



**PENERBIT
YAYASAN MMI**

MONOGRAP

ALGORITMA DECISION TREE UNTUK KLASIFIKASI TINGKAT KEPUASAN PELANGGAN

P E N U L I S

RIAMA PAKPAHAN
DECI IRMAYANI
VOLVO SIHOMBING

ALGORITMA DECISION TREE UNTUK KLASIFIKASI TINGKAT KEPUASAN PELANGGAN



**PENERBIT
YAYASAN MAM**

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).

ALGORITMA DECISION TREE UNTUK KLASIFIKASI TINGKAT KEPUASAN PELANGGAN

Riama Pakpahan;

Deci Irmayani;

Volvo Sihombing;



**PENERBIT
YAYASAN MMI**

ALGORITMA DECISION TREE UNTUK KLASIFIKASI TINGKAT KEPUASAN PELANGGAN

Riama Pakpahan; Deci Irmayani; Volvo Sihombing;

Desain Cover :
Musthafa Haris Munandar

Sumber :
<https://munandar.yayasanmmi.com/algoritma-decision-tree>

Tata Letak :
Deci Irmayani

Proofreader :
Volvo Sihombing

Ukuran :
Jml hal judul: 1, Jml hal isi naskah: 76, Uk: 15x23 cm

ISBN :
978-634-04-2403-4

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

PRAKATA

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga buku ini yang berjudul “Algoritma Decision Tree untuk Klasifikasi Tingkat Kepuasan Pelanggan” dapat disusun dan diselesaikan dengan baik. Buku ini hadir sebagai kontribusi dalam bidang data mining dan ilmu komputer terapan, khususnya dalam menjawab tantangan analisis data kepuasan pelanggan menggunakan pendekatan algoritma klasifikasi berbasis decision tree.

Dalam era digital saat ini, data telah menjadi aset yang sangat berharga. Kemampuan untuk menggali informasi tersembunyi dari data pelanggan dapat menjadi kunci keunggulan kompetitif bagi berbagai organisasi dan perusahaan. Melalui buku ini, penulis berusaha menjelaskan konsep, teknik, serta penerapan praktis dari algoritma decision tree seperti ID3, C4.5, dan CART dalam konteks klasifikasi tingkat kepuasan pelanggan. Penjelasan disertai dengan studi kasus sederhana agar pembaca dapat memahami tidak hanya teori, tetapi juga implementasi nyata dalam proses pengambilan keputusan berbasis data.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan buku ini. Oleh karena itu, kritik dan saran yang membangun dari para pembaca sangat diharapkan untuk penyempurnaan di masa yang akan datang.

Akhir kata, semoga buku ini dapat memberikan manfaat dan menjadi referensi yang berguna dalam pengembangan pengetahuan serta praktik pengolahan data untuk mendukung pengambilan keputusan yang lebih cerdas dan terarah.

Hormat Kami,

Penulis

DAFTAR ISI

PRAKATA	vi
DAFTAR ISI	vii
BAB 1: KONSEP DASAR KEPUASAN PELANGGAN	1
1.1 Pengertian dan Pentingnya Kepuasan Pelanggan	2
1.2 Faktor-faktor yang Mempengaruhi Kepuasan	6
1.3 Metode Pengukuran Kepuasan Pelanggan	11
1.4 Peran Data dalam Menganalisis Kepuasan Pelanggan	17
 BAB 2: PENGANTAR DATA MINING DAN KLASIFIKASI	 23
2.1 Definisi dan Tujuan Data Mining	24
2.2 Teknik-Teknik Data Mining	29
2.3 Klasifikasi sebagai Teknik Analitik	34
2.4 Algoritma Populer untuk Klasifikasi	39
2.5 Proses CRISP-DM dalam Data Mining	46
 BAB 3: ALGORITMA DECISION TREE	 53
3.1 Apa itu Decision Tree	53
3.2 Struktur dan Komponen Decision Tree	54
3.3 Algoritma ID3, C4.5, dan CART: Perbandingan dan Kegunaan	55
3.4 Keunggulan dan Kelemahan Decision Tree	56
3.5 Studi Kasus Sederhana dengan Decision Tree	57

BAB 4: PERSIAPAN DATA UNTUK KLASIFIKASI KEPUASAN	59
4.1 Desain Kuesioner dan Pengumpulan Data	59
4.2 Pembersihan dan Normalisasi Data	60
4.3 Seleksi Fitur dan Penentuan Target Variabel	61
4.4 Representasi Data dalam Bentuk Tabel dan CSV	63
DAFTAR PUSTAKA	64
BIOGRAFI PENULIS	66
SINOPSIS	68

BAB I

KONSEP DASAR KEPUASAN PELANGGAN

Kepuasan pelanggan merupakan elemen krusial dalam keberlangsungan dan kesuksesan suatu bisnis. Dalam era persaingan yang semakin ketat, perusahaan tidak hanya dituntut untuk menawarkan produk atau layanan yang berkualitas, tetapi juga harus mampu memahami dan memenuhi ekspektasi pelanggan secara menyeluruh. Kepuasan pelanggan mencerminkan sejauh mana produk atau layanan yang diberikan mampu memenuhi harapan pelanggan. Tingkat kepuasan ini secara langsung memengaruhi loyalitas pelanggan, citra merek, dan pertumbuhan bisnis jangka panjang.

Secara konseptual, kepuasan pelanggan diartikan sebagai perbandingan antara harapan pelanggan terhadap suatu produk atau layanan dengan persepsi mereka setelah mengalaminya. Jika persepsi pelanggan lebih tinggi dari harapan, maka pelanggan akan merasa sangat puas. Sebaliknya, jika kenyataan tidak sesuai harapan, maka pelanggan akan merasa kecewa. Oleh karena itu, pengukuran kepuasan pelanggan menjadi penting untuk dilakukan secara berkala sebagai bahan evaluasi dan peningkatan kualitas layanan.

Pengumpulan data kepuasan biasanya dilakukan melalui survei atau kuesioner yang dirancang untuk menangkap opini dan pengalaman pelanggan. Hasil dari survei ini dapat dianalisis menggunakan pendekatan statistik maupun teknik kecerdasan buatan, salah satunya adalah klasifikasi menggunakan algoritma Decision Tree. Algoritma ini memungkinkan perusahaan untuk mengelompokkan pelanggan berdasarkan tingkat kepuasannya, serta mengidentifikasi faktor-faktor dominan yang memengaruhinya.

Dengan mengklasifikasikan tingkat kepuasan pelanggan, perusahaan dapat lebih cepat merespons masalah layanan, menyesuaikan strategi pemasaran, dan meningkatkan kualitas produk atau layanan secara tepat sasaran. Oleh karena itu, penggabungan konsep kepuasan pelanggan dengan pendekatan analitik berbasis algoritma klasifikasi menjadi langkah strategis yang bernilai tinggi dalam pengambilan keputusan bisnis yang berbasis data.

1.1 Pengertian dan Pentingnya Kepuasan Pelanggan

Kepuasan pelanggan merupakan konsep fundamental dalam manajemen bisnis dan pemasaran modern. Istilah ini merujuk pada perasaan senang atau kecewa seseorang yang muncul setelah membandingkan antara harapan sebelumnya terhadap suatu produk atau layanan dengan kenyataan atau pengalaman aktual yang dirasakan setelah konsumsi. Jika produk atau layanan mampu memenuhi atau bahkan melampaui ekspektasi pelanggan, maka pelanggan akan merasa puas. Sebaliknya, jika harapan tidak terpenuhi, maka ketidakpuasan dapat terjadi, bahkan mengarah pada kehilangan loyalitas pelanggan.

Menurut Philip Kotler, seorang pakar pemasaran terkemuka, kepuasan pelanggan adalah "tingkat perasaan seseorang setelah membandingkan kinerja (atau hasil) yang dirasakan terhadap harapannya." Dalam hal ini, ekspektasi pelanggan bisa berasal dari pengalaman sebelumnya, rekomendasi dari orang lain, citra merek, atau janji-janji promosi yang disampaikan oleh perusahaan. Artinya, persepsi pelanggan tidak bersifat mutlak, melainkan relatif dan dipengaruhi oleh berbagai faktor psikologis dan sosial.

Konsep kepuasan pelanggan tidak hanya berlaku untuk perusahaan barang konsumsi, tetapi juga sangat relevan dalam sektor jasa, pemerintahan, pendidikan, dan pelayanan publik. Di

semua sektor tersebut, kualitas hubungan antara penyedia layanan dan penerima manfaat sangat dipengaruhi oleh seberapa besar pihak pemberi layanan mampu memahami dan memenuhi kebutuhan pengguna.

Dimensi Kepuasan Pelanggan

Untuk mengukur dan memahami kepuasan pelanggan secara lebih mendalam, para ahli telah mengembangkan beberapa dimensi atau aspek utama yang memengaruhi tingkat kepuasan, antara lain:

1. **Kualitas Produk atau Layanan:** Apakah produk bekerja sesuai dengan fungsi yang dijanjikan? Apakah layanan diberikan secara profesional, cepat, dan ramah?
2. **Harga:** Apakah harga yang ditawarkan sesuai dengan manfaat dan nilai yang diterima pelanggan?
3. **Pelayanan Purna Jual (after-sales service):** Apakah perusahaan menyediakan dukungan teknis, garansi, atau bantuan jika terjadi masalah setelah pembelian?
4. **Kemudahan Akses dan Proses Pembelian:** Apakah proses pembelian mudah, cepat, dan nyaman?
5. **Pengalaman Emosional:** Apakah pelanggan merasa dihargai, didengar, dan dilayani dengan baik?

Pemahaman terhadap dimensi-dimensi tersebut sangat penting untuk membangun sistem pengukuran yang objektif, sekaligus menjadi dasar dalam menyusun strategi peningkatan kualitas layanan atau produk.

Pentingnya Kepuasan Pelanggan

Kepuasan pelanggan memiliki dampak besar terhadap keberlangsungan dan pertumbuhan suatu bisnis. Berikut adalah beberapa alasan utama mengapa kepuasan pelanggan sangat penting:

1. Meningkatkan Loyalitas Pelanggan

Pelanggan yang puas cenderung akan kembali membeli produk atau menggunakan layanan dari perusahaan yang sama. Mereka merasa nyaman dan percaya terhadap kualitas yang ditawarkan. Loyalitas pelanggan ini merupakan aset berharga, karena biaya mempertahankan pelanggan yang sudah ada jauh lebih rendah dibandingkan dengan biaya untuk menarik pelanggan baru.

2. Mendorong Word-of-Mouth Positif

Pelanggan yang puas seringkali menjadi promotor alami bagi perusahaan. Mereka akan merekomendasikan produk atau layanan kepada teman, keluarga, atau rekan kerja. Dalam era digital dan media sosial, satu ulasan positif bisa menjangkau ribuan orang dalam waktu singkat. Sebaliknya, satu pengalaman buruk juga dapat menyebar luas dan merusak reputasi bisnis.

3. Meningkatkan Pendapatan dan Profitabilitas

Riset menunjukkan bahwa perusahaan yang secara konsisten menjaga tingkat kepuasan pelanggan yang tinggi cenderung memiliki kinerja keuangan yang lebih baik. Pelanggan yang puas bersedia membayar lebih untuk produk atau layanan yang mereka percaya. Mereka juga cenderung membeli lebih banyak, lebih sering, dan lebih lama.

4. Menjadi Umpan Balik untuk Inovasi

Umpan balik dari pelanggan—baik dalam bentuk pujian maupun keluhan—merupakan sumber informasi berharga bagi perusahaan. Dari sini, perusahaan dapat mengetahui kelemahan sistem, kekurangan produk, atau peluang untuk menciptakan solusi baru. Kepuasan pelanggan dapat menjadi indikator seberapa efektif inovasi yang telah dilakukan perusahaan.

5. Membangun Reputasi dan Citra Positif

Citra perusahaan sangat dipengaruhi oleh pengalaman pelanggan. Jika perusahaan dikenal karena memperlakukan pelanggannya dengan baik dan konsisten memberikan nilai lebih, maka reputasi perusahaan akan meningkat. Reputasi positif ini akan memberikan keunggulan kompetitif, terutama dalam pasar yang kompetitif dan sensitif terhadap persepsi publik.

6. Mengurangi Keluhan dan Biaya Operasional

Perusahaan yang secara aktif menjaga kepuasan pelanggan akan lebih jarang menerima keluhan atau komplain. Ini bukan hanya mengurangi tekanan pada tim layanan pelanggan, tetapi juga menekan biaya-biaya tidak langsung seperti pengembalian barang, perbaikan layanan, atau kerusakan merek akibat penyebaran ulasan negatif.

Kepuasan Pelanggan di Era Digital

Perkembangan teknologi digital telah mengubah cara pelanggan berinteraksi dengan bisnis. Saat ini, pelanggan memiliki akses lebih luas terhadap informasi, ulasan pengguna lain, dan berbagai alternatif produk. Mereka juga dapat menyampaikan keluhan atau pujian secara terbuka melalui platform media sosial. Hal ini menuntut perusahaan untuk lebih responsif, transparan, dan proaktif dalam menjaga pengalaman pelanggan.

