

Penerapan

MACHINE LEARNING

**Pada Pelanggan Perusahaan Air
Minum Daerah dengan Metode
Naive Bayes**

**FAHRI ANUGRAH NASUTION
IBNU RASYID MUNTHE, S.T., M.Kom
MASRIZAL, S.Kom., M.Kom.
MARNIS NASUTION, S.Kom., M.Kom**

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:

penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;

Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;

Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan

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

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).

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).

PENERAPAN MACHINE LEARNING PADA PELANGGAN PERUSAHAAN AIR MINUM DAERAH DENGAN METODE NAIVE BAYES

FAHRI ANUGRAH NASUTION
IBNU RASYID MUNTHE, S.T., M.Kom
MASRIZAL, S.Kom., M.Kom.
MARNIS NASUTION, S.Kom., M.Kom



**PENERAPAN MACHINE LEARNING PADA PELANGGAN PERUSAHAAN
AIR MINUM DAERAH DENGAN METODE NAIVE BAYES**

**FAHRI ANUGRAH NASUTION
IBNU RASYID MUNTHE, S.T., M.Kom
MASRIZAL, S.Kom., M.Kom.
MARNIS NASUTION, S.Kom., M.Kom**

Editor :
Sahat Parulian Sitorus

Desain Cover :
Nurhasanah

Sumber :
<https://isbn.jndi.my.id/products/penerapan-machine-learning-pada-pelanggan-perusahaan-air-minum-daerah-dengan-metode-naive-bayes>

Tata Letak :
Ida Royani Sitorus

Proofreader :

Ukuran :
Jml hal judul 125, Jml hal isi naskah 115, Uk: 14x20 cm

ISBN :
978-634-04-0887-4

Cetakan Pertama :
02 Juli 2025

Hak Cipta 2025, Pada Penulis

Isi diluar tanggung jawab percetakan

Copyright © 2025 by PT.JNDI
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 PT. JASA NIAGA DIGITAL INDONESIA

Jl. H. Maulana Link. VI. Labuhan Ruku, Talawi, Batu Bara, Sumatera Utara 21254

Telp/Wa : 082361274081

<https://isbn.jndi.my.id>

<https://perpusnas.jndi.my.id>

E-mail: office@jndi.my.id

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir ini yang berjudul *“Penerapan Machine Learning pada Pelanggan Perusahaan Air Minum Daerah dengan Metode Naïve Bayes”* sebagai bagian dari persyaratan akademik dalam menyelesaikan studi pada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu.

Penelitian ini dilatarbelakangi oleh kebutuhan untuk memahami tingkat kepuasan pelanggan terhadap layanan Perusahaan Daerah Air Minum (PDAM) secara lebih objektif dan terukur. Dalam era digital seperti saat ini, pendekatan teknologi melalui machine learning menjadi salah satu metode yang efektif dalam menganalisis data pelanggan dan membantu pengambilan keputusan berbasis

data. Dengan menggunakan metode Naïve Bayes, penelitian ini diharapkan mampu mengidentifikasi pola kepuasan pelanggan berdasarkan beberapa faktor seperti kualitas air, ketersediaan, harga, dan pelayanan.

Dalam proses penyusunan karya ilmiah ini, penulis menghadapi berbagai tantangan, baik secara teknis maupun non-teknis. Namun, berkat semangat dan ketekunan yang terus dijaga, serta dukungan dari berbagai pihak, akhirnya tugas akhir ini dapat diselesaikan. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah memberikan dukungan moral, motivasi, dan bantuan selama proses penelitian berlangsung.

Ucapan terima kasih secara khusus penulis sampaikan kepada keluarga tercinta yang selalu memberikan doa dan semangat, serta kepada rekan-rekan seperjuangan yang

senantiasa mendukung dan menjadi teman diskusi yang berarti selama penyusunan ini.

Rantauprapat, 27 Juni 2025

Penulis

Fahri Anugrah Nasution

DAFTAR ISI

KATA PENGANTAR	v
DAFTAR ISI.....	v
BAB I PENDAHULUAN.....	1
BAB II KONSEP DAN TEORI PENDUKUNG	7
2.1. Tingkat Kepuasan.....	7
2.2. Air Bersih PDAM.....	9
2.3. Knowledge Discovery in Database (KDD)	11
2.1.1. Machine Learning	13
2.1.2. Database dan Data Processing	15
2.1.3. Visualization	17
2.1.4. Statistik	19
2.1.5. Pattern Recognition.....	21
2.4. Metode Naïve Bayes	23
2.5. Model Klasifikasi	25
2.6. Alat Bantu Program Aplikasi Orange	27
2.7. Kelebihan dan Kekurangan Metode Naive Bayes	
29	
2.8. Evaluasi Metode Naïve Bayes.....	32
2.9. Penelitian Terdahulu	35
BAB III ANALISA DAN PERANCANGAN	45

3.1. Arsitektur Sistem.....	45
3.2. Pengumpulan Data	47
3.3. Langkah-Langkah Pengolahan pada Metode Naive Bayes	60
BAB IV HASIL DAN PEMBAHASAN	68
4.1 Analisa Data	68
4.2. Pengumpulan Data	69
4.3. Pembagian Data.....	73
4.4. Pembersihan Data.....	82
4.5. Perancangan Model Klasifikasi.....	83
4.6. Hasil Klasifikasi	85
4.7. Perancangan Model Evaluasi	94
4.8. Hasil Evaluasi.....	96
4.8.1. Hasil Test and Score	96
4.8.2. Hasil Confusion Matrix	98
4.8.3. Hasil Box Plot.....	100
DAFTAR PUSTAKA	103
TENTANG PENULIS	113

BAB I

PENDAHULUAN

Isi naskah bisa anda ketik mulai dari sini, ukuran dan jenis font bisa disesuaikan menurut keinginan anda. Selain itu, spasi naskah bisa anda setting sendiri atau anda percayakan tataletak naskah pada kami. Paragraph selanjutnya silahkan tekan enter.

Air PDAM (Perusahaan Daerah Air Minum) adalah air yang disalurkan oleh perusahaan milik pemerintah daerah untuk memenuhi kebutuhan air bersih masyarakat, terutama untuk keperluan konsumsi dan kegiatan rumah tangga. Air yang disalurkan PDAM biasanya berasal dari berbagai sumber, seperti sungai, danau, mata air, atau sumur dalam yang telah melalui serangkaian proses pengolahan untuk menjadikannya aman dan layak dikonsumsi. Proses pengolahan ini mencakup penyaringan, pengendapan, pengolahan kimia, serta sterilisasi dengan menggunakan

bahan-bahan seperti klorin untuk membunuh bakteri dan mikroorganisme berbahaya. Namun, kualitas air PDAM bisa bervariasi tergantung pada lokasi dan teknologi yang digunakan oleh masing-masing PDAM.

Air PDAM, meskipun disalurkan untuk memenuhi kebutuhan air bersih masyarakat, terkadang memiliki kualitas yang kurang baik di beberapa daerah. Hal ini dapat terjadi akibat berbagai faktor, seperti sumber air baku yang tercemar, proses pengolahan yang belum maksimal, atau kerusakan pada sistem distribusi yang menyebabkan terjadinya kebocoran atau pencemaran ulang selama perjalanan air. Di beberapa wilayah, kualitas air PDAM dapat terpengaruh oleh tingginya kandungan kotoran, endapan lumpur, atau zat kimia yang belum sepenuhnya terfilter selama proses pengolahan.

Oleh karena itu, tidak sedikit masyarakat yang merasa tidak puas dengan kualitas air PDAM yang mereka terima, terutama di daerah-daerah yang sering mengalami masalah terkait kebersihan dan kestabilan pasokan air. Ketidakpuasan ini muncul karena kualitas air yang kadang-kadang keruh, berbau, atau memiliki rasa yang tidak enak, yang disebabkan oleh berbagai faktor seperti pencemaran sumber air, kualitas pengolahan yang kurang optimal, atau infrastruktur distribusi yang rusak. Selain itu, keterbatasan pasokan air yang tidak merata juga menjadi masalah utama, di mana beberapa wilayah sering kekurangan air pada musim kemarau atau mengalami gangguan pasokan yang membuat warga harus mengandalkan sumber air lain, seperti air sumur atau air kemasan.

Terkadang, masih ada sebagian masyarakat yang merasa puas dengan layanan air PDAM, terutama di daerah-

daerah yang memiliki sistem distribusi yang lebih baik dan sumber air yang terjaga kebersihannya. Bagi mereka, air PDAM yang diterima sudah melalui proses pengolahan yang efektif, sehingga air yang sampai ke rumah mereka bersih, jernih, dan aman untuk digunakan baik untuk keperluan mandi, mencuci, maupun konsumsi langsung. Proses penyaringan yang baik, ditambah dengan pemeliharaan rutin pada sistem pipa dan sumber air baku yang berkualitas, membuat air yang diterima tetap memenuhi standar kesehatan yang ditetapkan.

Namun, dari fenomena tersebut, tidak dapat dipastikan bagaimana tingkat kepuasan masyarakat terhadap air PDAM secara keseluruhan, karena berbagai faktor yang mempengaruhi persepsi individu. Beberapa faktor yang membuat masyarakat tidak puas meliputi kualitas air yang keruh, berbau, atau tidak layak konsumsi, serta gangguan

pasokan air yang sering terjadi, terutama pada musim kemarau atau daerah-daerah yang memiliki infrastruktur yang kurang memadai. Di sisi lain, faktor yang membuat masyarakat merasa puas termasuk kualitas air yang bersih, jernih, dan aman untuk dikonsumsi langsung, serta kelancaran distribusi yang dapat diandalkan. Hal ini menunjukkan bahwa ada kebutuhan untuk mengidentifikasi dan menganalisis berbagai faktor yang mempengaruhi tingkat kepuasan masyarakat terhadap air PDAM. Oleh karena itu, penting untuk melakukan penelitian yang mendalam mengenai hal ini, yang dapat dilakukan dengan menggunakan pendekatan machine learning, khususnya dengan metode Naive Bayes. Metode ini dapat membantu dalam menganalisis data survei atau feedback dari masyarakat untuk mengidentifikasi pola-pola tertentu yang dapat mengungkapkan faktor-faktor yang berkontribusi pada

kepuasan atau ketidakpuasan mereka terhadap layanan air PDAM, sehingga hasil penelitian dapat memberikan rekomendasi yang lebih akurat bagi perbaikan kualitas dan layanan PDAM di masa depan.

BAB II

KONSEP DAN TEORI PENDUKUNG

2.1. Tingkat Kepuasan

Tingkat kepuasan masyarakat merupakan indikator penting dalam menilai kualitas layanan yang diberikan oleh suatu perusahaan atau instansi, termasuk dalam hal ini adalah PDAM. Kepuasan ini dipengaruhi oleh berbagai faktor, seperti kualitas produk atau layanan, kemudahan akses, responsivitas terhadap keluhan, serta harga yang wajar. Dalam konteks layanan air PDAM, faktor-faktor seperti kejernihan air, bau, kelayakan untuk konsumsi, serta keberlanjutan pasokan air menjadi aspek-aspek utama yang membentuk persepsi masyarakat. Jika layanan yang diterima sesuai dengan harapan masyarakat, mereka cenderung merasa puas, namun apabila ada ketidaksesuaian, seperti kualitas air yang buruk atau gangguan pasokan yang sering,

hal ini dapat menyebabkan ketidakpuasan yang berujung pada pencarian alternatif lain.

Tingkat kepuasan masyarakat terhadap air PDAM sangat relevan untuk dianalisis dalam upaya meningkatkan kualitas pelayanan. Dalam hal ini, pengukuran kepuasan dapat dilakukan dengan menganalisis data yang dikumpulkan dari masyarakat di wilayah tertentu, yang dapat mengungkapkan pola kepuasan atau ketidakpuasan yang ada. Melalui pendekatan ini, faktor-faktor yang mempengaruhi tingkat kepuasan dapat diidentifikasi secara lebih jelas, yang kemudian dapat digunakan untuk merancang kebijakan atau langkah-langkah perbaikan layanan PDAM. Penggunaan metode Naive Bayes dalam mengklasifikasikan tingkat kepuasan masyarakat akan membantu dalam memetakan persepsi masyarakat berdasarkan data yang ada, memberikan gambaran yang

lebih konkret tentang kualitas layanan yang diharapkan oleh pengguna, dan bagaimana PDAM dapat lebih baik memenuhi ekspektasi tersebut.

2.2. Air Bersih PDAM

Air bersih yang disalurkan oleh PDAM (Perusahaan Daerah Air Minum) merupakan salah satu kebutuhan dasar yang sangat penting bagi masyarakat. PDAM bertanggung jawab untuk mengolah, mengawasi, dan mendistribusikan air yang layak konsumsi ke rumah-rumah warga. Proses pengolahan air ini melibatkan berbagai tahapan seperti penyaringan, pengendapan, dan penambahan bahan kimia untuk memastikan air yang sampai di rumah tangga bebas dari kontaminasi dan aman untuk digunakan. Namun, kualitas air bersih yang disalurkan oleh PDAM sering kali dipengaruhi oleh berbagai faktor, termasuk kualitas sumber air baku, kondisi pipa distribusi, serta kemampuan sistem

pengolahan yang ada. Meskipun PDAM berusaha menyediakan air yang memenuhi standar kesehatan, banyak warga yang masih merasa kurang puas dengan kualitas air yang mereka terima.

