

BUKU AJAR



**PENAMBANGAN**

# **DATA** DAN INTELEGENSI **BISNIS**

(Teori dan Implementasi Menggunakan Pemrograman R)



Ulfa Khaira, S.Komp, M.Kom



**eureka**  
media akhara  
Anggota IKAPI  
No. 225/JTE/2021

0858 5343 1992  
eurekamediaaksara@gmail.com  
Jl. Banjaran RT.20 RW.10  
Bojongsari - Purbalingga 53362



EC00202320930



**BUKU AJAR  
PENAMBANGAN DATA DAN  
INTELEGENSI BISNIS  
(TEORI DAN IMPLEMENTASI  
MENGUNAKAN PEMROGRAMAN R)**

**Ulfa Khaira, S.Komp, M.Kom**



**eureka  
media aksara**

**PENERBIT CV.EUREKA MEDIA AKSARA**

**BUKU AJAR**  
**PENAMBANGAN DATA DAN INTELEGENSI BISNIS**  
**(TEORI DAN IMPLEMENTASI MENGGUNAKAN**  
**PEMROGRAMAN R)**

**Penulis** : Ulfa Khaira, S.Komp, M.Kom

**Desain Sampul** : Eri Setiawan

**Tata Letak** : Siwi Rimayani Oktora

**ISBN** : 978-623-487-830-1

**No. HKI** : EC00202320930

Diterbitkan oleh : **EUREKA MEDIA AKSARA, MARET 2023**  
**ANGGOTA IKAPI JAWA TENGAH**  
**NO. 225/JTE/2021**

**Redaksi:**

Jalan Banjaran, Desa Banjaran RT 20 RW 10 Kecamatan Bojongsari  
Kabupaten Purbalingga Telp. 0858-5343-1992

Surel : eurekaediaaksara@gmail.com

Cetakan Pertama : 2023

**All right reserved**

Hak Cipta dilindungi undang-undang  
Dilarang memperbanyak atau memindahkan sebagian atau seluruh  
isi buku ini dalam bentuk apapun dan dengan cara apapun,  
termasuk memfotokopi, merekam, atau dengan teknik perekaman  
lainnya tanpa seizin tertulis dari penerbit.

## KATA PENGANTAR

Pada awal 1984, dalam bukunya Megatrends, John Naisbitt menyebutkan bahwa “We are drowning in information but starved for knowledge”, kita kaya akan informasi tapi miskin akan pengetahuan. Manusia telah dibanjiri oleh data di berbagai bidang. Data-data ini bila diolah dan dianalisis tentunya akan menghasilkan pengetahuan yang dapat menjadi landasan dalam pengambilan keputusan.

Teknik yang dapat mengekstrak pengetahuan dari data yang besar adalah data mining atau dikenal juga penambangan data. Buku ini disusun agar dapat membantu mahasiswa dalam memahami konsep teori dan implementasi Penambangan Data Dan Inteligensi Bisnis, dalam buku ini juga dilengkapi dengan contoh kasus perhitungan manual dan penerapannya dengan pemrograman R. Akhir kata untuk penyempurnaan buku ini, maka kritik dan saran dari pembaca sangatlah berguna untuk kedepannya.

Jambi, Februari 2023

Penyusun

## DAFTAR ISI

<b>KATA PENGANTAR.....</b>	<b>iii</b>
<b>DAFTAR ISI.....</b>	<b>iv</b>
<b>DAFTAR TABEL .....</b>	<b>vii</b>
<b>DAFTAR GAMBAR.....</b>	<b>viii</b>
<b>BAB 1 KONSEP DASAR PENAMBANGAN DATA DAN</b>	
<b>INTELIGENSI BISNIS.....</b>	<b>1</b>
A. Tujuan Pembelajaran.....	1
B. Mengapa Perlu Mempelajari Penambangan Data dan Inteligensi Bisnis.....	1
C. Definisi Penambangan Data dan Intelegensi Bisnis .....	2
D. Proses Penambangan Data .....	5
E. Arsitektur Penambangan Data.....	5
F. Kebutuhan akan Penambangan Data .....	7
G. Daftar Pustaka .....	8
H. Latihan.....	8
<b>BAB 2 PEMODELAN DAN PERSIAPAN DATA .....</b>	<b>9</b>
A. Tujuan Pembelajaran.....	9
B. Pengertian Data .....	9
C. Kualitas Data.....	10
D. Praproses Data .....	10
E. Daftar Pustaka .....	13
F. Latihan.....	13
<b>BAB 3 TEKNIK DALAM PENAMBANGAN DATA.....</b>	<b>14</b>
A. Tujuan Pembelajaran.....	14
B. Penambangan Data Prediktif .....	14
C. Penambangan Data Deskriptif .....	18
D. Daftar Pustaka .....	26
E. Latihan.....	26
<b>BAB 4 TEKNIK ESTIMASI .....</b>	<b>27</b>
A. Tujuan Pembelajaran.....	27
B. Konsep Estimasi .....	27
C. Konsep Regresi Linier .....	28
D. Algoritma Regresi Linier Sederhana.....	29