Di sisi lain, kemajuan teknologi juga memberikan peluang bagi perusahaan untuk mengelola kepuasan pelanggan secara lebih efektif. Dengan bantuan sistem Customer Relationship Management (CRM), analitik data, dan algoritma klasifikasi seperti decision tree, perusahaan dapat memahami perilaku pelanggan, memprediksi tingkat kepuasan, serta menyesuaikan strategi layanan secara lebih personal.

Indikator dan Pengukuran Kepuasan

Untuk memastikan kepuasan pelanggan terukur dengan baik, banyak perusahaan menggunakan instrumen seperti:

- **Survei Kepuasan Pelanggan (Customer Satisfaction Survey):** Umumnya dilakukan melalui kuesioner yang mencakup berbagai dimensi pelayanan.
- **Net Promoter Score (NPS):** Mengukur sejauh mana pelanggan bersedia merekomendasikan produk/layanan kepada orang lain.
- **Customer Effort Score (CES):** Menilai seberapa mudah pelanggan mendapatkan layanan atau menyelesaikan masalah.
- **Customer Retention Rate dan Churn Rate:** Indikator loyalitas pelanggan dari waktu ke waktu.

Hasil dari pengukuran ini dapat digunakan sebagai dasar dalam menganalisis faktor-faktor yang memengaruhi kepuasan, yang selanjutnya dapat diproses lebih lanjut menggunakan teknik data mining, seperti klasifikasi decision tree.

1.2 Sejarah perkembangan komputer dan arsitekturnya

Kepuasan pelanggan merupakan hasil dari interaksi kompleks antara harapan dan persepsi mereka terhadap kualitas produk atau layanan yang diberikan oleh perusahaan. Tingkat kepuasan ini tidak muncul begitu saja, melainkan dipengaruhi oleh berbagai faktor yang saling berhubungan. Memahami faktor-faktor tersebut sangat penting bagi organisasi untuk mengelola kualitas layanan secara proaktif dan menyusun strategi peningkatan yang tepat sasaran. Dalam konteks bisnis modern yang sangat kompetitif, mengidentifikasi dan mengoptimalkan faktor-faktor ini menjadi salah satu kunci keberhasilan

perusahaan dalam mempertahankan loyalitas pelanggan dan membangun keunggulan kompetitif.

Berikut adalah faktor-faktor utama yang memengaruhi kepuasan pelanggan:

1. Kualitas Produk atau Layanan

Faktor yang paling mendasar dalam memengaruhi kepuasan pelanggan adalah kualitas dari produk atau layanan itu sendiri. Kualitas mencakup sejauh mana produk atau layanan tersebut mampu memenuhi kebutuhan, keinginan, dan harapan pelanggan. Dalam konteks barang, hal ini bisa berarti ketahanan, performa, estetika, dan kemudahan penggunaan. Dalam konteks layanan, kualitas mencakup kecepatan layanan, keakuratan informasi, sikap ramah staf, dan kemampuan penyelesaian masalah.

Produk atau layanan dengan kualitas tinggi cenderung memberikan pengalaman yang menyenangkan bagi pelanggan, yang pada akhirnya meningkatkan tingkat kepuasan. Sebaliknya, kualitas yang buruk atau tidak konsisten akan menyebabkan kekecewaan dan penurunan kepercayaan pelanggan terhadap merek.

2. Harga

Harga merupakan salah satu variabel penting yang memengaruhi persepsi pelanggan terhadap nilai yang mereka terima. Harga yang dianggap wajar dan sesuai dengan kualitas produk atau layanan akan menciptakan kepuasan. Sebaliknya, harga yang terlalu tinggi tanpa disertai peningkatan nilai akan menyebabkan ketidakpuasan. Pelanggan umumnya melakukan evaluasi terhadap manfaat yang diperoleh dibandingkan dengan biaya yang mereka keluarkan, yang dikenal sebagai konsep *value for money*.

Perusahaan harus cermat dalam menetapkan harga, karena harga yang terlalu murah bisa diasosiasikan dengan kualitas rendah, sedangkan harga yang terlalu tinggi bisa menghalangi keputusan pembelian. Strategi harga yang transparan, kompetitif, dan konsisten merupakan salah satu kunci penting dalam menjaga kepuasan pelanggan.

3. Pelayanan Pelanggan (Customer Service)

Pelayanan pelanggan berperan penting dalam membentuk kesan dan persepsi pelanggan terhadap perusahaan. Sikap ramah, kemampuan komunikasi yang baik, kesigapan dalam menangani masalah, serta kepedulian terhadap kebutuhan pelanggan merupakan aspek-aspek krusial dalam pelayanan. Pelanggan yang merasa diperlakukan dengan baik akan merasa dihargai dan cenderung lebih puas terhadap pengalaman mereka.

Sebaliknya, pelayanan yang buruk—seperti sikap kasar, lambat, atau kurang informatif—dapat merusak hubungan yang sudah terjalin dan menyebabkan pelanggan berpindah ke pesaing. Oleh karena itu, pelatihan sumber daya manusia dalam aspek layanan menjadi investasi yang sangat penting bagi perusahaan.

4. Citra Merek dan Reputasi

Citra merek (brand image) adalah persepsi keseluruhan yang dimiliki pelanggan terhadap suatu perusahaan atau produk. Merek yang memiliki reputasi positif akan menciptakan ekspektasi tinggi terhadap produk dan layanan yang ditawarkan. Jika pengalaman pelanggan sesuai atau bahkan melampaui ekspektasi tersebut, maka tingkat kepuasan akan meningkat.

Citra merek dibentuk dari berbagai faktor, seperti kualitas produk, pengalaman pelanggan sebelumnya, kampanye pemasaran, testimoni, dan ulasan online. Perusahaan yang secara konsisten menjaga reputasi positif akan lebih mudah mendapatkan kepercayaan pelanggan dan mempertahankan loyalitas mereka.

5. Harapan Pelanggan

Harapan pelanggan terbentuk dari berbagai sumber, termasuk pengalaman pribadi, rekomendasi dari orang lain, promosi iklan, atau ulasan daring. Harapan ini menjadi tolok ukur yang digunakan pelanggan untuk menilai apakah produk atau layanan memenuhi kebutuhan mereka. Apabila produk atau layanan yang diterima sesuai atau melampaui harapan tersebut, maka pelanggan akan merasa puas. Namun jika tidak sesuai, maka kekecewaan sangat mungkin terjadi.

Perusahaan harus mampu mengelola ekspektasi pelanggan dengan cara menyampaikan janji yang realistis dan konsisten, serta memastikan bahwa pengalaman nyata pelanggan tidak mengecewakan.

6. Kemudahan Akses dan Proses

Kemudahan dalam mengakses layanan atau membeli produk juga merupakan faktor penting dalam menentukan tingkat kepuasan. Proses pembelian yang cepat, sistem pembayaran yang aman, navigasi situs web yang intuitif, serta respons cepat dalam layanan pelanggan digital adalah contoh aspek yang meningkatkan kenyamanan pelanggan.

Dalam era digital saat ini, pelanggan mengharapkan pengalaman yang serba mudah dan cepat. Perusahaan yang mampu menyediakan platform pelayanan yang efisien dan bebas hambatan akan lebih disukai dan mampu menciptakan pengalaman pelanggan yang positif.

7. Kesesuaian dengan Kebutuhan Individu

Kepuasan juga sangat dipengaruhi oleh sejauh mana produk atau layanan disesuaikan dengan kebutuhan spesifik pelanggan. Dalam pasar yang semakin terfragmentasi, pendekatan personalisasi menjadi semakin penting. Produk atau layanan yang dapat disesuaikan dengan preferensi individu

memiliki peluang lebih besar untuk memberikan kepuasan yang tinggi.

Penggunaan data pelanggan untuk memberikan rekomendasi personal, pengalaman yang relevan, atau paket layanan yang sesuai merupakan strategi modern yang dapat meningkatkan kepuasan sekaligus loyalitas.

8. Layanan Purna Jual (After-Sales Service)

Pelayanan setelah penjualan menjadi indikator nyata komitmen perusahaan terhadap pelanggan. Dukungan teknis, garansi, kemudahan klaim, layanan konsultasi, atau pemeliharaan produk merupakan bentuk layanan purna jual yang sangat dihargai pelanggan. Ketika masalah muncul setelah pembelian, kemampuan perusahaan dalam memberikan solusi cepat dan memuaskan akan sangat memengaruhi persepsi dan kepuasan pelanggan.

Pelanggan yang merasa tidak dibiarkan sendiri setelah transaksi cenderung akan mempercayai perusahaan dalam jangka panjang.

9. Kenyamanan dan Lingkungan Fisik

Dalam bisnis berbasis layanan langsung, seperti restoran, hotel, atau pusat perbelanjaan, kenyamanan dan lingkungan fisik turut memainkan peran dalam menciptakan pengalaman pelanggan. Desain ruang, kebersihan, tata letak, suhu ruangan, hingga pencahayaan semuanya dapat memengaruhi suasana hati pelanggan, yang pada gilirannya berdampak pada kepuasan mereka.

Lingkungan yang bersih, aman, dan menyenangkan akan memberikan kesan positif dan memperkuat nilai pengalaman pelanggan secara keseluruhan.

10. Pengalaman Sebelumnya dan Loyalitas

Pengalaman sebelumnya memengaruhi ekspektasi dan sikap pelanggan terhadap interaksi selanjutnya. Jika pelanggan memiliki pengalaman positif di masa lalu, mereka cenderung akan memberikan penilaian yang lebih baik terhadap layanan yang sama di kemudian hari. Loyalitas yang dibangun dari waktu ke waktu menciptakan hubungan emosional antara pelanggan dan merek, yang dapat memperbesar toleransi pelanggan terhadap kekurangan kecil sekaligus memperkuat rasa puas.

Namun, loyalitas juga harus terus dijaga dan ditingkatkan. Pelanggan setia sekalipun bisa berubah menjadi tidak puas jika merasa diabaikan atau jika kualitas layanan menurun.

1.3 Perbedaan arsitektur dan organisasi computer

Mengukur kepuasan pelanggan merupakan langkah penting bagi perusahaan untuk memahami sejauh mana produk atau layanan yang ditawarkan mampu memenuhi atau melampaui harapan konsumen. Tanpa pengukuran yang tepat, perusahaan hanya akan bersandar pada asumsi atau dugaan semata, yang dapat mengarah pada keputusan bisnis yang tidak akurat. Oleh karena itu, diperlukan metode yang sistematis, terukur, dan dapat diandalkan untuk menilai tingkat kepuasan pelanggan secara objektif.

Pengukuran kepuasan pelanggan tidak hanya bertujuan untuk mengetahui apakah pelanggan puas atau tidak, tetapi juga untuk mengidentifikasi aspek-aspek yang perlu ditingkatkan, memahami perilaku pelanggan, serta merumuskan strategi layanan yang lebih efektif dan efisien. Dalam bab ini akan dibahas berbagai metode yang umum digunakan dalam mengukur kepuasan pelanggan, kelebihan dan kekurangannya, serta penerapannya dalam konteks organisasi modern.

1. Survei Kepuasan Pelanggan (Customer Satisfaction Survey)

Survei merupakan metode yang paling umum dan langsung digunakan untuk mengukur kepuasan pelanggan. Metode ini biasanya dilakukan melalui kuesioner yang dirancang secara khusus untuk menilai berbagai aspek layanan atau produk. Survei dapat dilakukan secara offline (menggunakan formulir cetak) atau online melalui email, website, atau aplikasi.

Kuesioner dalam survei kepuasan pelanggan umumnya mencakup sejumlah indikator seperti:

- Kualitas produk/layanan
- Harga
- Kecepatan layanan
- Sikap petugas
- Kemudahan akses
- Tingkat kenyamanan
- Kemungkinan merekomendasikan kepada orang lain

Pertanyaan dalam survei biasanya menggunakan **skala Likert** (misalnya dari 1 = sangat tidak puas hingga 5 = sangat puas) atau skala numerik lainnya yang memungkinkan analisis kuantitatif.

Kelebihan:

- Data kuantitatif mudah dianalisis secara statistik.
- Dapat menjangkau banyak responden.
- Biaya relatif rendah, terutama jika dilakukan secara daring.

Kekurangan:

- Terkadang tidak mencerminkan perasaan sebenarnya pelanggan.
- Jawaban bisa bias karena pelanggan ingin “terlihat baik”.
- Tingkat partisipasi bisa rendah jika tidak ada insentif.

2. Net Promoter Score (NPS)

Net Promoter Score adalah metode sederhana yang digunakan untuk mengukur loyalitas pelanggan terhadap suatu merek atau layanan. Pelanggan diberi pertanyaan inti: **“Seberapa besar kemungkinan Anda merekomendasikan [produk/jasa/perusahaan] kepada teman atau kolega Anda?”**

Skor diberikan dalam rentang 0 hingga 10, dan berdasarkan jawabannya, pelanggan dikategorikan menjadi:

- **Promoter (9–10):** Sangat puas dan kemungkinan besar akan merekomendasikan.
- **Passive (7–8):** Cukup puas tetapi tidak terlalu antusias.
- **Detractor (0–6):** Tidak puas dan cenderung memberi penilaian negatif.

