



Data Mining

dan Aplikasinya

Penulis:

Mustika - Yunita Ardilla - Abraham Manuhutu - Nazaruddin Ahmad
Imanuddin Hasbi - Guntoro - Melda Agnes Manuhutu - Mohamad Ridwan - Hozairi
Anindya Khrisna Wardhani - Syariful Alim - Ikhsan Romli - Yoga Religia
D Tri Octafian - Unggul Utan Sufandi & Iin Ernawati

Data Mining

dan Aplikasinya

Penulis:

Mustika - Yunita Ardilla - Abraham Manuhutu - Nazaruddin Ahmad
Imanuddin Hasbi - Guntoro - Melda Agnes Manuhutu - Mohamad Ridwan - Hozairi
Anindya Khrisna Wardhani - Syariful Alim - Ikhsan Romli - Yoga Religia
D Tri Octafian - Unggul Utan Sufandi & Iin Ernawati



DATA MINING DAN APLIKASINYA

Tim Penulis:

Mustika, Yunita Ardilla, Abraham Manuhutu, Nazaruddin Ahmad, Imanuddin Hasbi, Guntoro, Melda Agnes Manuhutu, Mohamad Ridwan, Hozairi, Anindya Khrisna Wardhani, Syariful Alim, Ikhsan Romli, Yoga Religia, D Tri Octafian, Unggul Utan Sufandi & Iin Ernawati.

Desain Cover:

Usman Taufik

Tata Letak:

Handarini Rohana

Editor:

N. Rismawati

ISBN:

978-623-5811-14-7

Cetakan Pertama:

Desember, 2021

Hak Cipta 2021, Pada Penulis

Hak Cipta Dilindungi Oleh Undang-Undang

Copyright © 2021

by Penerbit Widina Bhakti Persada Bandung

All Right Reserved

Dilarang keras menerjemahkan, memfotokopi, atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari Penerbit.

PENERBIT:

WIDINA BHAKTI PERSADA BANDUNG

(Grup CV. Widina Media Utama)

Komplek Puri Melia Asri Blok C3 No. 17 Desa Bojong Emas
Kec. Solokan Jeruk Kabupaten Bandung, Provinsi Jawa Barat

Anggota IKAPI No. 360/JBA/2020

Website: www.penerbitwidina.com

Instagram: @penerbitwidina

PRAKATA

Rasa syukur yang teramat dalam dan tiada kata lain yang patut kami ucapkan selain mengucap rasa syukur. Karena berkat rahmat dan karunia Tuhan Yang Maha Esa, buku yang berjudul “*Data Mining dan Aplikasinya*” telah selesai disusun dan berhasil diterbitkan, semoga buku ini dapat memberikan sumbangsih keilmuan dan penambah wawasan bagi siapa saja yang memiliki minat terhadap pembahasan tentang *Data Mining* dan Aplikasinya.

Akan tetapi pada akhirnya kami mengakui bahwa tulisan ini terdapat beberapa kekurangan dan jauh dari kata sempurna, sebagaimana pepatah menyebutkan “*tiada gading yang tidak retak*” dan sejatinya kesempurnaan hanyalah milik Tuhan semata. Maka dari itu, kami dengan senang hati secara terbuka untuk menerima berbagai kritik dan saran dari para pembaca sekalian, hal tersebut tentu sangat diperlukan sebagai bagian dari upaya kami untuk terus melakukan perbaikan dan penyempurnaan karya selanjutnya di masa yang akan datang.

Terakhir, ucapan terima kasih kami sampaikan kepada seluruh pihak yang telah mendukung dan turut andil dalam seluruh rangkaian proses penyusunan dan penerbitan buku ini, sehingga buku ini bisa hadir di hadapan sidang pembaca. Semoga buku ini bermanfaat bagi semua pihak dan dapat memberikan kontribusi bagi pembangunan ilmu pengetahuan di Indonesia.

Desember, 2021

Tim Penulis

DAFTAR ISI

PRAKATA	iii
DAFTAR ISI	iv
BAB 1 PENGANTAR DATA MINING (DEFINISI DAN RUANG LINGKUP)	1
A. Pendahuluan.....	2
B. Definisi <i>Data Mining</i>	3
C. Fungsi <i>Data Mining</i>	4
D. Ruang Lingkup <i>Data Mining</i>	5
E. Proses <i>Data Mining</i>	9
F. Penerapan <i>Data Mining</i>	10
G. Rangkuman Materi	11
BAB 2 KONSEP MACHINE LEARNING	13
A. Pendahuluan.....	14
B. Komponen <i>Machine Learning</i>	16
C. Proses <i>Machine Learning</i>	17
D. Jenis-Jenis <i>Machine Learning</i>	19
E. Contoh Penerapan <i>Machine Learning</i>	22
F. Tantangan Pembelajaran Mesin (<i>Machine Learning</i>)	23
G. Rangkuman Materi	25
BAB 3 KONSEP DASAR PROSES DATA MINING	27
A. Pendahuluan.....	28
B. Konsep Dasar Proses <i>Data Mining</i>	29
C. Karakteristik dan Fungsi <i>Data Mining</i>	33
D. Tujuan <i>Data Mining</i>	33
E. Manfaat <i>Data Mining</i>	36
F. Rangkuman Materi	37
BAB 4 ARSITEKTUR DAN MODEL DATA MINING	41
A. Pendahuluan.....	42
B. Pengantar <i>Data Mining</i>	42
C. Arsitektur <i>Data Mining</i>	45
D. Model <i>Data Mining</i>	47
E. Rangkuman Materi	58

BAB 5 DATASET	63
A. Pendahuluan.....	64
B. <i>Dataset</i>	65
C. Rangkuman Materi	75
BAB 6 PROSES KNOWLEDGE DAN DISCOVERY	77
A. Pendahuluan.....	78
B. Proses <i>Knowledge</i> dan <i>Discovery</i>	79
C. Gambaran Proses <i>Knowledge</i> dan <i>Discovery</i>	85
D. Model Proses <i>Knowledge</i> dan <i>Discovery</i>	86
E. Rangkuman Materi	92
BAB 7 NAÏVE BAYES CLASSIFIER	95
A. Konsep <i>Naïve Bayes Classifier</i>	96
B. Kegunaan <i>Naïve Bayes</i>	100
C. Kelebihan <i>Naïve Bayes</i>	100
D. Kekurangan <i>Naïve Bayes</i>	101
E. Rangkuman Materi	101
BAB 8 ALGORITMA DATA MINING ESTIMASI LINIER	
REGRESI SEDERHANA DAN GANDA	103
A. Pendahuluan.....	104
B. Regresi	105
C. Koefisien Regresi	106
D. Koefisien Korelasi dan Determinasi	107
E. Regresi Linier Berganda	108
F. Aplikasi Alogaritma Regresi Linier Sederhana	109
G. Alat Analisa Regresi Linier	109
H. Bahasa Pemrograman R	109
I. Contoh Langkah-Langkah Analisa Regresi Linier Sederhana pada R	113
J. Contoh Langkah-Langkah Analisa Regresi Linier Berganda pada R.....	116
K. Visualisasi Data dengan Model Linier	118
L. Rangkuman Materi	119
BAB 9 ALGORITMA DATA MINING KLASIFIKASI DECISION TREE	121
A. Pendahuluan.....	122
B. Algoritma <i>Data Mining</i> Klasifikasi – <i>Decision Tree</i>	122

C.	Algoritma C4.5	126
D.	Rangkuman Materi	136
BAB 10	ALGORITMA KLASTERING K-MEANS	139
A.	Pendahuluan	140
B.	Kelebihan, Kekurangan dan Permasalahan dalam Algoritma <i>Klastering K-Means</i>	142
C.	Karakteristik pada Algoritma <i>Klastering K-Means</i>	144
D.	Langkah <i>Klastering K-Means</i>	145
E.	Contoh Perhitungan	147
F.	Rangkuman Materi	156
BAB 11	ALGORITMA AGGLOMERATIVE HIERARCHICAL CLUSTERING	159
A.	Pendahuluan	160
B.	Algoritma <i>Data Mining - Agglomerative Hierarchical Clustering</i>	161
C.	Klasterisasi	162
D.	Syarat <i>Clustering</i>	163
E.	<i>Agglomerative Hierarchical Clustering</i>	164
F.	Studi Kasus	165
G.	Rangkuman Materi	170
BAB 12	ANALISIS ASOSIASI	173
A.	Pendahuluan	174
B.	Definisi Masalah	175
C.	Pembangkitan <i>Frequent Itemset</i>	177
D.	Pembangkitan Aturan	186
E.	Pemangkasan Berdasarkan <i>Confidence</i>	187
F.	Pembangkitan Aturan dalam Algoritma Apriori	187
G.	Evaluasi Pola Asosiasi	188
H.	Rangkuman Materi	189
BAB 13	EVALUASI MODEL DATA MINING	193
A.	Pendahuluan	194
B.	Evaluasi Model Klasifikasi	195
C.	Evaluasi Model <i>Clustering</i>	207
D.	Rangkuman Materi	209