	E. Implementasi Regresi Linier Sederhana Menggunakan R.....	38
	F. Daftar Pustaka.....	41
	G. Latihan.....	42
<b>BAB 5</b>	<b>TEKNIK KLASIFIKASI .....</b>	<b>44</b>
	A. Tujuan Pembelajaran.....	44
	B. Konsep Klasifikasi.....	44
	C. Algoritma Bayesien.....	44
	D. Implementasi Algoritma Bayesien Menggunakan R... ..	49
	E. Daftar Pustaka.....	51
	F. Latihan.....	52
<b>BAB 6</b>	<b>TEKNIK CLUSTERING .....</b>	<b>53</b>
	A. Tujuan Pembelajaran.....	53
	B. Konsep Clustering.....	53
	C. Algoritma K-Means Clustering.....	54
	D. Implementasi Algoritma K-Means Clustering Menggunakan R.....	58
	E. Algoritma Agglomerative Hierarchical Clustering .....	60
	F. Implementasi Algoritma Agglomerative Hierarchical Clustering Menggunakan R.....	66
	G. Daftar Pustaka.....	68
	H. Latihan.....	68
<b>BAB 7</b>	<b>TEKNIK ASOSIASI.....</b>	<b>70</b>
	A. Tujuan Pembelajaran.....	70
	B. Konsep Asosiasi.....	70
	C. Algoritma Apriori.....	72
	D. Implementasi Algoritma Apriori Menggunakan R .....	78
	E. Daftar Pustaka.....	80
	F. Latihan.....	80
<b>BAB 8</b>	<b>EVALUASI KINERJA TEKNIK DALAM DATA MINING .....</b>	<b>81</b>
	A. Tujuan Pembelajaran.....	81
	B. Kinerja Teknik Estimasi.....	81
	C. Implementasi Evaluasi Teknik Estimasi Menggunakan R.....	85
	D. Kinerja Teknik Klasifikasi.....	87

E. Implementasi Evaluasi Teknik Klasifikasi	
Menggunakan R .....	91
F. Kinerja Teknik Clustering.....	92
G. Implementasi Evaluasi Teknik Clustering	
Menggunakan R .....	95
H. Kinerja Teknik Asosiasi .....	98
I. Implementasi Evaluasi Teknik Asosiasi	
Menggunakan R .....	100
J. Daftar Pustaka .....	101
K. Latihan.....	101
<b>TENTANG PENULIS.....</b>	<b>103</b>



## DAFTAR TABEL

<b>Tabel 1.1.</b> Bukan Data Mining VS Data Mining .....	3
<b>Tabel 7.1.</b> Dataset Transaksi .....	71

## DAFTAR GAMBAR

<b>Gambar 1.1.</b>	Proses dalam Knowledge discovery (mining) in databases (KDD) .....	3
<b>Gambar 1.2.</b>	Gambaran Business Intelligence (BI).....	4
<b>Gambar 1.3.</b>	Proses Penambahan Data.....	5
<b>Gambar 1.4.</b>	Komponen Utama Arsitektur Data Mining.....	6
<b>Gambar 3.1.</b>	Data set klasifikasi .....	17
<b>Gambar 3.2.</b>	Pohon keputusan .....	17
<b>Gambar 3.3.</b>	Dataset clustering.....	20
<b>Gambar 3.4.</b>	Hasil clustering .....	21
<b>Gambar 3.5.</b>	Cluster plot .....	21
<b>Gambar 3.6.</b>	Cluster dendrogram.....	22
<b>Gambar 3.7.</b>	Dataset transaksi belanja .....	23
<b>Gambar 3.8.</b>	Association Rules .....	25
<b>Gambar 4.1.</b>	Grafik Persamaan Regresi Linier Sederhana (Bangdiwala, 2018).....	30
<b>Gambar 4.2.</b>	Diagram pencar antara nilai UTS dan nilai UAS ....	38
<b>Gambar 6.1.</b>	Proses K-Means Clustering.....	55
<b>Gambar 6.2.</b>	Visualisasi hasil clustering.....	60
<b>Gambar 6.3.</b>	Proses AGNES dan DIANA.....	61
<b>Gambar 6.4.</b>	Dendrogram cluster .....	66
<b>Gambar 6.5.</b>	Dendrogram 2 cluster.....	68
<b>Gambar 8.1.</b>	Diagram Pencar Regresi Linier .....	86
<b>Gambar 8.2.</b>	Confusion matrix .....	88
<b>Gambar 8.3.</b>	Metode elbow .....	93
<b>Gambar 8.4.</b>	Elbow method .....	96
<b>Gambar 8.5.</b>	Visualisasi k optimal metode Average Silhouette Method.....	97
<b>Gambar 8.6.</b>	Visualisasi k optimal metode gap statistic.....	98

