



Pengantar DATA MINING & STATISTIK

Sari Ningsih, S.Si, MM., Miftahul Jannah, S.Kom., M.Kom.,
Djarot Hindarto, S.Kom., M.Kom., Dr. Nur Putri Erdianti, SKM., MKM.,
Erni Rihyanti, S.Si, M.M., Ratih Titi Komala Sari, S.Si, M.MSI,
Ira Diana Sholihati, S. Si. , MMSI, Ir. Endah Tri Esthi Handayani, M.MSI,
Dr. Fauziah, S.Kom., MMSI, Panca Dewi Pamungkasari, S.T., M.T., Ph.D.,
Raden Muhamad Firzatullah, M.Kom., Loso Judijanto, S.Si, M.M., M.Stats.,
Arie Gunawan, S.Kom. MMSI, Dr. Ir. Arief Wibowo, M.Kom.

Pengantar **DATA MINING & STATISTIK**

Sari Ningsih, S.Si, MM., Miftahul Jannah, S.Kom., M.Kom.,
Djarot Hindarto, S.Kom., M.Kom., Dr. Nur Putri Erdianti, SKM., MKM.,
Erni Rihyanti, S.Si, M.M., Ratih Titi Komala Sari, S.Si, M.MSI.,
Ira Diana Sholihati, S. Si. , MMSI, Ir. Endah Tri Esthi Handayani, M.MSI.,
Dr. Fauziah, S.Kom., MMSI., Panca Dewi Pamungkasari, S.T., M.T., Ph.D.,
Raden Muhamad Firzatullah, M.Kom., Loso Judijanto, S.Si, M.M., M.Stats.,
Arie Gunawan, S.Kom. MMSI, Dr. Ir. Arief Wibowo, M.Kom.



Pengantar Data Mining & Statistik

Copyright© PT Penerbit Penamuda Media, 2024

Penulis:

Sari Ningsih., S.Si., MM., Miftahul Jannah, S.Kom., M.Kom.,
Djarot Hindarto, S.Kom., M.Kom., Dr. Nur Putri Erdianti, SKM., MKM.,
Erni Rihyanti, S.Si., M.M., Ratih Titi Komala Sari., S.Si, M.MSI.,
Ira Diana Sholihati, S. Si. , MMSI., Ir. Endah Tri Esthi Handayani., M.MSI.,
Dr. Fauziah, S.Kom., MMSI., Panca Dewi Pamungkasari, S.T., M.T., Ph.D., Raden
Muhamad Firzatullah, M.Kom., Loso Judijanto, S.Si., M.M., M.Stats.,
Arie Gunawan, S.Kom.. MMSI., Dr. Ir. Arief Wibowo, M.Kom.

ISBN:

978-623-8586-66-0

Desain Sampul:

Tim PT Penerbit Penamuda Media

Tata Letak:

Enbookdesign

Diterbitkan Oleh

PT Penerbit Penamuda Media

Casa Sidoarium RT 03 Ngentak, Sidoarium Dodeam Sleman Yogyakarta

HP/Whatsapp : +6285700592256

Email : penamudamedia@gmail.com

Web : www.penamuda.com

Instagram : @penamudamedia

Cetakan Pertama, Juni 2024

x + 267, 15x23 cm

Hak cipta dilindungi oleh undang-undang

*Dilarang memperbanyak sebagian atau seluruh isi buku
tanpa izin Penerbit*



Kata Pengantar

Seiring dengan pesatnya perkembangan teknologi informasi, penting bagi para pembaca untuk memahami konsep dari data mining dan statistik. Buku ini hadir sebagai panduan yang komprehensif untuk memahami kedua bidang ini, serta bagaimana penggabungannya dapat memberikan wawasan bernilai dalam pengambilan keputusan.

Daftar Isi

Kata Pengantar.....	v
Daftar Isi.....	vi
Bab 1 - Konsep Dasar Data Mining.....	1
A. Data Mining dan KDD (Knowledge Discovery in Database).....	3
B. Asal Usul Data Mining.....	3
C. Tugas Data Mining (<i>Data Mining Task</i>).....	5
D. Data.....	10
Bab 2 - Pendekatan Statistik dalam Data Mining.....	14
A. Statistik dan Data Mining	14
B. Algoritma pada Data Mining.....	20
Bab 3 - Preprocessing Data untuk Data Mining	23
A. Pengenalan preprocessing data.....	23
B. Pentingnya preprocessing data dan tujuan preprocessing data.....	25

C. Pemahaman Data.....	26
D. Deskripsi data.....	31
E. Eksplorasi data	32
F. Evaluasi kualitas data	36
G. Pembersihan Data (Data Cleaning).....	37
Bab 4 - Analisis Deskriptif dan Inferensial	72
A. Analisis Deskriptif	72
B. Analisis Inferensial.....	78
Bab 5 - Regresi dan Korelasi dalam Data Mining	85
A. Regresi	85
B. Korelasi.....	93
C. Konsep Dasar Tau Kendall.....	98
Bab 6 - Analisis Kluster dalam Data Mining	101
A. Konsep Dasar Analisis Kluster.....	102
B. Metodologi Analisis Kluster	108
C. Aplikasi dan Evaluasi Analisis Kluster	117
Bab 7 - Analisis Asosiasi dan Frekuensi	125
A. Analisis Asosiasi.....	126
B. Analisis Frekuensi	126

C. Aturan Asosiasi	127
D. Algoritma Apriori.....	130

Bab 8 - Penggunaan Statistik dalam Pengelompokan Data 139

A. Pengenalan terhadap pengelompokan data dan signifikansinya dalam statistik dan analisis data	140
B. Definisi pengelompokan data dan perbedaannya dengan klasifikasi.....	142
C. Konsep dasar statistik variabel, distribusi, ukuran tendensi sentral, dan dispersi	144
D. Metodologi Pengelompokan Data	151

Bab 9 - Penggunaan Statistik dalam Klasifikasi Data..... 160

A. Data dalam Statistik.....	160
B. Dasar Klasifikasi Data Statistik	164
C. Metode Klasifikasi	167

Bab 10 - Evaluasi Model dalam Data Mining 173

A. Kualitas Data	174
B. Model Evaluasi pada Klasifikas	175
C. Model Evaluasi Cluster.....	185

Bab 11 - Teknik Sampling dalam Data Mining..... 187

Bab 12 - Interpretasi Hasil *Data Mining* dengan Pendekatan Statistika..... 199

- A. Tahap Interpretasi Hasil *Data Mining*..... 199
- B. Teknik Statistika yang Lazim Digunakan..... 202
- C. Interpretasi dengan Teknik Statistika..... 204

Bab 13 - Visualisasi Data untuk Data Mining 218

- A. Pengenalan Visualisasi Data..... 218
- B. Peran Visualisasi dalam Proses Data Mining 220
- C. Jenis-Jenis Visualisasi Data..... 221
- D. Metode Visualisasi Data Univariate 223
- E. Metode Visualisasi Data Multivariate 223
- F. Penggunaan Visualisasi untuk Eksplorasi Data 224
- G. Teknik Visualisasi untuk Pemodelan Data..... 225
- H. Pemahaman Pola dengan Visualisasi..... 225
- I. Interaktif dan Dinamika dalam Visualisasi Data..... 226
- J. Penerapan Visualisasi pada Algoritma Data Mining.. 227

Bab 14 - Tantangan dan Peluang Integrasi Data Statistik dalam Data Mining..... 229

- A. Tantangan Integrasi Data Statistik dan Data Mining 230
- B. Peluang Integrasi Data Statistik dan Data Mining 236

C. Kesimpulan dan Rekomendasi	241
Daftar Pustaka	243
Tentang Penulis	258



Konsep Dasar Data Mining

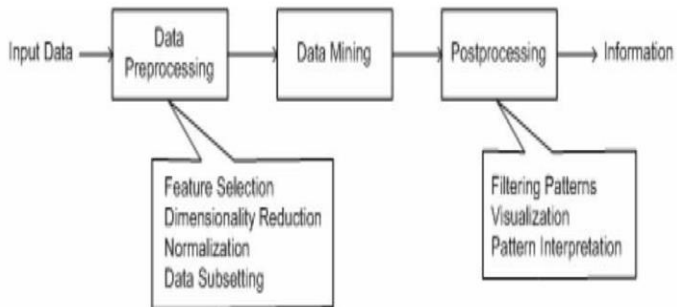
Sari Ningsih.,S.Si., MM.

Kemajuan pesat dalam pengumpulan dan teknologi penyimpanan data telah memungkinkan organisasi mengumpulkan data dalam jumlah besar. Namun, mengekstraksi informasi yang berguna terbukti sangat menantang. Seringkali, alat dan teknik analisis data tradisional tidak dapat digunakan karena ukuran kumpulan data yang sangat besar. Terkadang, sifat data yang non-tradisional menyebabkan pendekatan tradisional tidak dapat diterapkan meskipun kumpulan datanya relatif kecil. Dalam situasi lain, pertanyaan yang perlu dijawab tidak dapat dijawab dengan menggunakan teknik analisis data yang ada, sehingga perlu dikembangkan metode baru.

Data mining adalah penggunaan teknologi yang menggabungkan metode analisis data konvensional dengan algoritma yang canggih untuk mengolah volume data yang besar. Teknologi ini juga membuka kesempatan untuk menggali dan menganalisis jenis data baru serta menerapkan pendekatan baru dalam menganalisis data yang sudah ada.

A. Data Mining dan KDD (Knowledge Discovery in Database).

Data mining adalah komponen penting dari proses penemuan pengetahuan dalam database (KDD), yang merujuk pada proses keseluruhan untuk menginput data menjadi informasi yang berguna, melalui data preprocessing, data mining dan postprocessing, seperti yang tergambar dalam Gambar 1. Tahapan ini melibatkan serangkaian langkah transformasi, dimulai dari pra-pemrosesan data hingga pasca-pemrosesan hasil data mining.



