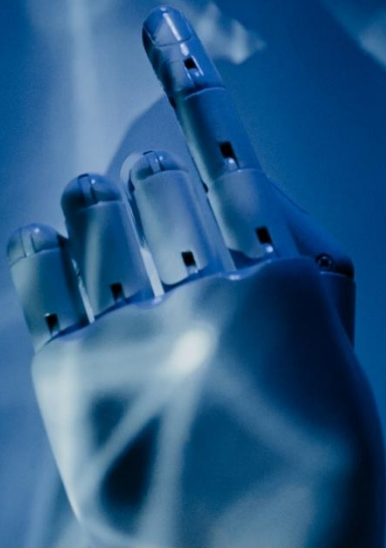




**PENERBIT
YAYASAN MMI**

MONOGRAP

DATA MINING UNTUK ANALISIS MINAT KONSUMEN



**NADIA PRISKA
DECI IRMAYANI
MILA NIRMALA SARI HASIBUAN**

DATA MINING UNTUK ANALISIS MINAT KOMSUMEN



**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).

DATA MINING UNTUK ANALISIS MINAT KOMSUMEN

Nadia Priska;

Deci Irmayani;

Mila Nirmala Sari Hasibuan;



**PENERBIT
YAYASAN MMI**

DATA MINING UNTUK ANALISIS MINAT KONSUMEN

Nadia Priska; Deci Irmayani; Mila Nirmala Sari Hasibuan;

Desain Cover :
Musthafa Haris Munandar

Sumber :
<https://munandar.yayasanmmi.com/data-mining-analisis-minat>

Tata Letak :
Deci Irmayani

Proofreader :
Mila Nirmala Sari Hasibuan

Ukuran :
Jml hal judul: 1, Jml hal isi naskah: 69, Uk: 15x23 cm

ISBN :
xxx xxx xxx

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 dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga buku berjudul "Data Mining untuk Analisis Minat Konsumen" ini dapat diselesaikan dengan baik. Buku ini disusun sebagai kontribusi dalam menjawab kebutuhan akan literatur yang menggabungkan teori dan praktik dalam pemanfaatan data mining, khususnya dalam konteks memahami perilaku dan preferensi konsumen.

Buku ini mengulas secara sistematis mulai dari konsep dasar data mining, proses kerja seperti CRISP-DM dan KDD, teknik-teknik analisis yang umum digunakan, hingga aplikasi langsung pada berbagai studi kasus nyata seperti e-commerce, retail, media sosial, dan survei daring. Selain itu, pembaca juga diperkenalkan dengan berbagai alat bantu seperti RapidMiner, Orange, Python, dan Weka yang dapat digunakan untuk mengimplementasikan proses data mining secara praktis.

Penulis berharap buku ini dapat menjadi referensi bermanfaat bagi mahasiswa, dosen, peneliti, praktisi bisnis, serta siapa pun yang tertarik untuk memahami bagaimana data mining dapat diterapkan dalam dunia pemasaran dan analisis konsumen. Kritik dan saran yang membangun sangat penulis harapkan demi penyempurnaan edisi mendatang.

Akhir kata, penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan inspirasi dalam penyusunan buku ini. Semoga buku ini memberikan manfaat dan mendorong pembaca untuk lebih aktif dalam mengembangkan wawasan di bidang analisis data dan perilaku konsumen.

Hormat Kami,

Penulis

DAFTAR ISI

PRAKATA	vi
DAFTAR ISI	vii
BAB 1: Dasar-dasar Data Mining	1
1.1 Definisi dan Konsep Data Mining	1
1.2 Proses Data Mining (CRISP-DM, KDD)	3
1.3 Teknik-teknik Umum dalam Data Mining	5
1.4 Data dan Sumber Data: Internal vs Eksternal	7
BAB 2: Karakteristik dan Perilaku Konsumen	10
2.1 Konsep Minat dan Preferensi Konsumen	10
2.2 Segmentasi Konsumen	12
2.3 Faktor-faktor yang Mempengaruhi Minat Konsumen	14
2.4 Data Konsumen sebagai Objek Analisis	16
BAB 3: Persiapan Data untuk Analisis	18
3.1 Pengumpulan Data Konsumen	18
3.2 Pembersihan dan Transformasi Data	20
3.3 Seleksi Fitur dan Penentuan Target	22
3.4 Representasi Data: Tabel, CSV, dan Database	24
BAB 4: Teknik Data Mining untuk Analisis Minat Konsumen	27
4.1 Clustering untuk Segmentasi Konsumen	27
4.2 Classification untuk Prediksi Minat	29
4.3 Association Rules untuk Pola Pembelian	32
4.4 Sentiment Analysis dari Review atau Sosial Media	34
4.5 Visualisasi Hasil Data Mining	36

BAB 5: Studi Kasus Analisis Minat Konsumen	39
5.1 Studi Kasus E-Commerce: Pola Pembelian Konsumen	39
5.2 Studi Kasus Retail: Produk Favorit Berdasarkan Lokasi	41
5.3 Studi Kasus Sosial Media: Analisis Ulasan Produk	42
5.4 Studi Kasus Kuesioner Online: Minat terhadap Produk Baru	45
BAB 6: Tools dan Platform untuk Data Mining Konsumen	48
6.1 RapidMiner	48
6.2 Orange	50
6.3 Python (pandas, scikit-learn, matplotlib)	52
6.4 Weka	53
6.5 Perbandingan dan Rekomendasi Penggunaan	56
DAFTAR PUSTAKA	58
BIOGRAFI PENULIS	59
SINOPSIS	61

BAB I

DASAR-DASAR DATA MINING

1.1 Definisi dan Konsep Data Mining

Data mining adalah suatu proses sistematis untuk menggali informasi yang sebelumnya tidak diketahui dari sejumlah besar data yang tersedia. Informasi ini berupa pola, tren, hubungan antar variabel, atau aturan yang tersembunyi di dalam data, yang dapat digunakan untuk mendukung pengambilan keputusan. Dalam era digital yang ditandai oleh ledakan data (data explosion), proses data mining menjadi sangat penting karena membantu menyaring dan menganalisis data dalam jumlah besar secara efisien dan efektif.

Secara umum, data mining merupakan bagian dari proses yang lebih besar yang disebut *Knowledge Discovery in Databases* (KDD), yaitu proses lengkap yang mencakup pemilihan data, pembersihan data, transformasi data, penambangan data, dan interpretasi hasil. Dalam konteks ini, data mining berfungsi sebagai tahap inti, di mana teknik-teknik statistik, matematika, dan kecerdasan buatan diaplikasikan untuk menemukan pola atau model yang relevan.

Definisi formal dari data mining menurut Han, Kamber, dan Pei (2012) adalah “proses mengekstraksi informasi yang sebelumnya tidak diketahui, dapat dipahami, dan berpotensi berguna dari data dalam jumlah besar.” Dari definisi tersebut, dapat dipahami bahwa tujuan utama data mining bukan sekadar menghasilkan data baru, melainkan mengubah data mentah menjadi wawasan bermakna yang dapat digunakan untuk memahami fenomena atau meramalkan kejadian.

Konsep dasar dari data mining mencakup beberapa elemen penting, yaitu:

1. **Data:** Merupakan sumber utama yang bisa berasal dari berbagai bentuk, seperti database relasional, data transaksi, data multimedia, log website, maupun data dari media sosial.
2. **Pola dan Model:** Data mining bertujuan untuk menemukan pola (patterns) atau model yang dapat menjelaskan perilaku atau kecenderungan tertentu, misalnya pola belanja konsumen berdasarkan waktu atau lokasi.
3. **Algoritma dan Teknik:** Proses ini menggunakan berbagai algoritma, mulai dari *classification*, *clustering*, *association rule mining*, hingga *regression*. Pemilihan teknik sangat bergantung pada jenis data dan tujuan analisis.
4. **Tujuan Analitik:** Secara umum terbagi dua, yaitu *descriptive* (untuk memahami data) dan *predictive* (untuk meramalkan hasil di masa depan).

Dalam analisis minat konsumen, data mining memungkinkan perusahaan untuk mengidentifikasi karakteristik pelanggan, memahami preferensi mereka, serta mengembangkan strategi pemasaran yang lebih personal dan efektif. Misalnya, dengan menerapkan algoritma *clustering*, perusahaan dapat mengelompokkan konsumen berdasarkan perilaku belanja mereka. Sementara itu, dengan *association rule mining*, dapat ditemukan hubungan antar produk yang sering dibeli bersamaan.

Penerapan data mining juga mendukung pengambilan keputusan berbasis data (data-driven decision making), yang lebih objektif dan akurat dibandingkan intuisi semata. Seiring dengan meningkatnya volume dan kompleksitas data konsumen, kemampuan untuk menambang informasi secara cepat dan tepat menjadi keunggulan kompetitif yang sangat bernilai dalam dunia bisnis modern.

1.2 Proses Data Mining (CRISP-DM, KDD)

Data mining bukanlah proses yang berdiri sendiri, melainkan bagian dari suatu rangkaian kegiatan sistematis yang bertujuan untuk mengubah data mentah menjadi pengetahuan yang berguna. Dua kerangka kerja (framework) yang paling banyak digunakan untuk menggambarkan proses data mining adalah **KDD (Knowledge Discovery in Databases)** dan **CRISP-DM (Cross-Industry Standard Process for Data Mining)**. Keduanya memberikan panduan terstruktur dalam menjalankan proyek data mining, mulai dari pemahaman data hingga interpretasi hasil.

A. KDD (Knowledge Discovery in Databases)

KDD adalah konsep awal yang diperkenalkan untuk menggambarkan seluruh siklus pengolahan data menjadi pengetahuan. Menurut Fayyad et al. (1996), proses KDD terdiri dari lima tahap utama:

1. **Selection (Seleksi):** Memilih data yang relevan dari berbagai sumber, baik internal seperti database pelanggan, maupun eksternal seperti data sosial media.
2. **Preprocessing (Pembersihan Data):** Menghapus data duplikat, mengatasi nilai kosong (missing values), dan memperbaiki kesalahan data. Tahap ini penting agar hasil mining tidak bias atau menyesatkan.
3. **Transformation (Transformasi Data):** Mengubah format atau struktur data agar sesuai dengan kebutuhan algoritma data mining, seperti normalisasi, encoding, atau penggabungan variabel.
4. **Data Mining:** Tahap inti di mana algoritma diterapkan untuk menemukan pola atau model dari data. Contohnya adalah klasifikasi untuk memprediksi jenis konsumen, atau asosiasi untuk menemukan produk yang sering dibeli bersamaan.

5. **Interpretation/Evaluation:** Mengevaluasi hasil data mining dan menginterpretasikannya dalam konteks bisnis untuk menghasilkan keputusan yang tepat.

B. CRISP-DM (Cross-Industry Standard Process for Data Mining)

CRISP-DM adalah model proses yang dikembangkan pada akhir 1990-an dan menjadi standar industri yang sangat populer karena fleksibilitas dan kelengkapannya. Proses CRISP-DM terdiri dari enam fase utama:

1. **Business Understanding:** Memahami tujuan bisnis dan kebutuhan yang ingin dicapai. Misalnya, memahami bahwa perusahaan ingin mengetahui minat konsumen terhadap produk baru.
2. **Data Understanding:** Mengumpulkan data awal, mengeksplorasi struktur data, dan mengidentifikasi masalah kualitas data.
3. **Data Preparation:** Menyiapkan dataset akhir yang akan digunakan untuk proses modeling, termasuk pembersihan, seleksi fitur, dan transformasi data.
4. **Modeling:** Menerapkan algoritma data mining, seperti decision tree, clustering, atau neural network, untuk menghasilkan model prediktif atau deskriptif.
5. **Evaluation:** Menilai kualitas dan kinerja model serta memastikan bahwa hasil sesuai dengan tujuan bisnis awal.
6. **Deployment:** Mengimplementasikan model ke dalam sistem operasional, atau menyajikan hasil dalam bentuk dashboard, laporan, atau rekomendasi.

Kedua kerangka kerja ini—KDD dan CRISP-DM—memiliki tujuan yang sama, yaitu membantu praktisi dalam menjalankan proses data mining secara sistematis dan dapat diulang. Dalam konteks analisis minat konsumen, proses ini

sangat penting untuk memastikan bahwa informasi yang dihasilkan benar-benar mencerminkan perilaku konsumen dan relevan dengan kebutuhan bisnis. Dengan mengikuti tahapan ini secara disiplin, perusahaan dapat menghindari kesalahan umum seperti misinterpretasi data, hasil yang bias, atau kesimpulan yang tidak akurat.