# BAB 1

## KONSEP DASAR PENAMBANGAN DATA DAN INTELIGENSI BISNIS

### A. Tujuan Pembelajaran

Setelah mempelajari bab ini mahasiswa diharapkan dapat :

1. Menjelaskan definisi penambangan data
2. Menjelaskan pentingnya penambangan data
3. Menjelaskan peranan penambangan data
4. Menjelaskan implementasi penambangan data
5. Menyebutkan bidang ilmu yang berkaitan dengan penambangan data

### B. Mengapa Perlu Mempelajari Penambangan Data dan Inteligensi Bisnis

Berdasarkan laporan dari sebuah konsultan pemasaran MarketsandMarkets, pasar big data global diperkirakan akan tumbuh dari 162,6 miliar USD pada tahun 2021 menjadi 273,4 miliar USD pada tahun 2026. Korporasi dan institusi di seluruh dunia sedang mempelajari data mining dan teknik prediksi untuk meningkatkan keuntungan. Perusahaan yang tidak menerapkan metode tersebut akan tertinggal dalam persaingan global ekonomi abad ke 21. Pada awal 1984, dalam bukunya Megatrends, John Naisbitt menyebutkan bahwa “We are drowning in information but starved for knowledge”, kita kaya akan informasi tapi miskin akan pengetahuan. Manusia telah dibanjiri oleh data di berbagai bidang. Banyak data yang sudah terkumpul dan tersimpan, baik data dari web (akses web), data

Algoritma yang digunakan : algoritma untuk klasifikasi **C4.5, SVM, KNN.**

- **Unsupervised Learning** : Teknik untuk menganalisis dan mengelompokkan kumpulan data yang tidak berlabel (unlabelled data). Teknik yang bersifat unsupervised learning : estimasi, clustering, asosiasi.

### **G. Daftar Pustaka**

Miranda, E. (2008). Pengembangan Business Intelligence Bagi Perkembangan Bisnis Perusahaan. *CommIT (Communication and Information Technology) Journal*, 2(2), 111-116.

Mishra, B. K., Hazra, D., Tarannum, K., & Kumar, M. (2016, November). Business intelligence using data mining techniques and business analytics. In 2016 International Conference System Modeling & Advancement in Research Trends (SMART) (pp. 84-89). IEEE.

### **H. Latihan**

Untuk memperdalam pemahaman anda mengenai materi pada Bab 1, kerjakanlah latihan berikut !

Berdasarkan konsep data mining yang telah anda pelajari, jelaskan apakah kegiatan menghitung total pendapatan suatu perusahaan adalah termasuk ke dalam data mining?

# BAB 2 | PEMODELAN DAN PERSIAPAN DATA

## A. Tujuan Pembelajaran

Setelah mempelajari bab ini mahasiswa diharapkan dapat :

1. Menjelaskan pengertian data, jenis data, dan kualitas data.
2. Menjelaskan Jenis-jenis set data beserta contoh-contohnya.
3. Menjelaskan proses yang dapat berpengaruh terhadap kualitas data.
4. Menjelaskan pengertian dan tujuan data preprocessing.
5. Menjelaskan berbagai cara data preprocessing beserta contoh penerapannya.

## B. Pengertian Data

Pemodelan dan persiapan data merupakan salah satu tahapan penting dalam proses data mining (Darmawan et al, 2018). Tujuannya untuk mendapatkan hasil pengetahuan atau informasi dari proses data mining yang berkualitas. Dapat dilihat Gambar 1.3, bahwa salah satu tahapan pada data mining adalah praproses data. Bukan data mining namanya jika tidak ada data yang diolah di dalamnya. Data adalah sekumpulan keterangan ataupun fakta yang dibuat dengan kata-kata, kalimat, simbol, angka, dan lainnya. Data disini didapatkan melalui sebuah proses pencarian dan juga pengamatan yang tepat berdasarkan sumber-sumber tertentu. Adapun pengertian lain dari data yaitu sebagai suatu kumpulan keterangan atau deskripsi dasar yang berasal dari obyek ataupun kejadian. Data adalah kumpulan objek dengan atribut-atribut tertentu.

Nama	Fitur 1	Fitur 2
X3	0,360	0,267
X4	-0,935	-0,267
X5	1,223	-1,336
X6	0,791	0,802

## E. Daftar Pustaka

Darmawan, A., Kustian, N., & Rahayu, W. (2018). Implementasi data mining menggunakan model svm untuk prediksi kepuasan pengunjung taman tabebuya. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 2(3), 299-307.

Simanjuntak, K. P., & Khaira, U. (2021). Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering: Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 7-16.