Gambar 1. Proses KDD (Tan,Pang-Nin,2006)

B. Asal Usul Data Mining.

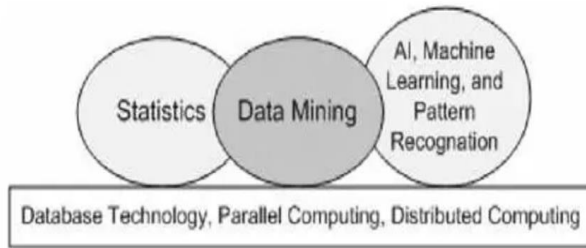
Terinspirasi oleh tujuan untuk mengatasi tantangan yang dibahas sebelumnya, para peneliti dari berbagai bidang ilmu mulai berkonsentrasi pada pengembangan

instrumen yang lebih efektif dan dapat diukur untuk menangani berbagai macam data.

Pekerjaan/Proyek-proyek yang mencapai puncak dalam ranah data mining ini, dibangun berdasarkan algoritma dan metodologi yang telah digunakan oleh peneliti sebelumnya. Lebih spesifik lagi, data mining memanfaatkan konsep-konsep (1) pengujian sampel data statistik, pendugaan statistic / estimasi, dan Uji hipotesis bidang statistik, dan (2) untuk kecerdasan buatan (AI), pengenalan pola, dan *machine learning* yaitu teknik pemodelan, algoritma pencarian, serta teori pembelajaran untuk kecerdasan buatan (AI), pengenalan pola, dan *machine learning*. Data Mining juga cepat mengadopsi gagasan dari disiplin ilmu lain, termasuk optimisasi, komputasi evolusioner, teori informasi, pemrosesan sinyal, visualisasi, dan pengambilan informasi.

Beberapa bidang lain juga berperan sebagai pendukung utama. Secara spesifik, sistem basis data diperlukan untuk menyediakan dukungan dalam hal penyimpanan yang efisien, pengindeksan dan pemrosesan query.

Teknik komputasi kinerja tinggi (paralel) seringkali penting dalam mengatasi ukuran besar dari beberapa kumpulan data. Teknik terdistribusi juga menyelesaikan permasalahan terkait skala dan memiliki kepentingan yang besar ketika data tidak dapat terkumpul di satu lokasi.



Gambar 2. Data Mining sebagai pertemuan banyak disiplin ilmu (Tan, Pang-Nin, 2006)

C. Tugas Data Mining (*Data Mining Task*)

Meliputi 2.

1. **Tugas Prediktif (*Predictive tasks*)**. Tujuan dari tugas ini adalah untuk meramal nilai suatu atribut tertentu variabel terikat (*dependent*) berdasarkan nilai atribut lainnya atau variabel bebas (*independent*).
2. **Tugas Deskriptif (*Descriptive tasks*)**. Tujuannya adalah untuk mendapatkan **pola**, diantaranya adalah **korelasi, tren, cluster, lintasan, dan anomali** yang merangkum hubungan mendasar dalam data. Kegiatan deskripsi data dalam data mining sering kali bersifat eksploratif dan umumnya membutuhkan teknik pasca-pemrosesan untuk memverifikasi dan menjelaskan temuannya.

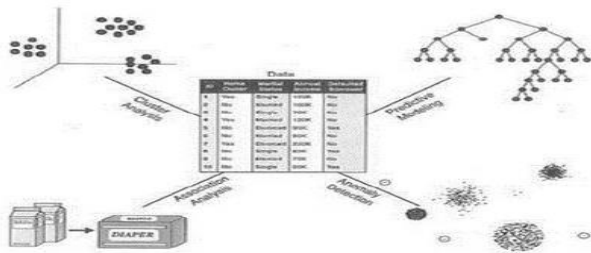


Figure 2: The four core of Data mining task

Gambar 3. Tugas Inti Data Mining Tan, Pang-Nin, 2006)

Pemodelan prediktif (*Predictive modeling*), mengacu pada tugas membangun model untuk variabel target sebagai fungsi dari variabel penjelas. Ada dua jenis tugas pemodelan prediktif: (1). Klasifikasi, yang digunakan untuk variabel diskrit, dan (2) Regresi, yang digunakan untuk variabel kontinu. Misalnya, memprediksi apakah pengguna web akan melakukan pembelian di toko buku online merupakan tugas klasifikasi karena variabel target bernilai biner. Di sisi lain, meramalkan harga saham di masa depan adalah tugas regresi karena harga adalah atribut yang bernilai kontinu. Tujuan dari kedua tugas tersebut adalah untuk mempelajari model yang meminimalkan kesalahan antara nilai prediksi dan nilai sebenarnya dari variabel target. Pemodelan prediktif dapat digunakan untuk mengidentifikasi pelanggan yang akan merespons kampanye pemasaran, memprediksi gangguan pada ekosistem bumi, atau menilai apakah seorang pasien mengidap penyakit tertentu berdasarkan hasil tes medis.

Contoh Kasus Pemodelan prediktif (*Predictive modeling*) : Memprediksi jenis bunga. Perhatikan tugas meramalkan suatu jenis bunga berdasarkan ciri-ciri bunga tersebut. Secara khusus, pertimbangkan untuk mengklasifikasikan bunga Iris berdasarkan apakah bunga tersebut termasuk dalam salah satu dari tiga Spesies Iris berikut: *Setosa*, *Versicolor*, atau *Virginica*. Untuk melakukan tugas ini, diperlukan kumpulan data yang berisi ciri-ciri berbagai bunga dari ketiga spesies tersebut. Kumpulan data dengan jenis informasi ini adalah kumpulan data Iris yang terkenal dari Repositori Pembelajaran Mesin (*machine learning*) UCI di <http://www.ics.uci.edu/mllearn> .

Selain spesies bunga, kumpulan data ini berisi empat atribut lainnya: lebar sepal, panjang sepal, panjang kelopak, dan lebar kelopak.

Menurut Ningsih S, (2021). Contoh analisis Klasifikasi tujuannya adalah membantu pemimpin mengorganisir data tentang kinerja dosen berdasarkan faktor-faktor seperti keahlian, tingkat pendidikan, pengalaman kerja, jabatan, serta kegiatan penelitian dan pengabdian masyarakat, yang merupakan elemen kunci dalam tugas mengajar di perguruan tinggi.

Analisis Asosiasi (*Associations analysis*) digunakan untuk menemukan pola yang menggambarkan fitur asosiasi yang kuat dalam data. Pola yang ditemukan biasanya direpresentasikan dalam bentuk aturan implikasi atau himpunan bagian fitur. Karena ukuran ruang pencariannya yang eksponensial, tujuan analisis asosiasi

adalah mengekstrak pola yang paling menarik dengan cara yang efisien.

Contoh Kasus Analisis Asosiasi (*Associations analysis*):

Analisis Keranjang Pasar (*Market Basket Analysis*).

Tabel1. Market basket data

No.	Kode Barang	Nama Barang
1.	A1	Paket 5 tan dendro spike, harga @40K /tan dan 5K/spike (tdk biasa memilih ID)
2.	A2	Paket 5 tan dendro paket remaja, harga @40K/tan (bisa memilih ID)
3.	A3	Paket 5 tan dendro paket dewasa, harga @65K/tan (bisa memilih ID)
4.	B1	Paket anggrek bulan spike, harga sesuai ID
5.	B2	Paket anggrek bulan dewasa tidak spike, harga sesuai ID (bisa memilih ID)
6.	B3	Paket anggrek bulan double spike , harga sesuai ID
7.	C1	Paket Vanda spike harga sesuai ID
8.	C2	Paket Vanda dewasa harga sesuai ID
9.	C3	Paket Anggrek Catleya harga Sesuai ID
10	D	Paket anggrek khusus sesuai pilihan ID coustumer

Transaksi	Data Penjualan/Hari (Kode Barang)	Transaksi	Data Penjualan/Hari (Kode Barang)	Transaksi	Data Penjualan/Hari (Kode Barang)
1.	A1,A3,B1	15.	A1, A2, B3, C1, C3	29.	B2,C2,D
2.	A1,A2,C1	16.	A1, D	30.	D
3.	A2,D	17.	B1, C2	31.	A1, C1,D
4.	A1,C2,D	18.	B2,C2,D	32.	B2,C3
5.	A1,A2,A3	19.	A1,A2,B1,B3,C1,C2	33.	A1, A3, B2, D
6.	B1, C3	20.	A3,B2,C3,D	34.	A2, B3,C3
7.	A2, D	21.	A1,A3,B1	35.	A1, A2, B3, C1, C3
8.	A1, B1,C1,D	22.	A1,A2,C1	36.	A1, D
9.	B2,C2,D	23.	A2,D	37.	B1, C2
10.	D	24.	A1,C2,D	38.	B2,C2,D
11.	A1, C1,D	25.	A1,A2,A3	39.	A1,A2,B1,B3,C1,C2
12.	B2,C3	26.	B1, C3	40.	A3,B2,C3,D
13.	A1, A3, B2, D	27.	A2, D	41.	A1,A3,B1
14.	A2, B3,C3	28.	A1, B1,C1,D	42.	A1,A2,C1
43.	A2,D	69.	B2,C2,D	95.	A1, A2, B3, C1, C3
44.	A1,C2,D	70.	D	96.	A1, D
45.	A1,A2,A3	71.	A1, C1,D	97.	B1, C2
46.	B1, C3	72.	B2,C3	98.	B2,C2,D
47.	A2, D	73.	A1, A3, B2, D	99.	A1,A2,B1,B3,C1,C2
48.	A1, B1,C1,D	74.	A2, B3,C3	100.	A3,B2,C3,D
49.	B2,C2,D	75.	A1, A2, B3, C1, C3	101.	A1,A3,B1
50.	D	76.	A1, D	102.	A1,A2,C1
51.	A1, C1,D	77.	B1, C2	103.	A2,D
52.	B2,C3	78.	B2,C2,D	104.	A1,C2,D
53.	A1, A3, B2, D	79.	A1,A2,B1,B3,C1,C2	105.	A1,A2,A3
54.	A2, B3,C3	80.	A3,B2,C3,D	106.	B1, C3
55.	A1, A2, B3, C1, C3	81.	A1,A3,B1	107.	A2, D
56.	A1, D	82.	A1,A2,C1	108.	A1, B1,C1,D
57.	B1, C2	83.	A2,D	109.	B2,C2,D
58.	B2,C2,D	84.	A1,C2,D	110.	D
59.	A1,A2,B1,B3,C1,C2	85.	A1,A2,A3	111.	A1, C1,D
60.	A3,B2,C3,D	86.	B1, C3	112.	B2,C3
61.	A1,A3,B1	87.	A2, D	113.	A1, A3, B2, D
62.	A1,A2,C1	88.	A1, B1,C1,D	114.	A2, B3,C3
63.	A2,D	89.	B2,C2,D	115.	A1, A2, B3, C1, C3
64.	A1,C2,D	90.	D	116.	A1, D
65.	A1,A2,A3	91.	A1, C1,D	117.	B1, C2
66.	B1, C3	92.	B2,C3	118.	B2,C2,D
67.	A2, D	93.	A1, A3, B2, D	119.	A1,A2,B1,B3,C1,C2
68.	A1, B1,C1,D	94.	A2, B3,C3	120.	A3,B2,C3,D