1.3 Teknik-teknik Umum dalam Data Mining

Data mining mencakup berbagai teknik yang dirancang untuk menggali pola, struktur, dan pengetahuan tersembunyi dari data. Teknik-teknik ini bervariasi tergantung pada tujuan analisis, jenis data, dan pendekatan yang digunakan—baik deskriptif maupun prediktif. Dalam konteks analisis minat konsumen, pemilihan teknik yang tepat menjadi sangat krusial untuk menghasilkan wawasan yang relevan dan dapat ditindaklanjuti. Berikut adalah beberapa teknik data mining yang umum digunakan:

1. Klasifikasi (Classification)

Klasifikasi adalah teknik prediktif yang digunakan untuk mengelompokkan data ke dalam kategori yang telah ditentukan sebelumnya. Tujuan utama klasifikasi adalah mempelajari model dari data historis dan menggunakannya untuk memprediksi label dari data baru. Contoh penggunaannya dalam bisnis adalah memprediksi apakah seorang konsumen kemungkinan akan membeli produk atau tidak, berdasarkan riwayat pembelian dan karakteristik pribadi.

Algoritma yang sering digunakan dalam klasifikasi antara lain Decision Tree, Naïve Bayes, Support Vector Machine (SVM), dan k-Nearest Neighbor (k-NN). Hasil klasifikasi dapat membantu perusahaan dalam menyusun strategi pemasaran yang lebih terarah.

2. Klastering (Clustering)

Berbeda dengan klasifikasi, teknik klastering bersifat deskriptif dan digunakan untuk mengelompokkan data berdasarkan kesamaan tertentu tanpa label yang telah ditentukan. Dalam analisis minat konsumen, klastering dapat digunakan untuk mengidentifikasi segmen pelanggan berdasarkan pola belanja, preferensi produk, atau frekuensi kunjungan.

Algoritma populer dalam klastering antara lain K-Means, Hierarchical Clustering, dan DBSCAN. Hasil klastering sangat bermanfaat untuk strategi segmentasi pasar dan personalisasi layanan.

3. Asosiasi (Association Rule Mining)

Teknik ini bertujuan untuk menemukan hubungan atau asosiasi antar item dalam satu dataset. Salah satu aplikasi paling terkenal adalah dalam analisis keranjang belanja (market basket analysis), di mana sistem dapat mengidentifikasi produk-produk yang sering dibeli bersamaan, seperti “Jika seseorang membeli roti, maka kemungkinan besar juga membeli mentega.”

Algoritma yang sering digunakan termasuk Apriori dan FP-Growth. Hasil dari teknik asosiasi biasanya disajikan dalam bentuk aturan IF-THEN yang mudah dipahami dan diterapkan dalam penataan produk atau penawaran promosi.

4. Regresi (Regression)

Regresi digunakan untuk memprediksi nilai numerik berdasarkan satu atau lebih variabel input. Dalam konteks minat konsumen, regresi dapat digunakan untuk memprediksi berapa banyak produk yang akan terjual dalam periode tertentu, atau memperkirakan nilai pembelian rata-rata per pelanggan.

Regresi linier adalah bentuk paling sederhana, namun tersedia juga regresi non-linier dan regresi logistik tergantung pada kompleksitas hubungan antar variabel.

5. Deteksi Anomali (Anomaly Detection)

Teknik ini digunakan untuk mengidentifikasi data yang menyimpang dari pola umum. Dalam dunia bisnis, deteksi anomali dapat berguna untuk menemukan transaksi mencurigakan, perilaku pelanggan yang tidak biasa, atau kesalahan data yang perlu ditindaklanjuti.

1.4 Data dan Sumber Data: Internal vs Eksternal

Dalam proses data mining, data merupakan komponen paling mendasar dan menentukan kualitas hasil analisis yang dihasilkan. Oleh karena itu, pemahaman terhadap sumber data sangat penting agar proses eksplorasi dan penggalian informasi dapat berjalan secara efektif. Secara umum, sumber data untuk analisis minat konsumen dibagi menjadi dua kategori utama, yaitu data internal dan data eksternal. Keduanya memiliki karakteristik, kelebihan, serta tantangan masing-masing dalam konteks penggunaannya.

A. Data Internal

Data internal adalah data yang berasal dari dalam organisasi atau perusahaan itu sendiri. Data ini biasanya dihasilkan dari aktivitas operasional sehari-hari dan tersimpan dalam sistem informasi perusahaan, seperti basis data pelanggan, sistem kasir (point of sale), log transaksi, laporan keuangan, atau data hasil survei internal.

Contoh data internal meliputi:

- Riwayat pembelian pelanggan
- Frekuensi kunjungan ke toko atau website
- Respon terhadap promosi atau diskon
- Data demografis pelanggan yang dikumpulkan saat pendaftaran akun

Keunggulan utama dari data internal adalah tingkat akurasi dan relevansi yang tinggi terhadap kebutuhan bisnis. Karena data ini dikumpulkan langsung dari interaksi pelanggan dengan layanan atau produk perusahaan, maka informasi yang diperoleh sangat kontekstual. Selain itu, data internal biasanya lebih mudah diakses, aman, dan bebas dari biaya tambahan.

Namun, tantangan dari data internal adalah keterbatasan cakupan. Data ini hanya mencerminkan perilaku pelanggan dalam konteks interaksi dengan perusahaan, sehingga tidak menggambarkan pengaruh eksternal seperti tren pasar atau aktivitas kompetitor.

B. Data Eksternal

Sebaliknya, data eksternal adalah data yang berasal dari luar organisasi dan dikumpulkan dari berbagai sumber independen. Jenis data ini dapat digunakan untuk melengkapi data internal agar analisis menjadi lebih komprehensif. Sumber data eksternal mencakup media sosial, data sensus, survei pasar, laporan industri, data dari pihak ketiga, hingga ulasan konsumen di platform publik.

Contoh data eksternal antara lain:

- Review produk di marketplace (seperti Tokopedia, Shopee, atau Amazon)
- Komentar dan reaksi konsumen di media sosial
- Laporan tren industri dari lembaga riset
- Data demografis dari badan statistik nasional

Kelebihan data eksternal adalah kemampuannya untuk memperluas cakupan analisis. Dengan menggabungkan data eksternal, perusahaan dapat memahami bagaimana persepsi publik terhadap merek mereka dibandingkan dengan pesaing, atau bagaimana tren global mempengaruhi minat konsumen

lokal. Selain itu, data eksternal memungkinkan perusahaan untuk melakukan benchmarking dan prediksi tren masa depan.

Namun, penggunaan data eksternal juga memiliki tantangan, terutama dari sisi validitas dan akurasi. Karena data ini sering kali berasal dari pihak ketiga, maka perlu ada proses verifikasi dan pembersihan sebelum digunakan dalam analisis. Selain itu, beberapa data eksternal memerlukan biaya berlangganan atau pembelian, serta mempertimbangkan aspek legal dan etika dalam pengumpulan dan penggunaannya.

BAB II

KARAKTERISTIK DAN PERILAKU KONSUMEN

2.1 Konsep Minat dan Preferensi Konsumen

Minat dan preferensi konsumen merupakan dua konsep kunci dalam pemasaran modern dan menjadi landasan utama dalam analisis perilaku pelanggan. Dalam konteks data mining, memahami minat dan preferensi konsumen sangat penting untuk menggali pola-pola tersembunyi yang dapat digunakan dalam pengambilan keputusan bisnis, seperti pengembangan produk, strategi promosi, hingga personalisasi layanan.

Minat konsumen mengacu pada ketertarikan individu terhadap suatu produk, layanan, atau merek. Minat ini muncul sebagai hasil dari rangsangan internal maupun eksternal, seperti kebutuhan pribadi, pengaruh sosial, pengalaman masa lalu, maupun promosi yang dilihat atau didengar. Minat merupakan tahap awal dalam proses pengambilan keputusan pembelian; ketika seseorang memiliki minat terhadap suatu produk, ia akan cenderung mencari informasi lebih lanjut dan mempertimbangkan pembelian.

Sedangkan **preferensi konsumen** merujuk pada pilihan atau kecenderungan individu untuk memilih satu produk dibandingkan yang lain, berdasarkan pertimbangan rasional maupun emosional. Preferensi menunjukkan peringkat nilai subjektif yang diberikan konsumen terhadap berbagai alternatif yang tersedia di pasar. Misalnya, seseorang mungkin memiliki minat terhadap beberapa merek kopi, tetapi lebih memilih merek tertentu karena rasa, harga, atau citra merek yang sesuai dengan kepribadiannya.

Perbedaan antara minat dan preferensi terletak pada kedalaman keterlibatan konsumen. Minat mencerminkan ketertarikan awal yang bisa bersifat umum dan belum tentu berujung pada tindakan, sedangkan preferensi sudah mengarah pada tindakan memilih dan membandingkan alternatif. Dalam praktiknya, minat dan preferensi sering kali saling memengaruhi dan berjalan bersamaan dalam membentuk perilaku konsumen.

Dalam era digital, minat dan preferensi konsumen dapat dilacak dan dianalisis melalui berbagai sumber data, seperti riwayat pencarian, aktivitas di media sosial, interaksi di aplikasi, hingga pola pembelian. Teknologi data mining memungkinkan perusahaan untuk menangkap sinyal-sinyal minat konsumen secara real time dan mengubahnya menjadi wawasan strategis. Misalnya, jika seorang pengguna sering mencari informasi tentang gadget terbaru dan mengklik iklan tertentu, maka sistem dapat mengidentifikasi minatnya dan merekomendasikan produk yang relevan.

Lebih lanjut, analisis preferensi konsumen dapat dilakukan dengan berbagai teknik, seperti survei kepuasan pelanggan, analisis ulasan produk, atau melalui model prediktif berbasis machine learning. Dengan mengetahui preferensi konsumen, perusahaan dapat merancang produk yang lebih sesuai, menawarkan harga yang kompetitif, atau menciptakan pengalaman pengguna yang lebih memuaskan.

Penting untuk dicatat bahwa minat dan preferensi bersifat dinamis. Mereka dapat berubah seiring waktu, tergantung pada tren pasar, pengaruh lingkungan, maupun pengalaman pribadi konsumen. Oleh karena itu, pemantauan yang berkelanjutan dan respons yang cepat terhadap perubahan perilaku konsumen sangat dibutuhkan agar perusahaan tetap relevan dan kompetitif di tengah persaingan yang ketat.

2.2 Segmentasi Konsumen

Segmentasi konsumen adalah proses membagi pasar atau basis pelanggan menjadi kelompok-kelompok yang memiliki karakteristik, kebutuhan, perilaku, atau preferensi yang serupa. Tujuan utama dari segmentasi adalah agar perusahaan dapat memahami perbedaan antar konsumen dan menyesuaikan strategi pemasaran, produk, serta layanan untuk setiap kelompok secara lebih spesifik dan efektif. Dalam era data digital saat ini, segmentasi konsumen telah berkembang dari pendekatan tradisional menjadi lebih canggih dengan memanfaatkan teknik data mining dan analitik.

Secara umum, segmentasi dilakukan berdasarkan sejumlah kriteria, yang secara tradisional dikategorikan ke dalam empat tipe utama:

1. **Segmentasi Demografis**, yang meliputi usia, jenis kelamin, pendapatan, tingkat pendidikan, pekerjaan, dan status pernikahan. Segmentasi ini mudah diakses dan sering digunakan karena data demografis biasanya tersedia dan terstruktur.
2. **Segmentasi Geografis**, berdasarkan lokasi tempat tinggal atau bekerja konsumen, seperti negara, kota, atau bahkan kode pos. Dalam banyak kasus, lokasi berpengaruh terhadap preferensi dan kebutuhan, terutama dalam bisnis ritel atau layanan lokal.
3. **Segmentasi Psikografis**, yang memperhatikan gaya hidup, nilai-nilai, kepribadian, dan minat konsumen. Segmentasi ini lebih kompleks karena memerlukan pendekatan kualitatif atau analisis perilaku mendalam, namun dapat memberikan wawasan yang sangat berharga dalam membangun hubungan emosional dengan pelanggan.