Kualitas air bersih yang disediakan PDAM sangat memengaruhi tingkat kepuasan masyarakat. Faktor-faktor seperti kejernihan air, bau, dan kestabilan pasokan sangat menentukan bagaimana masyarakat menilai layanan yang diberikan. Berdasarkan hal tersebut, penting untuk mengevaluasi seberapa besar pengaruh kualitas air terhadap kepuasan pengguna di wilayah tertentu, terutama di Kota Rantauprapat, dengan menggunakan pendekatan yang berbasis pada data. Dengan menganalisis pola-pola kepuasan yang ada melalui metode machine learning, seperti Naive Bayes, dapat diperoleh wawasan lebih dalam mengenai faktor-faktor utama yang memengaruhi kepuasan terhadap

air PDAM. Temuan ini akan sangat berguna untuk meningkatkan kualitas pelayanan air bersih yang disediakan oleh PDAM, sekaligus memberi gambaran yang lebih akurat tentang kondisi air yang diterima oleh masyarakat, serta memberikan solusi yang lebih tepat guna untuk memperbaiki pelayanan di masa mendatang.

2.3. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah proses untuk menemukan pengetahuan yang berguna dari data besar yang tersimpan dalam database [1]. KDD melibatkan serangkaian langkah mulai dari pemilihan data, pembersihan data, transformasi data, hingga penerapan algoritma data mining untuk mengeksplorasi pola dan hubungan yang tersembunyi dalam data [2]. Proses ini dimulai dengan tahap pemahaman dan pemilihan data yang relevan, diikuti dengan preprocessing untuk mengatasi data

yang tidak lengkap atau tidak konsisten [3]. Selanjutnya, teknik-teknik data mining seperti klasifikasi, clustering, atau asosiasi digunakan untuk menggali wawasan yang tersembunyi, yang akhirnya dapat digunakan untuk membuat keputusan yang lebih baik atau mengidentifikasi tren yang dapat memberikan keuntungan kompetitif [4].

Menggunakan pendekatan KDD dalam analisis kepuasan masyarakat terhadap air PDAM memungkinkan untuk mengidentifikasi faktor-faktor yang mempengaruhi kualitas layanan dan kepuasan pengguna secara lebih sistematis. Proses ini dimulai dengan pengumpulan data yang relevan dari masyarakat, yang kemudian melalui tahapan pembersihan dan transformasi agar dapat digunakan dalam model klasifikasi seperti Naive Bayes. Melalui teknik ini, data yang dikumpulkan dapat diolah untuk menemukan pola kepuasan masyarakat dan klasifikasi tingkat kepuasan

berdasarkan variabel-variabel yang mempengaruhi, seperti kejernihan air, ketersediaan, atau kelancaran distribusi. Dengan memanfaatkan KDD, hasil yang diperoleh bisa lebih akurat dan memberikan gambaran yang jelas mengenai faktor-faktor yang mempengaruhi kepuasan masyarakat terhadap layanan air PDAM.

2.1.1. Machine Learning

Machine Learning (ML) adalah cabang dari kecerdasan buatan (AI) yang memungkinkan komputer untuk belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit [5]. Proses dalam machine learning melibatkan pengumpulan data, pemrosesan data tersebut, dan kemudian melatih algoritma untuk mengenali pola dalam data [6]. Dengan kemampuan untuk memproses data dalam jumlah besar dan menemukan pola yang tidak dapat dilihat dengan mata manusia, ML telah diterapkan

dalam berbagai bidang, mulai dari analisis bisnis, pengenalan suara, hingga diagnosis medis [7]. Berbagai teknik dalam ML, seperti supervised learning, unsupervised learning, dan reinforcement learning, memungkinkan pemecahan masalah yang lebih kompleks dengan akurasi yang lebih tinggi dibandingkan metode tradisional [8].

Penggunaan machine learning, khususnya dengan metode Naive Bayes, memberikan pendekatan yang efektif untuk menganalisis tingkat kepuasan masyarakat terhadap layanan air PDAM. Melalui algoritma ini, data yang diperoleh dari survei atau feedback masyarakat dapat dianalisis untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap kepuasan pengguna. Naive Bayes, dengan kemampuannya dalam mengklasifikasikan data berdasarkan probabilitas, memungkinkan pengelompokan masyarakat yang puas atau tidak puas terhadap kualitas air

PDAM. Dengan demikian, ML tidak hanya membantu dalam memproses data besar tetapi juga memberi pemahaman yang lebih dalam tentang bagaimana faktor-faktor tertentu memengaruhi pengalaman masyarakat terhadap layanan air, yang pada gilirannya dapat digunakan untuk meningkatkan kualitas pelayanan yang diberikan.

2.1.2. Database dan Data Processing

Database merupakan kumpulan data yang tersusun secara terstruktur dan dapat diakses, dikelola, dan diperbarui secara efisien [9]. Database berfungsi sebagai tempat penyimpanan data yang memungkinkan pengguna untuk menyimpan, mengambil, dan mengolah informasi dengan mudah [10]. Sistem manajemen database (DBMS) menyediakan berbagai fitur untuk mendukung pengelolaan data, seperti keamanan, konsistensi, dan integritas data. Jenis-jenis database yang umum digunakan meliputi

database relasional, non-relasional, dan terdistribusi, yang masing-masing memiliki kelebihan dan kekurangan sesuai dengan kebutuhan aplikasi [11]. Database mempermudah proses pengolahan informasi yang besar dan kompleks, sekaligus memastikan bahwa data dapat diakses dengan cepat dan akurat untuk menghasilkan keputusan yang lebih baik [12].

Proses pengolahan data dalam konteks kualitas air PDAM melibatkan pengumpulan data dari masyarakat untuk menganalisis tingkat kepuasan terhadap pelayanan yang diberikan. Data yang dikumpulkan melalui survei akan disimpan dalam database, yang kemudian diproses untuk mendapatkan informasi yang relevan mengenai faktor-faktor yang mempengaruhi kualitas air dan kepuasan masyarakat. Melalui pengolahan data ini, pola-pola yang signifikan dapat diidentifikasi dengan menggunakan algoritma seperti Naive

Bayes, yang mengklasifikasikan data berdasarkan kriteria yang telah ditentukan. Data yang telah diproses dan dianalisis kemudian digunakan untuk menghasilkan rekomendasi bagi perbaikan layanan PDAM, sehingga meningkatkan kualitas air dan kepuasan masyarakat secara keseluruhan.

2.1.3. Visualization

Visualization adalah teknik yang digunakan untuk menyajikan data dalam bentuk grafis atau visual yang memudahkan pemahaman, analisis, dan interpretasi [13]. Melalui visualisasi, data yang kompleks dapat disajikan secara lebih sederhana dan mudah dipahami, mengubah angka dan statistik menjadi gambar atau grafik yang lebih mudah diakses oleh audiens [14]. Penggunaan grafik, diagram batang, pie chart, atau heatmap membantu untuk menggambarkan tren, pola, dan hubungan antar variabel

yang sulit terlihat hanya dengan data mentah. Selain itu, visualisasi juga memungkinkan penggunaanya untuk mengidentifikasi outlier, korelasi, dan distribusi data secara cepat, sehingga memudahkan pengambilan keputusan yang lebih tepat dan efektif [15].

Dalam hal ini, visualisasi dapat digunakan untuk memperjelas hasil evaluasi kualitas air dan tingkat kepuasan masyarakat terhadap layanan PDAM. Melalui grafik atau diagram, data yang dikumpulkan dari survei kepuasan dapat divisualisasikan untuk menggambarkan hubungan antara kualitas air dan tingkat kepuasan masyarakat di wilayah Rantauprapat.

Dengan menggunakan metode visualisasi, pola-pola tertentu dalam data, seperti variabilitas kepuasan berdasarkan kualitas air atau faktor-faktor yang mempengaruhi kepuasan, dapat terlihat dengan jelas,

sehingga membantu pihak terkait dalam merumuskan kebijakan yang lebih efektif. Visualisasi ini juga membantu dalam mengevaluasi performa model machine learning yang digunakan untuk klasifikasi kepuasan, memberikan gambaran yang lebih terstruktur tentang bagaimana setiap faktor berkontribusi terhadap hasil akhir.

2.1.4. Statistik

Statistik adalah cabang ilmu yang berkaitan dengan pengumpulan, analisis, interpretasi, penyajian, dan pengorganisasian data [16]. Dengan menggunakan metode statistik, kita dapat membuat kesimpulan yang lebih tepat tentang populasi yang lebih besar berdasarkan sampel yang diambil. Dalam statistik, terdapat dua jenis utama: statistik deskriptif yang menggambarkan karakteristik dasar dari suatu dataset, dan statistik inferensial yang digunakan untuk membuat prediksi atau generalisasi tentang populasi

berdasarkan sampel yang dianalisis [17]. Statistik juga memainkan peran penting dalam penelitian ilmiah dan pengambilan keputusan, karena dengan analisis yang tepat, kita bisa memperoleh informasi yang berguna untuk memecahkan berbagai masalah [18].

Dalam penelitian ini, statistik digunakan untuk menganalisis tingkat kepuasan masyarakat terhadap layanan air PDAM di Kota Rantauprapat. Data yang dikumpulkan melalui survei atau kuesioner akan dianalisis secara statistik untuk mengidentifikasi faktor-faktor utama yang mempengaruhi kepuasan masyarakat.

Melalui penerapan statistik inferensial, kita dapat menentukan apakah ada hubungan signifikan antara kualitas air, distribusi, dan tingkat kepuasan yang dirasakan oleh Masyarakat [19]. Hasil analisis statistik ini nantinya akan memberikan gambaran yang lebih jelas mengenai tingkat

kepuasan, serta membantu dalam merumuskan solusi yang lebih efektif untuk meningkatkan pelayanan PDAM [20].

2.1.5. Pattern Recognition

Pattern recognition adalah proses untuk mengidentifikasi pola atau hubungan yang tersembunyi dalam data yang kompleks [21]. Dalam bidang kecerdasan buatan dan machine learning, pattern recognition berfungsi untuk mengklasifikasikan dan memprediksi pola berdasarkan fitur-fitur yang ada dalam dataset [22]. Algoritma yang digunakan dalam pattern recognition mampu mempelajari data dari pengalaman sebelumnya untuk mengenali pola atau perilaku yang serupa. Salah satu pendekatan yang populer adalah menggunakan teknik-teknik statistik dan probabilistik, seperti Naive Bayes, Support Vector Machines (SVM), atau jaringan syaraf tiruan (neural networks), yang memungkinkan sistem untuk mengenali

pola dan mengklasifikasikan data dengan akurat [23]. Selain itu, aplikasi pattern recognition sangat luas, mulai dari pengenalan wajah, analisis suara, hingga klasifikasi gambar dan teks [24].

Dalam hubungan dengan penelitian ini, pattern recognition digunakan untuk menganalisis tingkat kepuasan masyarakat terhadap kualitas layanan air PDAM. Data yang dikumpulkan dari masyarakat yang menggunakan air PDAM dianalisis untuk menemukan pola-pola yang menunjukkan faktor-faktor yang mempengaruhi kepuasan mereka. Dengan menerapkan metode Naive Bayes, model mampu mengklasifikasikan data berdasarkan karakteristik tertentu, seperti kebersihan air, distribusi yang lancar, dan faktor lainnya yang berhubungan dengan pengalaman pengguna. Melalui pendekatan ini, dapat ditemukan pola-pola tertentu yang menggambarkan kecenderungan tingkat kepuasan atau

ketidakpuasan masyarakat, yang selanjutnya dapat digunakan untuk memberikan rekomendasi perbaikan bagi PDAM dalam meningkatkan pelayanan dan kualitas air yang disalurkan.

2.4. Metode Naïve Bayes

Metode Naive Bayes adalah salah satu teknik dalam machine learning yang sering digunakan untuk klasifikasi data berdasarkan teori probabilitas [25]. Metode ini berasumsi bahwa fitur-fitur yang ada dalam dataset saling independen satu sama lain, meskipun dalam kenyataannya, ini jarang terjadi. Meskipun demikian, Naive Bayes tetap dapat menghasilkan kinerja yang cukup baik, bahkan pada dataset dengan ketergantungan antar fitur [26]. Keunggulan utama dari metode ini adalah kesederhanaannya dalam implementasi dan efisiensinya dalam mengolah dataset besar dengan waktu komputasi yang relatif cepat [27]. Naive

Bayes sering digunakan dalam berbagai aplikasi, seperti klasifikasi email spam, analisis sentimen, dan pengenalan pola. Untuk perhitungan pada metode Naive Bayes menggunakan rumus sebagai berikut [28].

Selanjutnya silahkan dimulai dari sini

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

$P(A|B)$ = Probabilitas A bersyarat yang diberikan oleh B

$P(B|A)$ = Probabilitas B bersyarat yang diberikan oleh A

$P(A)$ = Probabilitas kejadian A

$P(B)$ = Probabilitas kejadian B

Naive Bayes dapat diterapkan untuk menganalisis tingkat kepuasan masyarakat terhadap kualitas air PDAM

berdasarkan data survei yang dikumpulkan dari masyarakat Kota Rantauprapat [29]. Dengan mengklasifikasikan tingkat kepuasan pengguna sebagai variabel target, fitur-fitur seperti kejernihan air, ketersediaan air, bau, dan distribusi dapat digunakan sebagai input untuk model. Melalui model Naive Bayes, dapat dianalisis hubungan antara fitur-fitur tersebut dengan kepuasan pengguna, serta mengidentifikasi pola-pola yang mempengaruhi persepsi masyarakat terhadap layanan air PDAM [30]. Hasil klasifikasi ini akan memberikan wawasan yang lebih terperinci tentang faktor-faktor yang berkontribusi pada tingkat kepuasan masyarakat, serta membantu dalam merancang kebijakan atau perbaikan layanan yang lebih tepat sasaran.