Menurut Ningsih S, (2022). Contoh analisis Asosiasi menggunakan algoritma *Apriori* adalah salah satu algoritma yang dapat diterapkan dalam berbagai konteks, termasuk dalam ranah bisnis yaitu pada sistem persediaan dan penjualan tanaman anggrek secara online.

Analisis kluster (*Cluster analysis*) berupaya menemukan kelompok observasi yang berkaitan erat sehingga observasi yang termasuk dalam kluster yang sama lebih mirip satu sama lain dibandingkan observasi yang termasuk dalam kluster lain. *Clustering* telah digunakan untuk mengelompokkan kumpulan pelanggan terkait, menemukan wilayah lautan yang memiliki dampak signifikan terhadap iklim bumi, dan mengompresi data.

Menurut Triayudi et.al.,(2024), Clustering adalah teknik dalam analisis data yang bertujuan untuk mengelompokkan objek-objek serupa menjadi kelompok-kelompok yang saling berbeda.

Contoh Kasus Analisis kluster (*Cluster analysis*): Pengelompokan dokumen. Kumpulan artikel berita dapat dikelompokkan berdasarkan topiknya masing-masing. Setiap artikel direpresentasikan sebagai kumpulan pasangan frekuensi kata (w,c), dimana w adalah sebuah kata dan c adalah berapa kali kata tersebut muncul dalam artikel. Ada dua cluster alami dalam kumpulan data. Cluster pertama terdiri dari empat artikel pertama yang berisi berita mengenai perekonomian, sedangkan cluster kedua berisi empat artikel terakhir yang berisi berita mengenai layanan kesehatan. Algoritma clustering yang baik harus dapat mengidentifikasi kedua cluster tersebut berdasarkan kemiripan kata yang muncul pada artikel.

Deteksi anomali (*Anomaly detection*) adalah tugas mengidentifikasi observasi yang karakteristiknya berbeda signifikan dengan data lainnya. Anomali atau *outlier* disebut sebagai pengamatan semacam ini. Algoritma deteksi anomali bertujuan untuk mengidentifikasi anomali yang sebenarnya dan mencegahnya pemberian label palsu pada objek normal sebagai anomali. Dengan kata lain, pendeteksi anomali yang baik harus memperoleh tingkat sensitivitas yang tinggi dan tingkat false alarm yang rendah. Penerapan deteksi anomali mencakup pengidentifikasian penipuan, pemrosesan jaringan, pola penyimpanan yang tidak biasa, dan gangguan ekosistem.

Contoh Kasus Deteksi Anomali : Deteksi Penipuan Kartu Kredit. Perusahaan kartu kredit mencatat transaksi yang dilakukan oleh setiap pemegang kartu kredit, beserta informasi pribadi seperti batas kredit, usia, pendapatan tahunan, dan alamat. Karena jumlah kasus penipuan relatif kecil dibandingkan dengan jumlah transaksi yang sah, teknik deteksi anomali dapat diterapkan untuk membangun profil transaksi yang sah bagi pengguna. Ketika transaksi baru masuk, transaksi tersebut dibandingkan dengan profil pengguna. Jika karakteristik transaksi sangat berbeda dengan profil yang dibuat sebelumnya, maka transaksi tersebut ditandai sebagai berpotensi penipuan.

D. Data

Di sini kita membahas beberapa masalah terkait data yang penting untuk keberhasilan data mining: Jenis data, Kualitas data, Langkah-Langkah pemrosesan awal agar

data cocok untuk data mining, dan analisis data dalam hubungannya.

1. Jenis Data.

Jenis data menentukan alat dan teknik mana yang dapat digunakan untuk menganalisis data; seringkali kita perlu mengakomodasi area aplikasi baru dan tipe data barunya. Suatu kumpulan data seringkali dapat dipandang sebagai kumpulan objek data (pengamatan) yang digambarkan oleh sejumlah atribut (*variabel*).

2. Atribut dan Pengukuran.

Kami mendeskripsikan data dengan mempertimbangkan tipe atribut apa yang digunakan untuk mendeskripsikan objek data.

3. Atribut.

Merupakan suatu sifat atau ciri suatu benda yang dapat berbeda-beda, baik dari satu benda ke benda lainnya, atau dari waktu ke waktu. Untuk menganalisis karakteristik suatu objek, kami memberikan nomor atau simbol pada objek tersebut.

4. Skala pengukuran.

Merupakan fungsi yang mengaitkan nilai numerik atau simbolik dengan atribut suatu objek. Ob. Nilai yang digunakan untuk mewakili suatu atribut mungkin memiliki properti yang bukan merupakan properti dari atribut itu sendiri, dan sebaliknya.

5. Berbagai jenis atribut.

Ada dua cara utama untuk mendefinisikan tipe; bergantung pada properti dan operasi yang dimiliki atribut tersebut, atau dalam hal transformasi yang diizinkan.

Kita juga dapat mendeskripsikan atribut berdasarkan jumlah nilainya. Seperti diskrit atau kontinu.

6. Atribut Asimetris.

Hanya kehadiran yang dianggap penting; kita mengabaikan keberadaan angka nol misalnya jika menyiratkan tidak adanya sesuatu.

Jenis Kumpulan Data. Pertama-tama kita melihat karakteristik umum kumpulan data:

Dimensi (jumlah atribut; bagian penting dari pra-pemrosesan data adalah pengurangan dimensi),

Ketersebaran (ketika kita harus mengabaikan sebagian besar nilai karena nol, jadi hanya nilai bukan nol yang perlu) disimpan dan dimanipulasi),

Resolusi (seringkali dimungkinkan untuk memperoleh data pada tingkat resolusi berbeda, dan propertinya berbeda pada resolusi berbeda).

Tabel 2. *Different Attribute Type*

	Attribute Type	Description	Examples	Operations
Categorical Qualitative	Nominal	Nominal attribute values only distinguish. ($=$, \neq)	zip codes, employee ID numbers, eye color, sex: { <i>male</i> , <i>female</i> }	mode, entropy, contingency correlation, χ^2 test
	Ordinal	Ordinal attribute values also order objects. ($<$, $>$)	hardness of minerals, { <i>good</i> , <i>better</i> , <i>best</i> }, grades, street numbers	median, percentiles, rank correlation, run tests, sign tests
Numeric Quantitative	Interval	For interval attributes, differences between values are meaningful. ($+$, $-$)	calendar dates, temperature in Celsius or Fahrenheit	mean, standard deviation, Pearson's correlation, <i>t</i> and <i>F</i> tests
	Ratio	For ratio variables, both differences and ratios are meaningful. ($*$, $/$)	temperature in Kelvin, monetary quantities, counts, age, mass, length, current	geometric mean, harmonic mean, percent variation

Dari tabel 2. Terlihat 2 tipe data, yaitu data kualitatif (kategori) terdiri dari data Nominal dan data Ordinal, sedangkan data kuantitatif terdiri dari data Interval dan data Ratio, terlihat deskripsi contoh serta operations nya dari setiap jenis data.



Pendekatan Statistik dalam Data Mining

Miftahul Jannah, S.Kom., M.Kom.

A. Statistik dan Data Mining

Data mining adalah proses pengumpulan (penambangan) data dalam jumlah besar dan kemudian mencari informasi dari data tersebut. Ini adalah bagian dari seni dan ilmu analisis data pintar yang bertujuan untuk mendapatkan informasi dari suatu data. Informasi ini sering disajikan ke dalam bentuk model, sehingga data mining juga dapat disebut sebagai proses membangun model data. Model ini dapat digunakan untuk mendapatkan informasi dari data (Williams, 2011).

Statistik adalah seni dan ilmu yang berfokus pada pengumpulan data, presentasi, dan interpretasi dalam upaya menguji teori dan membuat kesimpulan tentang semua fenomena. Tidak diragukan lagi, peran statistik dalam data mining sangat penting, mulai dari metode



Preprocessing Data untuk Data Mining

Djarot Hindarto, S.Kom., M.Kom.