4. **Segmentasi Perilaku**, yang didasarkan pada tindakan nyata konsumen seperti frekuensi pembelian, loyalitas merek, waktu berbelanja, respons terhadap promosi, atau penggunaan produk. Ini adalah pendekatan yang semakin penting dalam era digital karena aktivitas konsumen dapat dilacak secara real-time melalui website, aplikasi, atau media sosial.

Dengan kemajuan teknologi informasi dan ketersediaan data besar (*big data*), segmentasi kini dapat dilakukan secara otomatis dan lebih presisi menggunakan teknik data mining, seperti clustering. Misalnya, dengan algoritma K-Means atau DBSCAN, sistem dapat secara otomatis mengelompokkan pelanggan ke dalam segmen berdasarkan pola belanja atau perilaku browsing mereka, tanpa perlu ditentukan sebelumnya kriteria segmentasinya.

Manfaat dari segmentasi konsumen sangat besar dalam praktik bisnis. Perusahaan dapat menyusun kampanye pemasaran yang lebih personal, mengembangkan produk sesuai kebutuhan segmen tertentu, serta mengoptimalkan sumber daya dengan menargetkan konsumen yang paling potensial. Misalnya, segmen “milenial urban” mungkin lebih responsif terhadap kampanye digital dan produk berbasis teknologi, sementara segmen “orang tua konservatif” mungkin lebih menghargai keamanan dan layanan pelanggan yang personal.

Namun, penting juga untuk diingat bahwa segmentasi bukanlah proses satu kali. Perilaku dan minat konsumen dapat berubah seiring waktu, sehingga segmentasi perlu dievaluasi dan diperbarui secara berkala. Selain itu, perusahaan juga harus berhati-hati agar segmentasi tidak menjadi terlalu sempit atau eksklusif, yang dapat menyebabkan kehilangan peluang di pasar lain.

2.3 Faktor-faktor yang Mempengaruhi Minat Konsumen

Minat konsumen terhadap suatu produk atau layanan tidak terbentuk secara kebetulan, melainkan dipengaruhi oleh berbagai faktor internal dan eksternal yang saling berkaitan. Dalam konteks bisnis dan pemasaran, memahami faktor-faktor ini sangat penting untuk mengembangkan strategi yang mampu menarik perhatian dan membangun keinginan konsumen. Dengan bantuan data mining, faktor-faktor tersebut dapat dianalisis lebih mendalam untuk mengidentifikasi pola-pola yang memicu minat dan kecenderungan pembelian.

Secara umum, faktor-faktor yang mempengaruhi minat konsumen dapat dikategorikan ke dalam lima kelompok utama, yaitu:

1. Faktor Psikologis

Faktor ini berkaitan dengan kondisi mental dan emosional individu, termasuk motivasi, persepsi, pembelajaran, dan sikap. Misalnya, seseorang yang merasa kebutuhan akan status sosial lebih tinggi cenderung tertarik pada produk premium atau merek ternama. Persepsi konsumen terhadap kualitas, keamanan, atau manfaat produk juga sangat menentukan tingkat ketertarikannya. Sikap positif yang terbentuk dari pengalaman masa lalu akan memperbesar peluang munculnya minat terhadap produk serupa di masa depan.

2. Faktor Sosial

Minat konsumen juga sangat dipengaruhi oleh lingkungan sosial di sekitarnya. Faktor-faktor seperti keluarga, teman, kelompok referensi, serta status sosial memainkan peran besar dalam membentuk preferensi. Misalnya, seorang remaja mungkin tertarik pada suatu tren fashion tertentu karena dipengaruhi oleh teman sebaya atau figur publik di media sosial. Dalam era digital, pengaruh sosial diperkuat oleh keberadaan platform online, di

mana ulasan, rating, dan rekomendasi konsumen lain dapat mempengaruhi minat calon pembeli.

3. Faktor Pribadi

Faktor pribadi mencakup karakteristik individu seperti usia, jenis kelamin, pekerjaan, pendapatan, gaya hidup, dan tahap siklus hidup. Setiap individu memiliki kebutuhan dan keinginan yang berbeda pada tiap tahapan kehidupannya. Misalnya, minat seorang mahasiswa terhadap produk teknologi akan berbeda dengan minat seorang ibu rumah tangga terhadap produk kebutuhan rumah tangga. Segmentasi berdasarkan faktor ini sering digunakan dalam analisis pasar dan pemetaan target audiens.

4. Faktor Budaya

Budaya membentuk sistem nilai dan norma yang menjadi acuan perilaku masyarakat, termasuk dalam hal konsumsi. Faktor budaya dapat berasal dari latar belakang etnis, agama, tradisi, atau kebiasaan kolektif suatu komunitas. Produk atau layanan yang sesuai dengan nilai budaya cenderung lebih menarik bagi konsumen. Misalnya, makanan halal lebih diminati oleh masyarakat Muslim, atau kampanye ramah lingkungan menarik perhatian konsumen yang menjunjung nilai keberlanjutan.

5. Faktor Ekonomi dan Lingkungan Pasar

Kondisi ekonomi secara langsung mempengaruhi daya beli dan minat konsumen. Ketika ekonomi stabil dan pendapatan meningkat, konsumen cenderung lebih terbuka terhadap produk baru atau berharga tinggi. Sebaliknya, dalam kondisi krisis, konsumen akan lebih tertarik pada produk yang hemat dan fungsional. Selain itu, faktor eksternal seperti tren pasar, iklan, kemudahan akses, serta promosi juga memiliki dampak besar dalam membentuk minat.

2.4 Data Konsumen sebagai Objek Analisis

Dalam era digital yang ditandai dengan aktivitas konsumen yang terekam secara masif, **data konsumen** telah menjadi aset strategis bagi perusahaan. Data ini mencerminkan perilaku, preferensi, kebutuhan, dan bahkan ekspektasi pelanggan terhadap produk atau layanan yang ditawarkan. Oleh karena itu, dalam konteks data mining, data konsumen menjadi objek utama yang dianalisis untuk menggali pola-pola tersembunyi yang dapat mendukung pengambilan keputusan bisnis secara lebih akurat dan terukur.

Data konsumen adalah seluruh informasi yang berkaitan dengan individu atau kelompok pelanggan yang berinteraksi dengan suatu produk, layanan, atau platform bisnis. Informasi ini dapat mencakup data identitas (nama, usia, jenis kelamin), data demografis (lokasi, pekerjaan, penghasilan), data perilaku (riwayat transaksi, frekuensi pembelian, jenis produk yang dibeli), serta data psikografis (minat, gaya hidup, preferensi). Bahkan interaksi konsumen dengan iklan, email promosi, media sosial, atau chatbot juga menjadi sumber penting yang menyumbangkan insight berharga dalam proses analisis.

Data konsumen dikumpulkan dari berbagai titik kontak (touchpoints), baik secara online maupun offline. Di ranah digital, aktivitas konsumen saat menjelajah situs web, melakukan pencarian produk, memberikan ulasan, atau mengklik tautan iklan dapat dilacak dan direkam secara otomatis. Sementara dalam interaksi langsung, data dapat dikumpulkan melalui formulir keanggotaan, kartu loyalitas, survei pelanggan, atau sistem kasir.

Mengapa data konsumen penting sebagai objek analisis? Jawabannya terletak pada kemampuannya dalam merepresentasikan perilaku nyata pelanggan. Alih-alih hanya mengandalkan asumsi atau intuisi, data memungkinkan perusahaan melihat tren aktual, seperti produk apa yang paling diminati pada waktu tertentu, segmen mana yang paling responsif

terhadap diskon, atau bagaimana ulasan negatif berdampak pada penjualan.

Melalui proses data mining, data konsumen dapat dianalisis untuk berbagai tujuan, antara lain:

- **Segmentasi pasar**, yaitu mengelompokkan pelanggan berdasarkan kesamaan tertentu, sehingga strategi pemasaran bisa lebih terarah.
- **Prediksi perilaku pembelian**, seperti memprediksi pelanggan yang akan berhenti berlangganan atau yang berpotensi membeli produk tertentu.
- **Analisis pola asosiasi**, untuk mengetahui produk apa saja yang sering dibeli bersamaan.
- **Analisis loyalitas dan churn**, yaitu mengidentifikasi faktor-faktor yang membuat pelanggan setia atau meninggalkan brand.

Namun, penting untuk dicatat bahwa penggunaan data konsumen harus memperhatikan aspek **etika dan privasi**. Perusahaan wajib menjaga kerahasiaan data pelanggan dan mematuhi regulasi perlindungan data, seperti GDPR (General Data Protection Regulation) atau UU Perlindungan Data Pribadi di berbagai negara. Transparansi dalam pengumpulan dan penggunaan data adalah kunci untuk membangun kepercayaan konsumen.

Dalam praktiknya, kualitas data sangat menentukan hasil analisis. Data yang tidak lengkap, tidak akurat, atau tidak relevan dapat menghasilkan wawasan yang menyesatkan. Oleh karena itu, proses pembersihan, validasi, dan pemeliharaan data menjadi langkah penting sebelum dilakukan penggalian informasi melalui teknik data mining.

BAB III

PERSIAPAN DATA UNTUK ANALISIS

3.1 Pengumpulan Data Konsumen

Pengumpulan data konsumen merupakan tahap awal dan sangat krusial dalam proses data mining. Tanpa data yang tepat dan relevan, analisis yang dilakukan akan menjadi kurang akurat, bahkan bisa menyesatkan pengambilan keputusan. Oleh karena itu, proses ini tidak boleh dilakukan secara sembarangan. Setiap jenis data yang dikumpulkan harus memiliki tujuan yang jelas dan disesuaikan dengan kebutuhan analisis minat konsumen.

Data konsumen dapat dikumpulkan dari berbagai sumber, baik secara langsung maupun tidak langsung. Pengumpulan secara **langsung** dilakukan melalui interaksi dengan konsumen, seperti mengisi formulir keanggotaan, survei, wawancara, atau feedback setelah pembelian. Cara ini memungkinkan perusahaan memperoleh data primer yang lebih spesifik dan sesuai dengan kebutuhan. Namun, metode ini juga memiliki tantangan tersendiri, seperti tingkat partisipasi yang rendah atau adanya bias dalam jawaban konsumen.

Sementara itu, pengumpulan data **tidak langsung** terjadi melalui pelacakan aktivitas konsumen, baik secara online maupun offline. Di dunia digital, teknologi telah memungkinkan perusahaan mengumpulkan data secara otomatis melalui log aktivitas di situs web (web tracking), cookies, aplikasi mobile, hingga interaksi di media sosial. Misalnya, sistem dapat mencatat produk apa saja yang dilihat pelanggan, berapa lama mereka berada di suatu halaman, atau kapan mereka meninggalkan keranjang belanja. Semua informasi ini menjadi petunjuk penting tentang minat dan preferensi pelanggan.

Penggunaan **sistem Point of Sale (POS)** di toko fisik juga menjadi sumber data penting. Dari sistem ini, perusahaan bisa

mengetahui produk apa yang paling sering dibeli, waktu pembelian paling ramai, hingga pola pembelian berulang dari pelanggan tertentu. Integrasi antara data offline dan online (omnichannel data) akan memberikan gambaran yang lebih menyeluruh tentang perilaku konsumen.

Selain dari interaksi langsung, data konsumen juga dapat dikumpulkan melalui pihak ketiga, seperti lembaga riset pasar, penyedia data demografis, atau platform media sosial. Data eksternal ini biasanya digunakan untuk melengkapi informasi internal, misalnya dengan menambahkan data geografis, tren industri, atau sentimen publik terhadap suatu merek.

Namun, dalam proses pengumpulan data konsumen, aspek **etika dan legalitas** harus menjadi perhatian utama. Perusahaan wajib menjelaskan kepada konsumen bahwa data mereka dikumpulkan, serta bagaimana data tersebut akan digunakan. Prinsip transparansi, persetujuan eksplisit (consent), dan perlindungan data pribadi harus ditegakkan sesuai dengan regulasi yang berlaku, seperti **UU Perlindungan Data Pribadi (PDP)** di Indonesia atau **GDPR** di Eropa.

Selain itu, perusahaan perlu memastikan bahwa data yang dikumpulkan benar-benar relevan dan tidak berlebihan. Pengumpulan data yang terlalu banyak tanpa tujuan jelas tidak hanya membebani sistem, tetapi juga meningkatkan risiko keamanan data.