2.5. Model Klasifikasi

Model klasifikasi adalah metode dalam machine learning yang digunakan untuk mengelompokkan data ke

dalam kategori tertentu berdasarkan fitur atau atribut yang dimilikinya [31]. Proses ini melibatkan penggunaan algoritma yang dilatih untuk mempelajari pola dan hubungan antara data input dengan kategori output yang sudah ditentukan [32]. Model klasifikasi dapat diterapkan dalam berbagai jenis masalah, seperti pengenalan pola, deteksi spam, prediksi penyakit, dan banyak lagi. Beberapa algoritma yang sering digunakan dalam model klasifikasi adalah Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, dan

Support Vector Machine (SVM) [33]. Keberhasilan model ini sangat bergantung pada kualitas data yang digunakan serta pemilihan fitur yang relevan untuk memaksimalkan akurasi dalam proses klasifikasi.

Metode Naive Bayes akan digunakan untuk mengklasifikasikan tingkat kepuasan masyarakat terhadap

air PDAM, dengan mempertimbangkan berbagai faktor seperti kualitas air, kelancaran distribusi, dan kondisi infrastruktur. Naive Bayes merupakan algoritma yang berdasarkan pada teorema probabilitas Bayes, yang mengasumsikan bahwa semua fitur bersifat independen. Model ini cocok digunakan untuk mengolah data survei masyarakat, di mana berbagai faktor yang mempengaruhi kepuasan dapat dianalisis secara terpisah untuk menentukan tingkat kepuasan keseluruhan. Dengan mengaplikasikan model klasifikasi Naive Bayes, akan diperoleh hasil yang dapat mengidentifikasi pola-pola yang mendasari kepuasan atau ketidakpuasan masyarakat terhadap layanan air PDAM, yang nantinya dapat menjadi bahan evaluasi untuk perbaikan layanan air.

2.6. Alat Bantu Program Aplikasi Orange

Orange adalah sebuah perangkat lunak open-source yang digunakan untuk analisis data dan visualisasi yang menggabungkan antarmuka grafis dengan kemampuan scripting [34]. Dirancang untuk kemudahan penggunaan, Orange menyediakan berbagai alat untuk analisis data seperti pemodelan prediktif, klasifikasi, regresi, clustering, dan pengolahan data. Dengan berbagai widget yang dapat disambungkan dalam alur kerja yang intuitif, pengguna dapat mengimplementasikan berbagai algoritma machine learning dan teknik analisis data tanpa harus menulis kode secara manual [35]. Orange mendukung banyak format data, seperti file CSV, Excel, dan SQL, serta menawarkan integrasi dengan berbagai pustaka Python untuk analisis lanjutan, menjadikannya alat yang sangat berguna baik untuk pemula maupun ahli data.

Menggunakan Orange dalam menganalisis kepuasan masyarakat terhadap air PDAM memungkinkan untuk melakukan pemodelan prediktif yang akurat melalui berbagai metode klasifikasi, termasuk Naive Bayes [36]. Dengan mengimpor data survei atau hasil wawancara ke dalam Orange, berbagai faktor yang mempengaruhi kepuasan dapat dianalisis secara lebih sistematis. Widget analisis yang ada memungkinkan untuk mengevaluasi faktor-faktor seperti kualitas air, distribusi layanan, dan respon masyarakat, serta mengklasifikasikan tingkat kepuasan berdasarkan data yang ada [37]. Hasil dari proses klasifikasi ini akan memberikan gambaran yang lebih jelas mengenai pola-pola kepuasan masyarakat yang kemudian bisa digunakan untuk merancang kebijakan perbaikan yang lebih tepat sasaran.

2.7. Kelebihan dan Kekurangan Metode Naive Bayes

Metode Naive Bayes memiliki sejumlah kelebihan yang membuatnya populer dalam aplikasi data mining dan klasifikasi. Salah satu keunggulan utamanya adalah kesederhanaannya dalam implementasi dan kemampuannya untuk bekerja dengan cepat meskipun pada dataset yang besar. Naive Bayes juga dapat menghasilkan model yang cukup akurat meskipun dengan asumsi independensi yang sederhana antar fitur, yang sering kali tidak sepenuhnya benar di dunia nyata. Keunggulan lainnya adalah kemampuannya untuk menangani data yang tidak seimbang dengan cukup baik, karena model ini dapat memberikan probabilitas prediksi untuk setiap kelas, sehingga memungkinkan untuk analisis lebih mendalam. Selain itu, Naive Bayes juga efektif dalam menangani data dengan variabel kategorikal dan numerikal, serta relatif mudah diinterpretasikan.

Namun, meskipun memiliki banyak keunggulan, Naive Bayes juga memiliki kekurangan, terutama dalam hal asumsi independensi antar fitur yang sering kali tidak tercapai dalam banyak kasus. Hal ini dapat mengurangi akurasi model ketika ada korelasi kuat antara fitur-fitur yang digunakan. Selain itu, Naive Bayes cenderung kurang efektif bila diterapkan pada dataset dengan banyak fitur atau variabel yang saling berhubungan. Mengingat hal ini, penerapan metode Naive Bayes dalam mengklasifikasikan tingkat kepuasan masyarakat terhadap air PDAM dapat menghadapi tantangan tersendiri, terutama jika terdapat banyak faktor yang saling terkait dalam mempengaruhi kepuasan. Meskipun demikian, dengan data yang relatif terstruktur dan memiliki kategori yang jelas, Naive Bayes masih dapat memberikan hasil yang cukup baik dalam memprediksi kepuasan masyarakat, terutama dalam

memberikan gambaran probabilitas yang berguna bagi pihak pengelola PDAM untuk mengetahui faktor-faktor yang paling berpengaruh.

2.8. Evaluasi Metode Naïve Bayes

Evaluasi metode Naive Bayes menggunakan confusion matrix adalah salah satu cara yang efektif untuk menilai kinerja model klasifikasi. Confusion matrix memberikan gambaran yang jelas tentang bagaimana model memprediksi kelas-kelas yang ada dalam dataset, dengan membandingkan prediksi model dengan label sebenarnya. Terdapat empat elemen utama dalam confusion matrix, yaitu True Positives (TP), False Positives (FP), True Negatives (TN), dan False Negatives (FN). Dari elemen-elemen ini, kita dapat menghitung berbagai metrik penting seperti akurasi, precision, recall, dan F1-score, yang masing-masing memberikan informasi berbeda tentang kinerja model dalam

mengklasifikasikan data. Penggunaan confusion matrix memudahkan untuk mengidentifikasi masalah pada model, seperti ketidakseimbangan kelas yang dapat memengaruhi akurasi dan interpretasi hasil.

		Kelas Prediksi	
		Benar	Salah
Kelas Atribut	Kelas Benar	True Positive (TP)	False Positive (FP)
	Salah	False Negative (FN)	True Negative (TN)

1. TP (True Positive), yaitu jumlah data positif yang memiliki nilai benar.

2. TN (True Negative), yaitu jumlah data negatif yang memiliki nilai benar.
3. FN (False Negative), yaitu jumlah data negatif tetapi yang memiliki nilai salah.
4. FP (False Positive), yaitu jumlah data yang positif tetapi yang memiliki nilai salah.

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \times 100\% \quad [38]$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad [39]$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad [40]$$

Evaluasi dengan menggunakan confusion matrix sangat relevan untuk menilai kinerja metode Naive Bayes dalam mengklasifikasikan tingkat kepuasan masyarakat terhadap air PDAM. Dengan memanfaatkan confusion matrix, dapat dilihat seberapa baik model dalam memprediksi kategori kepuasan masyarakat, baik yang puas maupun yang tidak puas, berdasarkan data survei yang telah

dikumpulkan. Hasil dari confusion matrix ini akan menunjukkan seberapa tepat Naive Bayes dalam mengklasifikasikan data yang ada, serta mengungkapkan apakah model cenderung menghasilkan kesalahan prediksi pada salah satu kategori, seperti mengklasifikasikan orang yang puas sebagai tidak puas atau sebaliknya. Hal ini penting untuk mengukur keandalan model dalam memberikan hasil yang akurat, yang pada gilirannya dapat digunakan untuk perbaikan sistem PDAM atau kebijakan yang lebih tepat sasaran.

2.9. Penelitian Terdahulu

Penelitian terdahulu telah banyak mengkaji penggunaan metode Naive Bayes dalam klasifikasi data, baik dalam konteks analisis sentimen, deteksi spam, maupun

prediksi kepuasan pelanggan. Beberapa penelitian menunjukkan bahwa Naive Bayes mampu memberikan hasil yang memuaskan dalam klasifikasi dengan tingkat akurasi yang cukup tinggi, meskipun terkadang dipengaruhi oleh ketidakseimbangan data. Metode ini juga sering dibandingkan dengan algoritma lain, seperti SVM dan KNN, dalam konteks penelitian yang sama, untuk melihat keunggulan dan keterbatasannya. Hasil-hasil tersebut memberikan landasan yang kuat untuk menerapkan metode ini dalam analisis tingkat kepuasan masyarakat terhadap layanan air PDAM.

Penelitian	1
Judul	Analysis of the Naïve Bayes Method for Determining Social Assistance Eligibility

	Public
Nama	Adinda Pratiwi Siregar1)* , Deci Irmayani2) , Mila Nirmala Sari3)
Tahun	2023
Hasil	Penelitian ini bertujuan untuk mengklasifikasikan data masyarakat yang layak menerima bantuan sosial menggunakan metode Naïve Bayes. Dari 62 data masyarakat yang dikumpulkan, melalui proses pemfilteran dan pra-pemrosesan, hasil klasifikasi menunjukkan 14 data masyarakat layak menerima bantuan sosial, sementara 48 lainnya tidak layak. Hasil ini dapat dijadikan acuan untuk menentukan kelayakan penerima bantuan sosial [41].
Penelitian	2

Judul	Application Of The Naïve Bayes Algorithm In Determining Sales Of The Month
Nama	Hendra Supendar 1)* , Rusdiansyah2), Nining Suharyanti3), Tuslaela4)
Tahun	2023

Hasil	<p>Penelitian ini bertujuan untuk mengklasifikasikan data masyarakat yang layak menerima bantuan sosial menggunakan metode Naïve Bayes. Dari 62 data masyarakat yang dikumpulkan, melalui proses pemfilteran dan pra-pemrosesan, hasil klasifikasi menunjukkan 14 data masyarakat layak menerima bantuan sosial, sementara 48 lainnya tidak layak. Hasil ini dapat dijadikan</p>
-------	--

	acuan untuk menentukan kelayakan penerima bantuan sosial [41].
Penelitian	2
Judul	Application Of The Naïve Bayes Algorithm In Determining Sales Of The Month
Nama	Hendra Supendar 1)* , Rusdiansyah2), Nining Suharyanti3), Tuslaela4)
Tahun	2023
Hasil	Penilaian karyawan dalam mencapai target penjualan setiap bulan sangat penting bagi pertumbuhan perusahaan, namun hal ini menjadi sulit ketika jumlah karyawan sangat banyak. Tanpa adanya otomatisasi data mining,

	<p>manipulasi data bisa terjadi karena pengambilan keputusan yang rumit. Salah satu cara untuk meningkatkan prediksi penjualan adalah dengan memanfaatkan algoritma Naive Bayes, yang menghitung probabilitas berdasarkan atribut seperti absensi, target penjualan, dan retur penjualan. Penelitian ini menunjukkan bahwa penerapan metode Naive Bayes pada data target penjualan barang dapat mencapai tingkat optimasi 95,78%, dengan absensi karyawan yang sangat mempengaruhi kinerja mereka dalam mencapai target penjualan setiap bulan [42].</p>
Penelitian	3
Judul	Implementation of the Naïve Bayes Method to determine the Level of Consumer Satisfaction

Nama	Fitri Febriyani Hasibuan1)*, Muhammad Halmi Dar2), Gomal Juni Yanris3
Tahun	2023
Hasil	Kepuasan adalah perasaan senang yang diperoleh dari barang dan layanan, yang menjadi penilaian penting bagi penjual. Penelitian ini berfokus pada mengukur tingkat kepuasan konsumen dalam berbelanja menggunakan metode Naïve Bayes dengan sampel 49 konsumen. Penelitian ini bertujuan untuk mengidentifikasi faktor yang mempengaruhi kepuasan, di mana beberapa konsumen tidak puas karena kualitas produk yang buruk, sementara yang lain puas dengan produk yang baru dan berkualitas baik. Proses penelitian mencakup analisis data, pra-pemrosesan,

	penerapan algoritma Naïve Bayes, dan uji sistem. Hasil klasifikasi menunjukkan bahwa 47 konsumen
	puas berbelanja, sedangkan 2 konsumen tidak puas, yang menunjukkan bahwa sebagian besar konsumen puas, menandakan bahwa tempat berbelanja tersebut sangat disukai [43].
Penelitian	4
Judul	Digital Forensic Investigates Sexual Harassment on Telegram using Naïve Bayes
Nama	Meyti Eka Apriyani1)*, Rahmad Alfian Maskuri2), Muhammad Hasyim Ratsanjani3), Agung Nugroho Pramudhita4), Rawansyah5)