NPS dihitung dengan rumus:

$$\text{NPS} = \% \text{ Promoter} - \% \text{ Detractor}$$

Skor bisa berkisar dari -100 hingga +100. Semakin tinggi nilai NPS, semakin tinggi tingkat loyalitas pelanggan terhadap merek tersebut.

Kelebihan:

- Sederhana dan mudah diimplementasikan.
- Fokus pada loyalitas jangka panjang.

- Cocok untuk benchmarking antar periode atau perusahaan.

Kekurangan:

- Tidak menjelaskan alasan di balik skor.
- Butuh data tambahan untuk analisis mendalam.

3. Customer Satisfaction Score (CSAT)

CSAT adalah ukuran langsung dari tingkat kepuasan pelanggan terhadap pengalaman atau interaksi tertentu. Pertanyaan utama biasanya berbunyi:

“Seberapa puas Anda dengan pengalaman Anda hari ini?”

Jawaban diberikan dalam bentuk skala, seperti:

- Sangat Tidak Puas
- Tidak Puas
- Netral
- Puas
- Sangat Puas

CSAT dihitung sebagai persentase responden yang memilih "Puas" atau "Sangat Puas".

Rumus:

CSAT (%) = (Jumlah pelanggan yang puas / Total responden) × 100

Kelebihan:

- Spesifik untuk pengalaman tertentu (misal: pembelian atau kontak layanan).
- Mudah dikumpulkan setelah interaksi selesai.

Kekurangan:

- Tidak menggambarkan loyalitas jangka panjang.
- Bisa dipengaruhi oleh emosi sesaat.

4. Customer Effort Score (CES)

CES digunakan untuk mengukur seberapa besar upaya yang harus dikeluarkan pelanggan untuk menyelesaikan suatu masalah atau menyelesaikan transaksi. Pertanyaan kunci: **“Seberapa besar upaya yang Anda butuhkan untuk menyelesaikan permintaan Anda hari ini?”**

Biasanya menggunakan skala 1 (sangat mudah) hingga 7 (sangat sulit).

CES sangat berguna untuk mengukur efektivitas layanan pelanggan atau proses bisnis internal. Tujuannya adalah mengurangi “beban” pelanggan dan menciptakan pengalaman yang efisien.

Kelebihan:

- Fokus pada kemudahan proses.
- Relevan untuk layanan pelanggan atau sistem digital.
- Dapat digunakan untuk mendeteksi hambatan dalam sistem layanan.

Kekurangan:

- Tidak mencakup kepuasan secara keseluruhan.
- Perlu dikombinasikan dengan metode lain untuk analisis menyeluruh.

5. Analisis Umpan Balik Kualitatif

Selain pendekatan kuantitatif, pengukuran kepuasan pelanggan juga dapat dilakukan secara kualitatif melalui:

- Wawancara mendalam
- Focus group discussion
- Analisis komentar pelanggan (email, media sosial, forum)

Metode ini bertujuan menggali **motivasi, persepsi, dan emosi** pelanggan yang tidak selalu terlihat dalam data kuantitatif.

Kelebihan:

- Memberikan pemahaman yang lebih dalam.
- Dapat mengungkap hal-hal yang tidak terduga.
- Berguna untuk inovasi dan perbaikan layanan.

Kekurangan:

- Proses lebih lama dan mahal.
- Tidak bisa digeneralisasi secara statistik.

6. Analisis Data Transaksional dan Perilaku Pelanggan

Di era digital, perusahaan memiliki akses ke data perilaku pelanggan dalam jumlah besar, seperti:

- Riwayat pembelian
- Frekuensi kunjungan
- Durasi penggunaan layanan
- Aktivitas di platform online

Dengan memanfaatkan teknik analitik, machine learning, dan algoritma klasifikasi seperti **Decision Tree**, perusahaan dapat mengidentifikasi pola kepuasan pelanggan berdasarkan data

transaksional tersebut, bahkan tanpa perlu bertanya secara langsung.

Kelebihan:

- Minim intervensi langsung pada pelanggan.
- Dapat diproses otomatis dan real-time.
- Cocok untuk sistem berskala besar.

Kekurangan:

- Membutuhkan sistem data dan analitik yang baik.
- Tidak bisa menangkap aspek emosional atau subjektif secara langsung.

7. Penggunaan Benchmarking

Pengukuran kepuasan pelanggan juga dapat dikaitkan dengan benchmarking, yaitu membandingkan skor kepuasan terhadap standar industri atau pesaing. Ini berguna untuk menilai posisi perusahaan dalam kompetisi pasar dan mengidentifikasi praktik terbaik.

Benchmarking bisa dilakukan secara internal (membandingkan antar unit dalam perusahaan) maupun eksternal (dibandingkan dengan perusahaan lain di industri serupa).

1.4 Peran Data dalam Menganalisis Kepuasan Pelanggan

Dalam era digital dan informasi seperti sekarang ini, data telah menjadi aset penting bagi perusahaan dalam memahami perilaku dan kepuasan pelanggan. Tanpa data yang akurat dan relevan, perusahaan akan sulit untuk mengidentifikasi kebutuhan pelanggan, mengevaluasi kualitas layanan, maupun merancang strategi bisnis yang efektif. Data memungkinkan perusahaan untuk tidak hanya menilai pengalaman pelanggan secara objektif,

tetapi juga meramalkan tren dan mengambil keputusan yang lebih tepat sasaran.

Kepuasan pelanggan adalah konsep yang bersifat subjektif, namun dapat dianalisis secara kuantitatif maupun kualitatif melalui pendekatan berbasis data. Dengan mengumpulkan, menyimpan, memproses, dan menganalisis data pelanggan secara sistematis, perusahaan mampu memperoleh wawasan mendalam tentang seberapa baik mereka memenuhi harapan pelanggan, serta aspek mana yang perlu ditingkatkan.

1. Data sebagai Representasi Pengalaman Pelanggan

Data adalah representasi digital dari berbagai interaksi yang dilakukan pelanggan dengan perusahaan. Setiap titik kontak (touch point) yang dilalui pelanggan, baik secara online maupun offline, meninggalkan jejak data yang bisa dikumpulkan dan dianalisis. Mulai dari pembelian produk, pengisian survei, panggilan ke layanan pelanggan, kunjungan ke situs web, hingga ulasan di media sosial—semuanya mengandung informasi yang bernilai.

Dengan mengelola data ini secara cermat, perusahaan dapat membangun profil pelanggan yang komprehensif dan melacak perubahan dalam tingkat kepuasan secara waktu nyata. Misalnya, penurunan frekuensi pembelian bisa menjadi sinyal awal ketidakpuasan, atau lonjakan komplain bisa menunjukkan adanya penurunan kualitas layanan.

2. Sumber-Sumber Data Kepuasan Pelanggan

Terdapat berbagai sumber data yang dapat dimanfaatkan untuk menganalisis kepuasan pelanggan, antara lain:

- **Data Survei:** Data yang diperoleh melalui kuesioner kepuasan pelanggan, Net Promoter Score (NPS), Customer Satisfaction Score (CSAT), dan lainnya. Data ini bersifat langsung dan mencerminkan opini pelanggan.

- **Data Transaksional:** Informasi yang berasal dari aktivitas pembelian, histori pembayaran, penggunaan produk, atau keanggotaan pelanggan.
- **Data Interaksi Layanan Pelanggan:** Rekaman percakapan, email, tiket dukungan teknis, serta waktu tanggapan dan penyelesaian masalah.
- **Data Digital Behavior:** Aktivitas pelanggan di situs web atau aplikasi, seperti halaman yang dikunjungi, durasi sesi, klik, atau pengabaian keranjang belanja.
- **Data Media Sosial dan Ulasan Publik:** Komentar, testimoni, rating produk, dan keluhan di platform publik.

Semua jenis data ini dapat dikombinasikan untuk memperoleh gambaran menyeluruh mengenai pengalaman dan kepuasan pelanggan.

3. Analisis Data untuk Mendeteksi Pola Kepuasan

Melalui teknik analisis data, perusahaan dapat menggali informasi tersembunyi (*hidden insights*) dari kumpulan data pelanggan. Beberapa pendekatan analitik yang umum digunakan antara lain:

a. Descriptive Analytics

Digunakan untuk menjawab pertanyaan: *Apa yang terjadi?*

Contoh: Berapa persen pelanggan yang menyatakan puas tahun ini dibandingkan tahun sebelumnya? Berapa banyak keluhan yang masuk per bulan?

b. Diagnostic Analytics

Digunakan untuk menjawab pertanyaan: *Mengapa hal itu terjadi?*

Contoh: Mengapa terjadi penurunan kepuasan pada segmen

pelanggan tertentu? Faktor apa yang paling banyak disebut dalam komplain?

c. Predictive Analytics

Digunakan untuk menjawab pertanyaan: *Apa yang kemungkinan terjadi ke depan?*

Dengan memanfaatkan teknik machine learning seperti decision tree, regresi, atau clustering, perusahaan dapat memprediksi apakah pelanggan berpotensi puas atau tidak berdasarkan pola data sebelumnya.

d. Prescriptive Analytics

Digunakan untuk menjawab pertanyaan: *Apa yang harus dilakukan?*

Contoh: Apa intervensi terbaik untuk mempertahankan pelanggan yang berisiko tidak puas?

Melalui kombinasi teknik-teknik ini, perusahaan tidak hanya bisa memahami masa lalu, tetapi juga mengambil langkah proaktif untuk masa depan.

4. Decision Tree dalam Analisis Kepuasan Pelanggan

Salah satu metode yang banyak digunakan dalam analisis data kepuasan pelanggan adalah **algoritma Decision Tree**. Algoritma ini bekerja dengan membuat struktur pohon keputusan yang menunjukkan jalur logika dari fitur-fitur data ke hasil akhir (misalnya: puas, tidak puas).

Dengan menggunakan decision tree, perusahaan dapat:

- Mengidentifikasi variabel-variabel yang paling berpengaruh terhadap kepuasan (misalnya: kecepatan layanan, harga, atau keramahan staf).
- Membagi pelanggan ke dalam kelompok berdasarkan karakteristik dan tingkat kepuasannya.

- Menyusun aturan-aturan keputusan yang dapat digunakan sebagai panduan dalam meningkatkan kualitas layanan.

Misalnya, pohon keputusan dapat menunjukkan bahwa pelanggan yang dilayani lebih dari 10 menit dan menerima respons yang kurang ramah cenderung tidak puas. Informasi ini bisa menjadi dasar dalam pelatihan karyawan atau perbaikan SOP layanan.

5. Manfaat Analisis Data dalam Pengelolaan Kepuasan

Menggunakan data dalam menganalisis kepuasan pelanggan memberikan sejumlah manfaat strategis, antara lain:

- **Meningkatkan akurasi pengambilan keputusan:** Keputusan tidak diambil berdasarkan intuisi atau asumsi, melainkan berdasarkan bukti nyata dari data.
- **Deteksi dini terhadap potensi masalah:** Data memungkinkan perusahaan mengenali tren negatif sebelum menjadi krisis besar.
- **Personalisasi layanan:** Dengan profil pelanggan yang lebih lengkap, perusahaan dapat menyesuaikan pendekatan layanan berdasarkan kebutuhan individu.
- **Efisiensi operasional:** Dengan mengidentifikasi area yang menyebabkan ketidakpuasan, perusahaan bisa mengalokasikan sumber daya secara lebih efektif.
- **Peningkatan loyalitas dan retensi:** Pelanggan yang merasa diperhatikan cenderung akan lebih setia terhadap merek atau produk.

6. Tantangan dalam Pemanfaatan Data

Meskipun potensinya besar, pemanfaatan data dalam menganalisis kepuasan pelanggan juga memiliki tantangan:

- **Kualitas dan kelengkapan data:** Data yang tidak lengkap, tidak akurat, atau tidak terstruktur akan mengganggu hasil analisis.

- **Privasi dan etika penggunaan data:** Perusahaan harus memastikan penggunaan data pelanggan sesuai dengan peraturan perlindungan data pribadi.
- **Kapasitas analitik yang terbatas:** Tidak semua organisasi memiliki sumber daya manusia dan teknologi yang cukup untuk melakukan analisis lanjutan.
- **Integrasi antar sistem:** Data sering tersebar di berbagai platform (CRM, e-commerce, layanan pelanggan), sehingga perlu integrasi sistem yang baik.

Mengatasi tantangan ini membutuhkan investasi dalam teknologi, pengembangan SDM, dan tata kelola data yang kuat.

7. Integrasi Data ke Dalam Pengambilan Keputusan Strategis

Data tidak hanya berfungsi sebagai alat analisis, tetapi juga harus diintegrasikan ke dalam proses pengambilan keputusan di semua level organisasi. Manajemen puncak dapat menggunakan laporan analisis kepuasan pelanggan untuk menetapkan strategi jangka panjang, sementara tim operasional dapat menggunakan dashboard data untuk melakukan perbaikan harian.

Perusahaan modern juga mulai mengembangkan sistem **Customer Experience Management (CEM)** dan **Customer Data Platform (CDP)** yang memungkinkan pemantauan kepuasan pelanggan secara waktu nyata dan respons otomatis terhadap perubahan perilaku pelanggan.