Dengan pengumpulan data yang tepat, sistematis, dan etis, perusahaan akan memiliki fondasi yang kuat untuk menjalankan proses data mining secara efektif. Data yang akurat dan berkualitas tinggi akan menghasilkan wawasan yang lebih dalam, memungkinkan analisis yang lebih tajam terhadap minat konsumen, dan pada akhirnya mendukung pengambilan keputusan yang lebih cerdas dan berbasis bukti.

3.2 Pembersihan dan Transformasi Data

Setelah data konsumen berhasil dikumpulkan, langkah penting berikutnya dalam proses data mining adalah **pembersihan dan transformasi data**. Kedua tahap ini sangat vital untuk memastikan bahwa data yang digunakan dalam analisis adalah akurat, konsisten, dan siap untuk diolah lebih lanjut. Tanpa proses ini, kualitas analisis dapat menurun drastis, karena model yang dibangun di atas data yang kotor atau tidak terstruktur akan menghasilkan wawasan yang tidak valid.

Pembersihan Data (Data Cleaning)

Pembersihan data adalah proses mengidentifikasi dan memperbaiki kesalahan atau inkonsistensi dalam dataset. Kesalahan umum yang sering ditemukan meliputi data duplikat, nilai yang hilang (missing values), kesalahan penulisan, format yang tidak seragam, atau nilai yang tidak masuk akal (outlier). Misalnya, dalam data konsumen, kita bisa saja menemukan nama pelanggan yang sama ditulis dalam berbagai cara, atau tanggal lahir yang tidak logis seperti tahun 1800.

Beberapa teknik yang biasa digunakan dalam pembersihan data antara lain:

- **Penghapusan duplikat:** Menghapus entri yang muncul lebih dari sekali untuk pelanggan yang sama.
- **Pengisian nilai kosong (imputation):** Mengisi data yang hilang dengan rata-rata, median, atau metode prediktif.
- **Validasi format:** Memastikan bahwa data mengikuti format yang konsisten, seperti format tanggal atau kode pos.
- **Deteksi outlier:** Menandai dan mengevaluasi data ekstrem yang mungkin tidak mencerminkan realitas.

Proses pembersihan ini tidak hanya meningkatkan akurasi data, tetapi juga mencegah terjadinya bias atau kesalahan dalam proses analitik selanjutnya.

Transformasi Data (Data Transformation)

Setelah data dibersihkan, langkah selanjutnya adalah transformasi data. Transformasi bertujuan mengubah data ke dalam format atau struktur yang sesuai untuk kebutuhan analisis. Dalam konteks data mining, transformasi dapat mencakup beberapa proses, antara lain:

- **Normalisasi:** Mengubah data numerik ke dalam skala tertentu agar dapat dibandingkan secara adil, misalnya skala 0-1.
- **Pengkodean kategorikal (encoding):** Mengubah data teks atau kategori (seperti jenis kelamin: "Pria", "Wanita") menjadi bentuk numerik agar dapat diproses oleh algoritma.
- **Agregasi:** Menggabungkan beberapa entri data menjadi satu, seperti menghitung total pembelian per pelanggan dalam satu bulan.
- **Reduksi dimensi:** Mengurangi jumlah atribut atau fitur dalam data tanpa kehilangan informasi penting, misalnya dengan teknik PCA (Principal Component Analysis).
- **Pembuatan fitur baru (feature engineering):** Menghasilkan atribut baru dari data yang ada, contohnya membuat variabel “frekuensi pembelian per bulan” dari data transaksi harian.

Transformasi data memungkinkan sistem mengenali pola dan struktur yang tersembunyi dalam data, serta meningkatkan efisiensi algoritma data mining. Data yang telah ditransformasikan menjadi lebih mudah dianalisis dan

memberikan hasil yang lebih bermakna dalam proses eksplorasi maupun pemodelan.

Selain itu, penting untuk mencatat bahwa proses pembersihan dan transformasi bukanlah langkah sekali jalan. Dalam praktiknya, proses ini seringkali dilakukan secara iteratif. Ketika ditemukan masalah baru dalam tahap analisis atau pemodelan, data mungkin perlu dibersihkan atau ditransformasi ulang.

3.3 Seleksi Fitur dan Penentuan Target

Dalam proses data mining, khususnya yang bertujuan untuk menganalisis minat konsumen, tahap **seleksi fitur (feature selection)** dan **penentuan target** merupakan langkah krusial yang sangat memengaruhi keberhasilan model analitik yang dibangun. Kedua proses ini menentukan informasi apa yang paling relevan dari data yang tersedia, dan aspek apa yang ingin diprediksi atau dianalisis lebih lanjut.

Seleksi Fitur

Fitur atau atribut adalah kolom dalam dataset yang merepresentasikan karakteristik tertentu dari konsumen, seperti usia, jenis kelamin, riwayat pembelian, frekuensi kunjungan ke toko, atau preferensi produk. Tidak semua fitur yang ada akan berguna dalam proses analisis; beberapa bisa jadi tidak relevan, redundan, atau bahkan menimbulkan noise (gangguan) yang justru menurunkan kinerja model.

Oleh karena itu, dilakukan proses seleksi fitur untuk menyaring atribut yang paling berpengaruh terhadap minat konsumen. Tujuan utamanya adalah menyederhanakan model, meningkatkan akurasi, dan mengurangi waktu pemrosesan. Seleksi fitur dapat dilakukan secara manual, berdasarkan pengetahuan domain, atau secara otomatis dengan menggunakan algoritma statistik maupun machine learning, seperti:

- **Filter method**, yang menilai fitur berdasarkan korelasi statistik terhadap target variabel.
- **Wrapper method**, yang menguji subset fitur melalui iterasi dan evaluasi kinerja model.
- **Embedded method**, yang mengintegrasikan seleksi fitur dalam proses pelatihan model, seperti pada algoritma Lasso Regression atau decision tree.

Sebagai contoh, jika analisis bertujuan untuk memprediksi minat konsumen terhadap produk fashion, fitur-fitur seperti jenis kelamin, usia, dan histori pembelian kategori fashion akan lebih relevan dibanding fitur yang tidak terkait seperti lokasi pengambilan barang atau metode pembayaran.

Penentuan Target

Target variabel atau **label** adalah variabel utama yang ingin dianalisis atau diprediksi dalam proses data mining. Dalam konteks analisis minat konsumen, target bisa berupa:

- **Preferensi konsumen terhadap produk tertentu** (misalnya: suka/tidak suka, beli/tidak beli),
- **Tingkat kepuasan pelanggan**, atau
- **Kemungkinan konsumen melakukan pembelian ulang.**

Penentuan target harus disesuaikan dengan tujuan bisnis atau penelitian. Jika tujuannya adalah mengklasifikasikan pelanggan yang berpotensi membeli produk baru, maka target variabel yang digunakan bisa berupa status “membeli” atau “tidak membeli” pada data historis. Sedangkan jika tujuannya adalah mengukur intensitas minat, maka target bisa berupa skor atau rating yang diberikan pelanggan.

Penting untuk memastikan bahwa target variabel memiliki kualitas dan distribusi data yang memadai. Misalnya,

jika targetnya sangat tidak seimbang (seperti 95% tidak beli vs. 5% beli), maka perlu dilakukan penyesuaian seperti teknik oversampling atau undersampling untuk menghasilkan model yang lebih seimbang dan akurat.

3.4 Representasi Data: Tabel, CSV, dan Database

Dalam kegiatan data mining, representasi data memainkan peran penting dalam menentukan bagaimana data dapat diakses, diolah, dan dianalisis secara efisien. Representasi yang baik akan mempermudah proses pembersihan, transformasi, seleksi fitur, dan penerapan algoritma analitik. Tiga bentuk representasi data yang umum digunakan dalam analisis data konsumen adalah **tabel**, **file CSV (Comma-Separated Values)**, dan **sistem basis data (database)**. Masing-masing bentuk memiliki karakteristik, kelebihan, dan kegunaan yang berbeda sesuai dengan skala data dan kebutuhan proses analitik.

1. Tabel: Representasi Data Secara Visual dan Konseptual

Tabel merupakan bentuk representasi data yang paling sederhana dan intuitif. Setiap baris (row) merepresentasikan satu entitas atau catatan data, seperti seorang pelanggan, sedangkan setiap kolom (column) merepresentasikan atribut atau fitur dari entitas tersebut, seperti nama, usia, jenis kelamin, produk yang dibeli, dan sebagainya.

Dalam konteks analisis minat konsumen, tabel bisa menggambarkan hubungan antar atribut dengan sangat jelas. Misalnya, kita dapat menyusun sebuah tabel dengan kolom "ID Pelanggan", "Jenis Kelamin", "Kategori Produk Favorit", dan "Frekuensi Pembelian per Bulan". Dengan melihat tabel ini, analis bisa langsung memahami pola dan hubungan antar variabel.

Tabel juga menjadi dasar dalam aplikasi spreadsheet seperti Microsoft Excel atau Google Sheets, yang kerap digunakan untuk analisis awal pada skala data kecil hingga menengah.

2. CSV (Comma-Separated Values): Format Pertukaran Data yang Ringkas

CSV adalah format file berbasis teks yang digunakan untuk menyimpan data dalam bentuk tabular, di mana setiap baris mewakili satu record dan nilai-nilai dalam baris dipisahkan oleh tanda koma (atau delimiter lainnya). CSV sangat populer karena ringan, mudah dibaca manusia maupun mesin, serta dapat diimpor atau diekspor oleh hampir semua perangkat lunak analisis data.

Sebagai contoh, data pembelian konsumen yang diekspor dari sistem e-commerce biasanya disimpan dalam format CSV agar dapat dengan mudah dianalisis menggunakan Python, R, Orange, RapidMiner, atau Excel. CSV juga mendukung interoperabilitas tinggi antar sistem yang berbeda, menjadikannya pilihan utama dalam pertukaran data antar platform.

Namun, kelemahan CSV adalah tidak mendukung hubungan antar data (relational) dan tidak memiliki pengamanan atau validasi data yang kuat. Oleh karena itu, CSV lebih cocok digunakan untuk penyimpanan data datar (flat data) atau sebagai format pertukaran, bukan untuk manajemen data jangka panjang.

3. Database: Solusi untuk Skala Besar dan Kompleks

Basis data atau **database** adalah sistem yang dirancang untuk menyimpan, mengelola, dan mengakses data secara efisien dan terstruktur. Dalam konteks bisnis besar atau data konsumen skala besar, penggunaan database sangat penting untuk memastikan integritas, keamanan, dan efisiensi pengambilan data.

Sistem manajemen basis data seperti MySQL, PostgreSQL, dan MongoDB memungkinkan penyimpanan data

dalam skema relasional atau non-relasional (NoSQL), tergantung pada kompleksitas dan jenis data yang dikelola. Database mendukung query kompleks, transaksi, indexing, dan relasi antar tabel yang sangat berguna dalam proses analitik.

BAB IV

PERSIAPAN DATA UNTUK KLASIFIKASI KEPUASAN

4.1 Clustering untuk Segmentasi Konsumen

Salah satu pendekatan penting dalam data mining untuk memahami perilaku dan minat konsumen adalah melalui **clustering**, yaitu proses pengelompokan data ke dalam kelompok (klaster) yang memiliki karakteristik serupa. Dalam konteks analisis konsumen, clustering banyak digunakan untuk **segmentasi konsumen**, yaitu mengelompokkan pelanggan berdasarkan kesamaan atribut seperti preferensi, kebiasaan belanja, demografi, atau perilaku interaksi dengan produk.

Pengertian Clustering

Clustering adalah metode unsupervised learning, yang berarti tidak memerlukan label atau target tertentu dalam proses pelatihan model. Sistem akan secara otomatis menemukan pola atau struktur dalam data dan mengelompokkan entitas (dalam hal ini, konsumen) ke dalam klaster yang memiliki kesamaan tinggi secara internal namun berbeda signifikan dengan klaster lainnya.

Sebagai contoh, jika kita memiliki data pembelian konsumen yang mencakup frekuensi pembelian, total pengeluaran, dan jenis produk yang dibeli, algoritma clustering dapat mengidentifikasi kelompok konsumen “hemat”, “impulsif”, atau “loyal” berdasarkan pola perilaku mereka.