BAB 14 TOOLS DATA MINING	211
A. Pendahuluan.....	212
B. Aplikasi Penambangan Data	213
C. Praktik Operasi Dasar <i>Data Mining</i> Menggunakan Aplikasi <i>Orange</i>	214
D. Rangkuman Materi	227
BAB 15 PELUANG DAN TANTANGAN DALAM DATA MINING.....	231
A. Pendahuluan.....	232
B. Peluang Dunia Kerja.....	233
C. <i>Trends</i>	235
D. Area Penerapan	236
E. Isu Terkait	240
F. Tantangan.....	242
G. Rangkuman Materi	243
GLOSARIUM	246
PROFIL PENULIS	252



DATA MINING DAN APLIKASINYA

BAB 1: PENGANTAR *DATA MINING* (DEFINISI DAN RUANG LINGKUP)

Mustika, S.Kom., M.Kom

Universitas Muhammadiyah Metro

BAB 1

PENGANTAR *DATA MINING* (DEFINISI DAN RUANG LINGKUP)

A. PENDAHULUAN

Seiring dengan perkembangan zaman, kemajuan dalam proses pengumpulan data dan teknologi penyimpanan yang cepat dan akurat memungkinkan organisasi menghimpun jumlah data yang sangat luas, sehingga penggunaan alat dan teknik analisis data secara manual tentunya tidak dapat digunakan untuk mengekstrak informasi dari data yang sangat besar. Dalam menciptakan efisiensi pengumpulan data yang besar tentunya diperlukan metode baru yang dapat menjawab kebutuhan tersebut. *Data mining* merupakan sebuah teknologi yang dapat memproses data dalam volume besar yang digunakan oleh perusahaan untuk mengubah data mentah menjadi informasi yang berguna untuk membuat suatu keputusan bisnis yang sangat penting. Pada dasarnya *data mining* mempunyai 7 fungsi yaitu *Description*, *Classification*, *Clustering*, *Association*, *Sequencing*, *Forecasting*, dan *Prediction*. *Data mining* mempunyai tujuan sebagai *Explanatory* yaitu untuk menjelaskan beberapa kondisi terkait dengan suatu penelitian, *Confirmatory* digunakan sebagai mempertegas hipotesis, dan *Exploratory* yang berguna dalam menganalisis data untuk hubungan yang baru yang tidak diharapkan. Menurut (Prasetyo, 2014), bahwa ada 4 pokok bahasan (ruang lingkup) dalam keilmuan *data mining*, diantaranya model prediksi (*prediction modelling*), analisis kluster (*cluster analysis*), analisis asosiasi (*association analysis*) serta deteksi anomali (*anomaly detection*).

DAFTAR PUSTAKA

- A.North, Matthew (2012), *Data Mining for the Masses*, Edisi ke-1, Georgia : Global Text Project
- Baker, R. S. J. D. "Data mining for education." *International encyclopedia of education* 7.3 (2010): 112-118
- Baker, R. S. J. D. (2010). Data mining for education. *International encyclopedia of education*, 7(3), 112-118.
- Carneiro da Rocha, Bruno (2010), Identifying Bank Frauds Using CRISP-DM and Decision Trees, *International journal of computer science & information Technology (IJCSIT)* Vol.2, No.5.
- Chung, H. Michael, and Paul Gray. "Data mining." *Journal of management information systems* 16.1 (1999): 11-16.
- J. Zaki, Mohammed dan Wagner Meira JR (2014), *Data Mining and Analysis Fundamentals Concepts and Algorithms*, New York : Cambridge University Press
- Khormarudin, Agus Nur. "Teknik Data Mining: Algoritma K-Means Clustering." *J. Ilmu Komput* (2016): 1-12.
- Nicholson, Scott. "The bibliomining process: Data warehousing and data mining for library decision-making." (2003).
- Susanto, Sani, and Dedy Suryadi. "Pengantar data mining: mengagali pengetahuan dari bongkahan data." (2010).
- Wanto, Anjar, et al. *Data Mining: Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.



DATA MINING DAN APLIKASINYA

BAB 2: KONSEP MACHINE LEARNING

Yunita Ardilla, S.Kom., M.MT

UIN Sunan Ampel Surabaya

BAB 2

KONSEP *MACHINE LEARNING*

A. PENDAHULUAN

Machine learning merupakan salah satu bidang ilmu komputer yang paling pesat perkembangannya, dan merupakan salah satu teknologi paling berpengaruh di dunia pada saat ini. *Big data* merupakan salah satu pendorong atau penyebab mengapa *machine learning* mempunyai pengaruh yang besar. Karena dengan memanfaatkan *machine learning* kita bisa menganalisis dan menemukan pola yang tersembunyi didalam data tersebut. Dimana pola dan pengetahuan yang tersembunyi tentang suatu masalah yang dapat digunakan untuk memprediksi kejadian di masa depan dan melakukan semua jenis pengambilan keputusan yang kompleks. *Machine learning* adalah salah satu *tools* yang digunakan untuk mengubah informasi menjadi pengetahuan.

Sebagian besar dari kita tidak menyadari bahwa kita sudah berinteraksi dengan *machine learning* di kehidupan sehari-hari. Setiap kali kita mencari sesuatu di Google, mendengarkan lagu atau bahkan mengambil foto, *machine learning* menjadi bagian dari sistem-sistem tersebut. Pemanfaatan *machine learning* bisa kita rasakan di segala sektor, misalnya disektor kesehatan, pendidikan, otomotif dan lain sebagainya. Perusahaan seperti Netflix dan Amazon, telah membangun model *machine learning*nya dengan menggunakan *big data* untuk mengidentifikasi peluang yang menguntungkan dan menghindari resiko yang tidak diinginkan. Gambar 2.1 merupakan penyebab *machine learning* digunakan.

TUGAS DAN EVALUASI

1. Jelaskan macam-macam jenis *machine learning* beserta contohnya !
2. Jika data yang kita miliki adalah data yang tidak berlabel, pendekatan *machine learning* apa yang harus kita gunakan?
3. Analisis dampak yang akan timbul jika kecerdasan AI melebihi kecerdasan manusia!
4. Apakah yang dimaksud dengan *Machine learning* ?
5. Jelaskan tren dari penggunaan *machine learning* dimasa yang akan datang !

DAFTAR PUSTAKA

- Ardilla, Y. *et al.* (2014) 'Deteksi Penyakit Epilepsi dengan Menggunakan Entropi Permutasi, K-means Clustering, dan Multilayer Perceptron', *Jurnal Teknik POMITS*, 3(1), pp. 1–5.
- Ardilla, Y. (2019) 'Short Term Forecasting of Electricity Load: A Comparison of Methods to Paiton Subsystem East Java & Bali', *Journal of Information Technology and Computer Science*, 4(3), p. 284. doi: 10.25126/jitecs.201943159.
- Averweg, U. R. F. (2012) *Decision-making Support Systems - Theory & Practice*.
- Jamil, M. Rosihan. Fuad, A. (2016) *Cloud Computing Teori dan Aplikasi*. Yogyakarta: Deepublisher.
- Moisen, G. G. (2008) 'Classification and Regression Trees', *Encyclopedia of Ecology, Five-Volume Set*, (May), pp. 582–588. doi: 10.1016/B978-008045405-4.00149-X.
- Otero, F. E. B., Freitas, A. A. and Johnson, C. G. (2012) 'Inducing decision trees with an ant colony optimization algorithm', *Applied Soft Computing Journal*, 12(11), pp. 3615–3626. doi: 10.1016/j.asoc.2012.05.028.
- Rolliawati, D., Yusuf, A. and Hamdani, A. S. (2018) 'Desain Prototipe Sistem Informasi Penjaminan Mutu Internal Berbasis Standar Nasional Pendidikan Tinggi', *Seminar Nasional Inovasi dan Aplikasi Teknologi di Industri 2018*, 4(1), pp. 167–173.



