihitung sudah merupakan data yang memang benar benar layak untuk digunakan pada proses perhitungan. Proses ini dilakukan agar tidak ada kesalahan pada proses perhitungan data.

### **Perhitungan Data**

Pada perhitungan matematika untuk metode KNN dapat dilakukan dengan menggunakan metode sebagai berikut.

$$\text{Rumus Metode KNN, } d(x, y) = \sqrt{\sum_{i=1}^n (x_{training}^i - y_{testing}^i)^2}$$

Keterangan:

$d(x, y)$  : Jarak

$x_{training}^i$  : Data Training

$y_{testing}^i$  : Data Testing

$i$  : Variabel Data

$n$  : Dimensi Data

pada perhitungan metode KNN yaitu menggunakan rumus ecuilidean distance. Rumus tersebut merupakan rumus jarak (K). Untuk nilai K pada ini yaitu 5. untuk proses perhitungannya yaitu sebagai berikut.

## Uji Performa

Tahap uji performa dari metode ID3 digunakan untuk mengukur pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pola pola kebutuhan teknologi komunikasi dari metode ID3 ini digunakan untuk membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Hasil dari uji performa metode ID3 dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pola pola kebutuhan teknologi komunikasi dari metode ID3 dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Pada akhirnya, pola pola kebutuhan teknologi komunikasi dari metode ID3 dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pola pola kebutuhan teknologi komunikasi dari metode ID3 memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari

daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

## **BAB IV**

### **PENGENALAN METODE ID3**

Metode ID3 (Iterative Dichotomiser 3) merupakan salah satu algoritma klasifikasi yang digunakan untuk membentuk pola pola kebutuhan teknologi komunikasi dari berbagai atribut yang terkait dengan pola kebutuhan dan pola penerimaan teknologi komunikasi. Metode ini memungkinkan peneliti membuat pola pola kebutuhan teknologi komunikasi yang dapat digunakan sebagai landasan bagi berbagai pihak terkait guna membuat langkah langkah strategis dalam pengembangan teknologi komunikasi.

Metode ID3 bekerja dengan memanfaatkan nilai nilai atribut dari data untuk membentuk pola pola kebutuhan teknologi komunikasi yang dapat digunakan sebagai pola pola klasifikasi dari data yang belum pernah digunakan sebelumnya. Pola pola kebutuhan teknologi komunikasi dari metode ini digunakan untuk membuat pola pola kebutuhan teknologi komunikasi dari data yang belum pernah digunakan guna membuat pola pola kebutuhan teknologi komunikasi dari berbagai lapisan masyarakat. Dengan pola pola kebutuhan teknologi komunikasi ini, pola pola kebutuhan dari berbagai atribut terkait teknologi komunikasi dapat digunakan untuk membuat pola pola kebutuhan teknologi komunikasi dari berbagai daerah lainnya.

Pada penerapan metode ID3 ini, pola pola kebutuhan teknologi komunikasi digunakan sebagai landasan bagi pihak terkait guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah dengan pola pola kebutuhan teknologi komunikasi yang belum pernah digunakan sebelumnya. Pola pola kebutuhan teknologi komunikasi dari metode ID3 dapat digunakan untuk membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah dengan pola pola kebutuhan

teknologi komunikasi yang belum pernah digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan pola pola kebutuhan teknologi komunikasi dari metode ID3 ini, pola pola kebutuhan teknologi komunikasi dapat digunakan sebagai landasan bagi berbagai pihak terkait guna membuat pola pola kebutuhan teknologi komunikasi dari berbagai daerah dengan pola pola kebutuhan teknologi komunikasi yang belum pernah digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Hasil dari penerapan metode ID3 ini dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah dengan pola pola kebutuhan teknologi komunikasi dari berbagai daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pola pola kebutuhan teknologi komunikasi dari metode ID3 ini juga dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah dengan pola pola kebutuhan teknologi komunikasi yang belum pernah digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan pola pola kebutuhan teknologi komunikasi ini, pola pola kebutuhan teknologi komunikasi dapat digunakan sebagai landasan bagi berbagai pihak terkait guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

### **Pengertian Iterative Dichotomiser 3 (ID3)**

Iterative Dichotomiser 3 (ID3) merupakan salah satu metode klasifikasi dalam pembelajaran mesin yang digunakan untuk membuat pola pola kebutuhan teknologi komunikasi dari berbagai atribut data yang digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah tertentu. Metode ID3 memungkinkan pola pola kebutuhan teknologi komunikasi dari

berbagai daerah dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi yang belum pernah digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Metode ID3 bekerja dengan menghitung nilai entropi dari atribut atribut yang digunakan untuk membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah tertentu guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Nilai entropi digunakan sebagai ukuran untuk membuat pola pola kebutuhan teknologi komunikasi dari atribut atribut terkait teknologi komunikasi, guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya yang belum pernah digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Hasil dari penerapan metode ID3 memungkinkan pola pola kebutuhan teknologi komunikasi dari berbagai daerah digunakan untuk membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi yang belum pernah digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan metode ID3 ini, pola pola kebutuhan teknologi komunikasi dari berbagai daerah dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi yang belum pernah digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Metode ID3 (Iterative Dichotomiser 3) merupakan salah satu algoritma klasifikasi yang digunakan untuk membentuk pola pola kebutuhan teknologi komunikasi dari berbagai daerah

dengan memanfaatkan struktur data yang terdiri dari atribut-atribut spesifik terkait kebutuhan dan pola penerimaan teknologi. Metode ini bekerja dengan menghitung nilai entropi dan nilai information gain dari masing-masing atribut untuk menentukan atribut mana yang paling signifikan dalam membentuk pola pola kebutuhan teknologi komunikasi dari daerah yang diteliti. Dengan langkah ini, metode ID3 memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat diklasifikasi dengan tingkat akurasi dan relevansi yang tinggi, sehingga pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Pada penerapan metode ID3, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Proses klasifikasi diawali dengan menghitung nilai entropi dari masing-masing atribut kebutuhan teknologi komunikasi guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya, selanjutnya nilai information gain digunakan untuk menentukan atribut yang paling signifikan dalam membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan metode ini, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan untuk membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya yang dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna

membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Hasil dari klasifikasi pola pola kebutuhan teknologi komunikasi dengan metode ID3 ini memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan pola pola kebutuhan teknologi komunikasi dari metode ID3 ini, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya yang dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan penerapan metode ID3 ini, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya yang dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Tahapan seleksi data merupakan tahapan yang dilakukan untuk memilih data yang akan digunakan. Data dikumpulkan

menyesuaikan kebutuhan dari penelitian. Untuk data pada perhitungan metode ID3, hanya membutuhkan 1 data set, untuk data nya yaitu data testing. Hal ini karena pada perhitungan dengan menggunakan metode ID3, nantinya hasil yang akan diperoleh yaitu pohon keputusan.

### **Preprocessing Data**

Tahapan preprocessing data merupakan tahapan yang dilakukan untuk membersihkan data. Data yang dibersihkan juga merupakan data yang tidak layak digunakan. Hal ini dilakukan agar tidak adanya error pada proses perhitungan data. Kemudian setelah itu, data akan disusun berdasarkan setiap atribut nya masing-masing, hal ini dilakukan agar pada proses perhitungan data dapat dihitung dengan benar.