Algoritma Clustering Populer

Beberapa algoritma clustering yang sering digunakan dalam segmentasi konsumen antara lain:

- **K-Means Clustering:** Mengelompokkan data ke dalam k kluster berdasarkan jarak (biasanya Euclidean) ke pusat kluster (centroid). K-Means populer karena sederhana dan efisien, tetapi memerlukan penentuan jumlah kluster (k) di awal.
- **Hierarchical Clustering:** Membangun pohon kluster secara bertingkat tanpa perlu menentukan jumlah kluster sejak awal. Cocok untuk visualisasi struktur hubungan antar konsumen.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise):** Cocok untuk data yang memiliki bentuk kluster yang tidak beraturan. Dapat mengidentifikasi outlier (konsumen yang tidak termasuk dalam kluster mana pun).

Manfaat Clustering untuk Segmentasi

Segmentasi konsumen melalui clustering memberikan banyak manfaat strategis bagi perusahaan dan pemasar, antara lain:

1. **Personalisasi Pemasaran:** Setiap segmen dapat diberi pendekatan pemasaran yang disesuaikan dengan karakteristiknya. Misalnya, konsumen loyal dapat diberi program reward, sedangkan konsumen baru diberi diskon khusus.
2. **Pengembangan Produk:** Wawasan dari setiap segmen dapat membantu perusahaan dalam merancang produk yang lebih sesuai dengan kebutuhan spesifik kelompok tertentu.
3. **Efisiensi Sumber Daya:** Dengan mengetahui kelompok mana yang paling potensial, perusahaan dapat mengalokasikan anggaran pemasaran secara lebih efektif.

4. **Peningkatan Kepuasan Pelanggan:** Layanan dan komunikasi yang disesuaikan per segmen dapat meningkatkan pengalaman dan kepuasan pelanggan.

Tantangan dalam Clustering

Meski powerful, clustering memiliki tantangan tersendiri. Pemilihan jumlah kluster yang tepat tidak selalu mudah dan bisa memerlukan metode evaluasi seperti Elbow Method atau Silhouette Score. Selain itu, hasil clustering sangat bergantung pada kualitas data dan pemilihan fitur yang tepat. Oleh karena itu, proses pembersihan data dan seleksi fitur yang baik menjadi sangat penting sebelum melakukan clustering.

4.2 Classification untuk Prediksi Minat

Salah satu pendekatan utama dalam data mining untuk memahami dan memprediksi perilaku konsumen adalah melalui metode **classification**. Teknik ini termasuk dalam kelompok **supervised learning**, di mana model dilatih menggunakan dataset yang telah dilabeli sebelumnya, dengan tujuan memprediksi kategori atau kelas dari data baru berdasarkan pola yang telah dipelajari.

Dalam konteks **analisis minat konsumen**, classification digunakan untuk memprediksi apakah seorang konsumen tertarik atau tidak terhadap suatu produk, layanan, atau kampanye pemasaran tertentu. Misalnya, berdasarkan riwayat pembelian, demografi, dan perilaku online, perusahaan dapat memprediksi apakah konsumen kemungkinan besar akan membeli produk baru atau tidak.

Konsep Dasar Classification

Classification bekerja dengan membangun model prediktif berdasarkan data latih (training data) yang telah memiliki label atau kategori tertentu. Model ini kemudian

digunakan untuk mengklasifikasikan data baru yang belum diketahui kelasnya. Contoh penggunaan dalam dunia bisnis adalah mengklasifikasikan konsumen ke dalam kategori “berminat” dan “tidak berminat” terhadap suatu produk.

Proses classification melibatkan dua tahap utama:

1. **Training:** Model belajar dari data historis, memahami hubungan antara fitur (seperti umur, pendapatan, frekuensi pembelian) dan target label (misalnya, minat terhadap produk).
2. **Testing/Prediction:** Model diuji menggunakan data baru untuk memprediksi kelas dari data tersebut berdasarkan pola yang telah dipelajari.

Algoritma Classification Populer

Beberapa algoritma yang umum digunakan dalam classification untuk prediksi minat konsumen antara lain:

- **Decision Tree:** Membangun struktur pohon keputusan berdasarkan fitur-fitur penting. Sangat mudah dipahami dan divisualisasikan, cocok untuk interpretasi bisnis.
- **Naïve Bayes:** Berdasarkan prinsip probabilitas Bayes dan asumsi independensi antar fitur. Cepat dan efisien untuk dataset besar.
- **Support Vector Machine (SVM):** Mencari garis batas optimal (hyperplane) yang memisahkan kelas-kelas dalam ruang data. Cocok untuk data dengan margin yang jelas antar kelas.
- **K-Nearest Neighbors (KNN):** Menentukan kelas dari data baru berdasarkan kedekatannya dengan data lain dalam ruang fitur.

- **Random Forest:** Gabungan dari banyak pohon keputusan yang memberikan prediksi melalui voting. Akurat dan tahan terhadap overfitting.

Manfaat Classification dalam Analisis Minat

Menggunakan classification untuk prediksi minat konsumen memiliki banyak manfaat, antara lain:

- **Personalisasi Strategi:** Menyesuaikan promosi atau rekomendasi produk berdasarkan prediksi ketertarikan konsumen.
- **Efisiensi Biaya:** Menargetkan hanya pelanggan yang memiliki kemungkinan tinggi untuk tertarik, sehingga menghemat anggaran pemasaran.
- **Peningkatan Penjualan:** Dengan mengenali konsumen yang berpotensi berminat, perusahaan dapat meningkatkan konversi penjualan secara signifikan.
- **Pengambilan Keputusan Berbasis Data:** Classification memberikan dasar yang kuat untuk membuat keputusan bisnis yang lebih presisi.

Tantangan dalam Classification

Meski powerful, classification menghadapi beberapa tantangan, seperti ketidakseimbangan data (imbalance class), noise, dan pemilihan fitur yang tepat. Data konsumen yang tidak lengkap atau tidak relevan dapat menyebabkan model membuat prediksi yang salah. Oleh karena itu, penting untuk melakukan **pembersihan data** dan **seleksi fitur** secara menyeluruh sebelum menerapkan algoritma klasifikasi.

4.3 Association Rules untuk Pola Pembelian

Association rules atau aturan asosiasi merupakan salah satu teknik utama dalam data mining yang digunakan untuk menemukan hubungan tersembunyi atau pola menarik antar item dalam kumpulan data. Dalam konteks analisis minat konsumen, teknik ini sangat berguna untuk mengidentifikasi **pola pembelian konsumen**, seperti produk-produk apa saja yang sering dibeli bersamaan. Temuan ini tidak hanya memberikan wawasan terhadap perilaku konsumen, tetapi juga menjadi dasar dalam penerapan strategi bisnis seperti *cross-selling*, *product bundling*, dan penataan produk di rak toko (planogram).

Contoh paling terkenal dari penggunaan association rules adalah analisis keranjang belanja (*market basket analysis*). Misalnya, jika data menunjukkan bahwa konsumen yang membeli roti juga sering membeli mentega, maka aturan asosiasi seperti "*Jika membeli roti, maka kemungkinan membeli mentega*" dapat dibentuk. Aturan semacam ini membantu toko ritel dalam merancang promosi, mengatur tata letak produk, atau memberikan rekomendasi pembelian di platform e-commerce.

Konsep Dasar Association Rules

Aturan asosiasi berbentuk implikasi logis dari jenis:

Jika X, maka Y

Di mana X dan Y adalah itemset (kumpulan item). Dalam konteks data pembelian, X bisa berupa satu atau lebih produk yang dibeli, dan Y adalah produk lain yang kemungkinan dibeli bersamaan.

Tiga metrik utama yang digunakan untuk menilai kualitas aturan asosiasi adalah:

1. **Support:** Mengukur seberapa sering kombinasi item X dan Y muncul bersama dalam dataset. Ini menunjukkan relevansi aturan terhadap keseluruhan data.

2. **Confidence:** Menyatakan seberapa sering item Y muncul setelah X muncul. Ini mengukur keandalan aturan tersebut.
3. **Lift:** Mengukur kekuatan asosiasi antara X dan Y, dengan membandingkan confidence terhadap probabilitas Y terjadi secara umum. Nilai lift >1 menandakan adanya hubungan yang kuat.

Algoritma Populer

Algoritma yang paling umum digunakan dalam penerapan association rules adalah:

- **Apriori:** Menghasilkan aturan asosiasi dengan menggunakan prinsip bahwa subset dari itemset yang sering muncul juga sering terjadi. Ini membantu mempercepat pencarian pola dengan memotong kombinasi yang tidak relevan.
- **FP-Growth (Frequent Pattern Growth):** Alternatif yang lebih efisien daripada Apriori karena tidak perlu menghasilkan kandidat itemset secara eksplisit. Cocok untuk dataset besar.

Manfaat dalam Dunia Bisnis

Implementasi association rules memiliki berbagai manfaat strategis, antara lain:

- **Rekomendasi Produk:** Memberikan saran produk tambahan yang relevan saat konsumen berbelanja, meningkatkan nilai keranjang belanja.
- **Promosi dan Diskon:** Menawarkan diskon untuk produk yang sering dibeli bersamaan, guna meningkatkan volume penjualan.
- **Perencanaan Inventaris:** Menyesuaikan stok berdasarkan pola pembelian konsumen.

- **Tata Letak Toko:** Menempatkan produk-produk yang sering dibeli bersama di lokasi yang berdekatan.

Tantangan dan Pertimbangan

Meskipun powerful, association rules memiliki tantangan seperti:

- Banyaknya aturan yang mungkin tidak relevan jika tidak disaring dengan baik.
- Potensi overfitting jika hanya mengandalkan aturan tanpa mempertimbangkan konteks bisnis.
- Kebutuhan akan data transaksi yang lengkap dan akurat.

4.4 Sentiment Analysis dari Review atau Sosial Media

Sentiment analysis atau analisis sentimen merupakan salah satu pendekatan dalam data mining yang digunakan untuk memahami opini, emosi, dan sikap pengguna terhadap suatu produk, layanan, atau merek berdasarkan teks yang mereka tulis. Sumber utama data untuk analisis sentimen biasanya berasal dari **ulasan pelanggan (reviews)**, komentar di **media sosial**, forum diskusi, dan berbagai platform digital lainnya. Dalam konteks analisis minat konsumen, teknik ini berperan penting untuk mengungkap preferensi tersembunyi, kepuasan, dan ketidakpuasan konsumen secara real-time.

Di era digital, konsumen semakin aktif mengekspresikan pendapat mereka secara terbuka. Mereka menuliskan pengalaman belanja, menyatakan suka atau tidak suka terhadap produk, bahkan memberikan kritik terhadap pelayanan perusahaan. Semua teks tersebut adalah **sumber data tidak terstruktur** yang sangat kaya informasi. Tantangannya adalah bagaimana mengekstrak makna dari data tersebut secara sistematis dan terukur — inilah peran utama dari sentiment analysis.

Bagaimana Sentiment Analysis Bekerja

Proses sentiment analysis umumnya dimulai dengan **pengumpulan data teks** dari sumber digital. Setelah data dikumpulkan, dilakukan **pembersihan data (text preprocessing)**, termasuk penghapusan tanda baca, stop words, konversi ke huruf kecil, serta stemming atau lemmatization.

Langkah berikutnya adalah **klasifikasi sentimen**, di mana sistem berusaha mengidentifikasi apakah suatu teks bernada **positif, negatif, atau netral**. Teknik yang digunakan bisa berupa pendekatan berbasis **kamus (lexicon-based)** atau **pembelajaran mesin (machine learning)**. Pada pendekatan lexicon-based, sistem mencocokkan kata-kata dalam teks dengan daftar kata yang telah diberi label sentimen. Sementara pendekatan machine learning melatih model berdasarkan dataset contoh yang telah diberi label sentimen.

Model yang lebih canggih seperti **Deep Learning** (misalnya LSTM atau BERT) bahkan dapat memahami konteks kalimat yang lebih kompleks, seperti ironi atau sarkasme, yang biasanya sulit ditangkap oleh metode sederhana.

Manfaat dalam Analisis Minat Konsumen

Dengan menerapkan sentiment analysis, perusahaan dapat memperoleh **gambaran langsung tentang persepsi konsumen**, termasuk:

- **Mengukur kepuasan pelanggan** secara real-time dari ulasan produk.
- **Mendeteksi isu layanan atau produk** lebih cepat sebelum berkembang menjadi krisis.
- **Mengidentifikasi fitur produk yang disukai atau tidak disukai** pelanggan.
- **Mengamati tren opini publik** terhadap kampanye pemasaran atau peluncuran produk baru.

Data ini dapat diintegrasikan dengan teknik data mining lainnya seperti **clustering** atau **classification** untuk menyempurnakan segmentasi pasar atau prediksi perilaku konsumen.