DATA MINING DAN APLIKASINYA

BAB 3: KONSEP DASAR PROSES *DATA MINING*

Ir. Abraham Manuhutu, M.T

Politeknik Negeri Ambon

BAB 3

KONSEP DASAR PROSES *DATA MINING*

A. PENDAHULUAN

Memasuki era digital saat ini, tidak dapat dipungkiri bahwa terjadi peningkatan yang sangat besar pada jumlah data dan informasi yang disimpan pada media elektronik. Berbagai proses penyimpanan data tersebut ditampung dalam sebuah teknologi yang dinamakan database. Teknologi ini sangat menunjang proses penyimpanan sejumlah data bahkan dalam ukuran yang sangat besar. Namun, hal ini tidak hanya memberikan dampak positif sebab resiko yang dihasilkan pula cukup *drastic*. Salah satu hal yang ditakutkan adalah apabila terjadi persoalan ledakan data, oleh sebab terlalu banyaknya data yang tersimpan. Perlu dipahami bahwa pada prinsipnya data memang perlu disimpan namun lebih penting lagi untuk memahami proses dari data yang disimpan tersebut terutama mengenai proses *knowledge discovery* atau penemuan data. Data yang telah disimpan dalam gudang data atau data *warehouse* sangat perlu untuk dianalisa dan diteliti. Sebuah pemahaman singkat mengenai hal ini adalah ketika seorang pebisnis membutuhkan informasi yang dapat digunakan untuk menunjang operasi-operasi bisnis yang dilakukannya bahkan informasi yang dapat membantunya mengambil sebuah keputusan. Hal ini tidak dapat dilakukan dengan mengambil seluruh data tanpa memilah atau menganalisisnya, sebab sistem manajemen basis data atau database dapat saja memberikan akses terhadap data yang dibutuhkan namun tidak semua data tersebut berguna

DAFTAR PUSTAKA

- , (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Garaha Ilmu. Yogyakarta.
- B. Santoso. (2007). Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.
- Davies, B. (2004). Database Systems 3rd Edition. Palgrave, Basingstoke: United Kingdom
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. Jurnal Edik Informatika, 2(2), 213–219.
- Oskouei, R.J. and Phani Sajja. (2013). Perceived Internet Usage Behaviors as Predictor to Outlier Detection in Students' Communities in Academic Environments. Indian Journal of Science and Technology, Vol. 6, Issue 7, pp. 4923-4935.
- Pramudiono, I. (2007). Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data. <http://www.ilmukomputer.org/wp-content/uploads/2006/08/ikodatamining.zip>. 26 April 2017 (19:54).
- Santosa, Budi (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan. Bisnis. Yogyakarta : Graha Ilmu.
- Suyanto, (2017). Data Mining Untuk Klasifikasi dan Klasterisasi Data. Bandung: Informaatika.
- Y. G. Sucahyo. (2003). Data Mining : Mengenali Informasi Yang Terpendam. IlmuKomputer.com.



DATA MINING DAN APLIKASINYA

BAB 4: ARSITEKTUR DAN MODEL *DATA MINING*

Nazaruddin Ahmad, M.T

Universitas Islam Negeri Ar-Raniry Banda Aceh

BAB 4

ARSITEKTUR DAN MODEL *DATA MINING*

A. PENDAHULUAN

Banyaknya informasi yang masuk membuat kita kebanjiran informasi yang sering tidak dimanfaatkan kembali, atau kita tidak tahu bagaimana mengelola, mengeksplorasi dan mengolah informasi yang masuk menjadi informasi yang penting untuk pengambilan keputusan. Pengolahan informasi ini berfungsi untuk membersihkan data-data yang tidak penting sehingga kita mendapatkan data-data yang sudah disaring menjadi bentuk yang lebih sederhana.

Bab ini menguraikan arsitektur dari *data mining* sehingga akan memiliki gambaran bagaimana pengelolaan informasi atau data-data yang berjumlah besar yang akan diproses menggunakan model-model *data mining* yang dianggap mampu memberikan hasil yang baik dan tepat untuk menginterpretasikan data menjadi informasi yang berguna dalam menyusun kebijakan dan pengambilan keputusan. Dengan mengetahui model-model di dalam *data mining*, kita akan lebih mudah menggunakan model yang tepat untuk menggali informasi yang ada atau data yang ada di dalam database kita.

B. PENGANTAR *DATA MINING*

Salah satu elemen penting yang menjadi kebutuhan kehidupan sehari-hari pada era digital sekarang ini adalah informasi. Dengan penggunaan perangkat teknologi informasi sekarang ini, informasi akan didapatkan

DAFTAR PUSTAKA

- Badrul, M. (2016). Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan. *Jurnal Pilar Nusa Mandiri*, 12(2), 121–129.
- Budiyanto, U., & Fatimah, T. (2019). Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan Jaringan Syaraf Tiruan. *Seminar Nasional APTIKOM (SEMNASITIK)*, 152–160.
- Dahlia, & Andri. (2021). Implementasi Data Mining untuk Prediksi Persediaan Obat pada Puskesmas Kertapati menggunakan Regresi Linier Berganda. *Jurnal Sistem Dan Informatika*, 15(2), 95–103. <https://doi.org/10.30864/jsi.v15i2.331>
- Ginantra, N. L. W. S., Nur Arifah, F., Hadi Wijaya, A., Septarini, R. S., Ahmad, N., Yudhi Ardiana, D. P., Effendy, F., Iskandar, A., Hazriani, H., Sari, I. Y., Gustiana, Z., Prianto, C., Gustian, D., & Surya Negara, E. (2021). Data Mining dan Penerapan Algoritma. In *Yayasan Kita Menulis*.
- Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan. *Faktor Exacta*, 11(3), 266–274. <https://doi.org/10.30998/faktorexacta.v11i3.2777>
- Juni Arta, I. K., Indrawan, G., & Dantes, G. R. (2016). Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Stmik Denpasar Menggunakan Metode Technique for Others Reference By Similarity To Ideal Solution. *JST (Jurnal Sains Dan Teknologi)*, 5(2), 792. <https://doi.org/10.23887/jst-undiksha.v5i2.8549>
- Masrizal, M., & Hadiansa, A. (2019). Prediksi Jumlah Lulusan Mahasiswa Stmik Dumai Menggunakan Jaringan Syaraf Tiruan. *INFORMATIKA*, 9(2), 9–14. <https://doi.org/10.36723/juri.v9i2.98>
- Muningsih, E. (2016). Penentuan rekomendasi produk dengan metode data mining asosiasi generalized sequence pattern (gsp). *Seminar Riset Teknologi Informasi (SRITI)*, 218–224.

- Mustafidah, H. (2010). Model Regresi Data Mining Motivasi Belajar Pengaruhnya Terhadap Tingkat Kedisiplinan Mahasiswa (Data Mining Regression Models of Learning Motivation Effect on Disciplinary Level of Students). *JUITA Universitas Muhammadiyah Purwokerto*, 1(1), 1–4.
- Nasir, H. D., Dahlia, D., & Saharuna, Z. (2020). Prediksi Tingkat Polusi Udara Dengan Data Mining. *Prosiding Seminar Nasional Teknik Elektro Dan Informatika (SNTEI)*, 90–95.
- Purwadi, Ramadhan, P. S., & Safitri, N. (2019). Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang. *Sains Dan Komputer (SAINTIKOM)*, 18(1), 55–61.
- Rahma, I. N., & Setiadi, T. (2014). Penerapan Data Mining Untuk Memprediksi Jumlah Penumpang Bus Trans Jogja Menggunakan Time Series Data. *Jurnal Sarjana Teknik Informatika*, 2(3), 161–171. <https://doi.org/10.12928/jstie.v2i3.2886>
- Wibisono, Y. (2013). *Modul Praktik Data Mining Dengan Tools WEKA*. Ilmu Komputer Universitas Pendidikan Indonesia.



DATA MINING DAN APLIKASINYA

BAB 5: DATASET

Dr. Imanuddin Hasbi, S.T., M.M

Universitas Telkom

BAB 5

DATASET

A. PENDAHULUAN

Kala waktu kehidupan manusia sering digambarkan waktu lampau yang sudah dilalui, masa sekarang yang sedang dijalani, dan masa depan yang akan terjadi nanti. Dalam kurun waktu tersebut terkumpul peristiwa atau kejadian yang menjadi sebuah catatan. Catatan peristiwa ini dikumpulkan menjadi fakta atau data.

Data ialah fakta, statistik, atau item informasi individual, seringkali berupa angka. Data tunggal disebut datum (nilai tunggal dari variabel tunggal). Pengertian secara teknis, data adalah seperangkat nilai variabel kualitatif atau kuantitatif tentang satu atau lebih orang atau objek. (<https://en.wikipedia.org/wiki/Data>). Pemanfaatan data untuk berbagai kepentingan, pemrosesan data biasanya secara bertahap data diukur, dikumpulkan, dilaporkan, dan dianalisis, dan digunakan untuk membuat visualisasi data seperti gambar, peta, tabel, grafik, dan infografik. Konsep umum dari data mengacu pada fakta bahwa beberapa informasi direpresentasikan atau dikodekan dalam beberapa bentuk yang sesuai untuk pemanfaatan dengan pemrosesan yang lebih baik.

Data set (*dataset*) adalah sekumpulan data atau himpunan data. Dalam data tabular, kumpulan data sesuai dengan satu/lebih tabel basis data (database), pada setiap kolom tabel mewakili variabel tertentu, dan setiap baris sesuai dengan catatan tertentu dari kumpulan data yang bersangkutan. Kumpulan data mencantumkan nilai untuk setiap variabel. Kumpulan data terdiri dari kumpulan *file* atau dokumen. *Dataset* adalah kumpulan data yang menggambarkan suatu objek yang memiliki atribut.

DAFTAR PUSTAKA

"Joint Publication 2-0, Joint Intelligence" (PDF). *Joint Chiefs of Staff, Joint Doctrine Publications*. Department of Defense. 23 October 2013. pp. I-1. Retrieved July 17, 2018.