### **Uji Performa**

Tahap uji performa dari metode ID3 digunakan untuk mengukur sejauh mana pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pada tahap ini, data testing yang terdiri dari berbagai atribut kebutuhan teknologi komunikasi digunakan untuk mengevaluasi pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan melakukan pengujian pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya, dapat diketahui nilai akurasi dari pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya

dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya yang dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Hasil dari uji performa metode ID3 memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya ini, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Evaluasi pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya juga memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya yang dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Pada akhirnya, uji performa dari metode ID3 digunakan untuk memastikan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna

membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya ini, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya yang dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

## **BAB V**

### **ORANGE SEBAGAI ALAT BANTU PENGAPLIKASIAN**

Orange merupakan salah satu perangkat lunak open-source yang digunakan sebagai alat bantu pengaplikasian metode data mining dan machine learning. Aplikasi ini memungkinkan peneliti dan praktisi untuk membuat model klasifikasi dari berbagai data dengan antarmuka yang ramah pengguna dan visualisasi data yang lengkap. Dalam konteks ini, Orange digunakan untuk membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Kelebihan dari Orange ialah kemampuannya dalam mempermudah proses penerapan metode data mining bagi berbagai tingkat pengalaman pengguna, mulai dari pemula hingga praktisi data tingkat lanjut. Dengan Orange, pola pola kebutuhan teknologi komunikasi dapat dianalisis dari data yang belum pernah digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Aplikasi ini memungkinkan pola pola kebutuhan teknologi komunikasi dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Selain itu, Orange memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya ini, pola pola kebutuhan teknologi komunikasi dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Pada penerapan metode klasifikasi dengan Orange, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya ini dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

### **Pengertian Orange**

Orange merupakan perangkat lunak open-source untuk data mining dan pembelajaran mesin yang memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Aplikasi ini memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari

daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Fungsi dari Orange ialah membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Aplikasi ini memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Pada akhirnya, Orange digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya ini digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Berikut ini adalah panduan lengkap mengenai cara menggunakan aplikasi Orange, dimulai dari instalasi, hingga tahap evaluasi model.

### 1. Instalasi Orange

Sebelum menggunakan Orange, pastikan aplikasi telah terinstal di komputer Anda.

Langkah-langkah instalasi:

1. Unduh file instalasi, lalu jalankan dan ikuti petunjuk instalasi hingga selesai.

## 2. Antarmuka Orange

Setelah dibuka, Orange akan menampilkan antarmuka utama yang terdiri dari:

- a. Canvas Area (kanvas kerja): area kosong tempat Anda mengatur dan menghubungkan widget.
- b. Widget Toolbox (di sebelah kiri): berisi berbagai kategori widget seperti Data, Visualize, Model, Evaluate, dan lain-lain.
- c. Menu Bar dan Toolbar: untuk menyimpan proyek, membuka data, atau mengatur preferensi.
- d. Orange bekerja berdasarkan sistem drag and drop, di mana pengguna cukup menarik widget ke kanvas dan menghubungkannya untuk membuat alur kerja analisis data.

## 3. Mengimpor Data

Langkah pertama dalam penggunaan Orange adalah mengimpor data.

Cara mengimpor data:

- a. Tarik widget "File" dari kategori Data ke kanvas.
- b. Klik dua kali widget tersebut, lalu pilih file dataset dari komputer Anda, misalnya file .csv, .xlsx, atau file data bawaan Orange seperti iris.tab.
- c. Data akan dimuat dan bisa diteruskan ke widget lain untuk dianalisis.
- d. Alternatif lain, Anda bisa menggunakan widget "Datasets" untuk memuat dataset bawaan dari Orange.

#### 4. Melihat dan Memahami Data

Untuk memahami isi data, Anda dapat menggunakan widget:

“Data Table”: menampilkan data dalam bentuk tabel.

“Distributions”: melihat distribusi nilai atribut.

“Box Plot”, “Scatter Plot”, atau “Histogram”: untuk eksplorasi visual.

Langkah:

- a. Hubungkan widget File ke Data Table dengan menarik garis dari titik keluar File ke titik masuk Data Table.
- b. Klik dua kali Data Table untuk melihat data.
- c. Tambahkan visualisasi lain jika diperlukan untuk memahami struktur data.

#### 5. Mempersiapkan Data (Preprocessing)

Sebelum membuat model machine learning, data biasanya perlu diproses terlebih dahulu. Beberapa widget yang sering digunakan:

“Select Columns”: memilih kolom mana yang akan digunakan sebagai fitur dan target.

“Edit Domain”: mengubah tipe data atau label.

“Preprocess”: melakukan normalisasi, imputasi data hilang, dan lainnya.

Contoh:

Tambahkan widget “Select Columns”.

Hubungkan dari File → Select Columns, kemudian pilih atribut target (misalnya "species" pada data iris).

Lanjutkan ke pemodelan.

## 6. Membuat Model Machine Learning

Orange menyediakan berbagai algoritma ML seperti:

Naïve Bayes

Logistic Regression

Random Forest

k-Nearest Neighbors (kNN)

SVM (Support Vector Machine)

Neural Network (melalui add-on)

Langkah membuat model:

Tarik widget model, misalnya “Logistic Regression”, ke kanvas.

Hubungkan output dari preprocessing ke model.

Model akan otomatis dilatih jika ada data input.

Anda dapat menghubungkan lebih dari satu model untuk dibandingkan nantinya.

## 7. Menguji dan Mengevaluasi Model

Setelah model dibuat, penting untuk mengevaluasi kinerjanya.

Widget yang digunakan:

“Test & Score”: melakukan validasi silang (cross-validation).

“Confusion Matrix”: menampilkan hasil klasifikasi.

“ROC Analysis”: untuk mengukur kinerja klasifikasi biner.

Langkah evaluasi:

Tambahkan widget “Test & Score”.

Hubungkan dari data dan model ke widget ini.

Klik dua kali untuk melihat hasil metrik seperti akurasi, AUC, precision, dan recall.

Tambahkan widget “Confusion Matrix” untuk melihat kesalahan klasifikasi secara rinci.

## 8. Visualisasi Hasil

Orange menyediakan banyak widget visualisasi hasil:

“Scatter Plot”: menampilkan pemisahan kelas berdasarkan fitur.

“Tree Viewer”: menampilkan pohon keputusan.

“ROC Analysis”: grafik kurva ROC.

“Heat Map”, “Box Plot”, dll.

Visualisasi ini sangat membantu dalam memahami performa dan logika dari model yang dibuat.

## 9. Menyimpan dan Memuat Proyek

Untuk menyimpan pekerjaan:

Klik ikon Save di toolbar atau menu File → Save As.

Simpan file dengan ekstensi .ows (Orange Workflow Scheme).

Untuk memuat kembali:

Buka aplikasi Orange.

Pilih File → Open dan pilih file proyek yang disimpan.

## 10. Ekstensi dan Add-On

Orange dapat diperluas dengan berbagai add-on seperti:

Text Mining: untuk analisis teks.

Image Analytics: untuk pemrosesan citra.

Bioinformatics: untuk data biologis.

Geo: untuk data geospasial.

Cara menambah add-on:

Klik Options → Add-ons di menu atas.

Pilih add-on yang diinginkan dan klik Install.

Restart Orange jika diminta.

## 11. Contoh Alur Lengkap (Workflow)

Misalkan Anda ingin melakukan klasifikasi jenis bunga menggunakan dataset iris:

Tambahkan File → pilih iris.tab.

Hubungkan ke Data Table untuk melihat isi.