Tantangan dan Etika

Meskipun sangat berguna, sentiment analysis memiliki tantangan, terutama dalam **menghadapi bahasa informal, slang, emoji**, atau kalimat ambigu. Selain itu, perusahaan juga perlu menjaga **privasi data konsumen**, terutama saat mengambil data dari media sosial atau forum publik. Oleh karena itu, etika dalam pengumpulan dan analisis data harus menjadi perhatian utama.

4.5 Visualisasi Hasil Data Mining

Dalam proses data mining, penggalian pola dan pengetahuan dari data tidak akan optimal tanpa penyajian yang jelas dan mudah dipahami. Di sinilah **visualisasi data** memegang peranan penting. Visualisasi hasil data mining merupakan tahap krusial yang membantu pengguna—baik analis data, manajer, hingga pemangku kebijakan—memahami temuan secara intuitif, menarik, dan cepat. Tanpa visualisasi, hasil analisis yang kompleks hanya akan menjadi angka atau model matematis yang sulit diinterpretasikan.

Visualisasi data adalah teknik untuk menyajikan data atau informasi dalam bentuk grafis seperti grafik, diagram, peta, dan dashboard interaktif. Tujuan utamanya adalah **menerjemahkan data kuantitatif maupun kualitatif menjadi bentuk visual** yang mudah dimengerti oleh manusia, terutama dalam konteks pengambilan keputusan.

Peran Visualisasi dalam Data Mining

Setelah proses data mining seperti clustering, classification, association rule mining, atau sentiment analysis

dilakukan, hasilnya biasanya berupa model, pola, atau relasi kompleks. Dengan visualisasi, hasil tersebut bisa direpresentasikan secara lebih bermakna, misalnya:

- **Cluster hasil segmentasi konsumen** dapat divisualisasikan dalam bentuk scatter plot dua atau tiga dimensi untuk menunjukkan batas-batas antara kelompok.
- **Pohon keputusan** dari algoritma classification dapat ditampilkan dalam bentuk diagram hierarkis yang mudah diikuti.
- **Aturan asosiasi** dapat divisualisasikan dengan network graph atau heatmap untuk menunjukkan keterkaitan antara produk.
- **Sentimen konsumen** dari media sosial dapat disajikan dalam bentuk word cloud, grafik batang, atau garis tren waktu.

Visualisasi juga membantu **mengevaluasi kinerja model**. Misalnya, confusion matrix dan ROC curve sangat berguna dalam menilai model klasifikasi, sedangkan elbow method atau silhouette plot digunakan untuk mengevaluasi hasil clustering.

Alat dan Teknologi Visualisasi

Beberapa alat populer yang digunakan untuk visualisasi hasil data mining antara lain:

- **Tableau dan Power BI:** Cocok untuk visualisasi interaktif dan dashboard analitik.
- **Python (Matplotlib, Seaborn, Plotly):** Digunakan dalam lingkungan pemrograman untuk visualisasi kustom.
- **R (ggplot2, shiny):** Digunakan dalam analisis statistik dan presentasi grafis.
- **Orange Data Mining:** Platform visual data mining yang menyediakan visualisasi otomatis dari hasil analisis.

Manfaat Visualisasi dalam Konteks Konsumen

Dalam analisis minat konsumen, visualisasi hasil data mining memberikan banyak keuntungan, seperti:

- Menyajikan segmentasi pelanggan dalam bentuk yang mudah dipahami oleh tim pemasaran.
- Menunjukkan tren minat konsumen terhadap produk tertentu dari waktu ke waktu.
- Membantu mengidentifikasi anomali atau pola yang tidak biasa dalam perilaku pelanggan.
- Memfasilitasi komunikasi antar tim lintas departemen dalam menyusun strategi bisnis berbasis data.

BAB V

STUDI KASUS ANALISIS MINAT KONSUMEN

5.1 Studi Kasus E-Commerce: Pola Pembelian Konsumen

Dalam era digital saat ini, platform e-commerce menjadi salah satu sumber data konsumen paling kaya dan dinamis. Setiap klik, pencarian, penambahan produk ke keranjang, hingga transaksi pembelian menciptakan jejak data yang dapat dimanfaatkan untuk memahami perilaku pelanggan secara mendalam. Melalui penerapan data mining, perusahaan e-commerce dapat mengungkap **pola pembelian konsumen** yang berguna untuk meningkatkan pengalaman belanja, menyusun strategi pemasaran yang lebih efektif, serta meningkatkan loyalitas pelanggan.

Studi kasus ini mengangkat sebuah perusahaan e-commerce yang ingin mengetahui kebiasaan pembelian pelanggan selama satu tahun terakhir. Data yang digunakan mencakup: ID pelanggan, waktu dan frekuensi transaksi, kategori produk yang dibeli, nilai transaksi, lokasi pelanggan, serta waktu pembelian (jam dan hari).

Tahapan Analisis

Proses data mining diawali dengan **pembersihan dan transformasi data**, mengingat data yang diperoleh mengandung duplikasi, nilai kosong, dan format yang tidak seragam. Setelah dilakukan normalisasi dan penghapusan outlier, data kemudian dianalisis menggunakan teknik **association rule mining** untuk menemukan keterkaitan antara produk yang sering dibeli bersamaan. Misalnya, ditemukan bahwa pelanggan yang membeli "popok bayi" juga cenderung membeli "tisu basah" dan "susu formula" dalam transaksi yang sama.

Selain itu, digunakan teknik **clustering (pengelompokan)** untuk mengidentifikasi segmen pelanggan berdasarkan perilaku pembelian. Hasil analisis menunjukkan terdapat tiga segmen utama: (1) pelanggan setia yang melakukan pembelian rutin bulanan, (2) pelanggan musiman yang hanya aktif saat diskon besar, dan (3) pelanggan impulsif yang cenderung membeli dalam jumlah besar tetapi jarang.

Visualisasi data dilakukan menggunakan dashboard interaktif yang menampilkan heatmap waktu pembelian (terlihat bahwa puncak transaksi terjadi antara pukul 20.00–22.00 dan saat akhir pekan), serta grafik batang kategori produk terlaris per bulan.

Insight dan Manfaat Bisnis

Dari hasil data mining, diperoleh beberapa insight penting:

- **Produk bundling** dapat ditingkatkan berdasarkan pola pembelian bersama (misalnya, menawarkan paket popok + tisu + susu dengan diskon).
- **Segmentasi pelanggan** memungkinkan perusahaan merancang promosi berbeda untuk masing-masing segmen, seperti program loyalitas untuk pelanggan setia dan penawaran kilat untuk pelanggan impulsif.
- **Pengaturan stok barang** menjadi lebih akurat karena adanya prediksi permintaan berdasarkan waktu, kategori, dan wilayah.
- **Pengiriman notifikasi** dan iklan bertarget dapat disesuaikan berdasarkan perilaku dan preferensi pengguna yang terdeteksi melalui data historis.

5.2 Studi Kasus Retail: Produk Favorit Berdasarkan Lokasi

Dalam industri retail yang memiliki banyak cabang di berbagai daerah, memahami preferensi konsumen berdasarkan lokasi menjadi kunci strategi pemasaran yang efektif. Data mining memungkinkan perusahaan retail untuk menggali pola pembelian yang unik di tiap wilayah, sehingga dapat menyesuaikan stok, promosi, dan penempatan produk sesuai dengan karakteristik lokal. Studi kasus ini menggambarkan bagaimana sebuah jaringan retail nasional menggunakan data mining untuk mengetahui **produk favorit berdasarkan lokasi geografis**.

Perusahaan retail yang dikaji memiliki lebih dari 100 gerai yang tersebar di kota-kota besar dan kecil di seluruh Indonesia. Data yang dianalisis mencakup transaksi penjualan selama 12 bulan, dengan atribut: kode produk, nama produk, kategori, jumlah terjual, nilai penjualan, waktu transaksi, serta lokasi toko (provinsi dan kota/kabupaten).

Proses Analisis

Tahap pertama adalah **pembersihan data**, menghapus entri duplikat, mengisi nilai kosong, dan menyamakan format data lokasi agar seragam (misalnya, menyatukan ejaan yang berbeda dari nama kota). Selanjutnya, dilakukan agregasi data untuk menghitung **frekuensi dan volume penjualan per produk di setiap lokasi**.

Setelah data siap, digunakan teknik **clustering geografis** untuk mengelompokkan lokasi berdasarkan pola pembelian produk. Hasil clustering menunjukkan bahwa preferensi konsumen di daerah urban cenderung berbeda dari daerah suburban dan rural. Contohnya, produk makanan siap saji dan minuman instan menempati posisi teratas di Jakarta dan Surabaya, sementara di daerah seperti Padang dan Banjarmasin, produk sembako dan makanan tradisional lebih dominan.

Selain clustering, digunakan pula **frequent pattern mining** untuk mengidentifikasi kombinasi produk yang sering dibeli bersamaan di setiap wilayah. Di Yogyakarta, misalnya, ditemukan bahwa konsumen sering membeli teh kemasan bersama camilan ringan lokal, sedangkan di Bandung, pembelian pakaian sering dikaitkan dengan aksesoris seperti tas dan topi.

Visualisasi dan Insight

Hasil data mining kemudian divisualisasikan dalam bentuk **peta interaktif** dan **diagram batang**, yang menunjukkan produk terfavorit per kota. Warna dan ukuran lingkaran di peta menunjukkan volume penjualan, sementara grafik batang menampilkan produk terlaris per provinsi.

Beberapa insight utama yang diperoleh antara lain:

- **Distribusi produk bisa disesuaikan secara geografis**, menghindari penumpukan stok produk yang kurang laku di daerah tertentu.
- **Promosi lokal** dapat difokuskan pada produk yang sudah terbukti populer di wilayah tersebut, sehingga meningkatkan efektivitas kampanye pemasaran.
- **Pengembangan produk baru** bisa disesuaikan dengan tren lokal, seperti mengadaptasi kemasan atau rasa sesuai preferensi daerah.
- **Optimalisasi rantai pasok** dapat dicapai dengan memprioritaskan pengiriman produk yang paling dibutuhkan per lokasi.

5.3 Studi Kasus Sosial Media: Analisis Ulasan Produk

Di era digital, media sosial telah menjadi ruang publik utama bagi konsumen untuk menyampaikan pendapat, kritik, dan rekomendasi terhadap produk atau layanan. Informasi yang

terkandung dalam ulasan dan komentar tersebut menjadi sumber data yang sangat berharga bagi perusahaan. Dengan menerapkan teknik data mining, khususnya **sentiment analysis** dan **text mining**, perusahaan dapat menggali opini konsumen secara otomatis dan sistematis. Studi kasus ini mengeksplorasi bagaimana analisis ulasan produk di media sosial dapat memberikan wawasan mendalam terhadap persepsi konsumen dan membantu pengambilan keputusan bisnis.

Objek studi adalah sebuah merek produk perawatan kulit lokal yang sedang berkembang dan aktif memasarkan produknya di media sosial, khususnya Instagram dan Twitter. Perusahaan ingin mengetahui bagaimana respons publik terhadap peluncuran produk baru mereka—serum wajah berbahan alami—dengan cara menganalisis ribuan ulasan dan komentar yang ditinggalkan oleh pengguna di media sosial dalam periode dua bulan setelah peluncuran.

Langkah-langkah Analisis

1. Pengumpulan Data

Data dikumpulkan menggunakan **web scraping** dan **API media sosial** untuk menarik komentar, tweet, mention, dan tag yang mengandung nama produk atau merek. Hasilnya, terkumpul lebih dari 10.000 entri teks pendek yang terdiri dari komentar langsung, ulasan pengguna, dan percakapan.

2. Pembersihan dan Pra-pemrosesan

Teks ulasan dibersihkan dari karakter spesial, tautan, dan hashtag yang tidak relevan. Kemudian dilakukan tokenisasi, normalisasi kata (seperti mengubah “bgus” menjadi “bagus”), serta penghapusan stopwords untuk meningkatkan kualitas analisis.