<http://selfmotivator.web.id/2017/apa-itu-data-information-knowledge-dan-wisdom.html>

<https://en.wikipedia.org/wiki/Data>

<https://www.gurupendidikan.co.id/data-mining/>

<https://www.pengalaman-edukasi.com/2020/06/dataset-adalah-data-untuk-data-mining.html>

Leskovec, J., Rajaraman, A., dan Ullman, J. D. (2014) *Mining of Massive Datasets*, United Kingdom: Cambridge University Press.



DATA MINING DAN APLIKASINYA

BAB 6: PROSES KNOWLEDGE DAN DISCOVERY

Guntoro, S.T., M.Kom

Universitas Lancang Kuning

BAB 6

PROSES *KNOWLEDGE* DAN *DISCOVERY*

A. PENDAHULUAN

Setiap kegiatan yang dilakukan oleh manusia dalam kehidupan sehari-hari selalu menghasilkan sebuah data. Kegiatan tersebut dapat berbentuk transaksi komersial, data sensor, foto, video, *posting* di media sosial, lamaran pekerjaan, surat menyurat, kegiatan akademik, dan berbagai bentuk aktivitas lainnya. Jumlah data yang dihasilkan oleh manusia tidaklah sedikit, setiap hari semakin bertambah banyak. Data-data tersebut dibiarkan begitu saja dan dianggap sebagai sampah. Untuk itu perlu dikembangkan suatu sistem yang mampu mengekstraksi substansi informasi yang ada dan menghasilkan laporan secara otomatis. Data ini, bagaimanapun, sekarang mampu ditangani dengan sedemikian rupa sehingga menghasilkan sebuah informasi yang bermanfaat berkat munculnya *Knowledge Discovery in Databases* (KDD), *data mining*, dan *big data*. Kemajuan ilmu pengetahuan dan teknologi berpotensi mengubah persepsi manusia bahwa data hanyalah sampah yang tidak berarti menjadi sesuatu yang dapat dimanfaatkan dan digunakan sebagai pengambilan keputusan. Tumpukan-tumpukan data tersebut dapat diibaratkan seperti tambang emas yang dapat digunakan untuk memperkaya kehidupan manusia. Di era Industri 4.0, data dapat digunakan dalam berbagai masalah seperti bisnis, pembuatan kebijakan, dan pengambilan keputusan.

DAFTAR PUSTAKA

- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (Eds.), 1996. *Advances in Knowledge Discovery and Data Mining*, AAAI Press, Cambridge 8.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. 1996. From data mining to knowledge discovery: an overview. In Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (Eds.), *Advances in Knowledge Discovery and Data Mining*, 1–34, AAAI Press, Cambridge 9.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. 1996. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34 10.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. 1996. Knowledge discovery and data mining: towards a unifying framework. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 82–88, Portland, Oregon
- Agarwal, S., Kamber, M., & Pei, J. (2012). Data mining: Data mining Concepts and Techniques. In *British Library Cataloguing-in-Publication Data*. <https://doi.org/10.1109/ICMIRA.2013.45>
- Osei-Bryson, K. M., & Barclay, C. (2015). Knowledge Discovery Process and Methods: to Enhance Organizational Performance. In *Knowledge Discovery Process and Methods to Enhance Organizational Performance*. <https://doi.org/10.1201/b18231>



DATA MINING DAN APLIKASINYA

BAB 7: NAÏVE BAYES CLASSIFER

Melda Agnes Manuhutu, S.Kom., M.Cs

Universitas Victory Sorong

BAB 7

NAÏVE BAYES CLASSIFER

A. KONSEP NAÏVE BAYES CLASSIFER

Naive Bayes merupakan sebuah pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan. Algoritma menggunakan teorema bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. *Naive Bayes* juga didefinisikan sebagai pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Saleh, 2015).

Naive Bayes Classifier adalah salah satu algoritma pemecahan masalah yang termasuk dalam Metode Klasifikasi *Data Mining*. *Data mining* adalah proses menelusuri pengetahuan baru, pola dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam *repository* atau tempat penyimpanan dengan menggunakan teknik pengenalan pola serta statistik dan teknik matematika (Larose, 2005). Sedangkan, klasifikasi adalah salah satu masalah mendasar dan tugas utama dalam *data mining* (Zhang dkk., 2004), dalam klasifikasi sebuah pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah di tentukan sebelumnya. Performa pengklasifikasi biasanya diukur dengan ketepatan atau tingkat galat (Walpore dkk., 1995).

TUGAS DAN EVALUASI

1. Jelaskan konsep *naïve bayes classifier*?
2. Jelaskan kegunaan dari penerapan metode *naïve bayes*?
3. Sebutkan kelebihan dan kekurangan dari metode *naïve bayes classifier*?
4. Jelaskan kekurangan dari metode *naïve bayes classifier*?

DAFTAR PUSTAKA

- Aggarwal,C., and C. Zhai. (2012). Mining Text Data Chapter A Survey of Text Classification Algorithms. London : Kluwer Academic Publisher.
- Han, J., and M.Kamber. (2006). Data Mining Concept and Techniques Second Edition. San Frasisco : Elsevier.
- Larose, Daniel T. (2005). Data Mining Methods and Models. Canada : A Jhon Willey & Sons, Inc.,Publication.
- Prasetyo, Eko (2014). Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab
- Walpole, E. R., Myers, R. H. (1995). Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuan, Edisi ke-4. Bandung, ITB.
- Zhang, Harry. (2004). The Optimality of Naive Bayes. FLAIRS conference.



DATA MINING DAN APLIKASINYA

BAB 8: ALGORITMA *DATA MINING*

ESTIMASI LINIER REGRESI

SEDERHANA DAN GANDA

Mohamad Ridwan, M.T

Universitas Muhammadiyah Sukabumi (UMMI)

BAB 8

ALGORITMA *DATA MINING* ESTIMASI LINIER REGRESI SEDERHANA DAN GANDA

A. PENDAHULUAN

Kumpulan data yang sangat besar (*big data*) menjadi pemicu perkembangan *data mining*, dimana *data mining* adalah proses menemukan anomali, pola, dan korelasi dalam kumpulan data besar. Seorang analis data akan menjalani tahapan pengumpulan data, dilanjutkan dengan “*cleaning*”, menganalisa dan membuat model dari data tersebut (Lu, Rong, Bao-wen, Pei, & Su, 2018). *Data mining* salah satunya diaplikasikan sebagai alat prediksi. Prediksi didefinisikan sebagai proses perkiraan atau peramalan akan segala sesuatu yang akan terjadi dimasa depan berdasarkan informasi yang dikumpulkan di masa lalu dan sekarang (Ginting, Buulolo, & Siagian, 2019). Dengan demikian fungsi Prediksi sebagai bagian dari fungsi minor *data mining* merupakan salah satu cara untuk mencari kemungkinan hasil dari suatu hal pada masa akan datang dengan kesalahan atau ketidak tepatan antara hasil prediksi dan kenyataan yang tidak terlalu besar. Informasi hasil prediksi dapat digunakan untuk meningkatkan pendapatan, memangkas biaya, meningkatkan hubungan pelanggan, mengurangi risiko, dan banyak lagi.

Dalam data *data mining* atau *machine learning*, metode peramalan digunakan sebagai alat bantu memprediksi kemungkinan akan datang berdasarkan bukti-bukti temuan. Proses prediksi berhubungan erat dengan perhitungan matematik pada kumpulan data yang sangat besar,

TUGAS DAN EVALUASI

1. Menurut Maulud dan kawan-kawan ada 2 manfaat dari regresi. Sebutkan dan Jelaskan!
2. Mengapa Algoritma Regresi Linier salah satu algoritma yang sangat terkenal terutama penggunaannya pada *Data Mining* ?
3. Sebutkan macam-macam Regresi Linier dan Berikan masing-masing contoh implementasinya atau dalam penerapannya !
4. Buatlah contoh sederhana analisis regresi linier sederhana
5. Buatlah contoh sederhana analisis regresi linier berganda

DAFTAR PUSTAKA

- Ginting, F., Buulolo, E., & Siagian, E. R. (2019). Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang). *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3(1). doi:10.30865/komik.v3i1.1602
- Halili, F., & Rustemi, A. (2016). *Predictive Modeling : Data Mining Regression Technique Applied in a Prototype 1*.
- Kurniawan, D. (2008). Regresi Linier. Retrieved from <http://ineddeni.wordpress.com>
- Lu, C., Rong, S., Bao-wen, Z., Pei, X., & Su, J. (2018). The research of regression model in machine learning field. *MATEC Web of Conferences*, 176. doi:10.1051/mateconf/201817601033
- Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1(4), 140- 147. doi:10.38094/jastt1457



DATA MINING DAN APLIKASINYA

BAB 9: ALGORITMA DATA MINING KLASIFIKASI DECISION TREE

Dr. Hozairi, S.ST., M.T

Universitas Islam Madura

BAB 9

ALGORITMA DATA MINING KLASIFIKASI *DECISION TREE*

A. PENDAHULUAN

Pada umumnya metode *data mining* dapat dikelompokkan kedalam dua kategori yaitu deskriptif dan prediktif. Metode deskriptif bertujuan untuk mencari pola yang dapat dimengerti oleh manusia yang menjelaskan karakteristik dari data. Metode prediktif menggunakan ciri-ciri tertentu dari data. Pada umumnya metode *data mining* dapat dikelompokkan kedalam dua untuk melakukan prediksi.