Tambahkan Select Columns → pilih “species” sebagai target.

Hubungkan ke Logistic Regression dan Random Forest.

Tambahkan Test & Score → hubungkan dari data dan kedua model.

Lihat perbandingan akurasi dan performa.

Tambahkan Confusion Matrix untuk masing-masing model.

Simpan proyek untuk digunakan kembali.

## 12. Kelebihan Orange

Antarmuka visual memudahkan eksplorasi tanpa menulis kode.

Cocok untuk pendidikan dan prototyping.

Mendukung berbagai teknik machine learning dan visualisasi.

Open-source dan terus berkembang dengan komunitas aktif.

## **Implementasi Orange pada Metode kNN dan ID3**

Implementasi metode K-Nearest Neighbor (KNN) dan ID3 dengan bantuan perangkat lunak Orange memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan Orange, pola pola kebutuhan teknologi komunikasi dapat dianalisis dan diklasifikasi berdasarkan atribut-atribut terkait, mulai dari desain antarmuka, kualitas handphone, kualitas jaringan, performa mesin, hingga nilai

ekonomis dari teknologi komunikasi itu sendiri. Hal ini memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya, memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Pada penerapan metode K-Nearest Neighbor (KNN) dengan Orange, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Hasil dari penerapan metode K-Nearest Neighbor dengan Orange memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Pada penerapan metode ID3 dengan Orange, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Dengan penerapan metode ID3 ini, pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya, memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

Pada akhirnya, implementasi metode K-Nearest Neighbor dan ID3 dengan Orange memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya. Pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan

teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya, memungkinkan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya, sehingga pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dapat digunakan sebagai pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya guna membuat pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya dengan pola pola kebutuhan teknologi komunikasi dari daerah daerah lainnya.

## **Kelebihan dan Kekurangan Orange: Platform Visual untuk Data Science dan Machine Learning**

Aplikasi **Orange** adalah sebuah tools open-source yang menawarkan pendekatan visual untuk analisis data dan pembelajaran mesin (machine learning). Dengan antarmuka drag-and-drop dan konsep modular, Orange memberikan kemudahan bagi siapa saja—baik pemula maupun praktisi—untuk melakukan eksperimen data tanpa harus menulis banyak baris kode. Namun, meski memiliki banyak keunggulan, Orange juga tidak lepas dari keterbatasan dan kekurangan. Berikut ini adalah paparan mendalam mengenai kelebihan dan kekurangan Orange.

### **KELEBIHAN ORANGE**

#### **1. Antarmuka Visual yang Ramah Pengguna**

Salah satu keunggulan utama Orange adalah antarmukanya yang **sangat intuitif dan visual**. Dengan sistem drag-and-drop widget,

pengguna dapat membangun workflow analisis data dengan cara menghubungkan komponen secara grafis. Hal ini membuat Orange ideal bagi pemula yang belum menguasai bahasa pemrograman seperti Python atau R.

Misalnya, seseorang dapat memuat data, melakukan visualisasi, membangun model klasifikasi, dan menguji performanya hanya dengan menghubungkan widget seperti "File", "Scatter Plot", "Logistic Regression", dan "Test & Score". Semua ini bisa dilakukan tanpa menulis satu baris kode pun.

## 2. Cocok untuk Pembelajaran dan Pendidikan

Karena kesederhanaannya, Orange sangat populer digunakan dalam **dunia pendidikan**, baik di sekolah, universitas, maupun pelatihan mandiri. Mahasiswa dapat belajar konsep machine learning seperti klasifikasi, regresi, clustering, dan validasi model secara praktis.

Konsep visual yang ditawarkan memudahkan pengajar untuk menunjukkan **alur proses data science** mulai dari input, preprocessing, modeling, hingga evaluasi dalam satu tampilan yang mudah dipahami.

## 3. Dukungan Beragam Algoritma dan Teknik Analisis

Orange menyediakan banyak algoritma machine learning populer yang siap digunakan, seperti:

- Decision Trees
- Random Forest
- Logistic Regression

- Naïve Bayes
- SVM
- K-Means Clustering
- PCA
- Hierarchical Clustering
- t-SNE

Selain itu, pengguna juga dapat menggunakan teknik preprocessing seperti imputasi nilai hilang, normalisasi, encoding data kategori, dan lain-lain. Ini memungkinkan pengguna untuk melakukan end-to-end machine learning pipeline.

#### 4. Add-On dan Ekstensi yang Fleksibel

Orange memiliki **sistem add-on modular** yang memungkinkan pengguna menambahkan fitur sesuai kebutuhan. Beberapa add-on yang tersedia meliputi:

- **Text Mining:** untuk analisis teks dan NLP.
- **Image Analytics:** untuk memproses gambar dan visual recognition.
- **Time Series:** untuk data deret waktu.
- **Bioinformatics:** untuk data biologis dan genomik.

Pengguna dapat memasang add-on ini dengan mudah melalui menu Add-ons tanpa perlu konfigurasi tambahan yang rumit.

## 5. Kompatibel dengan Python dan Scikit-learn

Bagi pengguna tingkat lanjut, Orange bisa diintegrasikan dengan **Python scripting**, yang memungkinkan pemrograman kustom di balik antarmuka grafis. Orange dibangun dengan dasar **scikit-learn**, sehingga model dan pipeline-nya kompatibel dengan pustaka Python populer tersebut.

Hal ini membuat Orange fleksibel: pengguna pemula bisa memulai tanpa kode, lalu perlahan berpindah ke scripting saat keterampilan mereka meningkat.

## 6. Visualisasi Data yang Interaktif

Orange unggul dalam hal visualisasi data. Berbagai widget seperti:

- Scatter Plot
- Box Plot
- Distributions
- Confusion Matrix
- ROC Curve
- Tree Viewer

...memungkinkan pengguna mengeksplorasi data dan hasil analisis dengan lebih interaktif. Visualisasi ini tidak hanya cantik, tetapi juga informatif dalam mendukung pemahaman terhadap pola dan performa model.

## **7. Komunitas dan Dokumentasi yang Aktif**

Sebagai proyek open-source, Orange memiliki komunitas yang cukup aktif. Tersedia banyak dokumentasi resmi, video tutorial, dan contoh proyek yang membantu pengguna baru untuk belajar secara bertahap. Forum dan GitHub juga menjadi tempat berbagi solusi dan mengusulkan fitur baru.

## **KEKURANGAN ORANGE**

### **1. Terbatas untuk Dataset Sangat Besar**

Orange merupakan salah satu platform visualisasi dan pembelajaran mesin yang sangat populer, terutama dalam bidang pendidikan dan pelatihan data science. Antarmuka grafis yang intuitif dan sistem drag-and-drop yang user-friendly membuatnya sangat digemari oleh pengguna pemula hingga menengah. Namun demikian, di balik kemudahannya, Orange memiliki keterbatasan yang cukup signifikan dalam hal penanganan dataset berskala besar. Platform ini belum sepenuhnya dirancang untuk melakukan pemrosesan dan analisis data dalam skala jutaan baris atau ratusan ribu fitur dengan performa yang efisien.

#### **a) Kinerja Menurun pada Volume Data Tinggi**

Ketika pengguna mencoba memuat dataset yang sangat besar ke dalam Orange — misalnya file CSV berukuran lebih dari 1 GB atau dataset dengan jutaan entri — aplikasi ini cenderung mengalami penurunan kinerja yang signifikan. Hal ini terlihat dari proses loading yang sangat lambat, tingginya penggunaan memori RAM, dan bahkan kadang-kadang menyebabkan aplikasi freeze atau tertutup secara tiba-tiba (crash). Hal ini terjadi karena Orange memproses data sepenuhnya di memori (RAM), tanpa sistem manajemen data berbasis disk atau streaming data bertahap.