A. Pengenalan preprocessing data

Dalam dunia data mining, preprocessing data merupakan tahap awal yang kritikal dan memegang peranan penting dalam kesuksesan analisis yang akan dilakukan (Applied et al., 2003). Tahapan ini melibatkan serangkaian proses untuk mengubah data mentah menjadi format yang siap dianalisis yang pada akhirnya meningkatkan kualitas dan efektivitas pengambilan keputusan berdasarkan data tersebut. Preprocessing data tidak hanya membantu dalam membersihkan data dari ketidaksempurnaan seperti noise, data yang hilang, atau inkonsistensi, namun juga dalam mengintegrasikan dan mentransformasikan data sehingga menjadi lebih mudah diinterpretasikan oleh algoritma data mining.



Analisis Deskriptif dan Inferensial

Dr. Nur Putri Erdianti, SKM., MKM

Statistik deskriptif dan inferensial merupakan dua kategori utama dalam statistik. Statistik deskriptif menggambarkan kumpulan data untuk sekelompok objek, pengamatan, atau orang tertentu. Mereka tidak berusaha menggeneralisasi di luar serangkaian observasi. Statistik inferensial menggunakan kumpulan data untuk membuat kesimpulan populasi yang lebih besar dari mana sampel diambil. Statistik ini menggeneralisasi melampaui observasi spesifik yang ada dalam kumpulan data ke kelompok atau populasi yang lebih besar.

A. Analisis Deskriptif

Analisis deskriptif adalah metode analisis data yang bertujuan untuk menggambarkan, menampilkan, dan merangkum sekumpulan data. Jenis analisis ini menjadi populer karena kemampuannya dalam menjelaskan perubahan dari waktu ke waktu dan tujuannya untuk mengidentifikasi tren serta pola dalam data. Pada



Regresi dan Korelasi dalam Data Mining

Erni Rihyanti, S.Si., M.M

A. Regresi

1. Definisi Regresi

Analisis regresi adalah salah satu teknik statistik yang paling umum digunakan dalam ilmu sosial dan perilaku serta ilmu fisika yang melibatkan identifikasi dan evaluasi hubungan antara variabel terikat dan satu atau lebih variabel bebas, yang juga disebut variabel prediktor atau penjelas. Hal ini sangat berguna untuk menilai dan menyesuaikan perancu. Model hubungan dihipotesiskan dan estimasi nilai parameter digunakan untuk mengembangkan persamaan regresi estimasi. Berbagai pengujian kemudian dilakukan untuk menentukan apakah model tersebut memuaskan. Jika model dianggap memuaskan, persamaan regresi yang diestimasi dapat digunakan untuk memprediksi nilai variabel terikat mengingat nilai variabel bebas.



Analisis Kluster dalam Data Mining

Ratih Titi Komala Sari., S.Si, M.MSI

Analisis kluster dalam data mining merupakan teknik esensial yang digunakan untuk mengelompokkan set data menjadi beberapa kelompok atau kluster berdasarkan kesamaan atribut atau karakteristik di antara data. Tujuan utama dari teknik ini adalah untuk memastikan bahwa data dalam satu kluster memiliki karakteristik yang serupa satu sama lain dan berbeda secara signifikan dari data di kluster lain. Dengan demikian, analisis kluster memungkinkan penemuan pola dan hubungan tersembunyi dalam data yang luas dan kompleks, yang sering kali sulit dianalisis menggunakan metode tradisional.

Dalam konteks data mining, analisis kluster digunakan untuk berbagai aplikasi, mulai dari segmentasi pasar dalam bidang pemasaran, pengelompokan gen dalam studi biologi, hingga pengorganisasian dokumen atau informasi di internet. Metode ini tidak hanya meningkatkan keefektifitasan dalam pengambilan keputusan tetapi juga memberikan keuntungan signifikan dalam memperbaiki strategi bisnis dan keilmuan



Analisis Asosiasi dan Frekuensi

Ira Diana Sholihati, S. Si. , MMSI.

Data yang dimiliki oleh suatu organisasi adalah aset berharga yang dimiliki oleh organisasi tersebut. Data diperoleh dari kegiatan sehari-hari organisasi tersebut, sehingga semakin hari jumlah data akan semakin besar. Jika organisasi tersebut tidak dapat memanfaatkan data tersebut, maka data akan menimbulkan masalah bagi organisasi tersebut karena data tersebut tidak akan bermanfaat. Dengan demikian, semakin banyak data, semakin banyak usaha yang harus dilakukan oleh organisasi untuk mengubah data menjadi data bermanfaat. Usaha yang dilakukan adalah mengklasifikasikan data berdasarkan kemungkinan yang muncul secara bersamaan selama satu transaksi. Teknik analisis yang dapat dilakukan adalah **Analisis Asosiasi** dan **Analisis Frekuensi**.


Analisis asosiasi dan analisis frekuensi adalah dua teknik analisis yang umum digunakan dalam berbagai bidang, termasuk ilmu data, penelitian pasar, dan analisis bisnis. Analisis asosiasi (analisis aturan asosiasi) merupakan metode data mi-



Penggunaan Statistik dalam Pengelompokan Data

Ir. Endah Tri Esthi Handayani.,M.MSI

Pengelompokan data, atau yang sering dikenal sebagai clustering dalam bidang statistik dan analisis data, merupakan teknik penting yang digunakan untuk mengelompokkan kumpulan objek atau data points ke dalam beberapa kelompok berdasarkan kesamaan karakteristiknya. Tujuan utama dari pengelompokan ini adalah untuk memastikan bahwa data dalam satu kelompok memiliki kesamaan yang lebih besar dibandingkan dengan data di kelompok lain. Teknik ini sangat berguna dalam berbagai aplikasi, mulai dari segmentasi pasar, analisis sosial, hingga dalam bidang kedokteran dan biologi. Pengelompokan membantu dalam mengidentifikasi struktur yang tersembunyi dalam data yang besar atau kompleks, memungkinkan analis dan peneliti untuk membuat keputusan atau prediksi yang lebih informasi dan berbasis data.



Penggunaan Statistik dalam Klasifikasi Data

Dr. Fauziah, S.Kom., MMSI

A. Data dalam Statistik

1. Pengertian Data

Data merupakan gabungan dari karakter, angka, dapat juga berupa gambar yang kemudian dikumpulkan dan diolah sehingga menghasilkan informasi bagi pengguna (Rosalina et al., 2023). Pada Statistik, data dapat dipelajari, kemudian dianalisis dan disajikan kepada pengguna sehingga memiliki arti dan lebih bermanfaat.



Evaluasi Model dalam Data Mining

Panca Dewi Pamungkasari, S.T., M.T., Ph.D

Data mining adalah ilmu yang menggunakan teknik komputasi yang terdiri dari ilmu statistik, pembelajaran mesin/*machine learning* dan pengenalan pola untuk menganalisis kumpulan data atau database yang besar. Jenis analisis ini memiliki dua tujuan yaitu prediktif dan deskriptif. Metode prediksi menggunakan beberapa variabel untuk memprediksi satu atau lebih variabel seperti klasifikasi, regresi, deteksi anomali. Metode deskriptif mengidentifikasi pola-pola yang menggambarkan data seperti pengelompokan, penemuan aturan asosiasi, dan penemuan pola sekuensial. Setiap kali suatu pola ditemukan dalam data mining, pola tersebut harus dievaluasi untuk menentukan keandalannya dimana desain dapat dipertimbangkan menggunakan berbagai metrik, bergantung pada konteksnya. Algoritma untuk metode evaluasi pola dalam data mining dapat diuji dengan beberapa cara:



Teknik Sampling dalam Data Mining

Raden Muhamad Firzatullah, M.Kom

Data Mining merupakan proses kolektifitas dan pemrosesan data yang berfokus dalam mengekstrak informasi yang dapat diungkap pada koleksi data tersebut. Dalam prosesnya data direkap dan diproses dalam ukuran yang besar, dimana dalam pengolahannya diperlukan suatu metode untuk menggali informasi sehingga diperoleh pengetahuan dari koleksi data. Skalabilitas algoritma data mining dibatasi dalam beberapa aspek salah satunya adalah ukuran data. Dimana peningkatan ukuran data berimbas pada waktu komputasi dan kebutuhan sumber daya yang diperlukan, hal tersebut dikenal dengan istilah *Curse of Dimensionality*, sementara sebagian algoritma data mining memerlukan koleksi data yang besar untuk menghasilkan akurasi model yang baik. Berdasarkan hal tersebut maka diperlukan suatu metode yang dapat menentukan struktur kompleks dan estimasi data secara akurat yang dinamakan metode *sampling*.



Interpretasi Hasil *Data Mining* dengan Pendekatan Statistika

Loso Judijanto, S.Si., M.M., M.Stats.

A. Tahap Interpretasi Hasil *Data Mining*

Interpretasi *data mining* menggunakan statistika melibatkan beberapa langkah penting yang harus dilakukan secara sistematis untuk memastikan bahwa hasil yang diperoleh valid dan dapat diandalkan. Berikut adalah langkah-langkah yang umumnya diikuti dalam interpretasi data mining dengan pendekatan statistik:

1. **Pemahaman Tujuan dan Konteks Data:** Sebelum melakukan interpretasi adalah penting untuk memahami tujuan dari analisis data mining dan konteks di mana data tersebut dikumpulkan. Hal ini termasuk memahami pertanyaan penelitian,



Visualisasi Data untuk Data Mining

Arie Gunawan, S.Kom.. MMSI

A. Pengenalan Visualisasi Data

Pengenalan Visualisasi Data adalah konsep yang sangat penting dalam pemahaman dan analisis data. Secara sederhana, visualisasi data merujuk pada proses menggambarkan data dan informasi secara visual melalui grafik, diagram, atau representasi visual lainnya (Tufte, 2001). Tujuan utamanya adalah untuk membantu pemahaman yang lebih baik terhadap pola, hubungan, dan tren yang terdapat dalam dataset. Visualisasi data memungkinkan pengguna untuk menyajikan informasi kompleks dengan cara yang mudah dipahami dan dapat digunakan untuk mengambil keputusan yang lebih baik (Evergreen, 2012; Knaflic, 2015).