Tahun	2023
Hasil	<p>Penggunaan Telegram di Indonesia memberikan dampak positif dan negatif, salah satunya adalah menjadi platform bagi kejahatan digital, termasuk pelecehan seksual. Penelitian ini bertujuan untuk mengatasi kebutuhan analisis forensik yang efektif dengan menggunakan metodologi National Institute of Justice (NIJ) dan classifier Naïve Bayes untuk menganalisis percakapan di Telegram. Data yang dianalisis berasal dari percakapan yang mengandung pelecehan seksual, yang dikumpulkan melalui ekstraksi percakapan dan dianalisis menggunakan alat forensik seperti MOBIL edit Forensic Express dan FTK Imager. Hasil pengujian menunjukkan bahwa algoritma</p>

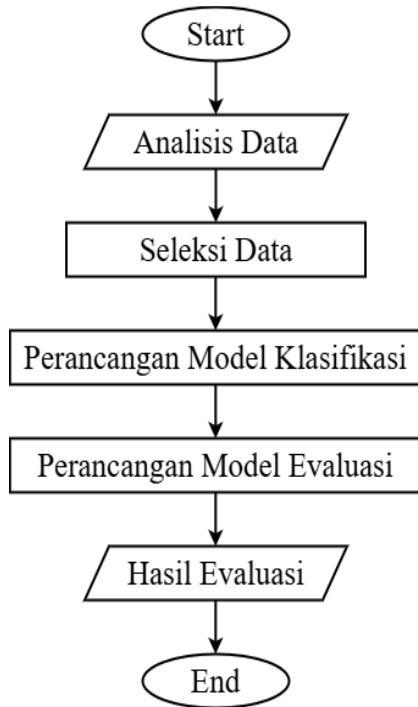
	Naïve Bayes mampu mengklasifikasikan percakapan menjadi positif dan negatif terkait pelecehan seksual, dengan akurasi 91,3%, Precision 100%, dan Recall 90% [44].
--	---

BAB III

ANALISA DAN PERANCANGAN

3.1. Arsitektur Sistem

Arsitektur sistem pada penelitian ini dirancang untuk menganalisis tingkat kepuasan masyarakat terhadap layanan air PDAM dengan menggunakan metode Naïve Bayes dalam machine learning. Metode ini dipilih karena kemampuannya dalam melakukan klasifikasi berdasarkan probabilitas dari masing-masing atribut yang digunakan dalam penelitian. Proses klasifikasi diawali dengan tahap prapemrosesan data untuk membersihkan serta menyiapkan dataset yang akan digunakan dalam pelatihan model. Untuk kerangka kerja yang digunakan yaitu sebagai berikut.



Dalam penelitian ini, terdapat dua dataset yang digunakan untuk membangun dan menguji model klasifikasi. Dataset pertama adalah data training yang terdiri dari 80 data masyarakat, yang digunakan untuk melatih model Naive Bayes agar dapat mengenali pola serta karakteristik utama dari masyarakat yang puas dan tidak puas terhadap layanan

PDAM. Dataset kedua adalah data testing yang berisi 20 data masyarakat, yang berfungsi untuk mengukur kinerja model yang telah dilatih dalam melakukan klasifikasi terhadap data baru. Dengan menggunakan pembagian data ini, penelitian dapat memastikan bahwa model yang dikembangkan memiliki kemampuan generalisasi yang baik dan dapat diterapkan untuk menganalisis kepuasan masyarakat secara lebih luas.

3.2. Pengumpulan Data

Pada tahapan pengumpulan data merupakan tahapan yang dilakukan untuk mengumpulkan data yang akan digunakan pada penelitian ini. Dengan adanya data, maka penelitian ini dapat dilakukan. Untuk data yang akan digunakan pada Penelitian ini terdapat 2 data set yang akan digunakan pada penelitian ini. Untuk data set yang pertama yaitu data training yang digunakan untuk membantu proses

perhitungan data. Untuk data set yang kedua yaitu data testing yang digunakan sebagai sampel Penelitian.

Tabel 3.1 Data Training

Nama	Kualita s Air	Ketersediaa n Air	Harga Air	Kualitas Pelayana n	Kategor i
Adi Prasetyo	Jernih	Cukup	Renda h	Lambat	Puas
Agus Santosa	Jernih	Cukup	Renda h	Cepat	Puas
Aisyah Yuliani	Jernih	Cukup	Renda h	Cepat	Puas
Andi Pratama	Jernih	Cukup	Renda h	Cepat	Puas
Anwar Alamsyah	Jernih	Cukup	Renda h	Cepat	Puas
Arif	Keruh	Tidak Cukup	Renda	Tinggi	Tidak

Susanto			h		Puas
Bella Oktaviani	Jernih	Cukup	Renda h	Cepat	Puas
Benny Mahendra	Jernih	Cukup	Renda h	Cepat	Puas
Bima Lestari	Jernih	Cukup	Renda h	Cepat	Puas
Budi Santoso	Jernih	Tidak Cukup	Renda h	Cepat	Puas
Citra Aryani	Jernih	Cukup	Renda h	Cepat	Puas
Damar Wijaya	Keruh	Tidak Cukup	Tinggi	Cepat	Tidak Puas
Dedi Haryanto	Jernih	Cukup	Tinggi	Cepat	Puas
Denny Hartono	Jernih	Cukup	Renda h	Cepat	Puas

Diah Sari	Jernih	Cukup	Rendah	Cepat	Puas
Dian Pratiwi	Jernih	Cukup	Rendah	Cepat	Puas
Dian Septyani	Keruh	Tidak Cukup	Tinggi	Cepat	Tidak Puas
Dini Widyanti	Jernih	Cukup	Rendah	Cepat	Puas
Edo Nugroho	Jernih	Tidak Cukup	Rendah	Cepat	Puas
Edo Supriyanto	Keruh	Cukup	Tinggi	Tinggi	Tidak Puas
Erni Suryani	Keruh	Cukup	Rendah	Cepat	Puas
Evi Kurniati	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas

Evi Nuraini	Jernih	Cukup	Rendah	Lambat	Puas
Fadil Kurniawan	Keruh	Cukup	Tinggi	Tinggi	Tidak Puas
Faisal Azhari	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Fajar Prabowo	Jernih	Cukup	Rendah	Cepat	Puas
Farah Nurul	Jernih	Tidak Cukup	Rendah	Cepat	Puas
Fauzi Ahnaf	Keruh	Cukup	Tinggi	Tinggi	Tidak Puas
Fenny Munir	Keruh	Tidak Cukup	Rendah	Tinggi	Tidak Puas
Fikri Ali	Jernih	Cukup	Rendah	Cepat	Puas

Gita Amalia	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Gita Rahman	Jernih	Cukup	Rendah	Cepat	Puas
Hafiz Ulya	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Hasan Agung	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Hendra Kurniawan	Jernih	Cukup	Rendah	Cepat	Puas
Imam Rafiq	Keruh	Tidak Cukup	Tinggi	Cepat	Tidak Puas
Imam Yulianto	Jernih	Cukup	Rendah	Cepat	Puas
Indri Kurniawan	Jernih	Cukup	Rendah	Cepat	Puas

ti					
Ira Hidayati	Jernih	Cukup	Rendah	Cepat	Puas
Ira Nurul	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Irfan Maulana	Jernih	Cukup	Tinggi	Cepat	Puas
Joko Susanto	Jernih	Cukup	Tinggi	Cepat	Puas
Junaedi Prakoso	Jernih	Cukup	Rendah	Cepat	Puas
Lestari Wulandari	Jernih	Cukup	Rendah	Cepat	Puas
Lila Puspita	Jernih	Cukup	Tinggi	Cepat	Puas
Lina	Jernih	Cukup	Rendah	Cepat	Puas

Hapsari			h		
Linda Ratnasari	Jernih	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Liza Zahra	Jernih	Cukup	Renda h	Cepat	Puas
Maria Melati	Jernih	Cukup	Renda h	Cepat	Puas
Maya Putri	Jernih	Tidak Cukup	Renda h	Cepat	Puas
Mira Adelia	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Mira Yulita	Jernih	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Nino Rahman	Jernih	Cukup	Renda h	Cepat	Puas
Rahmat Alamsyah	Keruh	Tidak Cukup	Tinggi	Lambat	Tidak Puas

Reni Amelia	Jernih	Cukup	Renda h	Cepat	Puas
Reza Fauzi	Jernih	Cukup	Renda h	Cepat	Puas
Reza Putra	Jernih	Cukup	Renda h	Cepat	Puas
Rina Marlina	Jernih	Cukup	Renda h	Lambat	Puas
Rina Yuliana	Jernih	Cukup	Renda h	Lambat	Puas
Rina Yuniarti	Jernih	Cukup	Renda h	Cepat	Puas
Rizal Fauzan	Jernih	Cukup	Renda h	Cepat	Puas
Rudi Setiawan	Jernih	Cukup	Tinggi	Cepat	Puas
Sandi	Keruh	Cukup	Renda	Cepat	Puas

Kurniawan			h		
Satria Surya	Jernih	Cukup	Rendah	Cepat	Puas
Sella Fitriani	Keruh	Tidak Cukup	Rendah	Tinggi	Tidak Puas
Siti Aminah	Keruh	Cukup	Rendah	Cepat	Puas
Siti Haryani	Jernih	Cukup	Rendah	Cepat	Puas
Siti Marliah	Jernih	Cukup	Rendah	Cepat	Puas
Siti Salamah	Keruh	Tidak Cukup	Rendah	Tinggi	Tidak Puas
Tika Fitria	Jernih	Cukup	Rendah	Cepat	Puas
Tono	Jernih	Cukup	Rendah	Cepat	Puas

Wicaksono			h		
Ulfa Saadiyah	Keruh	Cukup	Tinggi	Tinggi	Tidak Puas
Uli Farida	Jernih	Cukup	Rendah	Cepat	Puas
Vivi Kurniati	Jernih	Cukup	Rendah	Cepat	Puas
Wawan Kusuma	Jernih	Cukup	Rendah	Cepat	Puas
Wina Fariana	Jernih	Tidak Cukup	Rendah	Cepat	Puas
Yadi Purnama	Keruh	Cukup	Tinggi	Tinggi	Tidak Puas
Yudi Saputra	Keruh	Cukup	Rendah	Cepat	Puas
Yulia	Jernih	Cukup	Rendah	Cepat	Puas

Lestari			h		
Zainab Fitri	Jernih	Tidak Cukup	Tinggi	Tinggi	Tidak Puas

Pada table diatas merupakan data yang digunakan sebagai data training nantinya pada Penelitian ini. Untuk data training yang digunakan pada Penelitian ini terdapat 80 data Masyarakat yang digunakan pada Penelitian ini.

Tabel 3.2 Sata Traning

Nama	Kualitas Air	Ketersediaan Air	Harga Air	Kualitas Pelayanan
Aditya Setyawan	Keruh	Tidak Cukup	Tinggi	Tinggi
Arifandi Widiyanto	Jernih	Cukup	Rendah	Cepat
Aulia Maya	Keruh	Cukup	Tinggi	Tinggi
Azizah	Jernih	Cukup	Rendah	Cepat

Kamilah				
Deni Fadli	Keruh	Tidak Cukup	Tinggi	Tinggi
Dian Ariani	Jernih	Cukup	Rendah	Lambat
Dwi Lestari	Jernih	Tidak Cukup	Tinggi	Tinggi
Fadhil Mustopa	Jernih	Cukup	Rendah	Cepat
Feri Mulyadi	Jernih	Cukup	Rendah	Cepat
Fina Anggraini	Keruh	Cukup	Rendah	Cepat
Firdaus Ilham	Jernih	Cukup	Rendah	Cepat
Hani Setyowati	Keruh	Tidak Cukup	Tinggi	Tinggi
Kiki Setiawan	Keruh	Tidak Cukup	Tinggi	Lambat
Luthfi	Jernih	Cukup	Rendah	Cepat

Akbar				
Nabila Intan	Jernih	Tidak Cukup	Rendah	Cepat
Nita Cahyani	Jernih	Cukup	Rendah	Cepat
Rahmad Ginting	Jernih	Cukup	Rendah	Lambat
Susi Pratiwi	Keruh	Tidak Cukup	Tinggi	Tinggi
Titi Puspitasari	Jernih	Cukup	Tinggi	Cepat
Yuliana Pramesti	Jernih	Cukup	Rendah	Cepat

Pada table diatas merupakan data testing yang akan digunakan sebagai data sampel pada Penelitian ini. Untuk data testing yang digunakan sebanyak 20 data Masyarakat.

3.3. Langkah-Langkah Pengolahan pada Metode Naive Bayes

Pada tahapan ini terdapat beberapa tahapan yang akan dilakukan agar proses perhitungan data dapat dilakukan dengan baik dan memperoleh hasil yang kompleks.

1. Seleksi Data

Tada tahapan seleksi data merupakan tahapan yang dilakukan untuk memilih data yang akan digunakan pada penelitian ini. Data yang sudah dikumpulkan akan dimasukkan sebagai data yang akan digunakan pada penelitian ini. Tetapi data yang sudah diperoleh tidak langsung dapat digunakan untuk Penelitian ini.