Berbeda dengan platform analisis data lain seperti Apache Spark, Dask, atau KNIME, yang dirancang untuk skalabilitas dan pengolahan data dalam partisi atau batch, Orange belum memiliki sistem arsitektur yang mendukung distributed computing atau parallel processing dalam skala besar. Akibatnya, proses yang melibatkan pembelajaran mesin, seperti pelatihan model klasifikasi atau regresi, akan berjalan sangat lambat atau gagal sama sekali ketika dataset melebihi kapasitas sistem.

#### b) Keterbatasan Visualisasi untuk Data Besar

Salah satu fitur unggulan Orange adalah visualisasi data yang interaktif, seperti scatter plot, box plot, dan dendrogram. Namun, fitur ini pun tidak ideal jika digunakan pada dataset yang terlalu besar. Ketika pengguna mencoba memvisualisasikan data yang memuat ratusan ribu titik atau lebih, antarmuka Orange seringkali menjadi tidak responsif. Rendering grafik menjadi sangat lambat, dan manipulasi data secara visual menjadi tidak memungkinkan.

Masalah ini bukan hanya terkait dengan performa grafis, tetapi juga karena Orange tidak memiliki mekanisme agregasi otomatis atau sampling cerdas yang dapat menyederhanakan visualisasi ketika data terlalu banyak. Di platform lain, seperti Tableau atau Power BI, visualisasi berskala besar dapat diatasi dengan sistem ringkasan otomatis, namun Orange belum mengimplementasikan pendekatan serupa.

#### c) Tidak Tersedianya Opsi Streaming Data

Dalam era big data, banyak sistem yang bekerja dengan data stream, yaitu aliran data yang terus-menerus diperbarui secara real-time. Sayangnya, Orange tidak mendukung streaming data secara langsung, baik dari sumber seperti Kafka, MQTT, maupun

API real-time. Hal ini membuatnya tidak cocok untuk analisis log berskala besar, pemantauan sensor IoT, atau analisis keuangan yang mengandalkan pembaruan data detik per detik.

Orange bekerja dengan pendekatan batch data, yaitu semua data harus dimuat sekaligus sebelum dapat dianalisis. Pendekatan ini praktis untuk data kecil-menengah, tetapi menjadi hambatan besar bagi pengguna yang bekerja dengan data real-time atau data yang ukurannya terus tumbuh.

#### d) Terbatasnya Widget yang Mendukung Optimasi Performa

Sebagian besar widget di Orange tidak dirancang untuk menangani optimasi performa ketika dihadapkan pada data besar. Contohnya, widget seperti "Select Columns", "Data Table", dan "Scatter Plot" cenderung langsung memproses seluruh dataset yang dikirimkan ke dalamnya. Tidak tersedia pengaturan untuk memproses sebagian data (subset) atau menggunakan indeks sebagai acuan. Hal ini membuat pengguna tidak bisa mengontrol seberapa besar data yang akan dianalisis dalam setiap tahapan proses, sesuatu yang sangat penting dalam praktik data science profesional.

#### e) Solusi Sementara dan Workaround

Beberapa pengguna mencoba mengatasi keterbatasan ini dengan cara melakukan preprocessing data di luar Orange, seperti menggunakan Python (pandas, numpy) atau R, kemudian mengeksport subset data ke format CSV yang lebih kecil sebelum diimpor ke Orange. Strategi ini memang efektif untuk menghindari crash dan mempercepat proses, tetapi sekaligus mengurangi esensi visualisasi end-to-end yang menjadi kekuatan utama Orange. Selain itu, pengguna harus memiliki kemampuan

teknis tambahan di luar Orange, yang menjadikannya kurang ramah bagi kalangan non-programmer.

Alternatif lainnya adalah dengan menggunakan sampling data sebelum masuk ke Orange, tetapi ini pun memiliki risiko kehilangan informasi penting jika tidak dilakukan dengan tepat. Sampling tidak selalu mencerminkan populasi data sebenarnya, terutama dalam kasus dataset tidak seimbang atau memiliki distribusi yang kompleks.

#### f) Potensi Perkembangan ke Depan

Agar dapat bersaing di era big data dan AI modern, Orange perlu mengembangkan arsitektur yang lebih scalable dan modular, termasuk di antaranya:

- 1) Dukungan untuk pembacaan data bertahap (chunked reading)
- 2) Integrasi dengan basis data besar seperti PostgreSQL, Hadoop, atau BigQuery
- 3) Kemampuan visualisasi yang dapat menangani jutaan titik data dengan sistem summarization atau rendering cepat
- 4) Kompatibilitas dengan Python multiprocessing atau sistem komputasi paralel lainnya

Jika fitur-fitur ini dikembangkan, Orange akan dapat menjangkau segmen pengguna yang lebih luas, termasuk peneliti data, praktisi AI industri, dan analis yang bekerja dengan data real-time berskala besar.

## **2. Kustomisasi Terbatas di GUI**

Orange telah dikenal luas sebagai salah satu aplikasi open-source terbaik untuk analisis data dan pembelajaran mesin berbasis

antarmuka visual. Platform ini sangat disukai oleh kalangan pendidik, mahasiswa, dan peneliti pemula karena menyediakan pendekatan visual (drag-and-drop) untuk membangun workflow data science secara intuitif tanpa harus menulis kode secara langsung. Namun, di balik kemudahan ini, terdapat kelemahan mendasar yang seringkali menjadi hambatan dalam konteks penggunaan tingkat lanjut, yakni **kustomisasi yang terbatas pada antarmuka grafisnya (GUI)**.

Keterbatasan ini secara umum merujuk pada kemampuan pengguna dalam mengatur, menyesuaikan, atau memodifikasi elemen-elemen visual dan fungsional dari GUI Orange untuk memenuhi kebutuhan spesifik atau workflow yang kompleks. Bagi pengguna dengan kebutuhan fleksibilitas tinggi atau yang terbiasa dengan tingkat kendali yang lebih rinci (seperti di Jupyter Notebook, RStudio, atau IDE data science lainnya), kekakuan antarmuka Orange ini dapat menjadi kendala signifikan.

### **Minimnya Pengaturan Tampilan dan Interaktivitas**

Salah satu aspek kustomisasi yang paling terasa terbatas adalah pada tampilan visual antar-widget dan antarmuka keseluruhan. Orange tidak menyediakan fitur untuk mengubah tema secara bebas, merancang ulang layout, atau memodifikasi tampilan visual widget agar sesuai dengan preferensi pengguna. Hal ini membuat pengalaman pengguna terasa statis dan kurang personal, terutama bagi mereka yang menghabiskan banyak waktu dalam eksplorasi data atau presentasi hasil analisis kepada pihak lain.

Selain itu, beberapa widget visualisasi seperti **Scatter Plot**, **Box Plot**, dan **Heatmap**, hanya menyediakan pengaturan interaktif dalam jumlah terbatas. Pengguna tidak dapat mengatur palet warna secara detail, mengganti jenis grafik, menambahkan

anotasi khusus, atau menyesuaikan skala sumbu secara presisi. Bagi pengguna yang ingin menghasilkan visualisasi yang dipoles untuk keperluan publikasi atau presentasi profesional, keterbatasan ini seringkali memaksa mereka mengeksport data terlebih dahulu ke software visualisasi lain seperti Tableau, Power BI, atau matplotlib di Python.