Metode-metode yang ada dalam *data mining* antara lain: (1) Klasifikasi (*Classification*), (2) Pengelompokan (*Clustering*), (3) *Assosiation*, (4) *Regression*, (5) *Forecasting*, (6) *Sequence Analysis*, dan (7) *Deviation Analysis*. Bab ini membahas tentang penerapan metode *data mining* untuk klasifikasi dengan menggunakan metode pohon keputusan (*Decision Tree*), bab ini membahas tentang teori dasar metode, studi kasus, latihan, rangkuman materi serta tugas dan evaluasi.

B. ALGORITMA DATA MINING KLASIFIKASI – *DECISION TREE*

1. Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototipe untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan

DAFTAR PUSTAKA

- Atika & Priatna. (2020). Modul Perkuliahan Data Mining. In *Modul Perkuliahan Data Mining*. Fakultas Ilmu Komputer - Universitas Bhayangkara Jakarta Raya.
- Fauzi, R. (2017). Klasifikasi Siswa Yang Akan Mengikuti Lomba Olimpiade Sains Nasional (OSN) Menggunakan Algoritma C4.5. In *Universitas Muhammadiyah Gresik*.
- Hozairi, S. A. (2020). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor , Decision Tree Serta Naïve Bayes. *Ne*, 2(2), 1–11.
- Prabowo Pudjo Widodo, Herlawati, R. T. H. (2013). *Penerapan Data Mining dengan Matlab*. Bandung: Informatika.



DATA MINING DAN APLIKASINYA

BAB 10: ALGORITMA KLAUSTERING K-MEANS

Anindya Khrisna Wardhani, S.Kom., M.Kom

Politeknik Rukun Abdi Luhur

BAB 10

ALGORITMA KLAUSTERING K-MEANS

A. PENDAHULUAN

Algoritma *Klastering K-means* ditemukan oleh beberapa orang yaitu Lloyd pada tahun 1957 dan 1982, Forgey pada tahun 1965, Friedman dan Rubin pada tahun 1967, dan McQueen pada tahun 1967. Konsep dari pada awalnya ditemukan oleh Lloyd pada tahun 1957, tetapi hal tersebut baru dipublikasikan pada tahun 1982. Pada tahun 1965, Forgey mempublikasikan teknik yang sama dengan nama yang dikenal sebagai Lloyd-Forgey pada beberapa sumber.

Klastering K-means adalah algoritma dengan metode *unsupervised learning* yang paling sederhana dan populer. Terdapat dua jenis data yang sering dipergunakan dalam proses pengelompokan data yaitu *Hierarchical* dan *Non-Hierarchical*. *Klastering K-means* adalah suatu metode penganalisaan data atau metode *data mining* yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) (Huang *et al.*, 2018). Selain itu algoritma *klastering k-means* adalah salah satu metode yang mengelompokkan data dengan sistem partisi. *Klastering K-means* merupakan salah satu metode *klaster analysis non hirarki* yang berusaha untuk mempartisi objek yang ada dalam satu atau lebih kelompok atau kelompok objek yang berdasarkan karakteristiknya, maka objek yang berkarakteristik sama akan dikelompokkan dalam sebuah klaster yang sama serta objek yang memiliki karakteristik berbeda akan dikelompokkan kedalam klaster yang lain. (Wardhani, 2016)

DAFTAR PUSTAKA

- Huang, X., Zhang, L., Wang, B., Li, F., & Zhang, Z. (2018). Feature clustering based support vector machine recursive feature elimination for gene selection. *Applied Intelligence*, 48(3), 594–607. <https://doi.org/10.1007/s10489-017-0992-2>
- Özyurt, F., Sert, E., & Avc, D. (2020). *An expert system for brain tumor detection : Fuzzy C-means with super resolution and convolutional neural network with extreme learning machine*. 134(October 2019). <https://doi.org/10.1016/j.mehy.2019.109433>
- Wardhani, A. K. (2016). Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*, 14(1), 30–37.
- Wardhani, A. K. (2017). Penerapan Algoritma Partitioning Around Medoids Untuk Menentukan Kelompok Penyakit Pasien (Studi Kasus : Puskesmas Kajen Pekalongan). *Jurnal Kajian Ilmu Dan Teknologi (KILAT)*, 6(1), 6–10.
- Wardhani, A. K., Widodo, C. E., & Suseno, J. E. (2018). *Information System for Culinary Product Selection Using Clustering K-Means and Weighted Product Method*. 165(ICCSR), 18–22. <https://doi.org/10.2991/iccsr-18.2018.5>



DATA MINING DAN APLIKASINYA

BAB 11: ALGORITMA AGGLOMERATIVE HIERARCHICAL CLUSTERING

Syariful Alim, S.Kom., M.Cs

Universitas Bhayangkara Surabaya

BAB 11

ALGORITMA AGGLOMERATIVE HIERARCHICAL CLUSTERING

A. PENDAHULUAN

Clustering adalah salah satu metode untuk mengelompokkan data, yang sering dimasukkan sebagai salah satu metode dalam *Data Mining*, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu 'wilayah atau kelompok' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah atau kelompok' yang lain.

Beberapa pendekatan yang digunakan dalam mengembangkan metode *clustering* contohnya adalah pendekatan partisi dan hirarki. *Clustering* dengan pendekatan partisi sering disebut dengan *partition-based clustering* mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam *cluster-cluster* yang ada. *Clustering* dengan pendekatan hirarki sering disebut dengan *hierarchical clustering* mengelompokkan data dengan membuat suatu hirarki berupa dendogram dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan. Di samping kedua pendekatan tersebut, ada juga *automatic mapping (Self-Organising Map/SOM)*

Metode hirarki dibagi berdasarkan teknik pengukuran jarak antar kelompok yang digunakan untuk mengelompokkan obyek. Pengukuran jarak dalam metode hirarki antara lain *single linkage*, proses pengelompokan diawali dengan menemukan dua obyek yang mempunyai jarak minimum. Selanjutnya obyek tersebut menjadi satu kelompok, sedangkan pada metode *complete linkage* menggunakan jarak maksimum

DAFTAR PUSTAKA

- Alfi Fadliana (2015). Penerapan Agglomerative Hierarchical Clustering Untuk Klasifikasi Kabupaten/Kota Di Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana
- Binty, Tri (2019). <https://medium.com/@tribinty/algorithm-agglomerative-hierarchical-clustering-31d2cea14d9>
- Informatikalogi (2019). <https://informatikalogi.com/algorithm-hierarchical-clustering/>
- Maklin, Cory (2019), <https://towardsdatascience.com/machine-learning-algorithms-part-12-hierarchical-agglomerative-clustering-example-in-python-1e18e0075019>
- Sharma, Pulkit (2019). <https://www.analyticsvidhya.com/blog/2019/05/beginners-guide-hierarchical-clustering/>



DATA MINING DAN APLIKASINYA

BAB 12: ANALISIS ASOSIASI

Ikhsan Romli, S.Si., M.Sc

Universitas Pelita Bangsa

BAB 12

ANALISIS ASOSIASI

A. PENDAHULUAN

Analisis asosiasi (*association analysis*) dipakai untuk mengungkap interaksi yang menarik dan tersembunyi pada sebuah *dataset* besar. Hubungan yang dikatakan bisa diimplementasikan pada suatu anggaran asosiasi atau (*association rules*) yaitu himpunan item yang acapkali ada (*sets of frequent items*).