2. Preprocessing Data

Pada tahapan ini merupakan tahapan yang cukup penting agar penelitian yang dilakukan dapat berjalan dengan baik. Jadi pada proses ini merupakan proses yang dilakukan untuk membersihkan data yang sudah diperoleh sebelumnya. Data yang sudah diperoleh akan dibersihkan

dan dipilih data yang layak untuk digunakan dan data yang tidak layak untuk digunakan. Data yang layak akan digunakan dalam penelitian, dan data yang tidak layak akan dibuang agar tidak memberikan kesalahan pada saat penelitian dilakukan. Pada tahapan ini juga data akan diubah berdasarkan kebutuhan dari penelitian ini. Jadi untuk metode yang digunakan, kebutuhan penelitian untuk data training setiap masing-masing atribut akan dipisahkan di setiap table yang berbeda. Untuk datanya yaitu sebagai berikut.

Atribut	Partisi	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
Kualitas Air	Jernih	53	3	53/57	4/23
	Keruh	4	20	3/57	20/23
	Total	57	23	100%	100%

Atribut	Partisi	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
Harga	Rendah	52	4	52/57	4/23
	Tinggi	5	19	5/57	19/23
Air	Total	57	23	100%	100%

Atribut	Partisi	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
Kualitas	Cepat	53	3	53/57	4/23
	Lambat	4	20	3/57	20/23
Pelayanan	Total	57	23	100%	100%

Kategori		P (Puas) dan P (Tidak Puas)
Puas	57	57/80
Tidak Puas	23	23/80
Total	80	100%

Pada tabel diatas merupakan data disetiap atribut yang ada pada data training. Data pada tabel diatas yang nantinya akan membantu proses perhitungan data.

3. Perhitungan Data

Pada perhitungan metode Naive Bayes terdapat perhitungan yang dilakukan secara manual yaitu sebagai berikut.

Rumus Metode Naïve Bayes

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

Information:

A : hipotesis data A (kelas tertentu)

B : data dengan kelas yang tidak diketahui

$P(A|B)$: Probabilitas hipotesis berdasarkan

kondisi B $P(A)$: Kemungkinan hipotesis A

$P(B|A)$: Probabilitas B ketika kondisi A

P (B) : Probabilitas

Pada perhitungan yang akan dilakukan, pertama penulis akan menghitung data Adi Prasetyo. Adapun perhitungannya sebagai berikut.

$$P(\text{Kategori}) = P(\text{Kualitas Air|Jernih}) \times P(\text{Ketersediaan Air|Cukup}) \times P(\text{Harga Air|Rendah}) \times P(\text{Kualitas Pelayanan|Lambat}) \times$$

P(Kategori|Puas)

$$P(\text{Puas}) = P(\text{Jernih|Puas}) \times P(\text{Cukup|Puas}) \times P(\text{Rendah|Puas}) \times P(\text{Lambat|Puas}) \times P(\text{Kategori|Puas})$$

$$\begin{aligned} & P(\text{Lambat|Puas}) \times P(\text{Kategori|Puas}) \\ = & \left(\frac{53}{57}\right) \times \left(\frac{52}{57}\right) \times \left(\frac{52}{57}\right) \times \left(\frac{3}{57}\right) \times \left(\frac{57}{80}\right) \\ = & 0,029019 \text{ (Nilai Puas)} \end{aligned}$$

$$\begin{aligned} P(\text{Tidak Puas}) = & P(\text{Jernih|Tidak Puas}) \times P(\text{Cukup|Tidak Puas}) \times P(\text{Rendah|Tidak Puas}) \times P(\text{Lambat|Tidak Puas}) \times P(\text{Kategori|Tidak Puas}) \\ = & \left(\frac{4}{23}\right) \times \left(\frac{5}{23}\right) \times \left(\frac{4}{23}\right) \times \left(\frac{4}{23}\right) \times \left(\frac{23}{80}\right) \\ = & 0,000328 \text{ (Nilai Tidak Puas)} \end{aligned}$$

Hasil yang diperoleh dari data Adi Prasetyo bahwasanya nilai puas yang diperoleh yaitu sebesar 0,029019 dan untuk nilai tidak puas yang diperoleh yaitu sebesar 0,000328. Jadi dari hasil yang sudah diperoleh bahwa Adi Prasetyo puas pada PDAM Rantauprapat.

4. Evaluasi Metode

Evaluasi metode Naive Bayes pada penelitian ini menggunakan Confusion Matrix. Hal ini dilakukan agar nantinya dapat diperoleh hasil akurasi presisi dan recall dari Confusion Matrix. Hasil tersebut yang nantinya akan memberikan nilai efektivitas dari metode yang digunakan.

		Predicted		Σ
		Puas	Tidak Puas	
Tidak	Puas	13	0	
	Tidak Puas	7	1	6
Σ		14	6	20

Hasil True Positive (TP) adalah 13. True Negative (TN) adalah 6, False Positive (FP) adalah 0 dan False Negative (FN) adalah 1. Maka Nilai akurasi, presisi dan recall adalah sebagai berikut:

$$\begin{aligned}
 &= 95\% \quad \frac{13+6}{13+6+0+1} \text{ Accuracy} = \quad \times 100\% \quad \text{Then the Accuracy value} \\
 \text{Presisi} &\quad \frac{13}{13+0} = \times 100\% \quad \text{Then the Precision value} = 100\% \text{ Recall} = \times \\
 &100\% \quad \text{Then the Recall value} = 92\% \\
 &\quad \frac{13}{13+1}
 \end{aligned}$$

Hasil evaluasi yang diperoleh dari metode Naive Bayes bahwa metode ini cocok digunakan untuk Penelitian ini. Walaupun hasil yang diperoleh belum sempurna. Tetapi hasil yang diperoleh sudah bagus, hal ini karena hasil yang diberikan hamper mencapai 100%.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisa Data

Analisis data pada penelitian ini dilakukan dengan memanfaatkan beberapa atribut penting, yaitu kualitas air, ketersediaan air, harga air, dan kualitas pelayanan, yang semuanya diperoleh melalui data survei masyarakat. Setiap atribut dianalisis untuk melihat pengaruhnya terhadap tingkat kepuasan masyarakat terhadap layanan PDAM, di mana data terlebih dahulu dibersihkan dan diproses agar siap digunakan dalam pelatihan model machine learning menggunakan metode Naïve Bayes. Melalui analisis ini, pola-pola hubungan antaratribut dapat diidentifikasi, seperti apakah kualitas air yang baik selalu berkorelasi dengan kepuasan tinggi atau apakah harga air yang dianggap mahal menjadi faktor utama ketidakpuasan. Hasil analisis

kemudian digunakan untuk membangun model klasifikasi yang mampu memprediksi tingkat kepuasan masyarakat secara akurat berdasarkan kombinasi dari keempat atribut tersebut.

4.2. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan mengambil data dari PDAM yang berkaitan dengan tingkat kepuasan masyarakat terhadap layanan yang diberikan. Data yang berhasil dikumpulkan sebanyak 80 data, yang mencakup informasi penting seperti kualitas air, ketersediaan air, harga air, dan kualitas pelayanan yang dirasakan masyarakat. Setiap data mewakili persepsi dan pengalaman masyarakat dalam menggunakan layanan PDAM, sehingga memberikan gambaran yang cukup representatif untuk dianalisis lebih lanjut. Data ini kemudian digunakan sebagai dasar dalam membangun model

klasifikasi dengan metode Naïve Bayes untuk mengidentifikasi faktor-faktor yang memengaruhi tingkat kepuasan maupun ketidakpuasan pelanggan.

Tabel 4. 1. Data Sampel Penelitian

Nama	Kualitas Air	Ketersediaan Air	Harga Air	Kualitas Pelayanan
Adi Prasetyo	Jernih	Cukup	Rendah	Lambat
Agus Santosa	Jernih	Cukup	Rendah	Cepat
Aisyah Yuliani	Jernih	Cukup	Rendah	Cepat
Andi Pratama	Jernih	Cukup	Rendah	Cepat
Anwar Alamsyah	Jernih	Cukup	Rendah	Cepat
Arif Susanto	Keruh	Tidak Cukup	Rendah	Tinggi
Bella Oktaviani	Jernih	Cukup	Rendah	Cepat
Benny Mahendra	Jernih	Cukup	Rendah	Cepat

Bima Lestari	Jernih	Cukup	Rendah	Cepat
Budi Santoso	Jernih	Tidak Cukup	Rendah	Cepat
Citra Aryani	Jernih	Cukup	Rendah	Cepat
Damar Wijaya	Keruh	Tidak Cukup	Tinggi	Cepat
Dedi Haryanto	Jernih	Cukup	Tinggi	Cepat
Denny Hartono	Jernih	Cukup	Rendah	Cepat
Diah Sari	Jernih	Cukup	Rendah	Cepat
Dian Pratiwi	Jernih	Cukup	Rendah	Cepat
Dian Septyani	Keruh	Tidak Cukup	Tinggi	Cepat
Dini Widyanti	Jernih	Cukup	Rendah	Cepat
Edo Nugroho	Jernih	Tidak Cukup	Rendah	Cepat
Edo Supriyanto	Keruh	Cukup	Tinggi	Tinggi
Erni Suryani	Keruh	Cukup	Rendah	Cepat
Evi Kurniati	Keruh	Tidak Cukup	Tinggi	Tinggi
Evi Nuraini	Jernih	Cukup	Rendah	Lambat

Fadil Kurniawan	Keruh	Cukup	Tinggi	Tinggi
Faisal Azhari	Keruh	Tidak Cukup	Tinggi	Tinggi
Fajar Prabowo	Jernih	Cukup	Rendah	Cepat
Farah Nurul	Jernih	Tidak Cukup	Rendah	Cepat
Fauzi Ahnaf	Keruh	Cukup	Tinggi	Tinggi
Fenny Munir	Keruh	Tidak Cukup	Rendah	Tinggi
Fikri Ali	Jernih	Cukup	Rendah	Cepat
Gita Amalia	Keruh	Tidak Cukup	Tinggi	Tinggi
Gita Rahman	Jernih	Cukup	Rendah	Cepat

Pada tabel di atas merupakan tabel data yang diperoleh dari data PDAM, yang berisi informasi mengenai nama masyarakat beserta atribut-atribut penting seperti kualitas air, ketersediaan air, harga air, dan kualitas pelayanan yang mereka rasakan. Jumlah data yang berhasil dikumpulkan sebanyak 80 data, yang kemudian akan digunakan sebagai

bahan analisis untuk mengetahui tingkat kepuasan masyarakat terhadap layanan PDAM. Data ini mencerminkan variasi pengalaman masyarakat, mulai dari kondisi air yang jernih atau keruh, ketersediaan air yang cukup atau tidak, harga air yang rendah atau tinggi, hingga kualitas pelayanan yang cepat, lambat, atau tinggi. Data tersebut menjadi dasar penting untuk diolah menggunakan metode machine learning agar dapat mengidentifikasi pola kepuasan masyarakat secara lebih mendalam.

4.3. Pembagian Data

Pada tahapan pembagian data, dataset yang telah diperoleh dibagi menjadi dua bagian, yaitu data training dan data testing. Data training digunakan untuk melatih model machine learning agar dapat mengenali pola dari data, sedangkan data testing digunakan untuk menguji dan

mengevaluasi seberapa baik model yang telah dilatih dalam melakukan prediksi atau klasifikasi pada data baru.

Tabel 4. 2. Data Training

Nama	Kualita s Air	Ketersediaa n Air	Harga Air	Kualitas Pelayana n	Kategor i
Citra Aryani	Keruh	Cukup	Renda h	Cepat	Puas
Damar Wijaya	Keruh	Tidak Cukup	Tinggi	Cepat	Tidak Puas
Dini Widyanti	Jernih	Tidak Cukup	Renda h	Cepat	Puas
Edo Supriyant o	Keruh	Cukup	Tinggi	Tinggi	Tidak Puas
Fadil	Keruh	Cukup	Tinggi	Tinggi	Tidak

Kurniawan					Puas
Indra	Keruh	Cukup	Rendah	Cepat	Puas
Linda Ratnasari	Jernih	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Liza Zahra	Jernih	Cukup	Tinggi	Cepat	Puas
Maria Melati	Jernih	Cukup	Rendah	Lambat	Puas
Rahmat Alamsyah	Keruh	Tidak Cukup	Tinggi	Lambat	Tidak Puas
Reza Putra	Keruh	Cukup	Rendah	Cepat	Puas
Rina Yuniarti	Jernih	Tidak Cukup	Rendah	Cepat	Puas

Pada tabel di atas merupakan data training penelitian yang digunakan sebagai data latih untuk membangun model machine learning dalam penelitian ini, dengan total sebanyak 20 data. Data training ini mencakup atribut seperti kualitas air, ketersediaan air, harga air, kualitas pelayanan, serta kategori kepuasan masyarakat, yang semuanya akan membantu algoritma dalam mempelajari pola dan hubungan antarvariabel untuk menghasilkan prediksi yang akurat pada tahap pengujian.