### **Widget yang Tidak Sepenuhnya Terbuka untuk Modifikasi Visual**

Meskipun Orange bersifat open-source, dan secara teknis memungkinkan pengguna untuk membuat widget kustom, proses ini tidaklah mudah dan justru menuntut pemahaman teknis yang mendalam akan struktur internal Orange, termasuk bahasa pemrograman Python, Qt (untuk antarmuka), dan sistem plugin-nya. Artinya, **pengguna umum yang tidak berlatar belakang teknis akan kesulitan untuk menyesuaikan GUI atau menambah fitur visual secara mandiri.**

Sebagai contoh, pengguna tidak dapat menambahkan elemen tambahan seperti ikon, label khusus, atau tooltip interaktif tanpa melakukan perubahan pada kode sumber. Ini tentu bertolak belakang dengan tujuan awal Orange sebagai alat yang ramah bagi pengguna non-programmer. Dengan kata lain, meskipun tampilannya sederhana dan bersahabat, namun tidak tersedia opsi kustomisasi ringan yang bisa dilakukan langsung dari GUI-nya tanpa menyentuh kode.

### **Kurangnya Dukungan untuk Perancangan Workflow yang Lebih Fleksibel**

Secara visual, Orange mempresentasikan proses analisis data melalui “workflow canvas” yang terdiri dari node-node (widget)

yang saling terhubung. Namun, pengaturan layout canvas ini juga sangat terbatas. Pengguna tidak dapat mengelompokkan widget ke dalam subbagian (seperti grup atau frame), memberi warna atau penanda pada jalur proses tertentu, ataupun membuat anotasi yang menjelaskan alur analisis secara visual.

Hal ini cukup kontras jika dibandingkan dengan tools seperti KNIME atau RapidMiner yang menyediakan fitur semacam **container nodes**, **annotations**, dan **workflow groups** yang sangat membantu dalam dokumentasi serta komunikasi tim. Dalam proyek kolaboratif atau saat menyusun workflow kompleks, kustomisasi semacam ini sangat penting agar visualisasi proses analisis tetap rapi, terstruktur, dan mudah dipahami.

### **Terbatasnya Dukungan Multibahasa dan Aksesibilitas Visual**

Antarmuka Orange secara default hanya tersedia dalam bahasa Inggris, dan belum mendukung secara luas opsi multibahasa (multi-language UI). Ini menjadi keterbatasan tersendiri bagi pengguna di negara-negara non-Inggris yang ingin menerapkan Orange secara luas dalam pendidikan dasar atau pelatihan masyarakat umum. Selain itu, dari sisi aksesibilitas, GUI Orange juga belum sepenuhnya ramah bagi pengguna dengan kebutuhan khusus — misalnya, tidak ada opsi untuk high-contrast mode, pembaca layar (screen reader support), atau pengaturan ukuran teks dan ikon yang fleksibel.

Hal ini menunjukkan bahwa meskipun GUI Orange terlihat sederhana dan bersih, ia tidak dirancang untuk skenario penggunaan yang sangat beragam atau inklusif secara menyeluruh.

## **Kustomisasi Interaksi Terbatas pada Workflow Dinamis**

Selain keterbatasan estetika, antarmuka Orange juga belum mendukung **interaktivitas dinamis** antar-widget yang canggih. Misalnya, tidak tersedia fitur untuk mengaktifkan/menonaktifkan widget berdasarkan kondisi tertentu, mengatur parameter secara real-time melalui slider global, atau menyimpan beberapa konfigurasi eksperimental dalam satu canvas tanpa harus menduplikat seluruh proses.

Dalam praktik data science profesional, kebutuhan untuk menjalankan beberapa skenario uji coba secara paralel dan membandingkan hasilnya sangat penting. Namun, Orange mengharuskan pengguna untuk secara manual menduplikasi bagian-bagian workflow jika ingin membandingkan hasil model dengan parameter yang berbeda. Hal ini memperlihatkan keterbatasan GUI dalam menyokong fleksibilitas eksperimen dan pengambilan keputusan berbasis iterasi cepat.

## **Dampak Terhadap Pengguna Mahir dan Profesional**

Bagi pengguna pemula, keterbatasan GUI mungkin tidak menjadi masalah serius karena kebutuhan mereka biasanya sederhana dan lebih terfokus pada pemahaman konsep dasar. Namun, bagi **pengguna tingkat lanjut**, seperti peneliti, data scientist profesional, atau pengembang aplikasi analitik, keterbatasan ini justru menjadi titik lemah utama yang membatasi produktivitas dan kreativitas.

Ketika seorang pengguna ingin membangun solusi data science yang bersifat modular, fleksibel, dan dapat dipersonalisasi, GUI Orange tidak memberikan cukup ruang untuk eksperimen dan inovasi. Ini membuat banyak pengguna beralih ke platform lain yang lebih terbuka seperti Jupyter Notebook, KNIME, atau

bahkan membuat GUI kustom menggunakan Python dan library visualisasi seperti Streamlit atau Dash.

### 3. Fitur Advanced Masih Kurang Lengkap

Orange telah membuktikan dirinya sebagai salah satu alat pembelajaran mesin dan analisis data berbasis visual yang sangat efektif untuk pemula. Dengan antarmuka seret dan lepas (drag-and-drop) yang intuitif, Orange memberikan kemudahan dalam membangun alur kerja data science tanpa perlu menulis satu baris kode pun. Namun, di balik kekuatan tersebut, terdapat sebuah kekurangan fundamental yang kerap dirasakan oleh pengguna tingkat menengah dan lanjutan, yakni keterbatasan fitur **advanced** atau lanjutan yang belum sebanding dengan kebutuhan eksplorasi data yang lebih kompleks dan canggih.

Permasalahan ini menjadi semakin krusial ketika Orange digunakan dalam konteks profesional, riset ilmiah yang mendalam, atau proyek data science berskala besar. Di sinilah muncul keterbatasan-keterbatasan yang bersifat teknis namun berdampak signifikan terhadap fleksibilitas dan kedalaman analisis yang bisa dilakukan oleh platform ini. Berikut ini adalah ulasan mendalam tentang bagaimana dan mengapa keterbatasan fitur advanced menjadi kelemahan penting dari Orange.

### Kurangnya Kontrol Mendalam terhadap Algoritma Machine Learning

Salah satu aspek penting dalam penggunaan algoritma machine learning adalah kemampuan untuk melakukan penyesuaian terhadap parameter model. Orange memang menyediakan sejumlah algoritma populer seperti Naïve Bayes, K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine

(SVM), dan Neural Network. Namun, kontrol terhadap **hyperparameter tuning** dalam aplikasi ini masih sangat terbatas jika dibandingkan dengan lingkungan pemrograman seperti scikit-learn di Python atau caret di R.

Sebagai contoh, dalam modul Random Forest di Orange, pengguna hanya diberikan beberapa parameter dasar seperti jumlah pohon (trees) dan maksimum kedalaman (depth). Tidak tersedia pengaturan granular seperti kriteria pemisahan (gini vs entropy), pengaturan bootstrap, atau pengendalian atas jumlah fitur yang dipilih pada setiap node. Hal serupa terjadi pada SVM, yang tidak menyediakan fitur untuk memilih kernel secara custom atau menyesuaikan fungsi regularisasi secara rinci.