Tabel 12.1 Contoh transaksi belanja

	<i>Item</i>
1	{ <i>Bread, Milk</i> }
2	{ <i>Bread, Diapers, Beer, Eggs</i> }
3	{ <i>Milk, Diapers, Beer, Cola</i> }
4	{ <i>Bread, Milk, Diapers, Beer</i> }
5	{ <i>Bread, Milk, Diapers, Cola</i> }

Sebagai contoh, berikut ini merupakan aturan yang dapat diekstrak dari *dataset* yang transaksi belanja (*market basket transaction*) pada Tabel 12.1.

$$\{Diapers\} \Rightarrow \{Beer\}$$

☐ Data yang masih ada pada atas memperlihatkan bahwa keterikatan yang bertenaga dalam penjualan *diapers* menggunakan *beer* tadi, beberapa pembeli menentukan membeli *diapers* dan membeli *beer*. Data yang masuk dalam transaksi analisis asosiasi pula bisa pada

DAFTAR PUSTAKA

- Daniel T. Larose, *Discovering Knowledge in Data: an Introduction to Data Mining*, *John Wiley & Sons*, 2005
- Ethem Alpaydin, *Introduction to Machine Learning*, 3rd ed., *MIT Press*, 2014
- Florin Gorunescu, *Data Mining: Concepts, Models and Techniques*, *Springer*, 2011
- Ian H. Witten, Frank Eibe, Mark A. Hall, *Data mining: Practical Machine Learning Tools and Techniques* 3rd Edition, *Elsevier*, 2011
- Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques* Third Edition, *Elsevier*, 2012
- Markus Hofmann and Ralf Klinkenberg, *RapidMiner: Data Mining Use Cases and Business Analytics Applications*, *CRC Press Taylor & Francis Group*, 2014
- Oded Maimon and Lior Rokach, *Data Mining and Knowledge Discovery Handbook* Second Edition, *Springer*, 2010
- Warren Liao and Evangelos Triantaphyllou (eds.), *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications*, *World Scientific*, 2007



DATA MINING DAN APLIKASINYA

BAB 13: EVALUASI MODEL *DATA MINING*

Dr. Yoga Religia, S.Kom., M.Kom

Universitas Pelita Bangsa

BAB 13

EVALUASI MODEL DATA MINING

A. PENDAHULUAN

Evaluasi model pada *data mining* merupakan langkah penting yang diperlukan untuk memastikan bahwa pemodelan data yang digunakan mampu memberikan kinerja yang baik. Kinerja pemodelan data berkaitan langsung dengan tujuan dari teknik *data mining* yaitu menemukan pola-pola yang sebelumnya tidak diketahui pada data yang telah dikumpulkan untuk mendukung pengambilan keputusan. Apabila kinerja pemodelan yang digunakan tidak baik maka pola-pola informasi yang ditemukan dari kumpulan data yang digunakan dapat dikatakan tidak kuat untuk dijadikan dasar dalam mendukung pengambilan keputusan. Hal tersebut mengindikasikan bahwa tanpa adanya evaluasi model akan menjadikan pemodelan data yang dibangun menjadi sia-sia karena hasil pemodelannya menjadi tidak dapat digunakan.

Pentingnya tahap evaluasi model menjadikan tujuan dari evaluasi model *data mining* ada empat yang diantaranya adalah sebagai berikut:

- a. Memperoleh kualitas serta efektivitas dari model yang dibangun sebelum hasil pengolahannya disebarakan untuk digunakan.
- b. Memastikan bahwa model yang dibangun telah memenuhi tujuan pada fase awal.
- c. Memastikan bahwa sudah tidak terdapat permasalahan penting dari segi bisnis atau penelitian yang tidak bisa tertangani dengan baik.
- d. Sebagai sarana untuk mengambil keputusan yang berkaitan dengan penggunaan hasil dari model yang telah dibangun.

DAFTAR PUSTAKA

- Kartikasari, M. D. (2021). Self-Organizing Map Menggunakan Davies-Bouldin Index dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan. *Jambura Journal of Mathematics*, 3(2), 187-196.
- Novakovic, J. D., Veljovic, V., Ilic, S. S., Papic, Z., & Tomovic, M. (2017). Evaluation of Classification Models in Machine Learning. *Theory and Applications of Mathematics & Computer Science*, 7(1), 39-46.
- Nugroho, A., & Religia, Y. (2021). Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging. *Jurnal Resti*, 5(3), 504-510.
- Religia, Y., & Sunge, A. S. (2019). Comparison of Distance Methods in K-Means Algorithm for Determining Village Status in Bekasi District. *International Conference of Artificial Intelligence and Information Technology (ICAIT)*. Yogyakarta, Indonesia.



DATA MINING DAN APLIKASINYA

BAB 14: TOOLS DATA MINING

D Tri Octafian, M.Kom

STMIK PalComTech

BAB 14

TOOLS DATA MINING

A. PENDAHULUAN

Data mining atau penambangan data adalah prosedur membentuk dan menyajikan informasi dari kumpulan data yang besar. Penambangan data memakai metode perhitungan seperti statistik, matematika, dan *artificial intelligence* (kecerdasan buatan) dalam prosesnya untuk menghasilkan informasi yang bermanfaat. Istilah lain penambangan data adalah ekstraksi pengetahuan (*knowledge extraction*), analisis pola (*pattern analysis*), pemanenan informasi (*information harvesting*), dan arkeologi data (*data archaeology*). Informasi yang dihasilkan dari proses penambangan data sangat penting bagi sebuah perusahaan atau organisasi sebagai dasar untuk menentukan keputusan atau tindakan yang tepat di masa yang akan datang, selain itu juga hasil proses penambangan data dapat dipakai juga untuk memprediksi tren bisnis, dan melihat pola yang sulit diketahui (Tandika, 2020).

Alat penambangan data atau *data mining tools* adalah sebuah perangkat lunak yang digunakan untuk mentransformasi data dalam ukuran besar menjadi informasi yang terpilih. Pada perangkat lunak penambangan data mengimplementasikan satu atau lebih algoritma tertentu, analisis statistik, kecerdasan buatan dan basis data. Kegunaan dari perangkat lunak penambang data adalah mengekstraksi data dalam jumlah yang besar (*big data*) dan mengubahnya menjadi informasi yang dipahami (admin_sotech, 2020).

DAFTAR PUSTAKA

- Abdurrahman, G. (2019). Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, 4(1), 13. <https://doi.org/10.32528/justindo.v4i1.2418>
- admin_sotech. (2020). *Pilihan Data Mining Tools yang Penting untuk Perusahaan Anda*. <https://sotech.co.id/data-mining-tools/>
- Asmarani, N. G. C. (2020). *Apa Itu Data CSV?* https://news.ddtc.co.id/apa-itu-data-csv-24248?page_y=1000
- Binty, T. (2019). *Algorithm Agglomerative Hierarchical Clustering — and Practice with R*. <https://medium.com/@tribinty/algorithm-agglomerative-hierarchical-clustering-31d2cea14d9>
- Data analysis workflow-ACAPS draft*. (n.d.). 1–10. <http://www.coordinationtoolkit.org/wp-content/uploads/130907-Data-flow.pdf>
- dqlab. (2021). *Cara Membuat Visualisasi Data Scatter Plot dengan Excel*. <https://www.dqlab.id/cara-membuat-visualisasi-data-scatter-plot-dengan-excel>
- Fadliana, A., & Rozi, F. (2015). Penerapan Metode Agglomerative Hierarchical Clustering untuk Klasifikasi Kabupaten/Kota di Propinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana. *Cauchy*, 4(1), 35–40. <https://doi.org/10.18860/ca.v4i1.3172>
- Herlambang, M. (2019). *Machine Learning: Hierarchical Clustering*. <https://www.megabagus.id/machine-learning-hierarchical-clustering/2/>
- Tandika, B. (2020). *5 Aplikasi Data Mining Favorit Para Spesialis*. <https://glints.com/id/lowongan/aplikasi-data-mining/#.YTVzN50zBIU>
- Wurdianarto, S., Wurdianarto, S. R., Novianto, S., & Rosyidah, U. (2014). Perbandingan Euclidean Distance Dengan Canberra Distance Pada Face Recognition. *Techno.Com*, 13(1), 31–37. <https://publikasi.dinus.ac.id/index.php/technoc/article/view/539>



DATA MINING DAN APLIKASINYA

BAB 15: PELUANG DAN TANTANGAN DALAM *DATA MINING*

Unggul Utan Sufandi, S.Kom., M.Si¹

Iin Ernawati, S.Kom., M.Si²

Universitas Terbuka¹

UPN Veteran Jakarta²

BAB 15

PELUANG DAN TANTANGAN DALAM *DATA MINING*

A. PENDAHULUAN

World Economic Forum dalam laporannya menyampaikan bahwa secara umum (dunia) 3 peringkat teratas jenis pekerjaan yang tumbuh pesat hingga tahun 2025 adalah (1) “*Data Analyst dan Data Scientist*”, (2) “*AI dan Machine Learning Specialist*”, (3) “*Big Data Specialist*”. Pada area Indonesia 3 peringkat teratas jenis pekerjaan yang tumbuh pesat hingga tahun 2025 adalah (1) “*Data Analyst dan Data Scientist*”, (2) “*Big Data Specialist*”, (3) “*AI dan Machine Learning Specialist*”.

Peran yang akan dibutuhkan dan akan tumbuh pesat hingga tahun 2025 dalam area Data dan *Artificial Intelligence* antara lain *Artificial Intelligence Specialist, Data Scientist, Data Engineer, Big Data Developer, Data Analyst, Analytics Specialist, Data Consultant, Insights Analyst, Business Intelligence Developer, dan Analytics Consultant*.