3. Analisis Sentimen

Digunakan teknik **machine learning berbasis supervised learning** untuk mengklasifikasikan setiap ulasan menjadi tiga kategori: positif, negatif, dan netral. Model dilatih dengan dataset lokal berbahasa Indonesia yang telah dilabeli sebelumnya. Hasilnya, sekitar **68% ulasan bersentimen positif, 20% netral, dan 12% negatif.**

4. Ekstraksi Topik dan Kata Kunci

Dengan menggunakan algoritma seperti **TF-IDF** dan **LDA (Latent Dirichlet Allocation)**, diidentifikasi kata kunci yang paling sering muncul. Kata seperti "ringan", "tidak lengket", dan "lembab" sering muncul dalam ulasan positif. Sedangkan ulasan negatif sering mengandung kata seperti "iritasi", "bau menyengat", atau "tidak cocok".

Insight yang Didapat

- **Produk mendapatkan sambutan baik**, terutama karena bahan alami dan efek melembabkan yang dirasakan konsumen.
- **Keluhan utama berasal dari sebagian kecil pengguna yang mengalami iritasi kulit**, menandakan perlunya penyesuaian formula untuk jenis kulit sensitif.
- **Ulasan positif cenderung mendorong pembelian dari konsumen baru**, terbukti dari meningkatnya mention produk seiring dengan peningkatan penjualan.
- **Waktu terbaik untuk promosi ditemukan saat puncak aktivitas ulasan**, yaitu malam hari dan akhir pekan.

5.4 Studi Kasus Kuesioner Online: Minat terhadap Produk Baru

Kuesioner online merupakan salah satu metode paling praktis dan efisien dalam menjangkau opini konsumen, terutama ketika perusahaan ingin mengukur minat terhadap produk baru sebelum peluncuran resmi. Dalam studi kasus ini, akan dianalisis bagaimana data kuesioner yang dikumpulkan secara daring dapat dimanfaatkan melalui pendekatan data mining untuk mengidentifikasi preferensi konsumen, segmentasi pasar, dan potensi sukses produk baru di pasaran.

Perusahaan yang menjadi objek studi adalah produsen minuman kesehatan yang berencana meluncurkan varian baru jus probiotik. Sebelum peluncuran, tim pemasaran menyebarkan kuesioner secara online melalui media sosial, milis pelanggan, dan komunitas gaya hidup sehat. Kuesioner tersebut mencakup berbagai aspek: tingkat minat terhadap produk, kebiasaan konsumsi minuman sehat, preferensi rasa, harga yang diharapkan, serta demografi responden.

Langkah-langkah Analisis

1. Pengumpulan Data

Sebanyak 3.000 responden berhasil mengisi kuesioner. Data dikumpulkan dalam bentuk file CSV yang mencakup data kuantitatif (skala Likert untuk minat dan frekuensi konsumsi) serta data kualitatif (jawaban terbuka mengenai ekspektasi konsumen).

2. Pembersihan dan Transformasi Data

Langkah awal dilakukan untuk memastikan kualitas data, termasuk penghapusan entri duplikat, penyatuan format penulisan (misalnya untuk jenis kelamin atau usia), dan transformasi nilai kategori menjadi format numerik agar dapat diproses dalam model data mining.

3. Analisis Deskriptif dan Visualisasi Awal

Hasil awal menunjukkan bahwa 72% responden menyatakan ketertarikan terhadap produk baru tersebut. Preferensi rasa terbanyak adalah mangga dan jeruk. Mayoritas responden menyatakan bersedia membeli produk dalam rentang harga Rp10.000–Rp15.000. Segmentasi usia 20–35 tahun merupakan kelompok paling antusias terhadap produk ini.

4. Clustering dan Segmentasi Konsumen

Menggunakan algoritma **K-Means**, responden dikelompokkan menjadi tiga segmen utama:

- **Health Enthusiasts:** rutin mengonsumsi minuman sehat dan menunjukkan minat sangat tinggi.
- **Price Sensitive:** tertarik tapi sangat mempertimbangkan harga.
- **Casual Drinkers:** minat rendah dan konsumsi sesekali.

5. Prediksi Minat Menggunakan Classification

Model klasifikasi berbasis **Decision Tree** dilatih untuk memprediksi minat beli berdasarkan kombinasi variabel demografi dan kebiasaan konsumsi. Akurasi model mencapai 85%, menunjukkan bahwa pola minat cukup konsisten dan bisa diprediksi.

Insight dan Rekomendasi

Dari hasil analisis data kuesioner, perusahaan mendapat pemahaman yang jelas mengenai segmen konsumen utama yang menjadi target: kelompok usia muda dengan gaya hidup sehat dan preferensi rasa tropis. Strategi pemasaran disarankan untuk difokuskan pada platform digital yang populer di kalangan usia

tersebut. Selain itu, variasi harga dapat disesuaikan dengan sensitivitas segmen yang lebih memperhatikan aspek ekonomi.

BAB VI

TOOLS DAN PLATFORM UNTUK DATA MINING KONSUMEN

6.1 RapidMiner

RapidMiner adalah salah satu platform *open-source* terkemuka yang dirancang untuk analisis data, *machine learning*, dan data mining. Platform ini sangat populer di kalangan praktisi data karena kemampuannya yang kuat dalam melakukan berbagai tahapan pemrosesan data, mulai dari pengumpulan, pembersihan, transformasi, hingga analisis dan visualisasi hasil. RapidMiner memungkinkan pengguna untuk membangun model analitik secara cepat dan efisien melalui antarmuka berbasis *drag-and-drop*, tanpa harus menulis kode secara manual, meskipun tetap mendukung skrip jika diperlukan.

Salah satu keunggulan utama RapidMiner adalah arsitekturnya yang modular dan fleksibel. Pengguna dapat memilih dan menggabungkan berbagai operator (komponen fungsional) dalam sebuah proses untuk mencapai tujuan analisis yang diinginkan. Operator ini meliputi fungsi untuk *data preprocessing*, seleksi fitur, klasifikasi, klustering, regresi, validasi model, hingga pelaporan hasil. Proses ini divisualisasikan dalam bentuk diagram alur (*workflow*), yang memudahkan pemahaman logika analisis bahkan bagi pengguna pemula.

RapidMiner mendukung berbagai teknik *machine learning* dan algoritma data mining seperti Decision Tree, Naïve Bayes, k-NN, Support Vector Machine (SVM), K-Means Clustering, dan Association Rules. Tidak hanya itu, platform ini juga kompatibel dengan bahasa pemrograman seperti Python dan

R, sehingga pengguna tingkat lanjut dapat memperluas fungsionalitasnya dengan skrip tambahan. Hal ini menjadikan RapidMiner cocok digunakan baik oleh pemula yang ingin belajar data science, maupun profesional yang bekerja di industri dengan kebutuhan kompleks.

Dalam konteks bisnis dan pemasaran, RapidMiner sangat berguna untuk analisis minat dan perilaku konsumen. Misalnya, dengan menggunakan data historis transaksi, pengguna dapat membangun model prediksi untuk mengetahui preferensi pelanggan terhadap produk tertentu. Selain itu, analisis sentimen dari ulasan pelanggan di media sosial atau marketplace juga bisa dilakukan dengan mudah melalui modul-modul yang tersedia di RapidMiner.

RapidMiner juga menyediakan fitur untuk *cross-validation* dan pengukuran performa model, sehingga pengguna dapat membandingkan efektivitas berbagai teknik analisis dalam memecahkan masalah tertentu. Misalnya, dalam proyek klasifikasi minat konsumen, pengguna dapat membandingkan akurasi antara algoritma Decision Tree dan SVM, lalu memilih model terbaik berdasarkan metrik evaluasi seperti *precision*, *recall*, atau *ROC curve*.

Kelebihan lain dari RapidMiner adalah komunitas penggunanya yang besar dan dokumentasi yang lengkap. Tersedia banyak tutorial, video pembelajaran, dan forum diskusi yang dapat membantu pengguna dalam memahami dan mengembangkan proyek mereka. Selain versi gratis (RapidMiner Studio), tersedia pula versi komersial dengan fitur tambahan seperti konektivitas ke database besar, kolaborasi tim, dan kapasitas pemrosesan yang lebih tinggi.

6.2 Orange

Orange adalah perangkat lunak open-source untuk *data mining* dan *machine learning* yang menawarkan pendekatan visual dan intuitif dalam proses analisis data. Dikembangkan menggunakan bahasa pemrograman Python, Orange menjadi salah satu alat favorit dalam dunia data science, khususnya bagi pemula, pendidik, dan peneliti yang menginginkan eksplorasi data secara cepat dan interaktif tanpa harus menulis kode secara intensif.

Salah satu daya tarik utama dari Orange adalah **antarmuka grafis berbasis blok** (widget) yang memungkinkan pengguna membangun alur kerja analisis data (workflow) hanya dengan menyeret dan menghubungkan komponen visual. Tiap komponen mewakili langkah analitik tertentu seperti membaca data, membersihkan data, melakukan klasifikasi, klusterisasi, analisis asosiasi, dan visualisasi hasil. Dengan pendekatan ini, Orange memungkinkan pengguna untuk fokus pada pemahaman data dan interpretasi hasil, bukan pada aspek teknis pemrograman.

Orange mendukung berbagai **algoritma machine learning** yang umum digunakan, seperti *Naïve Bayes*, *Decision Tree*, *k-Nearest Neighbors (k-NN)*, *Support Vector Machines (SVM)*, *Random Forest*, dan *Logistic Regression*. Untuk segmentasi atau pengelompokan data, tersedia algoritma seperti *k-Means Clustering* dan *Hierarchical Clustering*. Selain itu, Orange juga menyediakan alat evaluasi model seperti *Confusion Matrix*, *ROC Curve*, dan *Cross-Validation*, yang memudahkan pengguna dalam menilai performa model yang dibangun.

Platform ini juga sangat cocok digunakan dalam konteks **analisis minat konsumen**, terutama jika datanya berasal dari survei, transaksi, atau ulasan media sosial. Misalnya, pengguna dapat memanfaatkan Orange untuk mengelompokkan konsumen berdasarkan minat, mengklasifikasikan pelanggan potensial, atau

mengidentifikasi pola perilaku pembelian. Orange juga mendukung pemrosesan data teks, sehingga bisa digunakan untuk *sentiment analysis* dari ulasan produk atau komentar di media sosial.

Selain kemudahan penggunaan, Orange juga didukung oleh ekosistem *add-on* (plugin) yang sangat berguna untuk memperluas fungsionalitas. Beberapa add-on populer antara lain:

- **Text Mining:** untuk analisis teks, termasuk ekstraksi kata kunci dan analisis sentimen.
- **Image Analytics:** untuk mengolah dan menganalisis data citra.
- **Time Series:** untuk analisis data berurutan atau berbasis waktu.
- **Educational Add-ons:** yang didesain untuk pembelajaran data mining dan statistik.

Orange tidak hanya populer di kalangan profesional, tetapi juga sangat efektif digunakan dalam **dunia pendidikan**. Banyak dosen dan instruktur data science menggunakan Orange dalam kelas untuk mengajarkan konsep-konsep dasar analisis data, karena sifatnya yang interaktif dan menyenangkan. Dengan melihat langsung dampak dari tiap proses dalam workflow, siswa dapat memahami proses analitik dengan lebih baik dan cepat.

Sebagai alat eksplorasi dan pembelajaran, Orange memiliki dokumentasi lengkap serta komunitas aktif yang menyediakan tutorial, buku, dan forum diskusi. Hal ini menjadikannya platform yang cocok baik untuk pemula maupun praktisi yang ingin melakukan eksplorasi data secara cepat.

Secara keseluruhan, Orange adalah pilihan tepat bagi siapa saja yang ingin melakukan analisis data secara visual, cepat, dan efisien tanpa harus menulis banyak kode. Dengan antarmuka yang ramah pengguna dan fungsionalitas yang kuat, Orange

membuktikan bahwa data mining dapat dilakukan oleh siapa saja—dari pelajar hingga profesional bisnis.

6.3 Python (pandas, scikit-learn, matplotlib)

Python telah menjadi bahasa pemrograman paling populer dalam bidang data science dan data mining berkat sintaksnya yang sederhana, komunitasnya yang luas, dan ketersediaan berbagai pustaka (library) yang kuat. Dalam konteks data mining, Python menawarkan ekosistem lengkap untuk seluruh siklus analitik, mulai dari pemrosesan data mentah hingga visualisasi dan pembuatan model prediktif. Tiga pustaka yang sering digunakan dalam alur kerja ini adalah **Pandas**, **Scikit-learn**, dan **Matplotlib**.