BAB II

PENGANTAR DATA MINING DAN KLASIFIKASI

Dalam era digital saat ini, data menjadi aset yang sangat berharga bagi perusahaan. Setiap interaksi dengan pelanggan, transaksi penjualan, atau umpan balik layanan menghasilkan jejak data yang kaya akan informasi. Namun, data yang melimpah ini tidak memiliki arti jika tidak diolah dan dianalisis dengan baik. Di sinilah peran **data mining** menjadi sangat penting. Data mining merupakan proses untuk menemukan pola tersembunyi, tren, atau hubungan dalam kumpulan data besar guna mendukung pengambilan keputusan yang lebih cerdas dan berbasis fakta.

Data mining terdiri dari berbagai teknik analitik, seperti klasifikasi, klustering, regresi, asosiasi, dan deteksi anomali. Salah satu teknik yang paling umum digunakan adalah **klasifikasi**, yaitu proses membagi data ke dalam kategori-kategori tertentu berdasarkan atribut-atribut yang dimiliki. Dalam konteks kepuasan pelanggan, klasifikasi memungkinkan kita untuk memetakan pelanggan ke dalam kelompok “puas”, “netral”, atau “tidak puas” berdasarkan data perilaku atau tanggapan survei.

Algoritma yang sering digunakan untuk klasifikasi meliputi Decision Tree, Naïve Bayes, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM). Di antara algoritma tersebut, Decision Tree banyak dipilih karena hasilnya mudah dipahami dan divisualisasikan dalam bentuk pohon keputusan.

Melalui pendekatan data mining, perusahaan tidak hanya dapat memahami kondisi saat ini, tetapi juga memprediksi perilaku pelanggan di masa depan. Dengan demikian, strategi

bisnis dapat dirancang lebih tepat sasaran, proaktif, dan berbasis analitik.

2.1 Definisi dan Tujuan Data Mining

Data mining merupakan salah satu teknologi utama yang berkembang pesat dalam bidang ilmu komputer dan analisis data. Istilah ini mengacu pada proses penggalian informasi atau pengetahuan yang bermanfaat dari sejumlah besar data. Di era digital seperti saat ini, hampir setiap aktivitas manusia—baik dalam dunia bisnis, pendidikan, kesehatan, pemerintahan, maupun media sosial—menyebabkan terjadinya akumulasi data yang sangat besar. Tanpa metode dan alat yang tepat, data ini akan tetap dalam bentuk mentah dan tidak berguna. Oleh karena itu, data mining hadir sebagai solusi untuk mengekstraksi makna yang tersembunyi dalam data tersebut.

Definisi Data Mining

Secara umum, **data mining** dapat didefinisikan sebagai proses otomatis atau semi-otomatis untuk menemukan pola, hubungan, dan tren dalam kumpulan data besar dengan menggunakan teknik statistik, matematika, pembelajaran mesin (machine learning), dan sistem basis data. Data mining sering kali dianggap sebagai tahap penting dalam proses Knowledge Discovery in Databases (KDD), yaitu proses lengkap yang mencakup pembersihan data, integrasi data, seleksi data, transformasi, data mining, evaluasi pola, dan presentasi hasil.

Menurut para ahli:

- Han dan Kamber (2006) mendefinisikan data mining sebagai "proses menemukan pola yang menarik dan pengetahuan dari jumlah besar data."
- Fayyad et al. (1996) menjelaskan bahwa data mining adalah "tahap spesifik dalam proses KDD yang berkaitan

dengan penerapan algoritma untuk mengekstrak pola dari data."

- IBM menyebut data mining sebagai "proses penggunaan teknologi untuk menemukan hubungan yang sebelumnya tidak dikenali dalam data."

Dengan demikian, data mining bukan sekadar pengambilan data, melainkan pencarian pengetahuan yang berguna dan bernilai dari data.

Perbedaan Data Mining dengan Statistik dan Pembelajaran Mesin

Meskipun teknik data mining sering melibatkan statistik dan pembelajaran mesin, terdapat perbedaan mendasar antara ketiganya:

- Statistik lebih berfokus pada model matematis dan teori probabilitas untuk menjelaskan data.
- Pembelajaran mesin menitikberatkan pada pembangunan model yang dapat belajar dari data untuk melakukan prediksi atau klasifikasi.
- Data mining mencakup penggunaan kedua bidang tersebut, tetapi lebih menekankan pada eksplorasi data skala besar untuk menemukan pola tersembunyi yang belum diketahui sebelumnya.

Data mining memiliki pendekatan yang lebih aplikatif dan eksploratif, dengan tujuan praktis untuk mendukung pengambilan keputusan berdasarkan data.

Tujuan Utama Data Mining

Tujuan dari data mining sangat erat kaitannya dengan kebutuhan untuk mengambil keputusan yang tepat, cepat, dan berbasis data. Secara umum, tujuan data mining dapat dirinci sebagai berikut:

1. Menemukan Pola dan Hubungan dalam Data

Tujuan pertama dan utama dari data mining adalah menemukan pola, asosiasi, urutan, dan keterkaitan yang tersembunyi di dalam kumpulan data. Misalnya, dalam bidang ritel, data mining dapat digunakan untuk mengidentifikasi bahwa pelanggan yang membeli produk A cenderung juga membeli produk B.

2. Klasifikasi dan Prediksi

Data mining memungkinkan pembentukan model klasifikasi untuk mengelompokkan data ke dalam kategori yang telah ditentukan, seperti pelanggan “puas” dan “tidak puas”, atau transaksi “normal” dan “fraud”. Selain itu, data mining juga digunakan untuk membuat prediksi, misalnya memprediksi kecenderungan pelanggan berhenti berlangganan (churn) berdasarkan pola perilaku mereka.

3. Segmentasi dan Profiling

Segmentasi pelanggan berdasarkan pola perilaku atau preferensi mereka adalah salah satu aplikasi data mining yang banyak digunakan dalam pemasaran. Dengan segmentasi yang tepat, strategi pemasaran dapat diarahkan secara lebih spesifik kepada target pasar yang sesuai.

4. Deteksi Anomali dan Outlier

Data mining dapat digunakan untuk mendeteksi penyimpangan atau outlier yang tidak sesuai dengan pola umum dalam data. Hal ini sangat bermanfaat dalam keamanan siber, perbankan (untuk deteksi penipuan), dan sistem kesehatan (untuk mendeteksi gejala tidak biasa).

5. Optimalisasi Proses Bisnis

Data mining mendukung pengambilan keputusan yang lebih efisien dengan memberikan insight terhadap performa proses, hambatan operasional, atau peluang efisiensi. Misalnya,

perusahaan logistik dapat menggunakan data mining untuk mengoptimalkan rute pengiriman berdasarkan histori keterlambatan atau cuaca.

6. Peningkatan Layanan dan Kepuasan Pelanggan

Dengan menganalisis feedback pelanggan, perilaku pembelian, dan histori interaksi, data mining dapat mengungkap faktor-faktor yang memengaruhi kepuasan pelanggan. Hasil analisis ini dapat digunakan untuk menyusun strategi peningkatan layanan secara lebih personal dan efektif.

Jenis-jenis Tugas dalam Data Mining

Untuk mencapai tujuannya, data mining biasanya melibatkan beberapa jenis tugas utama, yaitu:

1. **Klasifikasi:** Menetapkan label atau kategori pada data, seperti klasifikasi pelanggan berdasarkan tingkat kepuasan.
2. **Klastering:** Mengelompokkan data tanpa label berdasarkan kemiripan, seperti mengelompokkan pelanggan berdasarkan kebiasaan belanja.
3. **Asosiasi (Association Rule Mining):** Menemukan hubungan antar item dalam suatu dataset, seperti analisis keranjang pasar (market basket analysis).
4. **Prediksi:** Memperkirakan nilai masa depan berdasarkan data historis.
5. **Deteksi Outlier:** Mengidentifikasi data yang tidak biasa atau menyimpang.
6. **Rangkuman (Summarization):** Menyajikan data secara ringkas dan bermakna.

Manfaat Implementasi Data Mining

Implementasi data mining dalam berbagai bidang memberikan banyak manfaat yang signifikan, antara lain:

- **Dalam dunia bisnis:** membantu merancang strategi pemasaran yang lebih tepat, meningkatkan loyalitas pelanggan, dan mempercepat inovasi produk.
- **Dalam sektor keuangan:** memungkinkan deteksi dini terhadap transaksi mencurigakan dan risiko kredit.
- **Dalam bidang kesehatan:** mendukung diagnosis lebih akurat dan pengembangan pengobatan yang dipersonalisasi.
- **Dalam pendidikan:** mengidentifikasi pola prestasi siswa dan merancang metode pembelajaran adaptif.
- **Dalam pemerintahan:** membantu dalam pengawasan publik, pengelolaan data sensus, serta pembuatan kebijakan berbasis data.

Tantangan dalam Data Mining

Meskipun menawarkan banyak keuntungan, data mining juga menghadapi sejumlah tantangan, antara lain:

- **Kualitas Data:** Data yang tidak lengkap, tidak akurat, atau berisik dapat menghasilkan pola yang menyesatkan.
- **Privasi dan Etika:** Penggunaan data pelanggan harus mempertimbangkan aspek hukum dan etika.
- **Kompleksitas Komputasi:** Pemrosesan data dalam jumlah besar membutuhkan sumber daya yang tinggi.
- **Overfitting:** Ketika model terlalu menyesuaikan diri dengan data pelatihan dan tidak bekerja baik pada data baru.

2.2 Teknik-Teknik Data Mining

Dalam dunia yang didorong oleh data, teknik data mining menjadi fondasi utama dalam menggali informasi yang bernilai dari kumpulan data yang sangat besar. Teknik-teknik ini tidak hanya menjadi alat bantu dalam pengambilan keputusan, tetapi juga mampu mengungkap pola tersembunyi yang sebelumnya tidak terdeteksi. Dalam konteks bisnis, ilmu kesehatan, keuangan, pendidikan, hingga pemerintahan, data mining memberikan keunggulan strategis yang signifikan.

Pada subbab ini, akan dibahas secara mendalam berbagai teknik utama dalam data mining, mencakup pendekatan klasifikasi, regresi, klastering, asosiasi, deteksi anomali, serta teknik reduksi dimensi dan visualisasi data. Pemahaman atas teknik-teknik ini sangat penting untuk menerapkan algoritma data mining secara tepat sesuai tujuan analisis, termasuk dalam konteks klasifikasi tingkat kepuasan pelanggan.

1. Klasifikasi (Classification)

Klasifikasi adalah salah satu teknik paling umum dalam data mining. Teknik ini bertujuan untuk memetakan data ke dalam kategori yang telah ditentukan sebelumnya (label kelas). Contoh penggunaannya adalah mengelompokkan pelanggan menjadi “puas”, “netral”, dan “tidak puas” berdasarkan variabel perilaku dan demografi.

Algoritma yang sering digunakan dalam klasifikasi antara lain:

- **Decision Tree:** Menggunakan struktur pohon untuk memetakan aturan keputusan dari data pelatihan. Algoritma ini sangat populer karena interpretasinya yang mudah.
- **Naive Bayes:** Menggunakan probabilitas bersyarat berdasarkan Teorema Bayes untuk mengklasifikasikan data.

- **Support Vector Machine (SVM):** Mencari hyperplane terbaik yang memisahkan data dari dua kelas atau lebih.
- **K-Nearest Neighbor (KNN):** Mengklasifikasikan data baru berdasarkan kedekatan jarak ke data pelatihan yang sudah dilabeli.
- **Random Forest:** Ensembel dari beberapa decision tree yang digabungkan untuk hasil klasifikasi yang lebih stabil dan akurat.

Klasifikasi banyak digunakan dalam berbagai aplikasi seperti prediksi churn pelanggan, diagnosis medis, deteksi spam, dan sistem rekomendasi.

2. Regresi (Regression)

Regresi adalah teknik yang digunakan untuk memprediksi nilai numerik kontinu. Berbeda dengan klasifikasi yang memetakan ke dalam kategori, regresi berusaha menemukan hubungan antara variabel independen dan dependen.

Contoh penerapannya adalah memprediksi tingkat kepuasan pelanggan dalam skala 1–10 berdasarkan kecepatan layanan, harga, dan kualitas produk.

Beberapa jenis algoritma regresi dalam data mining meliputi:

- **Linear Regression:** Model paling dasar yang mengasumsikan hubungan linear antar variabel.
- **Polynomial Regression:** Menggunakan persamaan polinomial untuk menangkap hubungan non-linear.
- **Regression Tree:** Sama seperti decision tree namun digunakan untuk memprediksi nilai numerik.
- **Lasso dan Ridge Regression:** Teknik regularisasi untuk menghindari overfitting pada model regresi.

Regresi sangat bermanfaat dalam forecasting, penetapan harga, dan analisis performa operasional.

3. Klastering (Clustering)

Klastering adalah teknik pengelompokan data berdasarkan kemiripan tanpa label awal (unsupervised learning). Tujuannya adalah untuk menemukan struktur atau pola alami dalam data. Misalnya, perusahaan dapat menggunakan klastering untuk mengidentifikasi segmen pelanggan yang memiliki preferensi belanja yang sama.