Ketiadaan kontrol yang lebih dalam ini membatasi kemampuan pengguna untuk melakukan **fine-tuning model**, sesuatu yang krusial dalam kompetisi machine learning, eksperimen riset, atau produksi model real-world.

## **Minimnya Dukungan untuk Teknik-Teknik Data Science Modern**

Perkembangan ilmu data saat ini mencakup berbagai pendekatan lanjutan seperti **ensemble learning kompleks**, **gradient boosting**, **autoML**, dan **deep learning**. Sayangnya, Orange belum memiliki dukungan penuh terhadap teknik-teknik ini secara out-of-the-box. Meskipun plugin tambahan seperti Orange3-DeepLearning atau Orange3-Timeseries tersedia, pengguna masih dibatasi oleh kurangnya integrasi mendalam dengan pustaka machine learning terkini seperti XGBoost, LightGBM, TensorFlow, atau PyTorch.

Sebagai perbandingan, platform seperti Google Colab atau Kaggle Kernel memberikan akses langsung ke berbagai pustaka

tersebut, memungkinkan eksperimen canggih seperti stacked models, transfer learning, atau multi-task learning. Di Orange, eksperimen semacam ini tidak mungkin dilakukan tanpa modifikasi kode atau pembuatan widget khusus, yang justru bertentangan dengan semangat visualisasi intuitif yang menjadi kekuatan utama Orange.

### **Keterbatasan dalam Pemrosesan Data Teks dan Time Series**

Fitur lanjutan dalam analisis data tidak hanya mencakup algoritma, tetapi juga kemampuan dalam menangani berbagai tipe data, terutama **teks** (natural language processing/NLP) dan **time series**. Orange memang telah menambahkan plugin teks yang memungkinkan pengguna melakukan preprocessing sederhana seperti tokenisasi, stopwords removal, dan analisis frekuensi kata. Namun, fitur lanjutan seperti named entity recognition (NER), part-of-speech tagging, dependency parsing, word embeddings (seperti Word2Vec, GloVe, atau BERT), dan analisis sentimen berbasis model, masih belum tersedia secara menyeluruh.

Begitu pula pada data deret waktu (time series), Orange memiliki dukungan awal melalui plugin Orange3-Timeseries, tetapi kemampuannya masih terbatas pada visualisasi dasar, decomposition sederhana, dan beberapa model prediktif konvensional. Belum tersedia model-model lanjutan seperti ARIMA otomatis, Prophet dari Facebook, atau bahkan RNN/LSTM yang sering digunakan dalam prediksi data temporal. Ini menjadikan Orange kurang optimal untuk digunakan dalam konteks aplikasi seperti forecasting penjualan, analisis sensor IoT, atau pemodelan tren keuangan.

## **Tidak Mendukung Integrasi Langsung dengan Cloud dan API Eksternal**

Fitur advanced juga mencakup kemampuan aplikasi untuk **berinteraksi dengan layanan eksternal** seperti database cloud, API publik, dan platform komputasi awan. Dalam hal ini, Orange tertinggal dari beberapa pesaingnya. Orange tidak memiliki dukungan langsung untuk menghubungkan data ke Google Sheets, REST API, ataupun sistem manajemen basis data relasional (RDBMS) secara fleksibel. Padahal dalam praktik profesional, pengambilan data dari sumber eksternal secara otomatis adalah hal yang sangat penting.

Selain itu, Orange juga tidak memiliki fitur autentikasi atau token management yang memungkinkan pengguna berinteraksi dengan layanan cloud seperti Google Cloud Platform, AWS, atau Microsoft Azure. Hal ini menyulitkan pengguna yang ingin mengintegrasikan pipeline Orange ke dalam workflow data engineering modern.

## **Minimnya Dukungan untuk Automasi dan Reprodusibilitas**

Dalam lingkungan penelitian dan produksi, automasi dan reprodusibilitas adalah prinsip dasar yang tidak bisa diabaikan. Sayangnya, Orange masih belum menyediakan fitur scripting internal, export ke bahasa pemrograman, atau pengaturan pipeline otomatis yang dapat dijadwalkan. Hal ini berarti setiap kali pengguna ingin mengulangi workflow, mereka harus membuka aplikasi secara manual dan menjalankan proses satu per satu.

Tidak adanya **workflow automation** seperti yang tersedia di KNIME atau Apache Airflow menyebabkan Orange sulit diintegrasikan dalam proses deployment machine learning end-to-end. Akibatnya, pengguna tidak dapat menjalankan alur kerja

secara berkala untuk menangani data baru atau memperbarui model secara otomatis.

### **Kurangnya Dukungan Eksperimen Komprehensif dan Analisis Statistik Lanjutan**

Fitur advanced juga meliputi kemampuan untuk melakukan eksperimen secara komprehensif. Meskipun Orange menyediakan widget “Test & Score” dan “Cross Validation”, namun evaluasi model masih sangat terbatas. Belum tersedia fitur seperti **nested cross-validation**, **ROC analysis terperinci**, **permutation importance**, atau **partial dependence plots** yang umum digunakan dalam evaluasi mendalam sebuah model prediksi.

Selain itu, kemampuan Orange dalam hal **analisis statistik** juga belum menyeluruh. Beberapa uji statistik dasar memang tersedia (seperti ANOVA dan t-test), tetapi untuk analisis multivariat yang lebih kompleks, seperti MANOVA, regresi logistik multinomial, atau uji post-hoc, pengguna harus mencari alternatif lain. Hal ini cukup menghambat ketika Orange digunakan untuk kebutuhan penelitian kuantitatif yang memerlukan validasi statistik mendalam.

### **4. Kurangnya Integrasi Eksternal Secara Default**

Orange tidak secara default memiliki kemampuan untuk terhubung langsung ke API eksternal, database relasional (seperti MySQL atau PostgreSQL), atau sumber cloud. Semua data harus dimuat terlebih dahulu secara lokal atau dalam format tertentu (CSV, Excel, atau .tab).

Untuk pengguna enterprise yang ingin menghubungkan pipeline machine learning mereka ke sumber data real-time atau database besar, ini bisa menjadi kendala.

## **5. Belum Sepenuhnya Modular Secara Arsitektur**

Meskipun Orange dikenal sebagai salah satu platform visualisasi dan pembelajaran mesin yang sangat ramah pengguna, khususnya bagi pemula, namun dari sisi desain perangkat lunak, Orange masih memiliki kekurangan yang cukup mencolok, yaitu belum sepenuhnya modular secara arsitektur. Ketidaksempurnaan dalam modularitas ini dapat berdampak langsung pada fleksibilitas, skalabilitas, dan efisiensi kerja pengguna, terutama dalam penggunaan yang lebih kompleks dan berskala besar.

Apa yang Dimaksud dengan Modularitas Arsitektur?

Modularitas dalam konteks arsitektur perangkat lunak mengacu pada kemampuan sistem untuk dipecah menjadi komponen-komponen kecil atau modul yang dapat berdiri sendiri dan berfungsi secara independen namun tetap dapat saling terhubung. Modularitas yang baik memungkinkan pengguna untuk menyusun, memodifikasi, atau mengganti bagian-bagian sistem tanpa harus mengubah keseluruhan struktur. Dalam konteks aplikasi seperti Orange, modularitas seharusnya memungkinkan pengguna:

- a) Menyusun workflow secara fleksibel dan efisien
- b) Menguji satu bagian proses tanpa memengaruhi seluruh pipeline
- c) Menambahkan, menghapus, atau mengubah modul (widget) dengan efek terbatas dan terkontrol
- d) Mengisolasi kesalahan atau bug pada satu modul tanpa merusak keseluruhan analisis

Namun, pada kenyataannya, Orange belum sepenuhnya mencapai tingkat modularitas tersebut.