## F. Latihan

Apa tujuan dilakukannya normalisasi data?

Lakukan normalisasi linear pada data berikut :

Nama	Fitur 1	Fitur 2
X1	4	8
X2	2	3
X3	7	2
X4	3	7
X5	9	6
X6	6	9

# BAB 3

## TEKNIK DALAM PENAMBANGAN DATA

### A. Tujuan Pembelajaran

Setelah mempelajari bab ini mahasiswa diharapkan dapat :

1. Menjelaskan konsep teknik predictive data mining
2. Menyebutkan berbagai teknik yang digunakan pada teknik predictive data mining
3. Menjelaskan perbedaan antara teknik estimasi dan klasifikasi
4. Menyebutkan contoh penerapan teknik predictive data mining
5. Menyebutkan contoh algoritma yang digunakan pada teknik predictive data mining
6. Menjelaskan konsep teknik descriptive data mining
7. Menyebutkan berbagai teknik yang digunakan pada teknik descriptive data mining
8. Menjelaskan perbedaan antara teknik klusterisasi dan klasifikasi
9. Menyebutkan contoh penerapan teknik descriptive data mining
10. Menyebutkan contoh algoritma yang digunakan pada teknik descriptive data mining

### B. Penambangan Data Prediktif

Pemodelan data merupakan salah satu tahapan penting dalam proses penambangan data. Tujuannya untuk membuat model yang dapat mengidentifikasi pola-pola diantara atribut-

#### **D. Daftar Pustaka**

- Ginantra, N. L. W. S. R., Arifah, F. N., Wijaya, A. H., Septarini, R. S., Ahmad, N., Ardiana, D. P. Y., ... & Negara, E. S. (2021). Data mining dan penerapan algoritma. Yayasan Kita Menulis.
- Nofriansyah, D., Erwansyah, K., & Ramadhan, M. (2016). Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi). *Jurnal Saintikom*, 15(2).
- Ramady, G. D., & Wowiling, R. G. (2017). Analisa prediksi laju kendaraan menggunakan metode linear regresion sebagai indikator tingkat kemacetan. *Jurnal Online Sekolah Tinggi Teknologi Mandala*, 12(2), 22-28.

#### **E. Latihan**

Setelah memahami teknik dalam penambangan data, jawablah pertanyaan berikut :

Salah satu teknik dalam deskriptif data mining adalah klasterisasi (clustering)

Jelaskan apa itu clustering?, dan jelaskan juga sebuah contoh kasus penerapan clustering dalam kehidupan sehari-hari.



# BAB 4

# TEKNIK ESTIMASI

## A. Tujuan Pembelajaran

Setelah mempelajari bab ini, diharapkan mahasiswa dapat:

1. Menjelaskan konsep regresi linear sederhana
2. Menjelaskan tahapan pada teknik regresi linear sederhana beserta contoh penerapannya.

## B. Konsep Estimasi

Data mining merupakan proses analisa data untuk menemukan suatu pola dari sekumpulan data yang besar dan menerjemahkannya menjadi informasi yang berharga. Salah satu teknik yang ada pada data mining adalah estimasi. Teknik estimasi merupakan teknik untuk memperoleh nilai estimasi dalam sebuah data. Ada dua jenis variabel pada teknik estimasi yaitu variabel prediktor dan variabel target. Variabel prediktor dapat dianalogikan sebagai sebab, sementara variabel target merupakan akibat. Teknik estimasi menghasilkan pengetahuan berupa fungsi atau formula untuk menghitung estimasi nilai variabel target. Pada teknik estimasi, perkiraan nilai dari variabel target ditentukan berdasarkan dari nilai variabel prediktor.

Teknik estimasi telah banyak diterapkan di berbagai bidang. Sebagai contoh pada bidang pertanian, penerapan teknik estimasi untuk menentukan produktivitas tanaman kopi berdasarkan luas areal perkebunan, curah hujan, tekanan udara dan penyinaran matahari (Khumaidi, 2020). Pada bidang

Dari uji parsial dengan t-value sebesar 12,603 dan p-value 1,475e-06, jika p-value lebih kecil dari taraf signifikansi, di mana tingkat signifikansi yang digunakan adalah 0,05, memberikan hasil tolak  $H_0: \beta = 0$ . Artinya terdapat pengaruh yang signifikan antara variabel nilai UTS (X) dengan variabel nilai UAS (Y).