Salah satu keunggulan utama dari visualisasi data adalah kemampuannya untuk merangkum informasi yang kompleks menjadi bentuk yang lebih sederhana dan mudah dipahami. Misalnya, data tabular yang rumit dapat diubah menjadi grafik garis atau batang yang mudah



Tantangan dan Peluang Integrasi Data Statistik dalam Data Mining

Dr. Ir. Arief Wibowo, M.Kom.

Menurut Smith et al. (2018), data statistik dan data mining memiliki peran krusial dalam memecahkan masalah modern. Data statistik memberikan pemahaman mendalam tentang fenomena dalam populasi atau sampel, sementara data mining mengungkap pola, tren, dan hubungan tersembunyi dalam data besar dan kompleks (Johnson & Brown, 2015). Integrasi kedua jenis data ini memberikan manfaat besar dengan menggabungkan analisis statistik dan prediksi data mining (Chen & Li, 2019). Dengan memadukan data statistik dan teknik data mining, kita dapat menghasilkan wawasan yang lebih dalam dan solusi yang lebih efektif di berbagai bidang, termasuk bisnis, ilmu sosial, dan kesehatan (Wang et al., 2020). Integrasi ini mendukung pengambilan keputusan yang lebih baik, strategi yang lebih efisien, dan kinerja organisasi yang lebih optimal (Gupta & Kapoor, 2017). Dengan memahami peran dan kekuatan kedua

Daftar Pustaka

- Abror, K. (2023) Pengertian Interpretasi Data, Jenis, Cara Melakukan dan Contohnya, Materi Penelitian. Available at: <https://penelitianilmiah.com/interpretasi-data/> (Accessed: 12 April 2024).
- Adins (2023) 8 Tahapan Data Mining dalam Proses Pencarian Informasi, Data Mining. Available at: <https://www.ad-ins.com/id/our-story/kisah-adins/tahapan-data-mining/> (Accessed: 12 April 2024).
- Admin (2023) Uji Hipotesis Analisis Regresi (Konsep Dasar), Maglearning.id.
- AlFatih, M. (2022) Pengertian dan Cara Interpretasi Penyajian Data Statistik, Zenius Kampus. Available at: <https://www.zenius.net/blog/interpretasi-data> (Accessed: 12 April 2024).
- Amna et al. (2023) Data Mining. 1st edn. Edited by D. Ediana and A. Yanto. Padang: PT Global Eksekutif Teknologi. Available at: https://repository.uinjkt.ac.id/dspace/bitstream/123456789/71093/1/DATA_MINING.pdf.
- Amrin (2018) 'Data Mining dengan Regresi Linier Berganda untuk Peramalan Tingkat Inflasi', Jurnal Techno Nusa Mandiri, XIII(March 2016), pp. 74–79. Available at: <https://media.neliti.com/media/publications/227375-data-mining-dengan-regresi-linier-bergan-eb3e1657.pdf>.
- Anderson, B. (2023a) Koefisien Regresi, Statorials. Available at: <https://statorials.org/id/koefisien-regresi/> (Accessed: 12 April 2024).

- Anderson, B. (2023b) Korelasi, Statorials SPSS Indonesia. Available at: <https://statorials.org/id/korelasi/> (Accessed: 12 April 2024).
- Anderson, B. (2023c) Matrik Korelasi, Statorials SPSS Indonesia. Available at: <https://statorials.org/id/matriks-korelasi/> (Accessed: 12 April 2024).
- Anggreany, M.S. (2023) Confusion Matrix, Binus School of Computer Science Articles. Available at: <https://socs.binus.ac.id/2020/11/01/confusion-matrix/> (Accessed: 12 April 2024).
- Alpaydin E, 2020, Introduction to Machine Learning, 4th Edition, MIT Press, USA.
- Ahmed, M., 2020. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *mdpi Electronics*, Volume 9, pp. 1-12.
- Applied, P., Letters, M., Science, I., & Cercone, N. (2003). DB-HReduction : A Data Preprocessing Algorithm for Data Mining Appliqtions. *Applied Mathematics Letters*. [https://doi.org/https://doi.org/10.1016/S0893-9659\(03\)90013-9](https://doi.org/https://doi.org/10.1016/S0893-9659(03)90013-9)
- Adinugroho, S. and Arum Sari, Y. (2018) Implementasi Data Mining Menggunakan Weka. Universitas Brawijaya Press.
- Amalia, F.S., Setiawansyah, S. and Darwis, D. (2021) 'Analisis Data Penjualan Handphone Dan Elektronik Menggunakan Algoritma Apriori (Studi Kasus : Cv Rey Gasendra)', *Journal of Telematics and information Technology*, 2(1), pp. 1-6.
- Badrul, M. (2016) 'Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan', *Jurnal Pilar Nusa Mandiri*, 12(2), pp. 121-126.

- Bradley A.P., “The use of the area under the ROC curve in the evaluation of machine learning algorithm”, *Pattern Recognition*, 1997, vol. 30, No. 7, pp. 1145-1159.
- Bamber D., “The area above the ordinal dominance graph and the area below the receiver operating characteristic graph”, *Journal of Mathematical Psychology*, 1975, vol. 12, pp. 387 -415.
- Cahyana, N. and Aribowo, A. (2018) ‘Metode Data Mining K-Means untuk Klasterisasi Data Penanganan dan Pelayanan Kesehatan Masyarakat’, *Prosiding Seminar Nasional Informatika Medis*, (5), pp. 24–31. Available at:
<https://journal.uii.ac.id/snimed/article/download/11878/pdf/25907>.
- Chen, L., & Li, Y. (2019). Integrating Statistical Analysis and Data Mining for Improved Insights. *Journal of Data Science*, 17(3), 432-445.
- Christian, F. (2023) *Data Mining: Pengertian, Fungsi, Metode, dan Contohnya*, Toffeedev Resources. Available at:
<https://toffeedev.com/blog/website/data-mining-adalah/> (Accessed: 12 April 2024).
- Cherfi, A., 2018. Very Fast C4.5 Decision Tree Algorithm. *Applied Artificial Intelligence*, VOL. 32(NO. 2), pp. 19-137,.
- Evergreen, S. (2012). *Data Visualization: A Practical Guide for Analyzing and Presenting Data*. Pearson.
- Ernawati (2019) ‘Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang’, *FIBONACCI: Jurnal Pendidikan Matematika dan Matematika*, 5(2), p. 117. Available at:
<https://doi.org/10.24853/fbc.5.2.117-128>.

- Frost, J. (2019). *Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries*.
- Fu, T., & Zhang, C. (2024). Towards a generic model evaluation metric for non-normally distributed measurements in water quality and ecosystem models. *Ecological Informatics*, 80(January). <https://doi.org/https://doi.org/10.1016/j.ecoinf.2024.102470>
- Fawcett T., "An introduction to ROC analysis" *Pattern Recognition Letter*, 2006;, vol. 7: pp. 861 – 874.
- Fahrudin, N.F. (2019) 'Penerapan Algoritma Apriori untuk Market Basket Analysis', *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, 4(1), pp. 1–11.
- Falgenti, K. and Hambali, E. (2022) 'The analysis and design of independent smallholders FFB palm oil supply system to support traceability of CPO product derivatives', in *IOP Conference Series: Earth and Environmental Science*. Institute of Physics. Available at: <https://doi.org/10.1088/1755-1315/1063/1/012035>.
- Fathurahman, M. (2010) 'Pemilihan Model Regresi Terbaik Menggunakan Akaike's Information Criterion (The Best Regression Model Selection Using Akaike's Information Criterion)', *Eksponensial*, 1(2), pp. 26–33. Available at: <https://fmipa.unmul.ac.id/files/docs/4.Fathurrahman.pdf>.
- Fina Nasari and Surya Darma, S. (2015) 'Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru', *Prosiding Seminar Nasional Teknologi Informasi dan Multimedia 2015*, pp. 73–78. Available at: <https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/viewFile/837/801>.

- Ghozali, I. (2023) Memahami Analisis Regresi Linear Berganda, Binus Faculty Members Article.
- Gupta, A., & Kapoor, R. (2017). The Role of Data Integration in Decision Making: A Review. *International Journal of Computer Applications*, 169(5), 1-5.
- García, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems*, 98, 1–29. <https://doi.org/10.1016/j.knosys.2015.12.006>
- Hajer, B., Arwa, B., Lobna, H., & Khaled, G. (2020). Intention Mining Data preprocessing based on Multi-Agents System. *International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, 00. <https://doi.org/10.1016/j.procs.2020.09.084>
- Hindarto, D. (2023). A comparative study of sentiment classification: traditional nlp vs. neural network approaches. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat*, 8(2), 49–60. <https://doi.org/10.20527/jtiulm.v8i2>
- Herbert Stone, Rebecca N. Bleibaum, Heather A. (2012). *Descriptive Analysis. Sensory Evaluation Practices (Fourth Edition)*. Elsevier. <https://doi.org/10.1016/B978-0-12-382086-0.00006-6>
- Hindarto, D., & Djajadi, A. (2023). Android-manifest extraction and labeling method for malware compilation and dataset creation. *13(6)*, 6568–6577. <https://doi.org/10.11591/ijece.v13i6.pp6568-6577>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques (3rd ed.)*. Morgan Kaufmann.