Beberapa teknik klastering yang populer meliputi:

- **K-Means Clustering:** Mengelompokkan data ke dalam k kelompok berdasarkan centroid.
- **Hierarchical Clustering:** Membentuk hierarki klaster melalui pendekatan bottom-up atau top-down.
- **DBSCAN:** Mengelompokkan data berdasarkan densitas, cocok untuk data dengan bentuk klaster yang tidak teratur.
- **Gaussian Mixture Model (GMM):** Menggunakan distribusi probabilistik untuk menentukan keanggotaan klaster.

Klastering digunakan dalam segmentasi pasar, pengenalan pola, dan deteksi komunitas dalam media sosial.

4. Asosiasi (Association Rule Mining)

Teknik asosiasi digunakan untuk menemukan hubungan antar item dalam dataset, dikenal juga sebagai *market basket analysis*. Misalnya, dari data pembelian pelanggan dapat ditemukan bahwa "jika pelanggan membeli kopi, kemungkinan besar mereka juga membeli gula."

Algoritma utama dalam teknik ini adalah:

- **Apriori Algorithm:** Mencari frekuensi itemset dan menggenerasi aturan asosiasi berdasarkan dukungan (support) dan kepercayaan (confidence).
- **FP-Growth Algorithm:** Pendekatan efisien dengan menggunakan struktur pohon (frequent pattern tree).

Aturan asosiasi sangat efektif dalam rekomendasi produk, penjualan silang (cross-selling), dan perencanaan tata letak toko.

5. Deteksi Anomali (Anomaly Detection)

Deteksi anomali bertujuan untuk menemukan data yang tidak biasa atau menyimpang dari pola umum. Teknik ini berguna dalam konteks keamanan sistem, deteksi penipuan, dan pemeliharaan prediktif.

Beberapa metode untuk deteksi anomali antara lain:

- **Statistical Methods:** Menggunakan ambang batas statistik untuk mendeteksi nilai ekstrem.
- **Isolation Forest:** Membagi data secara acak dan mengidentifikasi titik yang mudah dipisahkan sebagai anomali.
- **Autoencoder (dalam deep learning):** Digunakan untuk membangun representasi data dan mendeteksi deviasi dari rekonstruksi.

Dalam konteks analisis pelanggan, anomali bisa berupa pola keluhan yang tiba-tiba meningkat atau perubahan mendadak dalam pola pembelian.

6. Reduksi Dimensi (Dimensionality Reduction)

Dalam banyak kasus, data yang dianalisis memiliki dimensi sangat tinggi, yang dapat mengganggu proses komputasi dan akurasi model. Teknik reduksi dimensi bertujuan untuk

menyederhanakan representasi data dengan tetap mempertahankan informasi penting.

Beberapa teknik reduksi dimensi:

- **Principal Component Analysis (PCA):** Mengubah data ke dalam komponen utama yang saling orthogonal untuk mengurangi dimensi.
- **t-Distributed Stochastic Neighbor Embedding (t-SNE):** Digunakan untuk visualisasi data berdimensi tinggi ke dalam 2D atau 3D.
- **Linear Discriminant Analysis (LDA):** Mengurangi dimensi sambil menjaga pemisahan antar kelas.

Teknik ini sangat penting sebelum visualisasi atau pemodelan machine learning agar tidak terjadi overfitting.

7. Visualisasi Data

Walaupun bukan teknik mining murni, visualisasi adalah bagian penting dalam proses eksplorasi data. Visualisasi membantu peneliti dan praktisi untuk memahami pola, tren, dan anomali dalam data dengan lebih intuitif.

Alat dan teknik visualisasi meliputi:

- Diagram batang, garis, pie chart untuk data agregat.
- Heatmap dan scatter plot untuk korelasi antar variabel.
- Tree map dan dendrogram untuk hierarki data.
- Dashboard interaktif (misalnya menggunakan Tableau atau Power BI) untuk penyajian data secara real-time.

Visualisasi sering kali menjadi langkah awal dalam eksplorasi sebelum dilakukan data mining yang lebih mendalam.

2.3 Klasifikasi sebagai Teknik Analitik

Dalam lanskap data yang semakin kompleks dan berlimpah, organisasi dituntut untuk mampu memahami dan merespons informasi secara cepat dan akurat. Salah satu pendekatan paling umum dan efektif dalam proses analitik data adalah **klasifikasi**. Klasifikasi merupakan teknik dalam data mining yang bertujuan untuk memetakan data ke dalam kategori-kategori tertentu berdasarkan fitur atau atribut yang dimilikinya. Teknik ini sangat berguna dalam proses pengambilan keputusan berbasis data, terutama ketika organisasi perlu memprediksi status, perilaku, atau preferensi entitas tertentu seperti pelanggan, transaksi, atau produk.

Dalam konteks kepuasan pelanggan, klasifikasi dapat digunakan untuk memprediksi tingkat kepuasan berdasarkan data historis, seperti kebiasaan pembelian, respons terhadap layanan, waktu pelayanan, atau data demografis. Dengan demikian, klasifikasi tidak hanya bersifat reaktif tetapi juga proaktif dalam membentuk strategi pemasaran dan peningkatan mutu layanan.

Pengertian dan Tujuan Klasifikasi

Klasifikasi merupakan proses pengelompokan data ke dalam **kelas** atau **kategori** yang telah ditentukan sebelumnya. Kategori ini bersifat diskret, contohnya: “puas” atau “tidak puas”, “risiko tinggi” atau “risiko rendah”, “setuju” atau “tidak setuju”, dan seterusnya.

Tujuan utama klasifikasi adalah untuk:

- Membangun **model prediktif** yang dapat digunakan untuk mengklasifikasikan data baru berdasarkan pola yang ditemukan dalam data historis.
- Menyederhanakan proses pengambilan keputusan dengan menyediakan **kerangka analitis** yang sistematis.

- Mengidentifikasi **hubungan tersembunyi** antara variabel yang memengaruhi hasil (kelas).
- Memberikan wawasan yang dapat ditindaklanjuti untuk **pengembangan strategi bisnis** atau operasional.

Dalam praktiknya, klasifikasi melibatkan dua fase utama:

1. **Training (pelatihan):** Model dibangun berdasarkan dataset yang telah dilabeli (diketahui kelasnya).
2. **Testing (pengujian):** Model diuji dengan data baru untuk mengevaluasi akurasi dalam mengklasifikasikan entitas ke dalam kategori yang benar.

Penerapan Klasifikasi dalam Dunia Nyata

Klasifikasi digunakan secara luas di berbagai domain. Beberapa contohnya meliputi:

- **Kesehatan:** Memprediksi penyakit berdasarkan gejala pasien.
- **Keuangan:** Mengklasifikasikan kredit menjadi “layak” atau “tidak layak”.
- **Pemasaran:** Menentukan apakah pelanggan berpotensi membeli produk atau tidak.
- **Pendidikan:** Mengidentifikasi mahasiswa yang berisiko gagal.
- **E-commerce:** Menentukan apakah ulasan pelanggan bersifat positif atau negatif.
- **Deteksi Penipuan:** Mengklasifikasikan transaksi sebagai “normal” atau “mencurigakan”.

Dalam konteks buku ini, fokus utamanya adalah pada klasifikasi **tingkat kepuasan pelanggan**, yang dapat dibagi

menjadi beberapa label seperti “sangat puas”, “puas”, “netral”, “tidak puas”, dan “sangat tidak puas”.

Komponen Kunci dalam Klasifikasi

Proses klasifikasi terdiri dari beberapa komponen penting:

1. Fitur (features)

Fitur adalah atribut atau variabel input yang digunakan untuk membentuk model klasifikasi. Dalam kasus kepuasan pelanggan, fitur dapat berupa waktu respon layanan, harga produk, kualitas layanan, pengalaman pengguna, dan lainnya.

2. Kelas (class)

Kelas adalah label yang ingin diprediksi. Contohnya adalah label kepuasan pelanggan.

3. Algoritma Klasifikasi

Berbagai algoritma digunakan dalam proses klasifikasi, di antaranya:

- Decision Tree
- Naive Bayes
- K-Nearest Neighbor
- Support Vector Machine
- Neural Network

4. Model

Model adalah hasil dari proses pelatihan algoritma pada data berlabel. Model ini digunakan untuk memprediksi kelas data baru.

5. Evaluasi

Model dikaji berdasarkan metrik seperti akurasi, presisi,

recall, F1-score, dan ROC-AUC untuk mengukur performanya.

Algoritma Klasifikasi Populer

1. Decision Tree

Decision Tree adalah algoritma yang paling intuitif dalam klasifikasi. Ia membentuk struktur pohon berdasarkan aturan keputusan yang secara iteratif membagi data berdasarkan fitur yang paling informatif. Keunggulan utama Decision Tree adalah kemudahan interpretasi dan visualisasi, membuatnya sangat cocok untuk digunakan dalam aplikasi bisnis, termasuk kepuasan pelanggan.

2. Naive Bayes

Berdasarkan teorema Bayes, algoritma ini menghitung probabilitas suatu kelas berdasarkan atribut yang dimiliki oleh entitas. Meski sederhana, Naive Bayes sangat efisien dan efektif dalam banyak kasus klasifikasi teks dan spam filtering.

3. K-Nearest Neighbor (KNN)

KNN bekerja dengan prinsip bahwa suatu data cenderung memiliki kelas yang sama dengan data terdekatnya. Ia tidak memerlukan model pelatihan yang kompleks, namun cukup berat saat melakukan klasifikasi karena perlu menghitung jarak ke semua data dalam dataset.

4. Support Vector Machine (SVM)

SVM mencari hyperplane terbaik yang memisahkan dua kelas secara optimal. Cocok untuk dataset dengan margin yang jelas antar kelas, meskipun memerlukan tuning parameter dan dapat menjadi sulit diinterpretasikan.

5. Neural Network

Model yang meniru cara kerja otak manusia ini sangat kuat dalam menangani klasifikasi yang kompleks, terutama

ketika fitur yang digunakan sangat banyak dan tidak linear. Namun, interpretasi model menjadi lebih sulit dibandingkan dengan Decision Tree.

Klasifikasi dan Kepuasan Pelanggan

Dalam konteks klasifikasi kepuasan pelanggan, teknik ini memungkinkan organisasi untuk:

- **Memprediksi pelanggan yang mungkin tidak puas** berdasarkan histori interaksi mereka.
- **Menyesuaikan pendekatan layanan pelanggan** untuk mencegah churn (kehilangan pelanggan).
- **Mengidentifikasi pola demografis** pelanggan yang cenderung merasa puas atau tidak puas.
- **Membantu manajemen menetapkan prioritas** dalam peningkatan mutu layanan berdasarkan hasil klasifikasi.

Sebagai contoh, sebuah perusahaan e-commerce menggunakan data histori pesanan, waktu pengiriman, dan ulasan pelanggan untuk melatih model klasifikasi. Hasilnya, perusahaan dapat mengantisipasi pelanggan mana yang berisiko memberikan ulasan negatif dan segera mengambil tindakan seperti memberikan kompensasi atau layanan tambahan.

Evaluasi Model Klasifikasi

Setelah model klasifikasi dibangun, penting untuk mengevaluasinya dengan menggunakan metrik-metrik tertentu:

- **Akurasi:** Persentase prediksi yang benar dari seluruh data uji.
- **Presi:** Seberapa banyak prediksi positif yang benar (true positive dibandingkan semua prediksi positif).
- **Recall (Sensitivity):** Seberapa banyak data aktual positif yang berhasil terprediksi benar.

- **F1 Score:** Rata-rata harmonik dari presisi dan recall.
- **Confusion Matrix:** Tabel 2x2 atau lebih yang menunjukkan distribusi prediksi benar dan salah.

Evaluasi ini penting untuk memastikan bahwa model yang digunakan tidak hanya akurat, tetapi juga seimbang dalam mengklasifikasikan berbagai kelas.

Tantangan dalam Klasifikasi

Meskipun klasifikasi sangat berguna, teknik ini juga memiliki beberapa tantangan:

- **Data yang tidak seimbang:** Ketika jumlah data dalam satu kelas jauh lebih banyak dari yang lain, model bisa cenderung bias.
- **Overfitting:** Model terlalu cocok dengan data pelatihan hingga kehilangan generalisasi.
- **Noise dalam data:** Data yang tidak akurat atau outlier dapat memengaruhi kualitas model.
- **Pemilihan fitur:** Tidak semua fitur relevan. Pemilihan fitur yang tepat sangat krusial.

2.4 Algoritma Populer untuk Klasifikasi

Dalam dunia data mining dan machine learning, **klasifikasi** adalah salah satu metode paling umum dan luas digunakan untuk menganalisis data. Klasifikasi memungkinkan kita untuk memetakan suatu objek ke dalam kategori atau kelas tertentu berdasarkan sejumlah atribut atau fitur yang dimilikinya. Untuk mencapai tujuan tersebut, dibutuhkan algoritma yang mampu mengenali pola dari data historis dan membuat prediksi yang akurat terhadap data baru.