Selanjutnya dalam bab ini akan dibahas beberapa hal terkait peluang dan tantangan yang ada dalam *data mining*, yaitu:

1. Peluang Dunia Kerja
2. *Trend Data Mining* pada Masa Depan
3. Area Penerapan *Data Mining*
4. Isu Terkait
5. Tantangan

DAFTAR PUSTAKA

- Albert Verasius Dian Sano, penerapan-penerapan data mining – seri data mining for business intelligence (4), <https://binus.ac.id/malang/2019/01/penerapan-penerapan-data-mining-seri-data-mining-for-business-intelligence-4/>
- Acharjee, Sauvik. 2021. Data Mining Challenges: A Comprehensive Guide(2021).
<https://www.jigsawacademy.com/blogs/data-science/data-mining-challenges>
- Geeks For Geeks - Applications of Data Mining,
<https://www.geeksforgeeks.org/applications-of-data-mining/>
- Ghorbian, Mohsen. 2019.
<https://www.researchgate.net/post/What-are-issues-in-data-mining>
- Mining for Meaning – A Peek inside the World of Data Analytics,
<https://nusbadata.wordpress.com/2017/09/11/mining-for-meaning/>
- Tutorialpoint – Data Mining - Applications & Trends,
https://www.tutorialspoint.com/data_mining/dm_applications_trends.htm
- World Economic Forum, The Future of Jobs Report, 2020
https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf

PROFIL PENULIS

Mustika, S.Kom., M.Kom



Penulis lahir di Palembang, pada tanggal 4 Maret 1983. Menyelesaikan pendidikan strata 1 bidang sistem informasi di STMIK PalComTech Palembang pada tahun 2005. Penulis kemudian melanjutkan studi Software Engineering dan menyelesaikan pendidikan strata 2 di Universitas Bina Darma Palembang pada tahun 2013. Sejak tahun 2008 sampai saat ini penulis masih aktif mengajar. Penulis merupakan dosen tetap pada program studi Ilmu Komputer di Universitas Muhammadiyah Metro sejak tahun 2018, sebelumnya penulis merupakan dosen tetap pada Politeknik Palcomtech Palembang. Fokus bidang keahlian penulis utamanya dibidang rekayasa perangkat lunak dan multimedia. akan tetapi penulis juga tertarik untuk mengamati dan membahas materi terkait bidang pemanfaatan Teknologi Informasi dan Komunikasi terutama di bidang pendidikan.

Yunita Ardilla, S.Kom., M.MT



Penulis lahir di kota Sidoarjo pada tanggal 04 Juni 1992. Riwayat pendidikannya dimulai sejak lulus dari program sarjana Teknik Informatika pada tahun 2014 dari Institut Teknologi Sepuluh Nopember, yang kemudian dilanjutkan sampai jenjang pasca sarjana (S2) Manajemen Teknologi Informasi di Institut Teknologi Sepuluh Nopember. Ia juga menjadi dosen tetap (*faculty member*) Fakultas Dakwah dan Komunikasi di Universitas Islam Negeri Sunan Ampel Surabaya sebagai dosen sistem informasi manajemen. Informasi lengkap dapat dilihat di web pribadi www.yunita.net | surel: yunita.ardilla@uinsby.ac.id

Ir. Abraham Manuhutu, M.T



Penulis menyelesaikan studi Strata-1 pada Universitas Pattimura Ambon dan Strata-2 pada Institut Teknologi Sepuluh Nopember. Penulis merupakan salah satu dosen pada Politeknik Negeri Ambon. Selain mengajar, Penulis juga aktif melakukan penelitian dan pengabdian kepada masyarakat sebagai bentuk implementasi tridarma pendidikan tinggi.

Nazaruddin Ahmad, M.T



Penulis berasal dari kota Banda Aceh, Aceh, lahir di Banda Aceh 5 Juni 1982, merupakan anak keempat dari lima bersaudara dari pasangan Drs. H. Ahmad Habib Lubis dan Hj. Rostina. Saat ini penulis bekerja sebagai dosen tetap pada Prodi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam negeri Ar-Raniry Banda Aceh sejak tahun 2014. Penulis menyelesaikan pendidikan Strata Satu (S1) Teknik Informatika di Universitas Jabal Ghafur Sigli – Aceh Tahun 2008, kemudian melanjutkan pendidikan Strata Dua (S2) di Universitas Atma Jaya Yogyakarta pada program studi magister Teknik Informatika dan mendapatkan gelar Magister Teknik pada tahun 2013. Penulis saat ini mengajar mata kuliah Basis Data, Manajemen Proyek Teknologi Informasi dan Jaringan Syaraf Tiruan. Penulis juga aktif dalam menulis *Bookchapter* untuk pengembangan ilmu dalam bidang Informatika.

Dr. Imanuddin Hasbi, S.T., M.M



Penulis menempuh pendidikan jenjang S1 Teknik Manajemen & Industri, Universitas Pasundan Bandung, S2 Magister Manajemen, Sekolah Tinggi Manajemen Bandung dan S3 Administrasi Pendidikan, Universitas Pendidikan Indonesia. Pendidikan dan pelatihan sertifikasi profesi Metodologi Asesor dari BNSP, sertifikasi profesi keahlian bidang Brand, Service, dan Selling dari BNSP & Markplus Institut, dan sertifikasi profesi *Certified Coaching Entrepreneurship* dari Coaching Indonesia. Penulis sebagai dosen

menjalankan misi tridharma perguruan tinggi pendidikan, penelitian, dan pengabdian kepada masyarakat bidang Manajemen, Administrasi dan Kewirausahaan, di Universitas Telkom. Penulis menulis buku 1] Manajemen Sumber Daya Manusia; 2] Rencana Pemasaran Efektif; 3] Riset Pemasaran; 4] *Entrepreneurship*; 5] Manajemen Sumber Daya Manusia (Sebuah Strategi, Perencanaan, Dan Pengembangan); 6] Manajemen Pariwisata; 7] Sosiologi Komunikasi; 8] Manajemen Risiko; 9] Perilaku Konsumen; 10] Komunikasi Organisasi; 11] Tata Hukum Indonesia; 12] Konsep Dasar Pendidikan Anak Usia Dini; 13] Perkembangan Peserta Didik; 14] Penganggaran Perusahaan; 15] Manajemen Pembiayaan Pendidikan; 16] Kebijakan Publik; 17] Pengantar Pariwisata; 18] Pembelajaran Digital; 19] Pengantar Manajemen dan Bisnis; 20] Psikologi Sosial; 21] Manajemen Perubahan, 22] Administrasi Pendidikan; 23] Pemasaran Kontemporer; dan 24] Karier.

Guntoro, S.T., M.Kom



Penulis lahir di Bangkinang, 12 Januari 1988. Lulus S1 di Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau tahun 2011. Lulus S2 di Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) IPB University tahun 2015. Saat ini adalah Dosen tetap di Program Studi Teknik Informatika Fakultas Ilmu Komputer, Universitas Lancang Kuning. Mengampu mata kuliah Forensika Digital, Sistem Operasi, Teknik Kompilasi, Metode Numerik dan Manajemen Proyek Perangkat Lunak. Saat ini diamanahkan sebagai Kepala Bagian Penelitian LPPM Universitas Lancang Kuning. Sebagai *Editor in Chief* Jurnal Dinamisia: Jurnal Pengabdian Kepada Masyarakat, Jurnal JITACS: *Journal of Information Technology and Computer Science* Universitas Lancang Kuning. Pernah menulis buku Pengantar Teknologi Informasi, buku Praktis Desain Dokumen dan Informasi dengan Canva Android : Membuat Dokumen dan Informasi Berbasis Digital dan Cetak. Ia bisa ditemui di www.guntoro.web.id

Melda Agnes Manuhutu, S.Kom., M.Cs



Penulis lahir di Ambon 21 Agustus 1990. Penulis menyelesaikan Pendidikan S1 dan S2 pada Universitas Kristen Satya Wacana. Saat ini penulis merupakan Dosen Program Studi Sistem Informasi pada Universitas Victory Sorong, mengampuh mata kuliah pengantar teknologi informasi, dasar-dasar pemrograman, jaringan komputer, sistem pendukung keputusan, sistem pakar, metode penelitian, rekayasa perangkat lunak, audit sistem informasi. Penulis aktif melakukan penelitian dan publikasi ilmiah pada skala nasional dan internasional, serta aktif melakukan pengabdian kepada masyarakat sebagai bentuk implementasi tridarma pendidikan tinggi. Saat ini penulis juga aktif dalam menulis buku.

Mohamad Ridwan, M.T



Penulis adalah pengajar di Program Studi Teknik Informatika Universitas Muhammadiyah Sukabumi. Penulis dilahirkan di kota Cirebon yang selanjutnya meneruskan kariernya sebagai pengajar dan menetap di Sukabumi. Lulus Pendidikan S1 dari Universitas Gunadarma Depok Jurusan Sistem Informasi dan meneruskan serta menamatkan kuliah S2 di Universitas Langlangbuana Bandung Jurusan Teknik Informatika dengan konsentrasi Bidang Keamanan Komputer. Penulis juga sudah menghasilkan beberapa karya publikasi dan buku yang sudah diterbitkan. Penulis aktif di beberapa organisasi seperti Forum Komunikasi Dosen, Aptikom, serta komunitas lainnya. Beberapa mata kuliah yang diajarkan seperti Sistem Informasi, Sistem Operasi, Keamanan *Cyber* dan *E- Government*. Penulis saat ini sedang mendalami Bidang Visualisasi Data dan Big Data. Penulis juga tertarik pada bidang jaringan komputer sebagai bagian implementasi dari sebuah *Enterprise System*.