Tabel 4. 3. Data Testing

Nama	Kualitas Air	Ketersediaan Air	Harga Air	Kualitas Pelayanan
Adi Prasetyo	Jernih	Cukup	Rendah	Lambat
Agus Santosa	Jernih	Cukup	Rendah	Cepat
Aisyah	Jernih	Cukup	Rendah	Cepat

Yuliani				
Andi Pratama	Jernih	Cukup	Rendah	Cepat
Anwar Alamsyah	Jernih	Cukup	Rendah	Cepat
Arif Susanto	Keruh	Tidak Cukup	Rendah	Tinggi
Bella Oktaviani	Jernih	Cukup	Rendah	Cepat
Benny Mahendra	Jernih	Cukup	Rendah	Cepat
Bima Lestari	Jernih	Cukup	Rendah	Cepat
Budi Santoso	Jernih	Tidak Cukup	Rendah	Cepat
Dedi Haryanto	Jernih	Cukup	Tinggi	Cepat
Denny Hartono	Jernih	Cukup	Rendah	Cepat
Diah Sari	Jernih	Cukup	Rendah	Cepat

Dian Pratiwi	Jernih	Cukup	Rendah	Cepat
Dian Septyani	Keruh	Tidak Cukup	Tinggi	Cepat
Edo Nugroho	Jernih	Tidak Cukup	Rendah	Cepat
Erni Suryani	Keruh	Cukup	Rendah	Cepat
Evi Kurniati	Keruh	Tidak Cukup	Tinggi	Tinggi
Evi Nuraini	Jernih	Cukup	Rendah	Lambat
Faisal Azhari	Keruh	Tidak Cukup	Tinggi	Tinggi
Fajar Prabowo	Jernih	Cukup	Rendah	Cepat
Farah Nurul	Jernih	Tidak Cukup	Rendah	Cepat
Fauzi Ahnaf	Keruh	Cukup	Tinggi	Tinggi
Fenny Munir	Keruh	Tidak Cukup	Rendah	Tinggi
Fikri Ali	Jernih	Cukup	Rendah	Cepat
Gita Amalia	Keruh	Tidak Cukup	Tinggi	Tinggi
Gita Rahman	Jernih	Cukup	Rendah	Cepat

Hafiz Ulya	Keruh	Tidak Cukup	Tinggi	Tinggi
Hasan Agung	Keruh	Tidak Cukup	Tinggi	Tinggi
Hendra Kurniawan	Jernih	Cukup	Rendah	Cepat
Imam Rafiq	Keruh	Tidak Cukup	Tinggi	Cepat
Imam Yulianto	Jernih	Cukup	Rendah	Cepat
Indri Kurniawati	Jernih	Cukup	Rendah	Cepat
Ira Hidayati	Jernih	Cukup	Rendah	Cepat
Ira Nurul	Keruh	Tidak Cukup	Tinggi	Tinggi
Irfan Maulana	Jernih	Cukup	Tinggi	Cepat
Joko Susanto	Jernih	Cukup	Tinggi	Cepat
Junaedi Prakoso	Jernih	Cukup	Rendah	Cepat

Lestari Wulandari	Jernih	Cukup	Rendah	Cepat
Lila Puspita	Jernih	Cukup	Tinggi	Cepat
Lina Hapsari	Jernih	Cukup	Rendah	Cepat
Maya Putri	Jernih	Tidak Cukup	Rendah	Cepat
Mira Adelia	Keruh	Tidak Cukup	Tinggi	Tinggi
Mira Yulita	Jernih	Tidak Cukup	Tinggi	Tinggi
Nino Rahman	Jernih	Cukup	Rendah	Cepat
Reni Amelia	Jernih	Cukup	Rendah	Cepat
Reza Fauzi	Jernih	Cukup	Rendah	Cepat
Rina Marlina	Jernih	Cukup	Rendah	Lambat
Rina Yuliana	Jernih	Cukup	Rendah	Lambat
Rizal Fauzan	Jernih	Cukup	Rendah	Cepat
Rudi Setiawan	Jernih	Cukup	Tinggi	Cepat
Sandi	Keruh	Cukup	Rendah	Cepat

Kurniawan				
Siti Aminah	Keruh	Cukup	Rendah	Cepat
Siti Marliah	Jernih	Cukup	Rendah	Cepat
Tika Fitria	Jernih	Cukup	Rendah	Cepat
Ulfa Saadiah	Keruh	Cukup	Tinggi	Tinggi
Vivi Kurniati	Jernih	Cukup	Rendah	Cepat
Wina Fariana	Jernih	Tidak Cukup	Rendah	Cepat
Yudi Saputra	Keruh	Cukup	Rendah	Cepat
Yulia Lestari	Jernih	Cukup	Rendah	Cepat

Pada tabel di atas merupakan data testing atau data uji yang akan digunakan sebagai sampel penelitian, dengan jumlah sebanyak 60 data. Data testing ini mencakup atribut-atribut penting seperti kualitas air, ketersediaan air, harga air, dan kualitas pelayanan, yang nantinya akan digunakan untuk menguji performa model yang telah dilatih sebelumnya. Tujuan penggunaan data testing ini adalah untuk

mengevaluasi sejauh mana model mampu memprediksi atau mengklasifikasikan data baru dengan tingkat akurasi yang baik, sehingga dapat diketahui efektivitas dan keandalan model dalam penelitian ini.

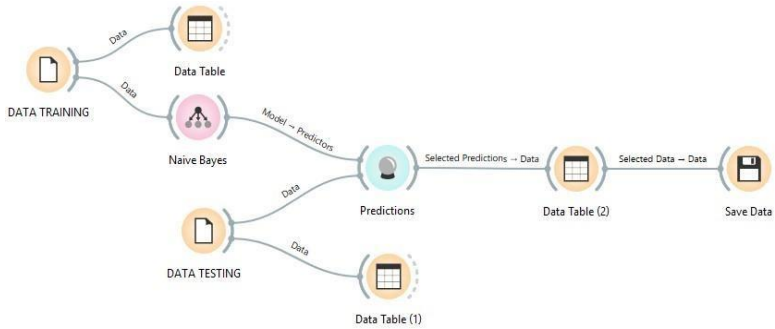
4.4. Pembersihan Data

Pada tahapan pembersihan data, dilakukan proses untuk memastikan bahwa data yang digunakan dalam penelitian benar-benar bersih, konsisten, dan bebas dari kesalahan yang dapat memengaruhi hasil analisis. Pembersihan data mencakup pengecekan terhadap adanya duplikasi, data yang hilang (missing value), kesalahan penulisan, atau inkonsistensi dalam format data, seperti perbedaan penulisan kategori atau angka. Jika ditemukan data yang tidak relevan atau tidak sesuai, maka data tersebut akan diperbaiki, disesuaikan, atau bahkan dihapus agar tidak mengganggu kualitas model. Tahapan ini penting agar

dataset yang dihasilkan menjadi rapi dan siap untuk diproses lebih lanjut, sehingga hasil penelitian dapat lebih akurat dan dapat diandalkan.

4.5. Perancangan Model Klasifikasi

Pada penelitian ini, perancangan model klasifikasi dilakukan dengan menyusun algoritma yang mampu memproses data atribut seperti kualitas air, ketersediaan air, harga air, dan kualitas pelayanan untuk memprediksi tingkat kepuasan pelanggan. Model dirancang agar dapat belajar dari data training dan diuji menggunakan data testing, sehingga mampu menghasilkan pola klasifikasi yang akurat dalam menentukan kategori puas atau tidak puas berdasarkan pola-pola yang teridentifikasi dari data



Gambar 4. 1. Perancangan Model Klasifikasi

Pada gambar di atas merupakan perancangan model klasifikasi yang dirancang di aplikasi Orange. Prosesnya dimulai dari pemisahan data menjadi data training dan data testing yang masing-masing dimasukkan melalui node Data Table. Data training digunakan untuk membangun model klasifikasi menggunakan algoritma Naive Bayes. Selanjutnya, model ini diterapkan pada data testing melalui node Predictions untuk menghasilkan prediksi kategori. Hasil prediksi ditampilkan melalui Data Table dan kemudian dapat disimpan menggunakan node Save Data. Rangkaian

ini menggambarkan alur lengkap mulai dari pelatihan model, pengujian, hingga penyimpanan hasil prediksi.

4.6. Hasil Klasifikasi

Hasil klasifikasi akan dipaparkan dalam bentuk tabel yang menampilkan data testing beserta prediksi kategori dari model Naive Bayes. Tabel ini memuat kolom seperti nama, atribut-atribut input, kategori aktual (jika ada), serta hasil prediksi dari model klasifikasi.

Tabel 4. 4. Hasil Klasifikasi

Nama	Kualita s Air	Ketersediaa n Air	Harg a Air	Kualitas Pelayana n	Katego ri
Adi Prasetyo	Jernih	Cukup	Rendah	Lambat	Puas

Agus Santosa	Jernih	Cukup	Rendah	Cepat	Puas
Aisyah Yuliani	Jernih	Cukup	Rendah	Cepat	Puas
Andi Pratama	Jernih	Cukup	Rendah	Cepat	Puas
Anwar Alamsyah	Jernih	Cukup	Rendah	Cepat	Puas
Arif Susanto	Keruh	Tidak Cukup	Rendah	Tinggi	Tidak Puas
Bella Oktaviani	Jernih	Cukup	Rendah	Cepat	Puas
Benny Mahendra	Jernih	Cukup	Rendah	Cepat	Puas
Bima Lestari	Jernih	Cukup	Rendah	Cepat	Puas
Budi	Jernih	Tidak	Rendah	Cepat	Puas

Santoso		Cukup			
Dedi Haryanto	Jernih	Cukup	Tinggi	Cepat	Puas
Denny Hartono	Jernih	Cukup	Rendah	Cepat	Puas
Diah Sari	Jernih	Cukup	Rendah	Cepat	Puas
Dian Pratiwi	Jernih	Cukup	Rendah	Cepat	Puas
Dian Septyani	Keruh	Tidak Cukup	Tinggi	Cepat	Tidak Puas
Edo Nugroho	Jernih	Tidak Cukup	Rendah	Cepat	Puas
Erni Suryani	Keruh	Cukup	Rendah	Cepat	Puas
Evi Kurniati	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Evi	Jernih	Cukup	Rendah	Lambat	Puas

Nuraini					
Faisal Azhari	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Fajar Prabowo	Jernih	Cukup	Rendah	Cepat	Puas
Farah Nurul	Jernih	Tidak Cukup	Rendah	Cepat	Puas
Fauzi Ahnaf	Keruh	Cukup	Tinggi	Tinggi	Tidak Puas
Fenny Munir	Keruh	Tidak Cukup	Rendah	Tinggi	Tidak Puas
Fikri Ali	Jernih	Cukup	Rendah	Cepat	Puas
Gita Amalia	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Gita Rahman	Jernih	Cukup	Rendah	Cepat	Puas
Hafiz	Keruh	Tidak	Tinggi	Tinggi	Tidak

Ulya		Cukup			Puas
Hasan Agung	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Hendra Kurniawan	Jernih	Cukup	Rendah	Cepat	Puas
Imam Rafiq	Keruh	Tidak Cukup	Tinggi	Cepat	Tidak Puas
Imam Yulianto	Jernih	Cukup	Rendah	Cepat	Puas
Indri Kurniawati	Jernih	Cukup	Rendah	Cepat	Puas
Ira Hidayati	Jernih	Cukup	Rendah	Cepat	Puas
Ira Nurul	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas

Irfan Maulana	Jernih	Cukup	Tinggi	Cepat	Puas
Joko Susanto	Jernih	Cukup	Tinggi	Cepat	Puas
Junaedi Prakoso	Jernih	Cukup	Rendah	Cepat	Puas
Lestari Wulandari	Jernih	Cukup	Rendah	Cepat	Puas
Lila Puspita	Jernih	Cukup	Tinggi	Cepat	Puas
Lina Hapsari	Jernih	Cukup	Rendah	Cepat	Puas
Maya Putri	Jernih	Tidak Cukup	Rendah	Cepat	Puas
Mira Adelia	Keruh	Tidak Cukup	Tinggi	Tinggi	Tidak Puas

Mira Yulita	Jernih	Tidak Cukup	Tinggi	Tinggi	Tidak Puas
Nino Rahman	Jernih	Cukup	Rendah	Cepat	Puas
Reni Amelia	Jernih	Cukup	Rendah	Cepat	Puas
Reza Fauzi	Jernih	Cukup	Rendah	Cepat	Puas
Rina Marlina	Jernih	Cukup	Rendah	Lambat	Puas
Rina Yuliana	Jernih	Cukup	Rendah	Lambat	Puas
Rizal Fauzan	Jernih	Cukup	Rendah	Cepat	Puas
Rudi Setiawan	Jernih	Cukup	Tinggi	Cepat	Puas
Sandi	Keruh	Cukup	Rendah	Cepat	Puas

Kurniawan					
Siti Aminah	Keruh	Cukup	Rendah	Cepat	Puas
Siti Marliah	Jernih	Cukup	Rendah	Cepat	Puas
Tika Fitria	Jernih	Cukup	Rendah	Cepat	Puas
Ulfa Saadiyah	Keruh	Cukup	Tinggi	Tinggi	Tidak Puas
Vivi Kurniati	Jernih	Cukup	Rendah	Cepat	Puas
Wina Fariana	Jernih	Tidak Cukup	Rendah	Cepat	Puas
Yudi Saputra	Keruh	Cukup	Rendah	Cepat	Puas
Yulia	Jernih	Cukup	Rendah	Cepat	Puas

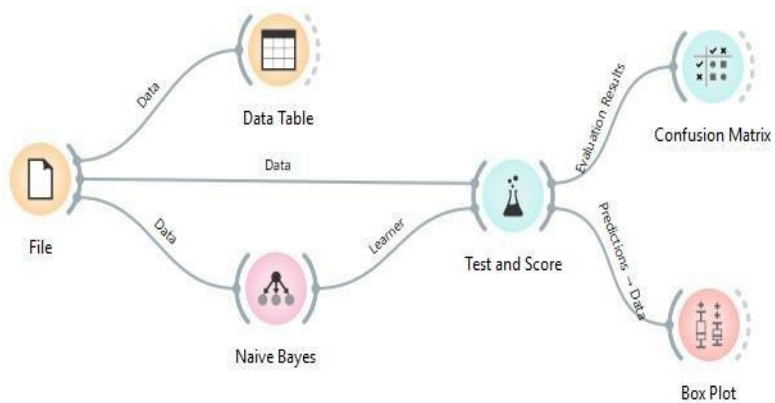
Lestari					
---------	--	--	--	--	--

Pada tabel di atas merupakan hasil klasifikasi yang sudah dilakukan sebelumnya dengan total sebanyak 60 data pelanggan yang digunakan sebagai sampel penelitian. Berdasarkan hasil klasifikasi tersebut, diperoleh bahwa sebanyak 46 pelanggan termasuk dalam kategori Puas, sedangkan 14 pelanggan lainnya termasuk dalam kategori Tidak Puas. Hal ini menunjukkan bahwa sebagian besar pelanggan merasa puas terhadap layanan yang diberikan, terutama pada aspek kualitas air, ketersediaan, harga, dan kecepatan pelayanan. Sementara itu, kelompok pelanggan yang merasa tidak puas umumnya disebabkan oleh kombinasi faktor seperti kualitas air yang keruh, ketersediaan air yang tidak cukup, harga yang tinggi, dan kualitas pelayanan yang kurang memadai. Temuan ini

penting untuk menjadi bahan evaluasi bagi penyedia layanan agar dapat meningkatkan kualitas pada aspek- aspek yang masih dirasakan kurang oleh sebagian pelanggan sehingga tingkat kepuasan secara keseluruhan dapat meningkat.