Seiring dengan meningkatnya volume dan keragaman data, banyak algoritma klasifikasi telah dikembangkan. Masing-masing algoritma memiliki pendekatan, kekuatan, kelemahan, dan skenario penggunaan yang berbeda. Dalam subbab ini, kita akan mengulas beberapa **algoritma klasifikasi yang paling populer**, terutama yang sering digunakan dalam analisis kepuasan pelanggan dan aplikasi bisnis lainnya. Pemahaman mendalam tentang algoritma ini akan membantu praktisi data dalam memilih metode yang paling sesuai dengan karakteristik data dan tujuan analisis.

1. Decision Tree

Decision Tree (Pohon Keputusan) merupakan salah satu algoritma klasifikasi yang paling intuitif dan mudah diinterpretasikan. Model ini bekerja dengan membagi dataset menjadi subset berdasarkan nilai atribut, menggunakan struktur pohon bercabang.

Cara Kerja:

- Decision Tree memulai dari node akar (root node).
- Setiap node merepresentasikan atribut tertentu dan setiap cabang menunjukkan hasil dari pengujian atribut.
- Proses ini berlanjut sampai mencapai *leaf node* yang mewakili kelas akhir.

Kelebihan:

- Mudah dipahami dan divisualisasikan.
- Tidak memerlukan normalisasi data.
- Dapat menangani data numerik maupun kategorikal.

Kekurangan:

- Rentan terhadap **overfitting**, terutama pada dataset yang besar dan kompleks.

- Performa dapat menurun jika data memiliki korelasi yang tinggi antar fitur.

Decision Tree sangat cocok untuk klasifikasi kepuasan pelanggan karena hasilnya bisa ditafsirkan oleh manajer non-teknis untuk pengambilan keputusan strategis.

2. Naive Bayes

Naive Bayes adalah algoritma klasifikasi berbasis probabilistik yang mengandalkan **teorema Bayes**. Algoritma ini mengasumsikan bahwa antar fitur dalam data saling independen (naive assumption).

Cara Kerja:

- Menghitung probabilitas suatu data masuk ke dalam kelas tertentu berdasarkan nilai-nilai fitur.
- Kelas dengan probabilitas tertinggi dipilih sebagai hasil klasifikasi.

Kelebihan:

- Cepat dan efisien, bahkan untuk dataset besar.
- Performa baik meskipun asumsi independensi tidak sepenuhnya valid.
- Sangat baik untuk klasifikasi teks seperti email spam detection dan analisis sentimen.

Kekurangan:

- Tidak akurat jika fitur-fitur saling bergantung.
- Tidak cocok untuk dataset dengan nilai numerik yang kompleks kecuali dilakukan preprocessing.

Naive Bayes dapat digunakan untuk menganalisis ulasan pelanggan atau survei untuk mengklasifikasikan respons sebagai “puas” atau “tidak puas”.

3. K-Nearest Neighbor (KNN)

K-Nearest Neighbor adalah algoritma **berbasis instance** yang sederhana namun kuat. Ia tidak membangun model selama pelatihan, tetapi menyimpan semua data latih dan memprediksi kelas berdasarkan tetangga terdekatnya.

Cara Kerja:

- Untuk data baru, hitung jarak (biasanya Euclidean) ke semua data latih.
- Pilih k tetangga terdekat.
- Kelas mayoritas dari tetangga tersebut dipilih sebagai hasil prediksi.

Kelebihan:

- Tidak memerlukan pelatihan eksplisit.
- Sederhana dan mudah diimplementasikan.
- Efektif untuk dataset kecil hingga menengah.

Kekurangan:

- Tidak efisien untuk dataset besar karena menghitung jarak ke semua data.
- Sensitif terhadap fitur tidak relevan dan skala data.
- Pemilihan nilai k yang salah bisa mengurangi akurasi.

KNN cocok digunakan saat data pelanggan tidak terlalu besar dan fitur-fitur telah dinormalisasi dengan baik.

4. Support Vector Machine (SVM)

SVM adalah algoritma klasifikasi yang mencari **hyperplane** terbaik yang memisahkan dua kelas secara optimal. SVM cocok untuk dataset dengan **margin besar** antar kelas.

Cara Kerja:

- Menentukan hyperplane yang memaksimalkan margin antara dua kelas.
- Menggunakan kernel untuk menangani klasifikasi non-linear.

Kelebihan:

- Akurat untuk data dengan dimensi tinggi.
- Kuat dalam menangani data non-linear dengan kernel trick.
- Tahan terhadap overfitting dalam banyak kasus.

Kekurangan:

- Kurang efisien pada dataset besar.
- Sulit diinterpretasikan oleh pengguna non-teknis.
- Memerlukan tuning parameter (seperti C, kernel) yang kompleks.

SVM sangat bermanfaat dalam prediksi perilaku pelanggan ketika pola klasifikasi cukup kompleks dan tidak linear.

5. Artificial Neural Network (ANN)

ANN meniru cara kerja otak manusia dengan menggabungkan sejumlah neuron buatan dalam lapisan (layers). Ia sangat kuat dalam menangani dataset kompleks, besar, dan tidak terstruktur.

Cara Kerja:

- Data dimasukkan ke input layer dan diproses melalui hidden layers dengan bobot tertentu.

- Output layer memberikan hasil prediksi berdasarkan fungsi aktivasi.

Kelebihan:

- Mampu menangkap hubungan non-linear yang kompleks.
- Sangat baik untuk data gambar, suara, dan teks.
- Dapat dilatih untuk klasifikasi multi-label.

Kekurangan:

- Proses pelatihan yang lama dan membutuhkan sumber daya komputasi besar.
- Sulit diinterpretasikan (black box).
- Rentan terhadap overfitting jika tidak dikontrol dengan benar.

ANN sangat cocok jika analisis kepuasan pelanggan melibatkan berbagai bentuk data seperti teks bebas, nilai numerik, dan data interaksi.

6. Random Forest

Random Forest adalah algoritma ensemble berbasis Decision Tree yang membangun banyak pohon keputusan dan menggabungkan hasilnya (voting) untuk prediksi akhir.

Cara Kerja:

- Membuat banyak decision tree menggunakan subset acak dari data dan fitur.
- Hasil klasifikasi diambil dari mayoritas prediksi semua pohon.

Kelebihan:

- Lebih stabil dan akurat daripada single Decision Tree.

- Mengurangi risiko overfitting.
- Dapat mengukur pentingnya fitur (feature importance).

Kekurangan:

- Kurang interpretatif dibandingkan satu pohon.
- Konsumsi memori dan waktu komputasi lebih besar.

Random Forest sangat efektif untuk proyek analisis kepuasan pelanggan yang melibatkan banyak fitur dan data beragam.

7. Logistic Regression

Logistic Regression adalah algoritma klasifikasi statistik yang menggunakan fungsi logistik (sigmoid) untuk memodelkan probabilitas suatu data masuk ke dalam kelas tertentu.

Cara Kerja:

- Menghitung probabilitas keluaran dengan fungsi sigmoid dari kombinasi linier fitur.

Kelebihan:

- Cepat dan efisien.
- Sangat interpretatif.
- Cocok untuk klasifikasi biner.

Kekurangan:

- Kurang efektif untuk hubungan non-linear antar fitur.
- Tidak bekerja baik jika ada multikolinearitas tinggi.

Logistic Regression sering digunakan untuk memprediksi kemungkinan kepuasan pelanggan (misalnya: puas vs tidak puas) dengan output dalam bentuk probabilitas.

Pemilihan Algoritma Klasifikasi

Tidak ada satu algoritma klasifikasi yang paling baik untuk semua kasus. Pemilihan algoritma tergantung pada berbagai faktor, seperti:

- **Ukuran dan jenis data** (numerik, kategorikal, teks, dll.)
- **Jumlah fitur dan dimensi data**
- **Tingkat noise dan outlier**
- **Tujuan bisnis** dan kebutuhan interpretabilitas
- **Sumber daya komputasi** yang tersedia

Dalam studi kepuasan pelanggan, apabila interpretabilitas tinggi dibutuhkan oleh tim manajemen, maka algoritma seperti Decision Tree atau Logistic Regression lebih disukai. Namun, jika akurasi model lebih penting dan data kompleks, maka Random Forest, SVM, atau Neural Network dapat menjadi pilihan utama.

2.5 Proses CRISP-DM dalam Data Mining

Dalam proses data mining, keberhasilan proyek tidak hanya bergantung pada teknik analisis yang digunakan, tetapi juga pada metode dan pendekatan sistematis dalam menangani data secara menyeluruh. Salah satu kerangka kerja yang paling banyak digunakan dalam praktik data mining adalah **CRISP-DM** (Cross-Industry Standard Process for Data Mining). CRISP-DM dikembangkan pada akhir 1990-an oleh konsorsium industri yang terdiri dari Daimler-Benz, ISL (Institut für Statistische und Mathematische Datenanalyse), dan NCR Corporation, dengan tujuan menciptakan standar metodologi data mining yang dapat digunakan lintas industri.

CRISP-DM telah terbukti efektif dan fleksibel karena dapat diterapkan pada berbagai jenis proyek data mining,

termasuk analisis kepuasan pelanggan. Model ini menyediakan pendekatan berbasis tahapan yang sistematis, iteratif, dan dapat disesuaikan, menjadikannya salah satu kerangka kerja yang paling banyak diadopsi di industri saat ini.

Model CRISP-DM terdiri dari **enam fase utama** yang saling berhubungan: *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*. Dalam subbab ini, akan dijelaskan masing-masing tahapan secara detail serta bagaimana penerapannya dalam konteks klasifikasi tingkat kepuasan pelanggan menggunakan algoritma Decision Tree.

1. Business Understanding

Tahap pertama dalam CRISP-DM adalah **pemahaman bisnis**, yaitu mengenali tujuan dan konteks bisnis dari proyek data mining yang sedang dilakukan. Pada tahap ini, analisis data harus bekerja sama dengan pemangku kepentingan untuk merumuskan masalah bisnis ke dalam bentuk yang bisa dijawab dengan data.

Dalam kasus kepuasan pelanggan, misalnya, tujuan bisnis mungkin adalah:

- Mengidentifikasi faktor-faktor yang memengaruhi tingkat kepuasan pelanggan.
- Memprediksi apakah seorang pelanggan merasa puas atau tidak berdasarkan data transaksi dan interaksinya.
- Meningkatkan loyalitas pelanggan melalui pemahaman yang lebih baik terhadap kebutuhan mereka.

Tahapan ini penting karena keberhasilan proyek data mining tidak hanya ditentukan oleh kecanggihan teknik analitik, tetapi juga oleh seberapa relevan solusi yang dihasilkan terhadap kebutuhan bisnis.

2. Data Understanding

Setelah tujuan bisnis ditentukan, langkah selanjutnya adalah **memahami data**. Proses ini mencakup:

- Pengumpulan data awal.
- Eksplorasi data untuk mengenali struktur dan isi.
- Identifikasi masalah kualitas data seperti nilai hilang, duplikat, atau outlier.
- Analisis statistik awal dan visualisasi.

Dalam studi kepuasan pelanggan, data yang dianalisis bisa berasal dari survei, catatan interaksi pelanggan, data transaksi pembelian, feedback terbuka, hingga data demografis. Analis harus mengidentifikasi fitur-fitur (atribut) yang potensial untuk klasifikasi, seperti umur pelanggan, frekuensi belanja, jumlah keluhan, atau hasil survei skor kepuasan.

Contohnya, jika ditemukan bahwa pelanggan dengan lebih dari tiga komplain dalam sebulan cenderung tidak puas, maka atribut “jumlah keluhan” menjadi variabel penting dalam model klasifikasi.

3. Data Preparation

Tahap ini sering disebut sebagai **tahap paling memakan waktu**, karena di sinilah data dibersihkan, disusun ulang, dan disiapkan agar siap digunakan dalam pemodelan. Proses ini mencakup:

- Pembersihan data (cleaning): menghapus atau memperbaiki data yang tidak konsisten, kosong, atau tidak valid.
- Transformasi data: mengubah data ke dalam format yang seragam atau menghitung variabel baru (misalnya, dari tanggal lahir menjadi usia).

- Seleksi atribut: memilih fitur yang relevan untuk model dan menghapus fitur yang tidak berguna.
- Pembuatan data training dan testing.

Dalam konteks klasifikasi kepuasan pelanggan, tahap ini bisa meliputi konversi data kategorikal (seperti jenis kelamin atau wilayah) menjadi angka (numerik), penghapusan data outlier seperti transaksi yang sangat tidak biasa, atau pengelompokan usia pelanggan ke dalam kategori.

Hasil akhir dari tahapan ini adalah dataset yang bersih dan siap digunakan oleh algoritma klasifikasi seperti Decision Tree.