Berdasarkan hasil uji signifikansi diketahui bahwa model regresi linier sederhana tersebut signifikan secara statistik, sehingga model regresi linier sederhana yang dibangun dapat digunakan untuk mengestimasi nilai variabel target. Berikut ini adalah sintaks untuk melakukan estimasi. Dari hasil eksekusi sintaks tersebut, didapatkan hasil bahwa jika nilai UTS 88 maka dapat diestimasi nilai UAS sebesar 89,31.

```
#Estimasi nilai UAS jika nilai UTS sebesar 88  
  
result <- predict(liniermodel, data.frame(UTS  
= 88))  
print(result)  
1  
89.31095
```

## F. Daftar Pustaka

- Bangdiwala, S. I. (2018). Regression: simple linear. *International journal of injury control and safety promotion*, 25(1), 113-115.
- Chatterjee, S., & Hadi, A. S. (2012). *Regression Analysis by Example*, 5 edn.. Hoboken.
- Khumaidi, A. (2020). Data mining for predicting the amount of coffee production using CRISP-DM method. *Techno Nusa Mandiri: Journal of Computing and Information Technology*, 17(1), 1-8. <https://doi.org/10.33480/techno.v17i1.1240>
- Frost, J. (2019). *Regression analysis: An intuitive guide for using and interpreting linear models*. Statistics By Jim Publishing.

- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences*. Houghton Mifflin College Division.
- Marigmen, J. L. D., & Addawe, R. C. (2022, August). Climatic influences on dengue incidence in Baguio city, Philippines: A multiple linear regression approach. In *AIP Conference Proceedings* (Vol. 2472, No. 1, p. 050017). AIP Publishing LLC.
- Mukaka, M. J. M. M. J. (2012). Statistics corner: a guide to appropriate use of correlation in medical research. *Malawi Med J*, 24(3), 69-71.
- Norouzian, R., & Plonsky, L. (2018). Correlation and simple linear regression in applied linguistics. In *The Palgrave handbook of applied linguistics research methodology* (pp. 395-421). Palgrave Macmillan, London.
- Singh, V., Poonia, R. C., Kumar, S., Dass, P., Agarwal, P., Bhatnagar, V., & Raja, L. (2020). Prediction of COVID-19 corona virus pandemic based on time series data using Support Vector Machine. *Journal of Discrete Mathematical Sciences and Cryptography*, 23(8), 1583-1597.
- Yang, Q., Shi, L., Han, J., Zha, Y., & Zhu, P. (2019). Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Research*, 235, 14

## G. Latihan

Sebuah perusahaan ingin memprediksi kinerja pegawainya. Kinerja pegawai diprediksi dengan memperhatikan tingkat stress para pegawai. Manajer perusahaan kemudian mengumpulkan data tingkat stress dan kinerja 10 orang pegawai.

1. Berdasarkan persoalan tersebut tentukan nilai  $a$  dan  $b$ , tuliskan model persamaan regresi, kemudian hitung berapa nilai kinerja jika tingkat stress sebesar 30?
2. Hitunglah nilai koefisien korelasi dan apa makna dari nilai tersebut?

**Tabel Tingkat Kinerja Pegawai**

No	Tingkat Stress	Kinerja Pegawai
1	18	26
2	25	22
3	20	24
4	22	20
5	19	27
6	18	28
7	24	24
8	22	22
9	21	27
10	23	22

# BAB 5

# TEKNIK KLASIFIKASI

## A. Tujuan Pembelajaran

Setelah mempelajari bab ini, diharapkan mahasiswa dapat:

1. Menyebutkan teknik-teknik yang termasuk teknik klasifikasi data mining
2. Menjelaskan pengertian dari algoritma Bayesian
3. Menjelaskan tahapan/langkah pada algoritma Bayesian beserta contoh penerapannya

## B. Konsep Klasifikasi

Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu (Syahfitri et al, 2021). Secara sederhana ada 3 proses pemecahan masalah klasifikasi diantaranya: Data latih atau disebut juga data historis yang tersedia, data historis akan diolah menggunakan algoritma klasifikasi, klasifikasi menghasilkan pengetahuan. Pada buku ini akan dibahas algoritma Bayesian.

## C. Algoritma Bayesian

Algoritma Bayesian adalah sebuah teknik klasifikasi yang menggunakan metode probabilitas dan statistik suatu class. Keuntungan menggunakan teknik ini adalah bahwa teknik ini hanya membutuhkan jumlah data training yang kecil untuk menemukan estimasi parameter yang diperlukan dalam proses

```

#Membuat Prediksi dari model yang telah
dihasilkan
y_pred <- predict(classifier_cl, newdata =
test_data)

#Confusion Matriks

matrik <- table(test_data$play, y_pred)

#Memanggil hasil Confusion matriks
matrik

```

Hasil dari Confusion matriks, Confusion matrix adalah sebuah tabel yang sering digunakan untuk mengukur kinerja dari model klasifikasi di machine learning.

```

y_pred
      No Yes
No    1  0
Yes   1  1

```

```

#Mengukur akurasi

akurasi<-sum(diag(matrik)) / sum(matrik)
print(paste('Akurasi dari pengujian data
sebesar', akurasi*100,'%'))

```

Kode program tersebut menampilkan akurasi dari data uji

```

"Akurasi    dari    pengujian    data    sebesar
66.66666666666667 %"

```

## E. Daftar Pustaka

Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. JURIKOM (Jurnal Riset Komputer), 8(6), 219-225.

Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes. *Walisongo Journal of Information Technology*, 1(1), 1-12.

## F. Latihan

Berikut ini adalah data dari database mahasiswa, lakukan klasifikasi dengan algoritma bayesian untuk menentukan kelas data baru.

Jenis Kelamin	Status Mhs	Status	IPK	Status Lulus
L	Reguler	Lajang	3,17	Tepat
L	Bekerja	Lajang	3,30	Tepat
P	Reguler	Lajang	3.01	Terlambat
P	Reguler	Menikah	3.25	Tepat
L	Bekerja	Menikah	3.01	Terlambat
P	Bekerja	Lajang	3.17	?

# BAB 6 | TEKNIK CLUSTERING

## A. Tujuan Pembelajaran

Setelah mempelajari bab ini, diharapkan mahasiswa dapat:

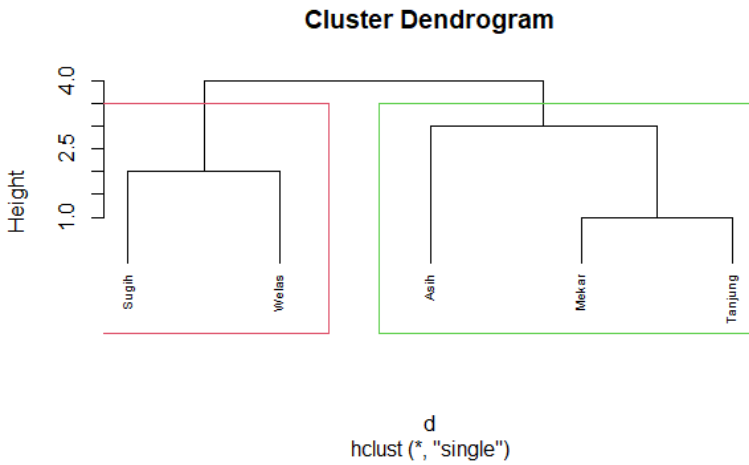
1. Menjelaskan algoritma clustering dalam data mining
2. Menjelaskan tahapan/langkah pada algoritma K-Means beserta contoh penerapannya
3. Menjelaskan tahapan/langkah pada algoritma Hierarchical Agglomerative Clustering beserta contoh penerapannya

## B. Konsep Clustering

Clustering adalah adalah satu teknik analisis data eskplorasi yang paling umum digunakan untuk mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain. Clustering mengelompokkan data ke dalam beberapa kelompok sehingga data dalam satu kelompok memiliki tingkat kemiripan yang maksimum dan data antar kelompok memiliki kemiripan minimum. Hasil clustering yang baik akan menghasilkan tingkat kesamaan yang tinggi dalam satu kelas dan tingkat kesamaan yang rendah antar kelas. Clustering merupakan metode unsupervised learning karena data yang akan dianalisis tidak terdapat target variabel. Dalam unsupervised learning lebih fokus dalam melakukan eksplorasi data seperti mencari pola dalam data.

Clustering banyak dimanfaatkan dalam berbagai bidang, misalnya business intelligence. Clustering bisa mengelompokkan





**Gambar 6.5.** Dendrogram 2 cluster

## G. Daftar Pustaka

- Simanjuntak, K. P., & Khaira, U. (2021). Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering: Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 7-16.
- Yustanti, W., Rahmawati, N., & Yamasari, Y. (2020). Klastering Wilayah Kota/Kabupaten Berdasarkan Data Persebaran Covid-19 Di Propinsi Jawa Timur dengan Metode K-Means. *JIEET (Journal of Information Engineering and Educational Technology)*, 4(1), 1-9.

## H. Latihan

Kelompokkan dataset berikut ini menggunakan metode Hierarchical Agglomerative Clustering Single Linkage menggunakan metric jarak Manhattan dan gambarkan dendogramnya!

No	Nama Mahasiswa	Matematika	B.Inggris	B.indonesia
1	Roy	89	85	75
2	Rudi	70	72	80
3	Justin	85	90	80
4	Ayu	73	65	75
5	Bayu	60	62	48
6	Cecep	52	61	60
7	Dani	76	80	83
8	Bagus	82	78	80
9	Ratna	71	68	62
10	Hafiz	58	55	60

# BAB 7

# TEKNIK ASOSIASI

## A. Tujuan Pembelajaran

Setelah mempelajari bab ini, diharapkan mahasiswa dapat:

1. Menjelaskan algoritma asosiasi dalam data mining
2. Menjelaskan tahapan/langkah pada algoritma Apriori beserta contoh penerapannya

## B. Konsep Asosiasi

Analisis Asosiasi adalah proses mendeteksi kumpulan atribut-atribut yang muncul bersamaan (co-occur) dalam frekuensi yang sering, dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut. Association rule mining adalah analisa dari kebiasaan belanja konsumen dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan konsumen dalam keranjang belanjanya. Contoh Association rule misalnya: "90% dari orang-orang yang membeli mie, juice dan saus akan membeli juga roti tawar". Beberapa penelitian terkait teknik asosiasi diantaranya: Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan (Badrul, 2016), Implementasi data mining untuk mengetahui pola pembelian obat menggunakan algoritma apriori (Ulfha & Amin, 2020).