4.7. Perancangan Model Evaluasi

Pada perancangan ini, dibuat model evaluasi untuk mengukur sejauh mana kemampuan metode yang digunakan dalam melakukan klasifikasi. Evaluasi ini penting untuk menilai akurasi, presisi, dan efektivitas model dalam memprediksi kategori dengan benar.



Gambar 4. 2. Perancangan Model Evaluasi

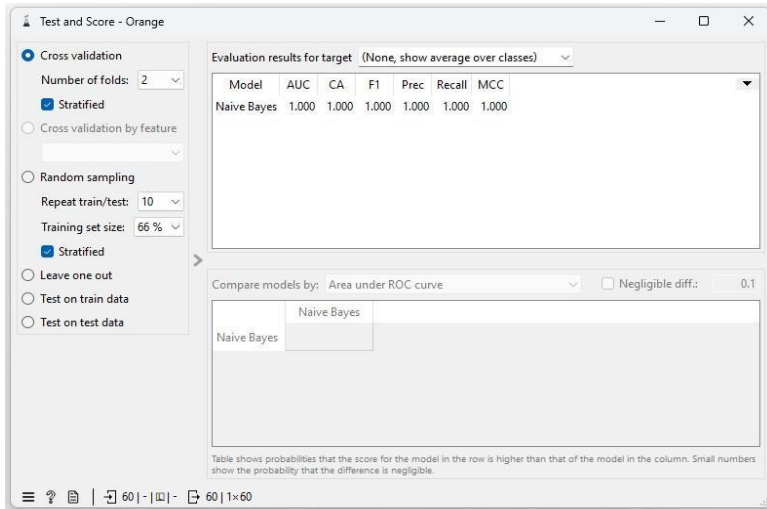
Pada gambar di atas merupakan model evaluasi yang dirancang pada aplikasi Orange untuk mengukur performa metode Naive Bayes dalam melakukan klasifikasi data secara menyeluruh. Proses ini diawali dengan memasukkan data melalui node File yang diteruskan ke Naive Bayes sebagai algoritma pembelajaran, kemudian diuji menggunakan Test and Score untuk menghitung metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Hasil evaluasi kemudian divisualisasikan melalui Confusion Matrix untuk melihat jumlah prediksi benar dan salah, serta Box Plot untuk menampilkan distribusi performa model secara grafis. Perancangan ini penting dilakukan untuk memastikan bahwa model yang dibangun mampu memberikan hasil klasifikasi yang akurat dan dapat diandalkan pada data yang diuji.

4.8. Hasil Evaluasi

Hasil evaluasi menunjukkan bahwa nilai akurasi dari Test and Score mencapai tingkat yang baik, menunjukkan kemampuan model dalam mengklasifikasi data dengan cukup tepat. Dari Confusion Matrix, terlihat bahwa sebagian besar data berhasil diprediksi dengan benar, meskipun terdapat sedikit kesalahan klasifikasi pada beberapa kasus.

4.8.1. Hasil Test and Score

Test and Score adalah proses evaluasi model yang digunakan untuk mengukur seberapa baik performa model klasifikasi dalam memprediksi data. Pada tahap ini, akan diperoleh metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk menilai kemampuan model secara menyeluruh.



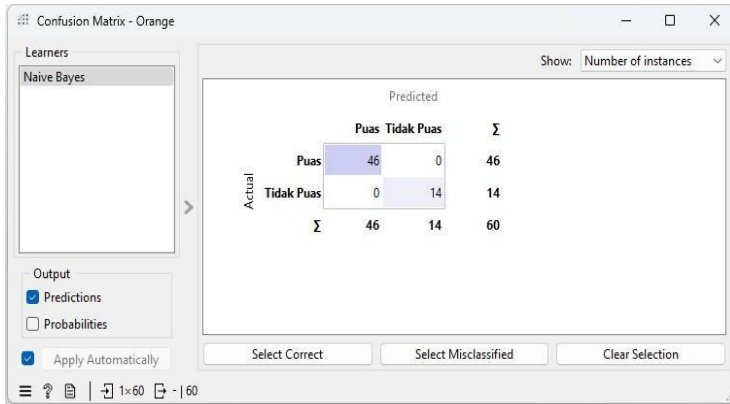
Gambar 4. 3. Hasil Evaluasi Test and Score

Pada gambar di atas merupakan hasil Test and Score dari metode Naive Bayes yang dirancang menggunakan aplikasi Orange. Hasil evaluasi menunjukkan bahwa nilai AUC (Area Under Curve), CA (Classification Accuracy), F1-score, Precision (Prec), Recall, dan MCC (Matthews Correlation Coefficient) semuanya mencapai angka sempurna yaitu 1.000 atau setara dengan 100%. Hal ini menunjukkan bahwa model Naive Bayes mampu

mengklasifikasikan data dengan sangat baik tanpa kesalahan, dengan tingkat akurasi, presisi, dan recall yang optimal pada data yang digunakan.

4.8.2. Hasil Confusion Matrix

Confusion Matrix adalah tabel yang digunakan untuk mengevaluasi performa model klasifikasi dengan menampilkan jumlah prediksi benar dan salah yang dilakukan oleh model pada setiap kelas. Matriks ini membantu mengidentifikasi kesalahan klasifikasi seperti false positive dan false negative sehingga memudahkan analisis akurasi dan metrik lainnya.



Gambar 4. 4. Hasil Evaluasi Confusion Matrix

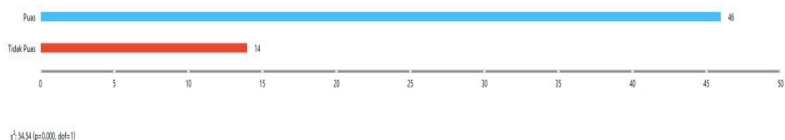
Pada gambar di atas ditampilkan hasil evaluasi model klasifikasi dalam bentuk Confusion Matrix untuk dua kelas, yaitu "Puas" dan "Tidak Puas". Dari total 60 data yang diuji, model berhasil mengklasifikasikan seluruh data dengan sangat baik, di mana 46 data yang termasuk kelas "Puas" terdeteksi dengan benar tanpa ada kesalahan, demikian juga 14 data kelas "Tidak Puas" berhasil diprediksi secara tepat tanpa satu pun kesalahan klasifikasi. Tidak terdapat false positive maupun false negative, yang berarti model tidak

salah mengkategorikan data manapun. Dengan kondisi ini, akurasi model dapat dihitung sebagai persentase jumlah prediksi benar dibandingkan total data, yaitu $((46 + 14) / 60) \times 100\% = 100\%$. Hasil ini menunjukkan bahwa model klasifikasi yang digunakan memiliki performa yang sangat optimal dan mampu membedakan kedua kelas dengan akurasi sempurna. Evaluasi Confusion Matrix ini sangat penting karena selain memberikan gambaran akurasi, juga memungkinkan analisis terhadap jenis kesalahan yang terjadi, namun dalam kasus ini tidak ada kesalahan yang ditemukan sehingga model sangat dapat diandalkan dalam pengambilan keputusan berdasarkan data tersebut.

4.8.3. Hasil Box Plot

Box Plot adalah grafik yang digunakan untuk menggambarkan sebaran data dengan menampilkan nilai minimum, kuartil pertama, median, kuartil ketiga, dan nilai

maksimum. Grafik ini memudahkan dalam mengidentifikasi penyebaran, simetri, dan adanya outlier dalam suatu kumpulan data.



Gambar 4. 5. Hasil Evaluasi Box Plot

Pada gambar diatas merupakan hasil evaluasi menggunakan Box Plot yang menunjukkan distribusi dua kategori kepuasan pelanggan, yaitu "Puas" dan "Tidak Puas". Balok berwarna biru menggambarkan kelompok pelanggan yang merasa puas dengan jumlah 46 orang, sedangkan balok berwarna merah mewakili kelompok yang tidak puas sebanyak 14 orang. Box Plot ini memperlihatkan sebaran data masing-masing kelompok serta membantu dalam memahami perbedaan dan variasi tingkat kepuasan pelanggan secara visual. Dengan demikian, grafik ini sangat

berguna untuk menganalisis karakteristik data kepuasan dan memudahkan dalam pengambilan keputusan berdasarkan distribusi tersebut

DAFTAR PUSTAKA

- [1] R. A. Pangestu and S. Noris, “Analisa Data Mining Prediksi Lelang Suku Cadang Dengan Metode K-NearestNeighbor (Studi Kasus PT. Parmud Jaya Perkasa),” *J. Inform. Multi*, vol. 1, no. 4, pp. 285–295, 2023.
- [2] Y. Pratiwi and M. Mulyawan, “Implementasi Algoritma K-Means untuk Menentukan Angka Harapan Hidup berdasarkan Tingkat Provinsi,” *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 284–294, 2023, doi: 10.56211/blendsains.v1i4.233.
- [3] T. Hartati, O. Nurdiawan, and E. Wiyandi, “Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi MaritimSuaka Bahari,” *J. Sains Teknol. Transp. Marit.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.51578/j.sitektransmar.v3i1.30.
- [4] S. A. A. Kharis, A. H. A. Zili, A. Putri, and A. Robiansyah, “Analisis Tren Minat Masyarakat Indonesia terhadap Artificial Intelligence dalam Menyongsong Society 5.0: Studi Menggunakan Google Trends,” *G-Tech J. Teknol. Terap.*, vol. 7, no. 4, pp. 1345–1354, 2023, doi: 10.33379/gtech.v7i4.3091.
- [5] Y. Lakhdari, E. Soldevila, J. Rezgui, and É. Renault, “Detection of Plant Diseases in an Industrial Greenhouse:

- Development, Validation & Exploitation,” 2023 Int. Symp. Networks, Comput. Commun. ISNCC 2023, 2023, doi: 10.1109/ISNCC58260.2023.10323932.
- [6] A. C. Study, “applied sciences Data Mining Techniques for Early Diagnosis of Diabetes :,” pp. 1–12, 2021.
- [7] F. Paquin, J. Rivnay, A. Salleo, N. Stingelin, and C. Silva, “Multi-phase semicrystalline microstructures drive exciton dissociation in neat plastic semiconductors,” *J. Mater. Chem. C*, vol. 3, pp. 10715–10722, 2015, doi: 10.1039/b000000x.
- [8] M. M. Arcinas, G. S. Sajja, S. Asif, S. Gour, E. Okoronkwo, and M. Naved, “Role of Data Mining in Education for Improving Students Performance for Social Change,” *Turkish J. Physiother. Rehabil.*, vol. 32, no. 3, pp. 6519–6526, 2021.
- [9] E. Fauzi, M. Visar Sinatrya, N. Daru Ramdhani, R. Ramadhan, and Z. Muhammad Rasid Safari, “Pengaruh kemajuan teknologi informasi terhadap perkembangan akuntansi,” *J. Ris. Pendidik. Ekon.*, vol. 7, no. 2, pp. 189–197, 2022, doi: 10.21067/jrpe.v7i2.6877.
- [10] M. I. Maliki, “Rancang Bangun Aplikasi Penjualan Grosir Sembako Pada Toko LA-RIS,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 3, pp. 304–311, 2021, doi: 10.33365/jatika.v2i3.1222.