4. Modeling

Tahap **pemodelan** adalah saat algoritma data mining mulai diterapkan. Di tahap ini, analis memilih satu atau lebih teknik analisis, mengatur parameter model, dan melatih model pada dataset yang telah disiapkan. Beberapa langkah penting dalam tahap ini antara lain:

- Memilih algoritma: dalam buku ini, fokus utama adalah pada *Decision Tree*, tetapi bisa juga dilakukan perbandingan dengan algoritma lain seperti Naive Bayes, KNN, atau Random Forest.
- Melatih model: proses pembelajaran dilakukan pada dataset training.
- Validasi model: menggunakan teknik seperti *cross-validation* untuk menghindari overfitting.
- Tuning parameter: mengatur parameter agar model lebih optimal.

Decision Tree, sebagai metode klasifikasi, memiliki keunggulan dalam menjelaskan aturan keputusan yang jelas dan dapat ditelusuri. Model akan membagi data secara iteratif

berdasarkan atribut yang paling signifikan hingga diperoleh aturan klasifikasi yang memisahkan data dengan akurasi tinggi.

Contohnya, hasil model bisa berupa aturan:

Jika **frekuensi belanja > 5 kali/bulan** dan **jumlah keluhan = 0**, maka pelanggan **sangat puas**.

5. Evaluation

Setelah model dibangun, penting untuk mengevaluasi sejauh mana model memenuhi tujuan bisnis yang telah dirumuskan di awal. Evaluasi tidak hanya berfokus pada akurasi teknis, tetapi juga pada:

- Relevansi hasil terhadap kebutuhan bisnis.
- Keterbacaan dan kemudahan interpretasi hasil.
- Kemampuan generalisasi model terhadap data baru.

Beberapa metrik evaluasi yang umum digunakan:

- **Akurasi**: persentase prediksi yang benar.
- **Precision dan Recall**: untuk melihat performa model dalam mengenali kelas tertentu.
- **Confusion Matrix**: menampilkan distribusi prediksi yang benar dan salah.
- **F1 Score**: gabungan dari precision dan recall.

Misalnya, jika model klasifikasi memiliki akurasi 90% dalam membedakan pelanggan puas dan tidak puas, maka manajemen dapat mulai mempertimbangkan tindakan berdasarkan hasil tersebut. Namun, evaluasi juga harus mempertimbangkan aspek lain seperti keberimbangan data—jangan sampai kelas “tidak puas” hanya sedikit sehingga model cenderung mengabaikannya.

6. Deployment

Tahapan akhir adalah **deployment**, yaitu mengimplementasikan model ke dalam sistem operasional agar dapat digunakan secara nyata. Proses ini dapat berupa:

- Laporan analitik untuk pengambilan keputusan manajerial.
- Integrasi model ke dalam sistem Customer Relationship Management (CRM).
- Otomatisasi proses klasifikasi pelanggan baru.

Deployment bisa dilakukan secara batch (misalnya bulanan) atau real-time (langsung saat data baru masuk). Dalam studi ini, jika model Decision Tree diterapkan secara langsung pada sistem manajemen pelanggan, maka setiap data pelanggan baru yang masuk akan otomatis diklasifikasikan ke dalam tingkat kepuasan tertentu.

Selain itu, deployment juga melibatkan monitoring berkala, karena performa model bisa menurun seiring berjalannya waktu akibat perubahan pola data (concept drift). Oleh karena itu, organisasi perlu menyiapkan mekanisme retraining model secara berkala.

Sifat Iteratif dan Fleksibel CRISP-DM

Perlu ditekankan bahwa proses CRISP-DM bersifat **iteratif**, artinya hasil dari satu fase bisa mempengaruhi fase sebelumnya. Misalnya, jika pada tahap evaluasi diketahui bahwa akurasi model rendah, maka analisis bisa kembali ke tahap data preparation untuk memperbaiki kualitas data, atau bahkan ke tahap business understanding untuk meninjau ulang tujuan proyek.

Selain itu, CRISP-DM bersifat fleksibel dan dapat diterapkan pada berbagai jenis proyek, baik di bidang perbankan, ritel, kesehatan, pendidikan, maupun pemerintahan. Dalam setiap

kasus, proses yang sama dapat digunakan untuk menggali wawasan dari data dan memberikan nilai tambah nyata bagi organisasi.

BAB III

ALGORITMA DECISION TREE

3.1 Apa itu Decision Tree

Decision Tree adalah salah satu algoritma dalam machine learning yang digunakan untuk melakukan proses klasifikasi atau prediksi berdasarkan keputusan logis yang disusun secara berurutan dalam bentuk struktur pohon. Struktur ini terdiri dari simpul akar (root node), simpul cabang (internal nodes), dan simpul daun (leaf nodes). Setiap simpul internal merepresentasikan pengujian terhadap atribut tertentu, setiap cabang menunjukkan hasil dari pengujian tersebut, dan setiap simpul daun menyatakan kelas atau nilai keputusan akhir.

Cara kerja Decision Tree menyerupai proses pengambilan keputusan manusia dalam kehidupan sehari-hari. Misalnya, dalam menentukan apakah pelanggan puas atau tidak, sistem dapat menilai atribut-atribut seperti kecepatan layanan, harga produk, kemudahan akses layanan, hingga frekuensi keluhan. Berdasarkan kombinasi nilai atribut-atribut tersebut, pohon akan “mengantar” data ke cabang yang paling sesuai hingga mencapai keputusan akhir di simpul daun.

Decision Tree memiliki kemampuan untuk menangani baik data kategorikal maupun numerik, dan sering digunakan karena sifatnya yang sangat intuitif dan mudah dijelaskan. Algoritma ini juga cocok untuk dataset besar dengan banyak variabel, serta dapat digunakan tanpa asumsi statistik yang rumit.

Dalam konteks kepuasan pelanggan, Decision Tree membantu perusahaan memahami faktor-faktor apa saja yang paling berkontribusi terhadap ketidakpuasan atau kepuasan pelanggan, serta menghasilkan aturan yang dapat digunakan untuk memperbaiki layanan atau menyusun strategi pemasaran yang lebih tepat sasaran.

3.2 Struktur dan Komponen Decision Tree

Sebuah **Decision Tree** dibangun dari struktur yang menyerupai pohon secara hierarkis, dengan berbagai komponen utama yang memiliki fungsi spesifik dalam proses pengambilan keputusan. Memahami setiap bagian dari struktur ini sangat penting agar kita dapat membaca, membangun, dan mengevaluasi model klasifikasi dengan lebih tepat.

Komponen utama dari Decision Tree terdiri dari:

1. **Root Node (Simpul Akar)**

Root node adalah titik awal dari pohon, yang merepresentasikan atribut pertama yang digunakan untuk memisahkan data. Atribut ini dipilih berdasarkan kriteria pemilihan terbaik, seperti *information gain* atau *Gini index*, tergantung pada algoritma yang digunakan. Root node menjadi dasar dari semua cabang pohon selanjutnya.

2. **Internal Nodes (Simpul Cabang)**

Internal nodes adalah simpul-simpul di antara root dan leaf node yang memuat pengujian terhadap atribut tertentu. Setiap node ini memiliki dua atau lebih cabang, tergantung pada jumlah nilai atau kondisi dari atribut yang diuji. Proses pengujian berlanjut hingga mencapai keputusan akhir.

3. **Branches (Cabang)**

Cabang menghubungkan simpul satu dengan simpul lainnya, menunjukkan hasil dari pengujian atribut. Cabang menggambarkan alur keputusan berdasarkan nilai atribut tertentu.

4. **Leaf Nodes (Simpul Daun)**

Leaf node adalah simpul akhir yang memuat hasil klasifikasi atau keputusan. Di sinilah data yang melalui berbagai

pengujian akhirnya diklasifikasikan ke dalam suatu kategori, misalnya “Puas” atau “Tidak Puas”.

Struktur Decision Tree sangat memengaruhi akurasi dan interpretabilitas model. Jika terlalu kompleks, model berisiko mengalami **overfitting**, sedangkan jika terlalu sederhana, bisa jadi informasi penting terabaikan. Oleh karena itu, keseimbangan dalam struktur sangat krusial dalam membangun model klasifikasi yang efektif.

3.3 Algoritma ID3, C4.5, dan CART: Perbandingan dan Kegunaan

Dalam pengembangan pohon keputusan (decision tree), terdapat beberapa algoritma utama yang banyak digunakan, yaitu **ID3 (Iterative Dichotomiser 3)**, **C4.5**, dan **CART (Classification and Regression Trees)**. Ketiganya memiliki prinsip dasar yang sama, yaitu membangun pohon berdasarkan pemilihan atribut terbaik untuk pemisahan data. Namun, masing-masing memiliki pendekatan dan keunggulan tersendiri.

ID3, yang dikembangkan oleh Ross Quinlan, merupakan algoritma awal yang menggunakan konsep *information gain* untuk memilih atribut terbaik. Atribut dengan nilai *information gain* tertinggi akan dipilih sebagai simpul pembagi. Kelemahan ID3 adalah tidak mampu menangani data dengan nilai numerik secara langsung, serta kurang optimal dalam menangani data dengan missing value.

C4.5 adalah pengembangan dari ID3 dan mengatasi banyak kelemahan pendahulunya. Algoritma ini menggunakan *gain ratio* sebagai kriteria pemilihan atribut, sehingga dapat menghindari bias terhadap atribut dengan banyak nilai. Selain itu, C4.5 mampu menangani atribut numerik, missing value, dan melakukan proses *pruning* (pemangkasan pohon) untuk mengurangi overfitting.

CART, dikembangkan oleh Breiman, berbeda dengan ID3 dan C4.5 karena menggunakan *Gini index* sebagai kriteria pemisahan dan membangun pohon biner—setiap simpul hanya memiliki dua cabang. CART juga dapat digunakan untuk tugas regresi selain klasifikasi, menjadikannya lebih fleksibel dalam berbagai aplikasi analitik.

Pemilihan algoritma yang tepat bergantung pada karakteristik data dan tujuan analisis. Untuk klasifikasi tingkat kepuasan pelanggan, C4.5 sering dipilih karena kemampuannya menangani berbagai tipe data serta hasil pohon yang mudah diinterpretasikan oleh pengguna non-teknis.

3.4 Keunggulan dan Kelemahan Decision Tree

Algoritma **Decision Tree** menjadi salah satu metode klasifikasi yang paling populer dalam data mining karena memiliki sejumlah keunggulan yang menjadikannya sangat praktis dan aplikatif. Namun demikian, seperti metode lainnya, Decision Tree juga memiliki keterbatasan yang perlu diperhatikan agar penggunaannya tepat guna.

Keunggulan Decision Tree:

1. **Mudah Dipahami dan Diinterpretasikan**
Struktur pohon yang visual dan hierarkis membuat Decision Tree mudah dipahami, bahkan oleh pengguna non-teknis. Setiap langkah dalam pengambilan keputusan dapat ditelusuri secara logis.
2. **Mendukung Data Kategorikal dan Numerik**
Algoritma seperti C4.5 dan CART mampu menangani kedua jenis data tersebut dengan efisien, sehingga fleksibel dalam berbagai konteks analisis.
3. **Sedikit atau Tanpa Pra-Pemrosesan Data**
Decision Tree tidak memerlukan normalisasi data atau

scaling seperti metode lain (misalnya SVM atau KNN), sehingga menghemat waktu dalam tahap praproses.

4. **Kemampuan Menangani Nilai yang Hilang (Missing Values)**

Beberapa algoritma Decision Tree mampu menangani data yang tidak lengkap tanpa harus membuang data secara keseluruhan.

Kelemahan Decision Tree:

1. **Mudah Mengalami Overfitting**
Jika pohon terlalu dalam dan kompleks, model dapat terlalu menyesuaikan dengan data pelatihan dan gagal melakukan generalisasi terhadap data baru.
2. **Rentan terhadap Perubahan Data**
Perubahan kecil pada data dapat menghasilkan struktur pohon yang sangat berbeda, yang dapat mempengaruhi kestabilan hasil.
3. **Tidak Optimal untuk Semua Kasus**
Pada dataset dengan hubungan yang sangat kompleks dan non-linear, metode lain seperti Random Forest atau Neural Network bisa memberikan performa yang lebih baik.

3.5 Studi Kasus Sederhana dengan Decision Tree

Untuk memberikan gambaran praktis tentang bagaimana algoritma Decision Tree bekerja, mari kita lihat sebuah studi kasus sederhana yang berkaitan dengan **klasifikasi tingkat kepuasan pelanggan**. Misalnya, sebuah perusahaan e-commerce ingin mengetahui faktor apa saja yang memengaruhi kepuasan pelanggan berdasarkan data transaksi yang dikumpulkan.

Dataset yang digunakan berisi atribut-atribut sebagai berikut:

- **Waktu Pengiriman (Cepat/Lambat)**

- **Kualitas Produk (Baik/Buruk)**
- **Respon Customer Service (Responsif/Tidak Responsif)**
- **Harga (Terjangkau/Mahal)**
- **Kepuasan Pelanggan (Puas/Tidak Puas)** ← label klasifikasi

Langkah pertama adalah membangun pohon keputusan dengan memilih atribut terbaik di akar (root node). Misalnya, dengan menggunakan algoritma **ID3**, atribut yang memiliki *information gain* tertinggi akan dipilih terlebih dahulu. Misalkan hasil analisis menunjukkan bahwa “Kualitas Produk” adalah atribut paling berpengaruh, maka atribut ini ditempatkan di simpul akar.