Pandas: Struktur Data yang Kuat dan Fleksibel

Pandas adalah pustaka Python yang sangat powerful untuk manipulasi dan analisis data. Dengan struktur data utamanya, yaitu DataFrame, Pandas memungkinkan pengguna untuk membaca, menulis, memfilter, mengelompokkan, dan menganalisis data dalam format yang efisien dan intuitif. Data dapat diimpor dari berbagai sumber, seperti file CSV, Excel, SQL, bahkan JSON.

Sebagai contoh, dalam analisis minat konsumen, Pandas dapat digunakan untuk:

- Membersihkan data responden dari kuesioner online.
- Menggabungkan data transaksi dari berbagai toko.
- Menghitung frekuensi pembelian produk tertentu berdasarkan segmen konsumen.

Dengan fitur `groupby()`, `merge()`, `pivot_table()`, dan banyak fungsi statistik bawaan, Pandas memberikan fleksibilitas

tinggi dalam tahap *data wrangling* sebelum diteruskan ke tahap modeling.

Scikit-learn: Mesin Utama untuk Machine Learning

Setelah data siap, Scikit-learn menjadi alat utama dalam melakukan *data mining* berbasis machine learning. Library ini menyediakan berbagai algoritma klasifikasi, regresi, klusterisasi, dan reduksi dimensi. Scikit-learn juga mendukung *preprocessing*, validasi model, dan optimisasi parameter.

Contoh penggunaannya dalam konteks bisnis:

- Menggunakan Decision Tree atau Logistic Regression untuk memprediksi minat beli konsumen terhadap produk baru.
- Menerapkan KMeans untuk segmentasi konsumen berdasarkan perilaku.
- Melakukan evaluasi model dengan *confusion matrix*, *cross-validation*, dan *ROC-AUC*.

Scikit-learn dirancang dengan API yang konsisten dan dokumentasi yang sangat baik, sehingga memudahkan pengguna dari level pemula hingga profesional.

Matplotlib: Visualisasi Data yang Informatif

Visualisasi merupakan bagian penting dalam analisis data. Matplotlib adalah pustaka dasar untuk membuat grafik dan plot di Python. Bersama dengan library pendukung seperti Seaborn, Matplotlib memungkinkan visualisasi data dalam bentuk histogram, scatter plot, heatmap, dan lain sebagainya.

6.4 Weka

Weka (Waikato Environment for Knowledge Analysis) adalah perangkat lunak open-source yang dirancang khusus untuk

keperluan *data mining* dan *machine learning*. Dikembangkan oleh University of Waikato, Selandia Baru, Weka telah menjadi salah satu alat paling populer di dunia pendidikan dan penelitian karena antarmukanya yang intuitif, fitur yang lengkap, dan kemudahan dalam mengimplementasikan berbagai teknik analisis data tanpa harus menulis kode pemrograman yang kompleks.

Salah satu keunggulan utama Weka adalah tampilannya yang berbasis GUI (Graphical User Interface), yang memungkinkan pengguna melakukan analisis data secara visual dan interaktif. Bagi pengguna yang tidak memiliki latar belakang pemrograman, Weka memberikan kemudahan untuk menerapkan berbagai algoritma klasifikasi, regresi, klusterisasi, asosiasi, serta fitur-fitur praproses data hanya dengan beberapa klik.

Fitur Utama Weka

Weka menyediakan berbagai fitur penting dalam proses *data mining*, antara lain:

- **Preprocessing Data:** Weka memiliki tools lengkap untuk membersihkan, mentransformasi, dan mengatur data sebelum dilakukan pemodelan. Fitur ini mencakup normalisasi, penghapusan atribut, dan imputasi nilai yang hilang.
- **Algoritma Machine Learning:** Weka menyediakan beragam algoritma populer seperti J48 (versi C4.5 dari decision tree), Naïve Bayes, k-Nearest Neighbors, Support Vector Machine, dan Random Forest. Semua algoritma dapat dijalankan dengan pengaturan parameter yang mudah diatur.
- **Evaluasi Model:** Pengguna dapat mengevaluasi kinerja model dengan menggunakan teknik validasi silang (*cross-validation*), confusion matrix, akurasi, precision, recall, dan ROC curve.

- **Visualisasi:** Weka juga memiliki fitur visualisasi data dan hasil klasifikasi, termasuk distribusi data, pohon keputusan, dan hasil klasterisasi dalam bentuk grafik 2D.

Aplikasi Weka dalam Analisis Minat Konsumen

Dalam konteks analisis minat konsumen, Weka dapat digunakan untuk berbagai keperluan seperti segmentasi pelanggan, prediksi perilaku konsumen, atau menemukan pola asosiasi dari data pembelian. Misalnya, data transaksi pelanggan dapat diolah dengan algoritma asosiasi (seperti Apriori) untuk menemukan aturan pembelian, misalnya “Jika membeli kopi, kemungkinan besar juga membeli roti.”

Begitu pula untuk klasifikasi, Weka memungkinkan pengguna untuk membangun model yang memprediksi apakah seorang konsumen tertarik pada produk tertentu berdasarkan atribut demografis atau riwayat belanja. Fitur *attribute selection* juga memudahkan pengguna dalam memilih fitur-fitur paling relevan untuk meningkatkan akurasi model.

Kelebihan dan Keterbatasan

Kelebihan Weka terletak pada kemudahannya diakses oleh pemula dan efisiensinya untuk prototipe cepat dalam penelitian. Karena berbasis Java, Weka dapat dijalankan di berbagai sistem operasi seperti Windows, Linux, dan macOS.

Namun, Weka juga memiliki keterbatasan, terutama dalam menangani dataset yang sangat besar. Karena proses dijalankan sepenuhnya di memori (RAM), pengguna perlu memperhatikan ukuran dataset agar tidak mengganggu kinerja komputer. Selain itu, untuk pengguna tingkat lanjut, Weka mungkin terasa kurang fleksibel dibandingkan platform berbasis bahasa pemrograman seperti Python.

6.5 Perbandingan dan Rekomendasi Penggunaan

Dalam dunia data mining, pemilihan alat atau platform yang tepat sangat penting untuk menunjang efektivitas dan efisiensi analisis, khususnya dalam menganalisis minat konsumen. Berbagai perangkat lunak seperti RapidMiner, Orange, Python (dengan pustaka seperti *pandas*, *scikit-learn*, *matplotlib*), dan Weka memiliki keunggulan dan kelemahan masing-masing. Pemilihan alat sebaiknya disesuaikan dengan kebutuhan analisis, tingkat keahlian pengguna, serta sumber daya yang tersedia.

RapidMiner dikenal sebagai platform berbasis GUI yang sangat intuitif dan cocok untuk pengguna non-programmer. Dengan fitur *drag-and-drop*, pengguna dapat membangun alur kerja data mining secara visual. RapidMiner sangat cocok digunakan untuk kebutuhan bisnis atau akademik yang memerlukan eksperimen cepat dan dokumentasi yang rapi. Namun, dalam versi gratisnya, ada batasan pada jumlah baris data dan fitur lanjutan tertentu.

Orange juga merupakan alat berbasis antarmuka visual yang sederhana dan edukatif. Sangat cocok untuk pembelajaran atau prototipe sederhana, Orange menyediakan banyak widget interaktif yang memungkinkan eksplorasi data secara intuitif. Meski tidak sekuat RapidMiner dalam hal skala proyek industri, Orange unggul dalam kemudahan penggunaan dan visualisasi hasil. Untuk kebutuhan analisis segmentasi atau asosiasi dasar dalam studi minat konsumen, Orange bisa menjadi pilihan tepat.

Python, dengan pustaka-pustaka populernya seperti *pandas*, *scikit-learn*, dan *matplotlib*, adalah pilihan ideal bagi pengguna yang mahir dalam pemrograman dan membutuhkan fleksibilitas tinggi. Python unggul dalam hal skalabilitas, integrasi dengan sistem lain, dan kustomisasi proses analitik. Selain itu, Python sangat cocok untuk mengembangkan model machine learning lanjutan, integrasi dengan API media sosial,

atau implementasi real-time prediksi minat konsumen. Namun, Python memerlukan pemahaman teknis yang lebih dalam dan waktu yang lebih lama dalam pembangunan proyek.

Weka, di sisi lain, adalah pilihan ideal untuk kalangan akademik yang ingin memahami dasar-dasar machine learning dan data mining tanpa perlu menulis kode. Weka sangat bermanfaat untuk eksplorasi awal, pengajaran, dan validasi algoritma. Meski terbatas dalam skala dan performa untuk dataset besar, Weka tetap menjadi alat yang andal dalam studi kasus berbasis file data kecil hingga menengah.

Dari perbandingan tersebut, rekomendasi penggunaannya dapat disesuaikan sebagai berikut:

- **Pemula dan edukasi dasar:** Gunakan **Orange** atau **Weka**, karena mudah dipahami dan tidak memerlukan keahlian pemrograman.
- **Analisis bisnis berskala menengah dengan dokumentasi kuat:** Gunakan **RapidMiner**.
- **Proyek lanjutan, integrasi sistem, dan analisis data besar:** Gunakan **Python** dengan pustaka terkait.

DAFTAR PUSTAKA

1. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). San Francisco: Morgan Kaufmann.
2. Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Amsterdam: Elsevier.
3. Berry, M. J. A., & Linoff, G. S. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (3rd ed.). Indianapolis: Wiley Publishing.
4. Tan, P.-N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining* (2nd ed.). Pearson Education.
5. Kotler, P., & Keller, K. L. (2016). *Marketing Management* (15th ed.). Pearson Education.
6. Sudaryono. (2012). *Perilaku Konsumen dalam Perspektif Pemasaran Modern*. Jakarta: Mitra Wacana Media.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York: Springer.
8. Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics* (2nd ed.). Hoboken, NJ: Wiley.
9. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1–135.
10. Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press.
11. Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.

BIOGRAFI PENULIS



Nadia Priska

Lahir di Kotapinang pada tanggal 22 Mei 2003, Nadia Priska adalah seorang mahasiswa program diploma (D3) Manajemen Informatika di universitas labuhanbatu. Memiliki ketertarikan pada bidang teknologi informasi, khususnya dalam pengelolaan data, sistem informasi, dan pengembangan perangkat lunak.

Selama kuliah, Nadia aktif dalam berbagai kegiatan kampus seperti, kegiatan himpunan mahasiswa (himpro) unit kegiatan mahasiswa (UKM) Teknologi Informasi dan menjadi asisten laboratorium untuk mata kuliah basis data. Ia juga pernah mengikuti program magang di sebuah perusahaan badan pusat statistik, di mana ia terlibat dalam pengelolaan sistem informasi kepegawaian.



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

DI ERA DIGITAL YANG DIPENUHI DENGAN LEDAKAN DATA, MEMAHAMI PERILAKU DAN MINAT KONSUMEN MENJADI TANTANGAN SEKALIGUS PELUANG BESAR BAGI PELAKU BISNIS, PENELITI, DAN ANALIS DATA. BUKU INI HADIR SEBAGAI PANDUAN KOMPREHENSIF YANG MENJEMBATANI ANTARA TEORI DAN PRAKTIK DALAM MENGUNGKAP WAWASAN KONSUMEN MELALUI PENDEKATAN DATA MINING.

BUKU INI MEMBAHAS KONSEP DASAR DATA MINING SECARA SISTEMATIS, MULAI DARI DEFINISI, PROSES CRISP-DM DAN KDD, HINGGA TEKNIK-TEKNIK UTAMA SEPERTI KLASIFIKASI, KLASSTERING, ASOSIASI, DAN ANALISIS SENTIMEN. LEBIH DARI SEKADAR TEORI, BUKU INI JUGA MENGULAS SECARA PRAKTIS BAGAIMANA DATA KONSUMEN DIKUMPULKAN, DIBERSIHKAN, DITRANSFORMASI, DAN DIOLAH MENJADI INFORMASI YANG BERMAKNA. PENEKANAN DIBERIKAN PADA PENERAPAN DATA MINING DALAM KONTEKS NYATA, SEPERTI E-COMMERCE, RETAIL, MEDIA SOSIAL, DAN SURVEI ONLINE.

DITUJUKAN BAGI MAHASISWA, AKADEMISI, ANALIS BISNIS, DAN SIAPA PUN YANG TERTARIK PADA PEMANFAATAN DATA UNTUK KEPERLUAN PEMASARAN DAN PENGEMBANGAN PRODUK, BUKU INI MENJADI SUMBER REFERENSI PENTING DALAM DUNIA YANG SEMAKIN DIGERAKKAN OLEH DATA DAN PEMAHAMAN PELANGGAN.