- Han, J. and Kamber, M. (2019) Data mining: concepts and techniques. London: Elsevier Inc.
- Han, J., Kamber, M. and Pei, J. (2019) Data mining: concepts and techniques. Elsevier.
- Hermawati, F.A. (2013) Data Mining. 1st edn. Edited by P. Christian. Yogyakarta: Penerbit Andi. Available at: https://www.researchgate.net/publication/317692865_Data_Mining.
- Hussein, S. (2021) Analisis Cluster: Pengertian, Contoh dan Metodenya, Geospasialis. Available at: <https://geospasialis.com/analisis-cluster/> (Accessed: 12 April 2024).
- Hossin M. and Sulaiman M.N, "A Review on evaluation metrics for data classification evaluations", International Journal of Data Mining & Knowledge Management Process, Vol. 5., No. 2, March 2015.
- Ichi, P. (2020) Metrik Evaluasi untuk Model Klasifikasi, Ichi.Pro Article.
- Ihwan, A. (2023) CLustering Analysis, LinkedIn Pulse. Available at: <https://id.linkedin.com/pulse/clustering-analysis-aris-ihwan> (Accessed: 12 April 2024).
- Krawczyk, B., & Herrera, F. (2017). A survey on data preprocessing for data stream mining : Current status and future directions. *Neurocomputing*, 239, 39–57. <https://doi.org/http://dx.doi.org/10.1016/j.neucom.2017.01.078>
- Kurgan, L., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review*, 21(1), 1-24.
- Kalaavathi, B. et al. (2023) 'Retail Shop Sales Forecast by Enhanced Feature Extraction with Association Rule Learning', *International Journal on Recent and*

Innovation Trends in Computing and Communication, 11(June), pp. 50–56. Available at: <https://doi.org/10.17762/ijritcc.v11i4s.6306>.

- Khadka, N. (2023) The Ultimate Guide to Association Rule Analysis, Data Science and Data Mining.
- Kurniyawati, F. et al. (2019) Analisis Regresi, Analisis Regresi. Pekanbaru. Available at: https://www.academia.edu/40296796/Analisis_Regresi.
- Kusrini (2021) Konsep dan Aplikasi Sistem Pendukung Keputusan. Penerbit Andi.
- Keogh, E. J. (2004). Visual Data Mining: Techniques and Applications. CRC Press.
- Kerren, A., Kohlhammer, J., & Michael, G. (2009). Visualizing Data Mining Results: Post-Processing and Interpretation. Springer.
- Knaflic, C. N. (2015). Storytelling with Data: A Guide to Communicating Effectively with Data. John Wiley & Sons.
- Johnson, M., & Brown, K. (2015). Unveiling Hidden Patterns: Data Mining Techniques. Journal of Information Science, 42(2), 256-270.
- Jones, P., & White, T. (2016). Harnessing the Power of Integrated Data Analysis. Journal of Business Analytics, 4(1), 78-92.
- Lutkevich, B. (2023) What are association rules in data mining?, Techtarget Business Analytics. Available at: [https://www.techtarget.com/searchbusinessanalytics/definition/association-rules-in-data-mining#:~:text=Association rule mining%2C at a basic level%2C involves,parts%3A an antecedent %28if%29 and a consequent %28then%29](https://www.techtarget.com/searchbusinessanalytics/definition/association-rules-in-data-mining#:~:text=Association rule mining%2C at a basic level%2C involves,parts%3A an antecedent %28if%29 and a consequent %28then%29.). (Accessed: 12 April 2024).

- Montgomery DC, Runger GC, 2023, Applied Statistics and Probability for Engineers, 7th Edition, Wiley, USA.
- Maliki, R. et al. (2022) 'Perbandingan Tingkat Pengangguran Terbuka Provinsi di Indonesia Berbasis Metode K-Means Clustering', Computer Science (CO-SCIENCE), 2(2), pp. 109-116. Available at: https://www.researchgate.net/publication/367597724_Penerapan_Metode_K-Means_Clustering_pada_Data_Tingkat_Pengangguran_Terbuka_Tahun_2016-2018_dan_2019-2021.
- Maulid, R. (2023) Kenali Karakteristik Metrik Model pada Data Science, DQLab Belajar Data Science di Rumah. Available at: <https://dqlab.id/kenali-karakteristik-metrik-model-pada-data-scientist> (Accessed: 12 April 2024).
- Meiryani (2023) Memahami Apa Itu Data Mining, Binus Faculty Members Article. Available at: <https://accounting.binus.ac.id/2019/10/03/memahami-apa-itu-data-mining/> (Accessed: 12 April 2024).
- Masruriyah, A. F. N., Sukmawati, C. E., & Dermawan, B. A. (2024). Memahami Data Mining dengan Python: Implementasi Praktis. In Eureka Media Aksara. Eureka Media Aksara.
- Much Aziz Muslim, Prasetyo, B., Mawarni, E. L. H., Herowati, A. J., Mirqotussa'adah, Rukmana, S. H., & Nurzahputra, A. (2019). Data Mining Algoritma C4.5 Disertai Contoh Kasus dan Penerapannya dengan Program Computer. Ilkom UNNES.
- Mustika, M., Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Guntoro, G., Manuhutu, M. A., Ridwan, M., Hozairi, H., Wardhani, A. K., Alim, S., Romli, I., Religia, Y., Octafian, D. T., Sufandi, U. U., & Ernawati, I. (2021). DATA MINING DAN APLIKASINYA. In N. Rismawati

(Ed.), CV WIDINA MEDIA UTAMA. CV WIDINA MEDIA UTAMA.

- Ni Luh Wiwik Sri Rahayu Ginantra, F. N. A., Anggi Hadi Wijaya, Ri Sabti Septarini, N. A., Dewa Putu Yudhi Ardiana, Faried Effendy, A. I., Hazriani, Ika Yusnita Sari, Z. G., & Cahyo Prianto, Dudih Gustian, E. S. N. (2021). Data Mining dan Penerapan Algoritma. Data Mining Dan Penerapan Algoritma, 20–12.
- Ningsih S ,2021, Algoritma K-Means Untuk Analisis Cluster Kinerja Dosen Pada Bidang Keahliannya Di Unit Penjaminan Mutu FTUKI Universitas Nasional , Jurnal Sistem Informatika (JUNIF), Vol. 2, No. 1.
- Nuryadi, Astuti, T. D., Utami, E. S., & Budiantara, M. (2017). Buku Ajar Dasar-dasar Statistik Penelitian. In Sibuku Media.
- Napizahni, M. (2023) Data Mining: Pengertian, Metode, Fungsi & Penerapannya, Dewaweb Blog Digital Marketing. Available at: <https://www.dewaweb.com/blog/apa-itu-data-mining/> (Accessed: 12 April 2024).
- Narkhede, S. (2018) Understanding Confusion Matrix, Towards Data Science. Available at: <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62> (Accessed: 12 April 2024).
- Nayoan, A. (2022) Apa Itu Data Mining? Manfaat, Proses, dan Contohnya, Insight Bisnis Online. Available at: <https://www.niagahoster.co.id/blog/data-mining-adalah/> (Accessed: 12 April 2024).
- Nurahman, N. and Susanto, J. (2023) 'Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma K-Means', JURIKOM (Jurnal Riset Komputer), 10(2), p. 461. Available at: <https://doi.org/10.30865/jurikom.v10i2.5807>.

- Nurul Sabrina, P., Kania Ningsih, A. and Kasyidi, F. (2022) 'Interpretasi dan Visualisasi Hasil Clustering Menggunakan K-Medoid untuk Identifikasi Penyebaran Virus Covid-19', *Jurnal ICT : Information Communication & Technology*, 21(1), pp. 1-7. Available at: <https://ejournal.ikmi.ac.id/index.php/jict-ikmi/article/download/4/1>.
- Ningsih S ,2022 , Algoritma Apriori Pada Sistem Persediaan Dan Penjualan Tanaman Anggrek Secara Online, *Jurnal Sistem Informasi Bisnis (JUNSIBI)*, Vol. 3, No. 2, 63-70.
- Purnamasari, S.M. (2011) Analisis Kelompok (Cluster Analysis), ITB Working Paper. Bandung. Available at: <http://informatika.stei.itb.ac.id/~rinaldi.munir/Probstat/2010-2011/Makalah2010/MakalahProbstat2010-030.pdf>.
- Purwanto, S. (2023) *Data Mining: Pengertian, Fungsi, Metode & Penerapannya, Jagoan Hosting Technology*.
- Prastyadi Wibawa Rahayu, S.Kom., M. K., I Gede Iwan Sudipa, S.Kom., M. C., Suryani, S.Kom., M. ., Arie Surachman, M. K., Achmad Ridwan, S.Kom., M. K., I Gede Mahendra Darmawiguna, S.Kom., M. S., Ir. Muh. Nurtanzis Sutoyo, S.Kom., M.Cs., I., Drs. Isnandar Slamet, M.Sc., P. ., Sitti Harlina, SE., M. K., & I Made Dendi Maysanjaya, S.Pd., M. E. (2024). *Buku Ajar Data Mining (Issue January)*.
- Pratama IS, Budiarti LY, Nurjaman S, 2021, Optimalisasi Pengelompokan Data Menggunakan Metode Hybrid K-Means-Gaussian Mixture Model, *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 8, No. 5.
- Riszky, A.R. and Sadikin, M. (2019) 'Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi

- Pelanggan ', *Jurnal Teknologi dan Sistem Komputer*, 7(3), pp. 103–108.
- Rosalina, L., Oktarina, R., Rahmiati, & Saputra, I. (2023). *Buku Ajar STATISTIKA*. FEBS Letters, 185(1), 4–8.
- Raharjo, S. (2023) *Panduan Lengkap Uji Analisis Regresi Linear Sederhana dengan SPSS, Tutorial SPSS Indonesia*. Available at: <https://www.spssindonesia.com/2017/03/uji-analisis-regresi-linear-sederhana.html> (Accessed: 12 April 2024).
- Rayward-Smith, V.J. (2007) 'Statistics to measure correlation for data mining applications', *Computational Statistics & Data Analysis*, 51(8), pp. 3968–3982. Available at: <https://doi.org/10.1016/j.csda.2006.05.025>.
- Rizaldi, D. and Adnan, A. (2021) 'Market Basket Analysis Menggunakan Algoritma Apriori: Kasus Transaksi 212 Mart Soebrantas Pekanbaru', *Jurnal Statistika dan Aplikasinya*, 5(1), pp. 31–40. Available at: <https://doi.org/10.21009/jsa.05103>.
- Romli, I. (2021) 'Penerapan Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit ISPA', *Indonesian Journal of Business Intelligence (IJUBI)*, 4(1), p. 10. Available at: <https://doi.org/10.21927/ijubi.v4i1.1727>.
- Shalev-Shwartz S, Ben-David S, 2021, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, UK
- Smith, J., et al. (2018). The Crucial Role of Data Analytics in Modern Problem Solving. *Journal of Applied Statistics*, 45(4), 789-802.
- Sari RTK, Dianto A, 2023, *Perbandingan Metode Evaluasi Kluster Berbasis Statistik untuk Optimalisasi*

Pengelompokan Data, Jurnal Sains dan Informatika, Vol. 9, No. 2.