Dengan kemajuan teknologi, data penjualan dapat disimpan dalam jumlah besar yang disebut dengan "*basket data*". Aturan asosiasi yang didefinisikan pada basket data tersebut, dapat digunakan untuk menganalisa data dalam rangka :

## E. Daftar Pustaka

- Badrul, M. (2016). Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan. *Jurnal Pilar Nusa Mandiri*, 12(2), 121-129.
- Ulfha, N. F., & Amin, R. (2020). Implementasi data mining untuk mengetahui pola pembelian obat menggunakan algoritma apriori. *Komputasi: Jurnal Ilmiah Ilmu Komputer dan Matematika*, 17(2), 396-402.

## F. Latihan

Berikut ini adalah data transaksi dari sebuah supermarket, transaksi diberikan berjumlah 5. Berdasarkan tabel tersebut tentukan aturan asosiasi dengan Algoritma FP Growth, Nilai minimum support 60% dan nilai confidencenya adalah 80%.

<b>Id_transaksi</b>	<b>Item</b>
T01	Susu (Su), Sereal (Se)
T02	Tepung (Te), Gula (Gu), Mentega (Me)
T03	Keju (Ke), Susu (Su), Sereal (Se)
T04	Susu (Su), Sereal (Se), Keju (Ke), Tepung (Te)
T05	Susu (Su), Sereal (Se), Gula (Gu), Keju (Ke), Tepung (Te)

# BAB 8 | EVALUASI KINERJA TEKNIK DALAM DATA MINING

## A. Tujuan Pembelajaran

Setelah mempelajari bab ini, diharapkan mahasiswa dapat:

1. Menjelaskan kinerja algoritma regresi linear
2. Menjelaskan kinerja algoritma klasifikasi
3. Menjelaskan kinerja algoritma clustering
4. Menjelaskan kinerja algoritma asosiasi

## B. Kinerja Teknik Estimasi

Evaluasi kinerja data mining ini digunakan untuk mengukur ketepatan atau jumlah error yang ada pada model yang sudah kita bangun. Hal ini bertujuan agar kita mengetahui seberapa optimal model kita untuk memecahkan suatu permasalahan. Selain itu disini kita bisa menguji kehandalan dataset yang kita punya. Apakah dataset tersebut relevan dan handal untuk diekstrak pengetahuanya.

Kualitas teknik estimasi model regresi dilihat dari seberapa baik prediksi yang cocok dengan nilai aktualnya. Regresi linier menciptakan model yang mengasumsikan hubungan linier antara input dan output. Residual adalah perbedaan antara nilai aktual dan nilai prediksi. Jika Residu kecil maka dapat dikatakan bahwa model bekerja dengan baik untuk menghasilkan output yang diharapkan. Untuk mengukur kualitas model regresi dapat menggunakan Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE).

```

> inspect(trans.rules)
  lhs      rhs      support confidence coverage lift      count
[1] {Baut}   => {Tang}  0.625  0.8333333  0.75  1.111111  5
[2] {Tang}   => {Baut}  0.625  0.8333333  0.75  1.111111  5
[3] {Tang}   => {Meteran} 0.625  0.8333333  0.75  1.111111  5
[4] {Meteran} => {Tang}  0.625  0.8333333  0.75  1.111111  5
[5] {Baut, Meteran} => {Tang} 0.500  1.0000000  0.50  1.333333  4

```

Kualitas dari suatu asosiasi dapat dilihat dari nilai lift rasio, pada rule di atas nilai lift paling tinggi 1,33 ada pada rule Jika pelanggan beli Baut dan Meteran maka pelanggan juga akan membeli Tang.

## J. Daftar Pustaka

- Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, 585, 609-629
- Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion matrix-based feature selection. *Maics*, 710(1), 120-127.

## K. Latihan

Sebuah perusahaan ingin memprediksi kinerja pegawainya. Kinerja pegawai diprediksi dengan memperhatikan tingkat stress para pegawai. Manajer perusahaan kemudian mengumpulkan data tingkat stress dan kinerja 10 orang pegawai. Berdasarkan algoritma regresi linear sederhana, diperoleh formula regresi linear adalah

$$Y = 40,97 - 0,79X$$

Berikut ini adalah data aktual dan prediksi tingkat kinerja pegawai berdasarkan dari formula regresi linear yang dihasilkan. Lakukan perhitungan Root Mean Squared Error (RMSE) dan Mean Absolute Error (MAE) untuk menentukan kualitas hasil algoritma regresi linear yang digunakan untuk menentukan tingkat kinerja pegawai.