- [11] D. Pakpahan, V. Siallagan, and S. Siregar, "Classification of E-Commerce Product Descriptions with The Tf-Idf and Svm Methods," *Sinkron*, vol. 8, no. 4, pp. 2130–2137, 2023, doi: 10.33395/sinkron.v8i4.12779.
- [12] N. Khasawneh, E. Faouri, and M. Fraiwan, "Automatic Detection of Tomato Diseases Using Deep Transfer Learning," *Appl. Sci.*, vol. 12, no. 17, 2022, doi: 10.3390/app12178467.
- [13] A. Sucipto, Q. J. Adrian, and M. A. Kencono, "Martial Art Augmented Reality Book (Arbook) Sebagai Media Pembelajaran Seni Beladiri Nusantara Pencak Silat," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 1, pp. 40–45, 2021, doi: 10.32736/sisfokom.v10i1.983.
- [14] S. I. Nurhafida and F. Sembiring, "ANALISIS TEXT CLUSTERING MASYARAKAT DI TWITER MENGENAI MCDONALD ' SXBTS MENGGUNAKAN ORANGE DATA MINING," pp. 28–35, 2021.
- [15] T. S. Rambe, M. N. S. Hasibuan, and M. H. Dar, "Sentiment Analysis of Beauty Product Applications using the Naïve Bayes Method," *SinkrOn*, vol. 8, no. 2, pp. 980–989, 2023, doi: 10.33395/sinkron.v8i2.12303.
- [16] I. maulana, H. Rizqi Sanjaya, F. Setiyansyah, D. Righel Wibowo, and F. Sinlae, "Sistem Operasi Pada Komputer Yang Paling Banyak Digunakan," *J. Pengabd.*

- Multidisiplin, vol. 2, pp. 9–17, 2024, [Online]. Available: <https://ejournal.cvrobema.com/index.php/aremben/article/view/49>
- [17] D. Montanesa and R. Ahmad, “Hubungan persepsi siswa tentang KIP kuliah dengan keputusan melanjutkan pendidikan ke perguruan tinggi,” *J. Educ. J. Pendidik. Indones.*, vol. 9, no. 1, p. 292, 2023, doi: 10.29210/1202322305.
- [18] S. Maryam, E. Bu’ulolo, and E. Hatmi, “Penerapan Metode Fuzzy Mamdani dan Fuzzy Tsukamoto Dalam Menentukan Harga Mobil Bekas,” *J. Informatics, Electr. Electron. Eng.*, vol. 1, no. 1, pp. 10–14, 2021, [Online]. Available: <https://djournals.com/jieeee/article/view/54%0Ahttps://djournals.com/jieeee/article/download/54/164>
- [19] H. Rusyada and M. Nasir, “Efektivitas Penerapan Hybrid Learning Pasca Pandemi Covid-19 di Sekolah Dasar,” *J. Basicedu*, vol. 6, no. 2, pp. 1714–1723, 2022, doi: 10.31004/basicedu.v6i2.2275.
- [20] A. Nursalma and H. Pujiastuti, “Pengaruh Waktu Belajar Dan Motivasi Belajar Terhadap Hasil Belajar Matematika,” *OMEGA J. Keilmuan Pendidik. Mat.*, vol. 2, no. 3, pp. 135–141, 2023, doi: 10.47662/jkpm.v2i3.479.

- [21] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [22] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma KMeans," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.
- [23] P. K. Samanta and N. K. Rout, "Convolutional Neural Network Using Convolutional Neural Network," *Springer*, vol. 2644, no. 2, pp. 747–749, 2016, [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4842-2845-6_6
- [24] Y. Lin, Y. Gou, Z. Liu, B. Li, J. Lv, and X. Peng, "COMPLETER: Incomplete Multi-view Clustering via Contrastive Prediction," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 11169–11178, 2021, doi: 10.1109/CVPR46437.2021.01102.
- [25] M. Nabeel, S. Majeed, M. J. Awan, H. Muslih-Ud-din, M. Wasique, and R. Nasir, "Review on effective disease prediction through data mining techniques," *Int. J. Electr. Eng. Informatics*, vol. 13, no. 3, pp. 717–733, 2021, doi: 10.15676/IJEEL.2021.13.3.13.

- [26] K. Anam, B. Nurhakim, and C. Juliane, “Komparasi Algoritma Klasifikasi Data Mining Menggunakan Optimize Selection untuk Peminatan Program Studi,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 2, pp. 606–613, 2022, doi: 10.47065/bits.v4i2.2160.
- [27] R. N. Juliadi and Y. Puspitarani, “Supervised Model for Sentiment Analysis Based on Hotel Review Clusters using RapidMiner,” *Sinkron*, vol. 7, no. 3, pp. 1059–1066, 2022, doi: 10.33395/sinkron.v7i3.11564.
- [28] M. Farid Naufal, A. Fernando Susanto, C. Nathaneil Kansil, S. Huda, and K. kunci, “Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Potensi Hilangnya Nasabah Bank Application of Machine Learning to Predict Potential Loss of Bank Customer,” *Februari*, vol. 22, no. 1, pp. 1– 11, 2023.
- [29] L. J. Muhammad, M. M. Islam, S. S. Usman, and S. I. Ayon, “Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients’ Recovery,” *SN Comput. Sci.*, vol. 1, no. 4, pp. 1–7, 2020, doi: 10.1007/s42979-020-00216-w.
- [30] Y. A. Singgalan, “Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Systematic Literature Review,” *J. Inf. Syst. Informatics*, vol. 3, no. 2, pp. 278–302, 2021, doi: 10.33557/journalisi.v3i2.125.

- [31] I. Lishania, R. Goejantoro, and Y. N. Nasution, "Perbandingan Klasifikasi Metode Naive Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda," *J. Eksponensial*, vol. 10, no. 2, pp. 135–142, 2019, [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/571>
- [32] S. Al Faraby and A. Romadhony, "Pengaruh Distribusi Panjang Data Teks pada Klasifikasi: Sebuah Studi Awal," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1501, 2022, doi: 10.30865/mib.v6i3.4259.
- [33] N. Hadianto, H. B. Novitasari, and A. Rahmawati, "Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 163–170, 2019, doi: 10.33480/pilar.v15i2.658.
- [34] A. D. Wibisono, S. Dadi Rizkiono, and A. Wantoro, "Filtering Spam Email Menggunakan Metode Naive Bayes," *TELEFORTECH J. Telemat. Inf. Technol.*, vol. 1, no. 1, 2020, doi: 10.33365/tft.v1i1.685.
- [35] D. A. Rabbani and F. U. Najicha, "Pengaruh Perkembangan Teknologi terhadap Kehidupan dan Interaksi Sosial Masyarakat Indonesia," *Researchgate.Net*, no. November, pp. 0–13, 2023, [Online].

- Available: https://www.researchgate.net/profile/Dana-Rabbani/publication/375525102_Pengaruh_Perkembangan_Teknologi_terhadap_Kehidupan_dan_Interaksi_Sosial_Masyarakat_Indonesia/links/654dc8dce88b87031d8db65/Pengaruh-Perkembangan-Teknologi-terhadap-Kehidupan-dan-Inte
- [36] K. D. Sisbintari and F. A. Setiawati, "Digital Parenting sebagai Upaya Mencegah Kecanduan Gadget pada Anak Usia Dini saat Pandemi Covid19," *Z J. Obs. J. Pendidik. Anak Usia Dini*, vol. 6, no. 3, pp. 1562–1575, 2021, doi: 10.31004/obsesi.v6i3.1781.
- [37] D. T. L. Sari, Khamdun, and M. A. Fardani, "Peran Orang Tua Dalam Mengontrol Penggunaan Gadget Pada Anak Usia SD di Desa Soco Kecamatan Dawe Kabupaten Kudus," *J. Pendidik. dan Konseling*, vol. 4, no. 4, pp. 1188–1194, 2022, [Online]. Available: <https://core.ac.uk/download/pdf/322599509.pdf>
- [38] P. Kamath, P. Patil, S. S, Sushma, and S. S, "Crop yield forecasting using data mining," *Glob. Transitions Proc.*, vol. 2, no. 2, pp. 402–407, 2021, doi: 10.1016/j.gltp.2021.08.008.
- [39] M. R. A. Yudianto, K. Kusriani, and H. Al Fatta, "Analisis Pengaruh Tingkat Akurasi Klasifikasi Citra Wayang dengan Algoritma Convolutional Neural Network," *J.*

- Teknol. Inf., vol. 4, no. 2, pp. 182–191, 2020, doi: 10.36294/jurti.v4i2.1319.
- [40] R. Rahman and F. Fauzi Abdulloh, “Performance of Various Naïve Bayes Using GridSearch Approach In Phishing Email Dataset,” *Sinkron*, vol. 8, no. 4, pp. 2336–2344, 2023, doi: 10.33395/sinkron.v8i4.12958.
- [41] A. P. Siregar, D. Irmayani, and M. N. Sari, “Analysis of the Naïve Bayes Method for Determining Social Assistance Eligibility Public,” *SinkrOn*, vol. 8, no. 2, pp. 805–817, 2023, doi: 10.33395/sinkron.v8i2.12259.
- [42] H. Supendar, R. Rusdiansyah, N. Suharyanti, and T. Tuslaela, “Application of the Naïve Bayes Algorithm in Determining Sales Of The Month,” *SinkrOn*, vol. 8, no. 2, pp. 873–879, 2023, doi: 10.33395/sinkron.v8i2.12293.
- [43] F. F. Hasibuan, M. H. Dar, and G. J. Yanris, “Implementation of the Naïve Bayes Method to determine the Level of Consumer Satisfaction,” *SinkrOn*, vol. 8, no. 2, pp. 1000–1011, 2023, doi: 10.33395/sinkron.v8i2.12349.
- [44] M. E. Apriyani, R. A. Maskuri, M. H. Ratsanjani, A. N. Pramudhita, and R. Rawansyah, “Digital Forensic Investigates Sexual Harassment on Telegram using Naïve Bayes,” *Sinkron*, vol. 8, no. 3, pp. 1409–1417, 2023, doi: 10.33395/sinkron.v8i3.12514.

TENTANG PENULIS



Fahri Anugrah Nasution

Lahir di Kota Aek Kanopan, Kabupaten Labuhanbatu Utara, pada tanggal 12 Mei 2003. Menempuh Pendidikan Sekolah Dasar di SDN 112139 Rantau Prapat, Kemudian Sekolah Menengah Pertama Di SMPN 3 Rantau utara,

Selanjutnya Sekolah Menengah Atas di SMAN 1 Rantau Utara, Sebelum akhirnya melanjutkan studi di jenjang perguruan tinggi. Pernah menjadi bagian dari mahasiswa SI Sistem Informasi di Universitas Labuhanbatu Sejak 2021-2025.



Ibnu Rasyid Munthe, S.T., M.Kom

Lahir di Kota Ranau 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 Magister Komputer (M.Kom). Beliau aktif dalam menulis artikel ilmiah serta buku akademik. Saa ini,

beliau berkiprah sebagai Dosen di Universitas Labuhanbau. Selain itu, beliau juga mengelola Kana YouTube @manjaddaajadda2022 sebagai media pembelajaran dan sarana berbagi ilmu pengeahuan.



Masrizal, S.kom., M.Kom.

NIDN 0105039401. Lahir Di Manggung 05 Maret 1994 Proses Dalam Menempuh Pendidikan dimulai Dari Sekolah Dasar SDN 13 Manggung, SMP.N 7 Pariaman, SMA N. 4 Pariaman Dan Melanjutit Peguruan Tinggi

Swasta dengan. Jurusan Sisem Informasi S1 (Sarjana) di Universitas Putra Indonesia “YPTK”Padang Tahun Lulusan 2016 Dan Melanjuti Magister Tahun 2016 Universitas Putra Indonesia “YPTK”Padang. Tahun Lulusan 2018 Dengan Jurusan Sisem Informasi, Fakultas Teknologi Informasi. Pada Tahun 2018. Saya Mengabdikan Diri Sebagai Salah Satu Dosen Dibidang Ilmu Komputer di Fakultas Sains dan Teknologi (FST) Prodi Sistem Informasi di Universias Labuhanbatu salah sau menjadi Dosen tetap diKampus tersebut. Syukur Allhamdulillah sekarang saya mengabdikan diri dibawah Wakil Rektor III diBidang Pengembangan, Kerja sama dan Alumni sebagai Ka.biro Bagian

Pengembangan , Kerjasama dan Alumnis serta Devisi Tracer Studi
Di Universitas Labuhanbatu.



Marnis Nasution, S.Kom., M.Kom

Lahir di Bengkulu 30 Maret 1990. Selama sekolah dasar sampai menengah di tempuh di kota Bengkulu. Melanjutkan Pendidikan tinggi strata-1 dan stratta-2 di Universitas Putra Indonesia “YPTK” Padang dari ahun

2008 sampai 2024 dengan jurusam Sisem Informasi. Saat ini aktif menjadi Dosen Yayan di Universitas Labuhanbau, Sumatera Utara dan menulis beberapa Karya Ilmiah dan Buku.

Buku ini membahas penerapan metode Naïve Bayes dalam mengklasifikasikan tingkat kepuasan pelanggan terhadap layanan PDAM di Kota Rantauprapat. Dengan pendekatan machine learning, penulis menganalisis data berdasarkan kualitas air, ketersediaan, harga, dan pelayanan. Hasil klasifikasi menunjukkan bahwa metode ini mampu memberikan prediksi akurat mengenai kepuasan masyarakat. Penelitian dilakukan menggunakan aplikasi Orange, sehingga proses visualisasi dan pengolahan data menjadi lebih efektif dan praktis. Buku ini cocok bagi mahasiswa, peneliti, dan praktisi yang ingin memahami bagaimana teknologi dapat digunakan untuk meningkatkan kualitas layanan publik. Dengan bahasa yang aplikatif dan berbasis studi kasus nyata, buku ini menjadi panduan strategis dalam mengolah data pelanggan secara cerdas dan efisien.