- Sharyanto, S. and Lestari, D. (2022) 'Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce', JURIKOM (Jurnal Riset Komputer), 9(4), p. 866. Available at: <https://doi.org/10.30865/jurikom.v9i4.4525>.
- Sholeh, M., Nurnawati, E.K. and Lestari, U. (2023) 'Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner', JISKA (Jurnal Informatika Sunan Kalijaga), 8(1), pp. 10–21. Available at: <https://doi.org/10.14421/jiska.2023.8.1.10-21>.
- S. Inayatus and A.F. Nabiilah (2023) Hierarchical Clustering: From Concept to Case Studies, Algoritma Technical Blog. Available at: <https://algotech.netlify.app/blog/introduction-to-hierarchical-clustering/> (Accessed: 12 April 2024).
- Setiawan, A. (2023a) Box-Plot: Memahami dan Membuat Box and Whisker Plots dalam Eksplorasi Data, Analisis Data Eksploratif. Available at: <https://www.smartstat.info/materi/statistika/eksplorasi-data/mengenal-box-plot-box-and-whisker-plots.html> (Accessed: 12 April 2024).
- Setiawan, A. (2023b) Regresi Linier Sederhana: Panduan Lengkap untuk Pemula dalam Statistika, Regresi Linier. Available at: <https://www.smartstat.info/materi/statistika/regresi/regresi-linier-sederhana.html> (Accessed: 12 April 2024).
- Sani, A. (2018) 'Penerapan metode K-means clustering pada perusahaan', Jurnal Ilmiah Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri, (May),

pp. 1-7. Available at:
https://www.researchgate.net/publication/326849650_PENERAPAN_METODE_K-MEANS_CLUSTERING_PADA_PERUSAHAAN.

Shodiq AA, 2023, Pemrograman Machine Learning dengan Python, PT Elex Media Komputindo, Jakarta.

Syarifudin MK, Sari RTK, 2022, Perbandingan Metode Euclidean Distance dan Haversine Distance pada Aplikasi Sistem PPDB dan algoritma K-Means Untuk Menentukan Kebijakan Peraturan Zonasi, Jurnal Faktor Exacta, Vol. 15, No. 4.

Sirisha, N., Gopikrishna, M., Ramadevi, P., Bokka, R., Ganesh, K. V. B., & Chakravarthi, M. K. (2023). IoT-based data quality and data preprocessing of multinational corporations. *Journal of High Technology Management Research*, 34(August). <https://doi.org/https://doi.org/10.1016/j.hitech.2023.100477>

Sousa, S. A. A., Magalhães, A., Miguel, M., & Ferreira, C. (2013). Optimized bucketing for NMR spectra : Three case studies. *Chemometrics and Intelligent Laboratory Systems*, 122, 93-102. <https://doi.org/http://dx.doi.org/10.1016/j.chemolab.2013.01.006>

Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2nd ed.). Graphics Press.

Tanti, L. (2021) *Metode Data Mining Clustering*, Binus Popular Articles.

Trivusi (2023) *Metrik Evaluasi Sistem Menggunakan Confusion Matrix*, Beranda Machine Learning. Available at: <https://www.trivusi.web.id/2022/04/evaluasi-sistem-dengan-confusion-matrix.html?m=1> (Accessed: 12

April 2024).

- Triayudi A, Pinastawa I, Ningsih S, et al. 2024 , Clustering , Penamuda Media, Yogyakarta.
- Tan Pang-Ning, Steinbach M, Kumar V,2006, Introduction To Data Mining , PEARSON Addison Wesley, Pearson Education,Inc . Boston, hal.1-44.
- Williams, A. S. W. A. S., 2011. Statistics for business and economics. CC Liu, JL Chen: A TRIZ.
- Wijaya RA, Suprpto, Suharto, 2022, Analisis Klaster Berbasis Statistik Multivariat untuk Segmentasi Pelanggan pada Industri Ritel, Jurnal Manajemen Bisnis dan Inovasi, Vol. 13, No. 3.
- Walpole RE, Myers RH, Myers SL, Ye K, 2022, Probability and Statistics for Engineers and Scientists, 10th Edition, Pearson, USA.
- Wang, H., et al. (2020). Leveraging Statistical Data and Data Mining Techniques for Enhanced Solutions. International Journal of Information Management, 50, 98-113.
- Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.
- Zaid, Mohamed Ahmed. (2015). Correlation and Regression Analysis. Statistical Economic And Social Research and Training Centre For Islamic Countries.
- Zhou, X., Niu, S., Li, X., Zhao, H., Gao, X., Liu, T., & Dong, J. (2023). Spatial – contextual variational autoencoder with attention correction for anomaly detection in retinal OCT images. Computers in Biology and Medicine, 152(July 2022). <https://doi.org/https://doi.org/10.1016/j.combiomed.2022.106328>

Zhu, J., Ge, Z., Song, Z., & Gao, F. (2018). Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. *Annual Reviews in Control*, 46, 107–133. <https://doi.org/https://doi.org/10.1016/j.arcontrol.2018.09.003>

Tentang Penulis



Sari Ningsih.,S.Si.,MM

Lahir pada tanggal 2 Juni 1967. Penulis merupakan dosen di Fakultas Teknologi Komunikasi dan Informatika Program Studi Sistem Informasi Universitas Nasional. Lulus Program S1 Matematika Komputasi Di FMIPA Universitas Indonesia dan Pasca Sarjana (S2) Ilmu Manajemen Keuangan Asuransi Universitas Gunadarma. Menjadi dosen sejak tahun 1991 sd sekarang. Pengalaman bekerja sebagai Kaprodi Manajemen Informatika Universitas Nasional dan sekarang menjabat Ka Unit Penjaminan Mutu Fakultas Teknologi Komunikasi Dan Informatika Universitas Nasional dan aktif mengajar di beberapa kampus PTS di Jakarta antara lain Universitas Gunadarma, Universitas Pancasila dan beberapa PTS lainnya. Penulis memiliki keahlian mengajar dalam bidang: Matematika: Aljabar Linier, Kalkulus, Statistika Dan Probabilitas, Matematika Diskrit, Riset Operasional, Data Mining dan Metode Numerik.



Miftahul Jannah, S.Kom., M.Kom. Lahir di Renda, 20 Juni 1996. Penulis menyelesaikan strata satu S-1 Teknik Informatika di Universitas Bumigora Mataram dan melanjutkan Megister Informatika di Universitas Islam Negeri Yogyakarta.

Penulis adalah Dosen Ilmu Komputer di Universitas Muhammadiyah Bima. Penulis juga menjadi Dosen terbang pada matakuliah Sistem Informasi Manajemen di Kampus STKIP Taman Siswa Bima di jurusan Pendidikan Teknologi Informasi. Penulis juga aktif melakukan riset di bidang system informasi, Rekayasa Perangkat Lunak dan Data Mining. Penulis dapat di hubungi melalui email mj2100343@gmail.com.



Djarot Hindarto.,S.Kom.,M.Kom

Lahir pada tanggal 17 Januari 1969. Penulis merupakan dosen di Fakultas Teknologi Komunikasi dan Informatika Prodi Informatika di Universitas Nasional. Lulus Program S1 Teknik Komputer di Institute Teknologi Sepuluh Nopember Surabaya dan Pasca Sarjana di Universitas Pradita. Menjadi dosen sejak tahun 2023 sd sekarang. Pengalaman bekerja sebagai Praktisi di bidang System Informasi, Artificial Intelligence, dan Internet of things, sejak 1994 sampai sekarang. Penulis memiliki keahlian mengajar dalam bidang: Data Mining, Deep

Learning, Simulasi dan Pemodelan, Kriptografi dan Internet of Things.



Dr. Nur Putri Erdianti adalah seorang pengajar di Universitas Gunadarma. Lahir di Jakarta pada tahun 1988, pada tahun 2010 dan 2017 telah menamatkan pendidikan S1 dan S2 nya di Universitas Indonesia. Lebih lanjut pada bulan Oktober 2023 beliau berhasil menyelesaikan pendidikan S3 nya di Universitas Gunadarma dalam waktu 3 tahun. Penelitian yang telah dilakukan berfokus pada beberapa bidang yaitu matematika, manajemen, dan ekonomi digital.



Erni Rihyanti adalah seorang pengajar di Universitas Gunadarma. Beliau lulus dari Universitas Indonesia pada tahun 1986 dan mulai mengajar di tahun tersebut. Fokusnya selalu pada bidang matematika sehingga beberapa matakuliah yang diampunya antara lain Matematika Sistem Informasi, Matematika Ekonomi, dan Algoritma Pemrograman. Tahun 2003 beliau menyelesaikan pendidikan Magister Manajemen Universitas Gunadarma.