No	Tingkat Stress	Kinerja Pegawai	Prediksi Kinerja Pegawai
1	18	26	26,75
2	25	22	21,22
3	20	24	25,17
4	22	20	23,59
5	19	27	25,96
6	18	28	26,75
7	24	24	22,01
8	22	22	23,59
9	21	27	24,38
10	23	22	22,8

## TENTANG PENULIS



**Ulfa Khaira, S.Komp, M.Kom.**, Lahir di Jambi pada tanggal 29 Desember 1989. Lulus S-1 dari Departemen Ilmu Komputer FMIPA Institut Pertanian Bogor (IPB) tahun 2012 dan pada tahun 2013 mendapatkan Beasiswa Pendidikan Pascasarjana Dalam Negeri (BPP-DN) DIKTI untuk melanjutkan studi S-2 di Sekolah Pasca Sarjana Ilmu Komputer IPB lulus tahun 2015. Sejak tahun 2016 bekerja sebagai dosen di Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Jambi. Sejak menjadi dosen, penulis aktif melaksanakan kegiatan Tri Dharma Perguruan Tinggi. Dalam bidang pengajaran, sejalan dengan bidang minatnya penulis mengampu Mata Kuliah yang terkait Data Science, antara lain mata kuliah Penambangan Data dan Intelegensi Bisnis, Statistika Dasar, Statistika Terapan, Riset Operasi, serta Analisis dan Visualisasi Data. Di bidang penelitian, sejak menempuh studi S-1 penulis memiliki minat riset di bidang penambangan data dan geoinformatika. Salah satu hasil penelitian tersebut telah dipublikasi di jurnal internasional terindeks scopus Q3 dengan judul *Detection and prediction of peatland cover changes using support vector machine and Markov Chain Model*. Di bidang pengabdian kepada masyarakat, penulis mengimplementasikan pengetahuan dan kemampuan dalam bidang teknologi informasi diantaranya dalam bentuk pendampingan pemanfaatan teknologi informasi untuk mendukung pembelajaran di sekolah.

Alamat e-mail : [ulfakhaira@unja.ac.id](mailto:ulfakhaira@unja.ac.id)





REPUBLIK INDONESIA  
KEMENTERIAN HUKUM DAN HAK ASASI MANUSIA

## SURAT PENCATATAN CIPTAAN

Dalam rangka perlindungan ciptaan di bidang ilmu pengetahuan, seni dan sastra berdasarkan Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta, dengan ini menerangkan:

Nomor dan tanggal permohonan : EC00202320930, 13 Maret 2023

**Pencipta**

Nama : **Ufa Khaira, S.Komp. M.Kom**

Alamat : Lrg Damai No 23 RT 18 RW 02 Kelurahan Kenali Besar Kecamatan Alam Barajo Kota Jambi, Jambi, JAMBI, 36129

Kewarganegaraan : Indonesia

**Pemegang Hak Cipta**

Nama : **Ufa Khaira, S.Komp. M.Kom**

Alamat : Lrg Damai No 23 RT 18 RW 02 Kelurahan Kenali Besar Kecamatan Alam Barajo Kota Jambi, Jambi, JAMBI, 36129

Kewarganegaraan : Indonesia

Jenis Ciptaan : **Buku**

Judul Ciptaan : **Buku Ajar Penambangan Data Dan Intelegensi Bisnis: Teori Dan Implementasi Menggunakan Pemrograman R**

Tanggal dan tempat diumumkan untuk pertama kali di wilayah Indonesia atau di luar wilayah Indonesia : 10 Maret 2023, di Purbalingga

Jangka waktu perlindungan : Berlaku selama hidup Pencipta dan terus berlangsung selama 70 (tujuh puluh) tahun setelah Pencipta meninggal dunia, terhitung mulai tanggal 1 Januari tahun berikutnya.

Nomor pencatatan : 000453853

adalah benar berdasarkan keterangan yang diberikan oleh Pemohon.  
Surat Pencatatan Hak Cipta atau produk Hak terkait ini sesuai dengan Pasal 72 Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta.



an Menteri Hukum dan Hak Asasi Manusia  
Direktur Jenderal Kekayaan Intelektual  
u.b.  
Direktur Hak Cipta dan Desain Industri

Anggoro Dasananto  
NIP.196412081991031002

Disclaimer:  
Dalam hal pemohon memberikan keterangan tidak sesuai dengan surat pernyataan, Menteri berwenang untuk mencabut surat pencatatan permohonan.