Setelah itu, pohon akan bercabang berdasarkan nilai dari atribut “Kualitas Produk”. Jika “Baik”, maka cabang selanjutnya mungkin mempertimbangkan “Waktu Pengiriman” atau “Harga”. Proses ini terus berlanjut hingga seluruh data diklasifikasikan ke dalam label “Puas” atau “Tidak Puas” di simpul daun.

Melalui Decision Tree ini, perusahaan dapat menyimpulkan bahwa misalnya, **pelanggan paling puas jika kualitas produk baik, pengiriman cepat, dan harga terjangkau**. Informasi ini sangat berguna untuk pengambilan keputusan strategis dalam meningkatkan layanan.

Studi kasus ini menunjukkan bagaimana Decision Tree dapat mengubah data mentah menjadi wawasan yang mudah dipahami dan langsung dapat ditindaklanjuti, menjadikannya alat yang sangat bermanfaat dalam analisis kepuasan pelanggan.

BAB IV

PERSIAPAN DATA UNTUK KLASIFIKASI KEPUASAN

4.1 Desain Kuesioner dan Pengumpulan Data

Tahapan awal dalam proses analisis kepuasan pelanggan adalah **pengumpulan data yang relevan dan representatif**. Salah satu metode paling umum dan efektif untuk mendapatkan data tersebut adalah melalui **kuesioner**. Desain kuesioner yang baik sangat krusial karena menentukan kualitas dan akurasi data yang dikumpulkan.

Desain kuesioner harus dimulai dengan **perumusan tujuan pengukuran**, dalam hal ini adalah mengidentifikasi tingkat kepuasan pelanggan dan faktor-faktor yang memengaruhinya. Oleh karena itu, kuesioner disusun dengan kombinasi pertanyaan **tertutup dan berskala likert** (misalnya: sangat tidak puas hingga sangat puas), agar data dapat diklasifikasikan dan dianalisis secara kuantitatif.

Pertanyaan dapat dibagi menjadi beberapa kategori, seperti:

- **Kualitas Produk/Jasa**
- **Kecepatan Layanan**
- **Harga dan Nilai**
- **Respon Customer Service**
- **Kemudahan Bertransaksi**
- **Kepuasan Umum**

Selain itu, bagian awal kuesioner dapat mencakup **data demografis** seperti usia, jenis kelamin, dan frekuensi pembelian untuk keperluan segmentasi pelanggan.

Setelah desain disetujui, proses **pengumpulan data** dapat dilakukan melalui berbagai saluran: survei online (Google Forms, SurveyMonkey), wawancara langsung, atau formulir kertas di titik layanan pelanggan. Pengumpulan data harus memperhatikan **etika penelitian**, termasuk transparansi tujuan, kerahasiaan responden, dan izin partisipasi.

Data yang terkumpul kemudian disaring, dibersihkan dari entri tidak valid, dan dikodekan ke dalam format numerik agar dapat diproses lebih lanjut oleh algoritma Decision Tree. Tahap ini merupakan pondasi penting dalam proses klasifikasi, karena kualitas output sangat bergantung pada kualitas input data.

4.2 Pembersihan dan Normalisasi Data

Setelah data dikumpulkan dari kuesioner atau sumber lain, langkah berikutnya yang sangat penting adalah proses **pembersihan (data cleaning)** dan **normalisasi (data normalization)**. Tahapan ini bertujuan untuk memastikan bahwa data yang akan dianalisis benar-benar bersih, konsisten, dan siap digunakan oleh algoritma klasifikasi seperti Decision Tree.

Pembersihan Data

Pembersihan data mencakup berbagai aktivitas, seperti:

- **Menghapus data duplikat**, yaitu entri yang sama muncul lebih dari sekali.
- **Menangani data yang hilang (missing values)**, dengan cara menghapus baris, mengganti nilai kosong dengan rata-rata atau modus, atau menggunakan teknik imputasi.

- **Memperbaiki data yang salah atau tidak logis**, misalnya usia pelanggan terisi “-5” atau kolom jenis kelamin berisi karakter asing.
- **Standarisasi format**, seperti menyatukan penulisan “Laki-laki” dan “Pria” agar terbaca sebagai satu kategori yang sama.

Proses ini penting untuk menghindari bias atau kesalahan hasil klasifikasi, karena Decision Tree sangat bergantung pada akurasi dan konsistensi data input.

Normalisasi Data

Walaupun Decision Tree tidak terlalu sensitif terhadap skala data seperti metode lain (misalnya KNN atau SVM), **normalisasi tetap penting** jika data akan digunakan bersama algoritma lain atau visualisasi. Normalisasi mengubah nilai data numerik ke dalam skala tertentu (biasanya 0–1) agar memudahkan interpretasi dan pengolahan selanjutnya.

Proses ini juga membantu dalam mendeteksi outlier, atau nilai ekstrem, yang bisa memengaruhi hasil analisis. Dengan data yang bersih dan terstruktur, proses klasifikasi tingkat kepuasan pelanggan dapat dilakukan dengan lebih akurat dan efisien, menghasilkan informasi yang valid dan dapat diandalkan untuk pengambilan keputusan.

4.3 Seleksi Fitur dan Penentuan Target Variabel

Dalam proses klasifikasi menggunakan algoritma Decision Tree, salah satu tahapan krusial adalah **seleksi fitur (feature selection)** dan **penentuan target variabel**. Kedua proses ini menentukan kualitas dan relevansi model yang dibangun.

Seleksi Fitur

Fitur adalah atribut atau kolom dalam dataset yang digunakan sebagai input dalam proses klasifikasi. Tidak semua fitur yang tersedia selalu relevan; beberapa bisa jadi redundan atau tidak memiliki kontribusi signifikan terhadap output. Oleh karena itu, seleksi fitur dilakukan untuk **memilih atribut yang paling informatif** dalam membedakan kelas atau kategori dalam data.

Beberapa metode yang umum digunakan dalam seleksi fitur adalah:

- **Information Gain**, yang mengukur seberapa besar informasi suatu fitur dalam mengurangi ketidakpastian (entropy).
- **Gain Ratio**, yang merupakan penyempurnaan dari Information Gain agar tidak bias terhadap jumlah kategori.
- **Chi-Square Test**, yang mengevaluasi hubungan antara fitur dan target variabel.

Dalam konteks analisis kepuasan pelanggan, fitur yang mungkin dipilih antara lain: waktu pengiriman, kualitas produk, harga, interaksi layanan pelanggan, dan kemudahan pemesanan.

Penentuan Target Variabel

Target variabel (class label) adalah atribut yang ingin diprediksi atau diklasifikasikan. Dalam studi ini, targetnya adalah **tingkat kepuasan pelanggan**, yang dapat dikategorikan secara biner (puas/tidak puas) atau multikategori (sangat puas, puas, netral, tidak puas, sangat tidak puas).

Pemilihan target variabel harus konsisten dan terdefinisi jelas, karena seluruh proses pembentukan pohon keputusan akan diarahkan untuk memprediksi nilai dari variabel ini.

4.4 Representasi Data dalam Bentuk Tabel dan CSV

Agar dapat dianalisis oleh algoritma Decision Tree, data hasil survei atau kuesioner perlu diubah ke dalam format yang terstruktur dan terkomputasi. Format yang paling umum digunakan adalah **tabel** dan **file CSV (Comma-Separated Values)**. Representasi ini tidak hanya memudahkan proses analisis, tetapi juga kompatibel dengan berbagai perangkat lunak analisis data seperti Python (pandas), R, RapidMiner, Orange, dan Weka.

Representasi Tabel

Dalam bentuk tabel, setiap baris merepresentasikan satu observasi atau responden, sedangkan setiap kolom mewakili satu atribut atau fitur. Contoh struktur tabel:

I D	Umur	Jenis Kelamin	Kualitas Produk	Harga	Kepuasan
1	25	Pria	Baik	Murah	Puas
2	34	Wanita	Cukup	Sedang	Tidak Puas

Kolom terakhir (“Kepuasan”) adalah target variabel yang akan diprediksi oleh model klasifikasi.

Format CSV

CSV adalah format berbasis teks yang memisahkan nilai antar kolom menggunakan koma atau delimiter lain (seperti titik koma). Format ini ringan, mudah dibaca, dan banyak digunakan dalam pertukaran data antar sistem.

Contoh format CSV:

```
ID,Umur,Jenis_Kelamin,Kualitas_Produk,Harga,Kepuasan
1,25,Pria,Baik,Murah,Puas
2,34,Wanita,Cukup,Sedang,Tidak Puas
```

CSV sangat ideal untuk proses data mining karena dapat langsung diimpor ke dalam perangkat lunak pemroses data. Selain itu, CSV mendukung otomatisasi dan integrasi, serta memudahkan dalam menyimpan dataset berukuran besar.

DAFTAR PUSTAKA

1. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier.
2. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. <https://doi.org/10.1007/BF00116251>
3. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann.
4. Kotler, P., & Keller, K. L. (2016). *Marketing Management* (15th ed.). Pearson Education.
5. Turban, E., Sharda, R., & Delen, D. (2014). *Decision Support and Business Intelligence Systems* (10th ed.). Pearson.
6. Hair, J. F., Wolfinbarger, M., Ortinau, D. J., & Bush, R. P. (2010). *Essentials of Marketing Research* (2nd ed.). McGraw-Hill Education.
7. Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining* (2nd ed.). Wiley.
8. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/15000000011>
9. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.
10. Orange Data Mining. (2023). *Data Mining Fruitful & Fun*. <https://orangedatamining.com/>
11. IBM. (2021). *The CRISP-DM methodology*. <https://www.ibm.com/docs/en/spss-modeler>
12. Suhartono, D., & Utami, M. I. D. (2018). Analisis tingkat kepuasan pelanggan menggunakan metode Decision Tree. *Jurnal Teknik Informatika dan Sistem Informasi*, 4(2), 45–52.
13. Nugroho, Y. (2017). *Kepuasan Pelanggan: Teori dan Implikasi untuk Strategi Pemasaran*. Andi Publisher.
14. Kurniawan, B. (2020). *Machine Learning dengan Python dan Scikit-Learn*. Elex Media Komputindo.

BIOGRAFI PENULIS



Riama Pakpahan

Lahir di Tanjung harapan pada tanggal 07 Maret 2001. Penulis menempuh Pendidikan di SD N 115858 PA II Tanjung harapan (2007-2013), SMP N2 SATAP Tanjung harapan (2013-2016), SMA N2 Pangkatan Tanjung harapan (2016-2019). Kemudian Penulis Melanjutkan Studinya di Fakultas Sains dan teknologi di Universitas Labuhanbatu-Rantau prapat. Bidang minat yang diambil penulis adalah Manajemen Informatika.



Deci Irmayani, S.Kom, M.Kom

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.



Volvo Sihombing, S.Kom, M.Kom

Lahir di Desa Durian Asahan Sumut, 15 Mei 1985. Telah menyelesaikan Pendidikan Sarjana (S1) di STMIK AKAKOM Yogyakarta 2008 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.

SINOPSIS

BUKU “ALGORITMA DECISION TREE UNTUK KLASIFIKASI TINGKAT KEPUASAN PELANGGAN” MERUPAKAN PANDUAN KOMPREHENSIF YANG MENGULAS BAGAIMANA TEKNIK DATA MINING, KHUSUSNYA ALGORITMA DECISION TREE, DAPAT DIGUNAKAN UNTUK MENGANALISIS DAN MENGLASIFIKASIKAN DATA TINGKAT KEPUASAN PELANGGAN SECARA SISTEMATIS DAN EFEKTIF.

DI TENGAH PERSAINGAN BISNIS YANG SEMAKIN KOMPETITIF, MEMAHAMI PERILAKU DAN KEPUASAN PELANGGAN MENJADI HAL YANG SANGAT PENTING BAGI KEBERLANGSUNGAN DAN PERTUMBUHAN PERUSAHAAN. BUKU INI MENGAWALI PEMBAHASAN DENGAN KONSEP DASAR KEPUASAN PELANGGAN, FAKTOR-FAKTOR YANG MEMENGARUHINYA, HINGGA CARA-CARA MENGUKURNYA DENGAN PENDEKATAN ILMIAH. SELANJUTNYA, BUKU INI MEMBAWA PEMBACA MEMASUKI DUNIA DATA MINING, MEMPERKENALKAN BERBAGAI TEKNIK ANALITIK DENGAN FOKUS UTAMA PADA METODE KLASIFIKASI.

MELALUI PEMBAHASAN MENDALAM TENTANG ALGORITMA DECISION TREE SEPERTI ID3, C4.5, DAN CART, PEMBACA AKAN DIAJAK MEMAHAMI CARA KERJA, STRUKTUR, SERTA KELEBIHAN DAN KEKURANGAN MASING-MASING ALGORITMA. TAK HANYA ITU, BUKU INI JUGA MENYAJIKAN PROSES PENGOLAHAN DATA MULAI DARI DESAIN KUESIONER, PEMBERSIHAN DATA, SELEKSI FITUR, HINGGA REPRESENTASI DATA DALAM FORMAT YANG SIAP DIANALISIS.



**PENERBIT
YAYASAN MMII**