Ratih Titi Komala Sari., S.T., M.M., M.MSI

Lahir pada tanggal 1 Maret 1983. Penulis merupakan dosen di Fakultas Teknologi Komunikasi dan Informatika Program Studi Informatika Universitas Nasional. Lulus Program S1 Teknik Informatika Di FTI Universitas Gunadarma dan Pasca Sarjana (S2) Program Manajemen dan Program Manajemen Sistem Informasi Universitas Gunadarma. Menjadi dosen sejak tahun 2000 sd sekarang. Pengalaman bekerja sebagai Kepala Lab Multimedia FTKI Universitas Nasional, Sekertaris Program Studi Informatika dan sekarang menjabat sebagai Ketua Program Studi Informatika Fakultas Teknologi Komunikasi Dan Informatika Universitas Nasional dan sedang menyelesaikan studi doktoral di Universitas Diponegoro . Penulis memiliki keahlian mengajar dalam bidang: Pemrograman, Kecerdasan Buatan: Algoritma dan Pemrograman 1, Algoritma dan Pemrograman 2, Mobile Programming, Pemrograman Game, Artificial Intelligence



Ira Diana Sholihati, lahir di Bogor pada 28 Maret 1973 dan saat ini menetap di Bogor. Menyelesaikan Pendidikan program Sarjana di Universitas Padjadjaran Bandung pada tahun 1996 dan program Magister di Universitas Gunadarma Jakarta pada tahun 2003. Saat ini menjadi dosen tetap pada Program Studi Sistem Informasi , Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional Jakarta.



Ir. Endah Tri Esthi Handayani.,M.MSI

Lahir pada tanggal 14 Oktober 1968 di Malang, Jawa Timur. Penulis adalah dosen dan peneliti di Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional Jakarta sejak tahun 2017.

Berlatarbelakang pendidikan sarjana di Teknologi Pangan Universitas Brawijaya , Magister Manajemen Sistem Informasi di Universitas Gunadarma dan saat ini sedang menempuh pendidikan doktoral sistem informasi di Universitas Diponegoro. Bidang pengajaran yang diampu adalah Statistika dan Probabilitas, Matematika Diskrit dan Sistem Informasi Manajemen. Sedangkan bidang penelitian yang ditekuni adalah data sains dan statistika.



Dr. Fauziah, S.Kom, MMSI

Dosen Universitas Nasional prodi Magister Teknologi Informasi, buku yang pernah ditulis antara lain informatika, TKJ, RPL, Pemrograman Gim.



Panca Dewi Pamungkasari, S.T., M.T., Ph.D merupakan salah satu dosen aktif di Fakultas Teknik Komunikasi dan Informatika (FTKI) Universitas Nasioanl yang lahir di Magelang, tanggal 5 Juli 1979. Menyelesaikan S1 di Fakultas Teknik Elektro Universitas Satya Wacana Salatiga, pada tahun 2005. Pada tahun 2005 sampai 2007 melanjutkan kuliah di Teknik Elektro Universitas Indonesia masih dalam bidang telekomunikasi, kemudian tahun 2012 melanjutkan di Keio University Japan dan mendapatkan gelar Ph.D di tahun 2017. Sebelum kembali ke Indonesia pernah menjadi Coperative Researcher di Josaphat Microwave Remote Sensing Laboratory (JMRS�) Chiba University Japan dari tahun 2017 sampai 2018. Pada tahun 2018 sampai 2023, menjadi Visiting Reseacher di Laboratorium Shinichi Shirakawa, Yokohama Nasional University (YNU) Japan dimana fokus penelitiannya dalam bidang Deep Learning khususnya mengenai image processing. Selama menjadi researcher mendapatkan penghargaan sebagai Young Researcher's Encouragement Award dari IEEE VTS Japan pada tahun 2015, Best paper award dalam International Conference on Samrt Computing and Communication (ICSCC) 2019 di Malaysia, beberapa kali menerbitkan International Conference, International Book chapter dan Jurnal dengan berkolaborasi dengan researcher dari Jepang, Malaysia dan Indonesia.



Raden Muhamad Firzatullah, lahir di Bogor pada 6 April 1994. Firzatullah menamatkan pendidikan sarjana di Universitas Negeri Jakarta (UNJ) pada jurusan Pendidikan Teknik Informatika dan Komputer. Selepas lulus dari UNJ, Firzatullah melanjutkan pendidikan di Institut Pertanian Bogor pada jurusan Ilmu Komputer dan meraih gelar M.Kom. Saat ini Firzatullah bekerja sebagai Pegawai Negeri Sipil (PNS) Dosen Komputer di salah satu sekolah kedinasan dibawah naungan Kementerian Perhubungan.



Loso Judijanto adalah peneliti yang bekerja pada lembaga penelitian **IPOSS Jakarta**. Penulis dilahirkan di Magetan pada tanggal 19 Januari 1971. Penulis menamatkan pendidikan *Master of Statistics* di *the University of New South Wales*, Sydney, Australia pada tahun 1998 dengan dukungan beasiswa ADCOS (*Australian Development Cooperation Scholarship*) dari Australia. Sebelumnya penulis menyelesaikan Magister Manajemen di Universitas Indonesia pada tahun 1995 dengan dukungan beasiswa dari Bank Internasional Indonesia. Pendidikan sarjana diselesaikan di Institut Pertanian Bogor pada Jurusan Statistika – FMIPA pada tahun 1993 dengan dukungan beasiswa dari KPS-Pertamina. Penulis menamatkan Pendidikan dasar hingga SMA di Maospati, Sepanjang karirnya, Penulis pernah ditugaskan untuk menjadi anggota Dewan Komisaris dan/atau Komite Audit pada beberapa perusahaan/lembaga yang bergerak di berbagai sektor antara

lain pengelolaan pelabuhan laut, telekomunikasi seluler, perbankan, pengembangan infrastruktur, sekuritas, pembiayaan infrastruktur, perkebunan, pertambangan batu bara, properti dan rekreasi, dan pengelolaan dana perkebunan. Penulis memiliki minat dalam riset di bidang kebijakan publik, ekonomi, keuangan, *human capital*, dan *corporate governance*. Penulis dapat dihubungi melalui e-mail di: losojudijantobumn@gmail.com.



Arie Gunawan

Lahir di kota Jakarta pada tanggal 10 April 1978. Penulis adalah Dosen di Fakultas Teknologi Komunikasi dan Informatika Program Studi Sistem Informasi Universitas Nasional dan juga sebagai Kepala Perpustakaan Universitas Nasional. Menamatkan pendidikan program Sarjana (S1) di Universitas Gunadarma Jakarta prodi Sistem Informasi dan menyelesaikan program Pasca Sarjana (S2) di Universitas Gunadarma prodi Sistem Informasi. Kemudian penulis menyelesaikan kuliah S3 di Universitas Pendidikan Bandung, Jawa Barat mengambil konsentrasi Sistem Informasi Manajemen.

Sebelum menjadi dosen pada tahun 2019, penulis sebelumnya adalah seorang praktisi yang bekerja di beberapa perusahaan, yaitu:

- ❑ PT Nusantara Surya Sakti sebagai MIS Dept Head
- ❑ PT Karunia Abadi Mandiri Persada sebagai IT Manager
- ❑ PT Semesta Finance sebagai IT Supervisor

Penulis memiliki keahlian dalam bidang:

- ❑ Database: MS SQL Server 2005/2008/2014, MySQL
- ❑ Programming: Visual Basic 6.0, Visual Basic.Net, ASP.Net VB, Java NetBeans, Android Studio, C++, Python, HTML 5, Bootstrap, CSS, AJAX Toolkit, Flutter with VS Code
- ❑ ETL Tools: SSIS (SQL Server Integration Services), Pentaho, Talend, DTS (Data Transformation System)
- ❑ BI Tools: Kyubit, Tableau, PowerBI
- ❑ MS Office: Word, Excel, PowerPoint
- ❑ Networking



Dr. Ir. Arief Wibowo, M.Kom., Doktor ilmu komputer lulusan dari FMIPA Universitas Gadjah Mada Yogyakarta. Berkarir sebagai dosen PNS Kementerian Pendidikan dan Kebudayaan Republik Indonesia sejak tahun 2005 dengan penempatan/dipekerjakan (dpk) di Universitas Budi Luhur. Tercatat sebagai anggota Ikatan Ahli Informatika Indonesia dan Persatuan Insinyur Indonesia (PII). Memiliki rekognisi sebagai Insinyur Informatika, dengan gelar profesi Insinyur (Ir.) dari Program Profesi Insinyur Universitas Negeri Yogyakarta. Berpengalaman dalam beberapa pekerjaan dan penugasan di lingkungan POLRI, Kementerian Komunikasi dan Informasi, maupun

Kementerian/Lembaga lainnya, sebagai narasumber pakar/akademisi. Fokus penelitian pada bidang Data Mining/Text Mining, IT User Behaviour, Kecerdasan Buatan dan Knowledge Management. Buku yang pernah diterbitkan adalah “Pemolisian Digital Dengan Artificial Intelligence” (2023), dan “Monograf Model Pengukuran Kualitas Layanan Publik Dengan Indikator PRESISI POLRI Berbasis Kecerdasan Buatan” (2023).

Pengantar DATA MINING & STATISTIK

Buku ini menjelaskan secara mendalam konsep-konsep kunci dalam data mining dan statistik, mulai dari dasar-dasar hingga aplikasi canggihnya dalam berbagai bidang. Dengan penekanan pada analisis data yang relevan dan teknik modeling yang efektif, pembaca akan dibimbing melalui studi kasus praktis yang mengilustrasikan kekuatan gabungan antara data mining dan statistik dalam mengungkap pola penting dan membuat prediksi yang akurat. Dengan membaca buku ini, diharapkan para pembaca dapat meningkatkan pemahaman mereka tentang cara menggali informasi berharga dari data dan mengoptimalkan penggunaannya dalam berbagai konteks.

ISBN 978-623-8586-66-0



9

786238

586660



PT Penerbit Penamuda Media
Godean, Yogyakarta
085700592256
@penamuda_media
penamuda.com