### 1. Ketergantungan Antar Widget Masih Tinggi

Salah satu contoh konkret dari ketidakmodularan Orange adalah ketergantungan antar-widget yang terlalu erat. Saat pengguna menghubungkan satu widget dengan widget lain dalam suatu pipeline, data dan hasil olahan dari widget sebelumnya secara langsung mengalir ke widget berikutnya. Meskipun ini memberikan kemudahan dalam eksekusi alur kerja yang linier dan cepat, namun kelemahannya adalah:

Jika terjadi kesalahan atau perubahan pada satu widget (misalnya, "Select Columns"), maka seluruh alur kerja setelah widget tersebut ikut terpengaruh.

Tidak ada mekanisme isolasi atau buffer data di antaranya, sehingga debugging menjadi lebih sulit.

Pengguna tidak dapat dengan mudah membekukan satu langkah (misalnya, hasil preprocessing) untuk digunakan secara independen di jalur workflow lainnya.

Dalam arsitektur yang modular, seharusnya ada pemisahan tanggung jawab yang jelas antar bagian, dan pengguna diberi opsi untuk mengatur aliran data atau menyimpan hasil antar-langkah agar tidak terjadi efek domino ketika ada kesalahan kecil.

### 2. Kesulitan Mengelola Workflow yang Kompleks

Dalam kasus alur kerja yang melibatkan banyak model, preprocessing, dan evaluasi paralel, tampilan kanvas Orange bisa menjadi sangat padat dan tidak tertata. Hal ini disebabkan karena Orange tidak mendukung pengelompokan modul, nested workflows, atau kompartemen logika secara visual yang bisa

membantu pengguna mengorganisasi proses kerja dengan lebih rapi.

Misalnya, jika pengguna ingin membandingkan tiga model klasifikasi dengan tiga jenis preprocessing berbeda, maka semua komponen harus disusun secara eksplisit di kanvas, tanpa ada cara untuk menyembunyikan, meringkas, atau mengelompokkan proses tertentu. Ini mengakibatkan:

- a) Tampilan kanvas menjadi penuh dan sulit dibaca
- b) Risiko kesalahan penghubungan antar-widget meningkat
- c) Sulit melakukan dokumentasi visual yang jelas terhadap proses

Dalam sistem modular yang lebih canggih, seharusnya pengguna dapat membuat subworkflow atau menyimpan blok proses sebagai template yang bisa digunakan kembali, seperti yang tersedia di platform seperti KNIME atau RapidMiner.

### 3. Keterbatasan dalam Reusabilitas dan Otomatisasi

Reusabilitas adalah aspek penting dari modularitas. Namun, Orange belum sepenuhnya mendukung konsep ini. Widget yang telah digunakan dalam satu proyek tidak dapat disimpan sebagai komponen yang bisa digunakan ulang dalam proyek lain, kecuali seluruh file proyek disalin dan dimodifikasi. Tidak tersedia fitur untuk menyimpan "blok kerja" sebagai fungsi atau modul pengguna.

Selain itu, Orange belum mendukung parameterisasi modul secara dinamis. Pengguna tidak bisa, misalnya, mendefinisikan parameter global (seperti nilai  $k$  untuk KNN) yang bisa digunakan di berbagai bagian workflow. Semua konfigurasi dilakukan secara manual melalui klik di masing-masing widget, yang menyulitkan proses otomatisasi dan uji coba berulang.

Padahal, dalam arsitektur modular modern, reusabilitas dan parameterisasi sangat penting, terutama dalam skenario berikut:

- a) Pengujian parameter hyperparameter tuning
- b) Eksperimen berulang dengan konfigurasi berbeda
- c) Replikasi proyek serupa di dataset lain

#### 4. Integrasi dan Ekstensi Terbatas oleh Struktur Widget Tertutup

Sebagai aplikasi berbasis GUI, Orange bergantung pada widget sebagai elemen dasar pembangunan workflow. Namun, tidak semua widget bersifat terbuka atau dapat dikustomisasi dengan mudah. Meskipun Orange menyediakan dokumentasi bagi pengembang untuk membuat widget kustom, prosesnya masih cukup rumit dan tidak semudah menyusun script Python biasa.

Keterbatasan ini menjadi penghambat bagi pengguna lanjutan yang ingin menambahkan fungsionalitas tertentu atau mengintegrasikan metode analisis dari luar Orange. Ketika widget yang dibutuhkan tidak tersedia, pengguna hanya memiliki dua pilihan: membuat sendiri dengan pemrograman tingkat lanjut atau meninggalkan Orange dan berpindah ke platform lain.

Dalam sistem yang benar-benar modular, seharusnya pengguna bisa menambahkan modul dari pustaka eksternal atau plugin hanya dengan antarmuka GUI, tanpa harus menulis ulang struktur backend yang kompleks.

#### 5. Tidak Ada Kontrol Versi Modular atau History pada Level Widget

Dalam pengembangan sistem modular, penting juga untuk menyediakan kontrol versi atau pencatatan histori pada level masing-masing modul. Di Orange, tidak tersedia cara untuk

melihat perubahan konfigurasi widget secara historis, atau untuk melakukan rollback pada bagian tertentu dari workflow.

Ini menjadi masalah serius dalam konteks eksperimen dan analisis yang melibatkan banyak iterasi. Jika terjadi kesalahan konfigurasi, satu-satunya cara untuk kembali ke kondisi sebelumnya adalah dengan membuka file proyek sebelumnya (jika sempat disimpan) atau mengubah ulang semua pengaturan secara manual. Sistem modular yang baik seharusnya memiliki kemampuan versi per modul atau checkpoints pada titik-titik kritis dalam workflow.

## 6. Pengujian dan Debugging Terbatas

Karena kurangnya isolasi antar widget dan keterbatasan logging internal, proses pengujian dan debugging di Orange menjadi tidak efisien. Saat terjadi kesalahan, tidak selalu jelas dari mana asal masalah tersebut. Pengguna hanya mendapatkan pesan kesalahan di satu titik, tanpa penelusuran logika dari sumber datanya.

Dalam arsitektur modular yang baik, seharusnya ada fasilitas seperti:

- a) Logging per modul
- b) Validasi input dan output
- c) Breakpoints atau simulasi per langkah

Fitur-fitur ini akan sangat membantu pengguna mendeteksi kesalahan, mempercepat iterasi, dan meningkatkan keandalan sistem analitik secara keseluruhan.

## **6. Ketergantungan pada Tampilan Desktop**

Salah satu kelemahan signifikan dari aplikasi Orange adalah ketergantungannya pada tampilan desktop sebagai lingkungan utama pengoperasian. Orange dirancang sebagai aplikasi standalone yang harus diinstal secara lokal pada sistem operasi seperti Windows, macOS, atau Linux. Meskipun pendekatan ini memberikan stabilitas dan kontrol penuh terhadap proses analisis data, hal ini sekaligus membatasi fleksibilitas pengguna dalam konteks kolaborasi daring, akses jarak jauh, dan komputasi berbasis cloud.

Dalam dunia kerja dan pendidikan modern yang semakin mengandalkan layanan berbasis web, kebutuhan akan platform yang dapat diakses melalui browser menjadi sangat penting. Banyak aplikasi data science kontemporer sudah menyediakan versi berbasis web atau integrasi dengan platform kolaboratif seperti Google Colab, JupyterHub, atau layanan cloud seperti AWS dan Azure. Namun, Orange belum menyediakan antarmuka web atau versi yang dapat dijalankan secara langsung di cloud tanpa instalasi lokal, sehingga membatasi penggunaannya dalam ekosistem digital yang dinamis dan terdistribusi.

Keterbatasan ini juga menyulitkan pengguna yang bekerja dalam tim atau proyek bersama, karena membagikan workflow hanya dapat dilakukan melalui file proyek lokal (.ows), bukan secara real-time atau berbasis cloud. Selain itu, pengguna dengan perangkat low-end atau sistem operasi yang tidak kompatibel juga akan kesulitan untuk menjalankan aplikasi ini secara optimal.

Dengan perkembangan teknologi dan kebutuhan kerja kolaboratif yang semakin tinggi, harapan ke depan adalah Orange dapat mengembangkan versi berbasis web atau menyediakan integrasi

langsung dengan platform cloud sehingga tidak lagi bergantung sepenuhnya pada tampilan desktop.

## BIOGRAFI PENULIS



### **Mayang Alvi Perdana Nasution**

Lahir di Ajamu pada tanggal 27 April 2002. Ia merupakan salah satu mahasiswi aktif di Universitas Labuhanbatu, yang saat ini sedang menempuh pendidikan pada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi. Ketertarikannya terhadap dunia teknologi informasi dan sistem komputer telah mendorongnya untuk terus mengembangkan pengetahuan dan keterampilan di bidang tersebut, baik melalui pembelajaran akademik maupun kegiatan di luar perkuliahan. Sejak awal menjalani pendidikan tinggi, Mayang dikenal sebagai pribadi yang tekun, bertanggung jawab, dan memiliki semangat belajar yang tinggi. Ia memiliki cita-cita besar untuk dapat berkarier di lingkungan Badan Usaha Milik Negara (BUMN), sebuah impian yang dilandasi oleh keinginannya untuk berkontribusi langsung dalam pembangunan bangsa melalui jalur profesional di sektor strategis negara. Bagi Mayang, bekerja di BUMN bukan hanya tentang prestise, tetapi juga tentang tanggung jawab dan kesempatan untuk memberikan dampak nyata bagi masyarakat luas. Oleh karena itu, ia terus mempersiapkan diri dengan memperkuat kemampuan teknis, memperluas wawasan, serta membangun sikap profesional yang dibutuhkan di dunia kerja. Dengan komitmen dan tekad yang kuat, Mayang Alvi Perdana Nst berusaha menjadikan masa kuliahnya sebagai fondasi yang kokoh untuk mencapai cita-citanya di masa depan.



### **Ibnu Rasyid Muntthe**

Lahir di Kota Rantau Prapat, Kabupaten Labuhanbatu, pada tahun 1987. Beliau memperoleh gelar Sarjana Teknik (S.T.) dari Universitas Nurtanio Bandung dan melanjutkan pendidikan Magister di Universitas Putra Indonesia (YPTK) Padang dengan gelar Magister Komputer (M.Kom.).

Beliau aktif dalam menulis artikel ilmiah serta buku akademik. Saat ini, beliau berkiprah sebagai dosen di Universitas Labuhanbatu. Selain itu, beliau juga mengelola kanal YouTube @manjaddawajada2022 sebagai media pembelajaran dan sarana berbagi ilmu pengetahuan.



### **Deci Irmayani**

Lahir di Rantauprapat pada tanggal 27 Mei 1986. Telah menyelesaikan Pendidikan Sarjana (S1) di STMIK Potensi Utama Medan dan melanjutkan Pendidikan Magister (S2) di UPI YPTK Padang. Saat ini bekerja sebagai Dosen di Universitas Labuhan Batu, Fakultas

Sains dan Teknologi, dengan spesialisasi di bidang komputer. berkomitmen untuk memberikan kontribusi dalam bidang pendidikan dan teknologi, serta membagikan pengetahuan kepada generasi mendatang.



### **Angga Putra Juledi**

Lahir di Kota Padang pada tanggal 19 Juli 1994. Perjalanan pendidikannya dimulai dari SDN 19 Padang yang ia tamatkan pada tahun 2006, kemudian melanjutkan ke SMPN 3 Padang dan lulus pada tahun 2009. Setelah itu, ia menyelesaikan pendidikan di SMA Pertiwi 2 Padang pada tahun 2012. Ia kemudian melanjutkan pendidikan tinggi jenjang Sarjana (S1) di Universitas Putra Indonesia “YPTK” Padang dan lulus pada tahun 2018 dengan jurusan Sistem Informasi. Selanjutnya, ia meneruskan studi jenjang Magister (S2) pada program Pascasarjana di universitas yang sama, yaitu Universitas Putra Indonesia “YPTK” Padang, dan lulus pada tahun 2019 melalui Program Studi Teknik Informatika dengan konsentrasi Sistem Informasi.

Setelah menyelesaikan pendidikannya, Angga Putra Juledi mengabdikan diri sebagai dosen di bidang Ilmu Komputer pada Fakultas Sains dan Teknologi, Program Studi Sistem Informasi di Universitas Labuhan Batu. Ia resmi menjadi dosen tetap sejak tahun 2020 dan hingga kini aktif mengajar di kampus tersebut. Selain mengajar, ia juga dipercaya mengemban amanah sebagai bagian dari struktur organisasi kampus, dengan menjabat sebagai Kepala Bagian Sumber Daya Manusia untuk periode 2023 hingga 2027.

Di bidang penulisan, ia telah menerbitkan beberapa buku. Buku pertamanya berjudul *Internetworking dan TCP/IP*, diterbitkan pada 31 Desember 2021. Sementara itu, buku keduanya terbit pada 17 Oktober 2023 dengan judul *Panduan Belajar HTML, CSS, dan JavaScript untuk Pemula*. Saat ini, ia masih terus aktif menulis buku secara rutin setiap tahunnya.

# SINOPSIS

Klasifikasi data merupakan salah satu teknik penting dalam data mining dan kecerdasan buatan yang digunakan untuk mengelompokkan data ke dalam kelas-kelas tertentu berdasarkan pola atau karakteristik yang dimiliki. Buku atau karya ini membahas secara mendalam dua metode populer dalam klasifikasi, yaitu K-Nearest Neighbor (KNN) dan ID3 (Iterative Dichotomiser 3).

Metode KNN bekerja dengan prinsip kedekatan jarak antara data uji dan data latih, di mana suatu data akan diklasifikasikan ke dalam kelas mayoritas dari tetangga terdekatnya. KNN dikenal karena kesederhanaan dan efektivitasnya dalam banyak kasus klasifikasi. Sementara itu, ID3 merupakan algoritma pembentukan pohon keputusan yang menggunakan konsep entropy dan information gain untuk menentukan atribut terbaik dalam membagi data.

Karya ini menyajikan penjelasan teoritis, algoritma dasar, kelebihan dan kekurangan masing-masing metode, serta studi kasus penerapan dalam berbagai bidang seperti diagnosis penyakit, klasifikasi nilai akademik, dan analisis data pelanggan. Selain itu, perbandingan performa kedua metode juga dibahas melalui eksperimen menggunakan dataset nyata.

ISBN 978-634-04-2085-2



9

786340

420852



**PENERBIT  
YAYASAN MMI**