Dr. Hozairi, S.ST., M.T



Penulis adalah pengajar dan peneliti di Universitas Islam Madura. Penulis menempuh D4 Sistem Informasi di Politeknik Elektronika Negeri Surabaya (PENS), S2 Teknik Sistem & Pengendalian Kelautan di Institut Teknologi Sepuluh Nopember (ITS), S3 Teknik Sistem & Pengendalian Kelautan di Institut Teknologi Sepuluh Nopember (ITS). Bidang penelitian penulis adalah *Marine Informatics*, *Data Sains*, *Artificial Intelligent*, *Machine Learning/Deep Learning* dan *Transportation*. Saat ini penulis juga menjabat sebagai Dekan Fakultas Teknik Universitas Islam Madura dan Ketua *Computer Society of Nahdlatul Ulama (CSNU)*.

Anindya Khrisna Wardhani, S.Kom., M.Kom



Penulis lahir di Kudus, Jawa Tengah pada tanggal 15 Agustus 1994. Saat ini penulis merupakan dosen pada Program Studi Rekam Medis dan Informasi Kesehatan pada Politeknik Rukun Abdi Luhur, Kudus, Jawa Tengah. Penulis menyelesaikan Program Sarjana Teknik Informatika di Fakultas Ilmu Komputer Universitas Dian Nuswantoro (Udinus) Semarang tahun 2016. Semangat untuk terus belajar itu membuat penulis terpilih sebagai penerima Beasiswa Unggulan Kementerian Pendidikan dan kebudayaan RI pada Program Magister Sistem Informasi di Universitas Diponegoro (Undip) Semarang dan selesai pada tahun 2018.

Syariful Alim, S.Kom., M.Cs



Penulis adalah pengajar dan peneliti di Program Studi Teknik Informatika Universitas Bhayangkara Surabaya. Matakuliah yang diampu oleh penulis adalah Sistem Optimasi, *Management Science*, Kriptografi, Matematika Diskrit, Metode Numerik dan Sistem Informasi. Bidang penelitian penulis adalah *Optimization*, *Data Sains*, *Artificial Intelligent*, *Machine Learning/Deep Learning* dan *Transportation*. Selain mengajar juga aktif

sebagai konsultan IT di PT. Piramida Teknologi Informasi dan PT Widyatech Latansa Teknologi yang bergerak di bidang rekayasa software untuk kebutuhan perkantoran dan perusahaan.

Ikhsan Romli, S.Si., M.Sc



Penulis lahir di Nganjuk, 13 Mei 1986. Sekarang bekerja sebagai dosen di Teknik Industri, Universitas Pelita Bangsa, Bekasi. Ia lulus S1 Matematika ITS (Institut Teknologi Sepuluh Nopember) Surabaya pada tahun 2009 dan mendapat *Master of Science* dari Universiti Teknikal Malaysia Melaka (UTeM), tahun 2013 di bidang *optimization*. Semasa menempuh S2 di Malaysia, ia bekerja sebagai peneliti di *UteM Research Center* yang didanai oleh FRGS (*Fundamental Research Grant Scheme*) Malaysia dengan menghasilkan dua publikasi internasional yaitu *Symmetric matrices properties to duality in linear programming* dan *Nonlinear Thermal Expansion Model for SiC/Al*. Selama berkarir menjadi dosen, Penulis juga pernah mendapatkan hibah penelitian dari DRPM Kemenristek antara lain:

- Skema Penelitian Dosen Pemula (PDP) dengan judul “Optimasi Penjadwalan Mata Kuliah Di Prodi Teknik Informatika STT Pelita Bangsa Dengan Menggunakan Algoritma Genetika”. (2017-2018)
- Skema Penelitian Dosen Pemula (PDP) dengan judul “Penerapan Algoritma *Naive Bayes* dalam Menentukan Merek Mobil Paling Laku”. (2018-2019)
- Skema Penelitian Dosen Pemula (PDP) dengan judul “Klasifikasi Data Karyawan Untuk Menentukan Jadwal *Overtime* Perusahaan Menggunakan Algoritma C4.5”. (2019-2020).
- Skema Penelitian Dosen Pemula (PDP) dengan judul “Implementasi *Internet Of Things* Dalam *Monitoring* Pembangkit Listrik Tenaga Surya”. (2020-2021).

Dr. Yoga Religia, S.Kom., M.Kom



Penulis merupakan dosen Teknik Informatika di Universitas Pelita Bangsa. Lulus dari jenjang S1 dan S2 Teknik Informatika melalui program Beasiswa Unggulan dan Fast Track di Universitas Dian Nuswantoro dan menyelesaikan S3 di Universitas Brawijaya pada usia 28 tahun. Aktif dalam kegiatan penelitian dengan melakukan publikasi penelitian di berbagai jurnal nasional beriputasi dan jurnal internasional bereputasi. Pernah menjadi data *analyst* di Kementerian Desa PDTT. Selain mengajar sebagai dosen, penulis juga membuat *belanjaukm.com* dan *bromobook.com*. Marketplace *belanjaukm.com* menjembatani pembinaan, pelatihan, pengembangan dan penyediaan layanan *e-commerce* agar UMKM dapat memasuki pasar digital. Toko buku *online* *Bromobook.com* bertujuan untuk menjadi toko buku *online* terkemuka sehingga mampu menciptakan literasi yang kuat di Indonesia, serta mampu membawa penulis dan penerbit ke pasar internasional. Info lebih lengkap dapat dilihat melalui yogareligia.com

D Tri Octafian, M.Kom



Penulis lahir di Palembang. Riwayat pendidikan penulis adalah menyelesaikan pendidikan D1 Analisis Komputer di Lembaga LPPMK Pranata Mulya Palembang pada tahun 2002, menyelesaikan pendidikan strata 1 program studi Teknik Informatika dengan konsentrasi *Software Engineering* di Universitas Indo Global Mandiri Palembang pada tahun 2009, menyelesaikan pendidikan strata 2 program studi Teknik Informatika dengan konsentrasi *Software Engineering* di Universitas Bina Darma Palembang pada tahun 2012. Sejak tahun 2008 sampai saat ini penulis adalah dosen tetap di program studi S1 Informatika STMIK PalComTech di Palembang. Fokus bidang keahlian penulis utamanya di bidang rekayasa perangkat lunak dan penulis adalah *founder channel* Youtube *kidimi* yang membahas seputar pemrograman komputer dan selain sebagai dosen, penulis merupakan *programmer* lepas.

Unggul Utan Sufandi, S.Kom., M.Si



Penulis kelahiran Kebumen, 11 September 1971 merupakan seorang dosen pada Program Studi Sistem Informasi sebuah perguruan tinggi negeri ke 45 yaitu Universitas Terbuka. Mengawali karir pada tahun 1993 sebagai *programmer* pada era *Novell Netware* menggunakan bahasa pemrograman Foxpro. Pada tahun 2000 beralih ke pemrograman berbasis desktop menggunakan *Power Builder* dengan database RDBMS, Sybase SQL *Anywhere*. Menyelesaikan pendidikan S1 Program Studi Sistem Informasi pada tahun 2003 di Universitas Pembangunan Nasional Veteran Jakarta, dan dilanjutkan dengan menamatkan Pendidikan S2 Program Studi Ilmu Komputer pada tahun 2007 pada Institut Pertanian Bogor. Buku yang pernah ditulis antara lain (1) Pemrograman Database Menggunakan Delphi; (2) Struktur Data serta *book chapter* pada buku (3) Universitas Terbuka di Era Informasi dengan judul *chapter "E-readiness Indonesia Untuk Pendidikan Pada Era digital"*. Pria yang berpengalaman dalam mengembangkan digital *printing system, contact center, student record system*, serta sebagai tim pendamping dalam pengembangan *enterprise architecture* ini memiliki minat pengajaran dan penelitian pada area AI dan *Machine Learning, Data Science, Human Computer Interaction, Software Engineering* dan *IT/IS Governance*.

lin Ernawati, S.Kom., M.Si



Penulis lahir di Madiun – Jawa Timur, pada tanggal 2 Januari 1976. Mengawali karir sebagai asisten laboratorium komputer pada Fakultas Ilmu Komputer – UPN Veteran Jakarta pada tahun 2001, khusus mata kuliah algoritma dan pemrograman, pemrograman C++, dan Basis data. Sejak tahun 2001 hingga saat ini aktif sebagai dosen Fakultas Ilmu Komputer UPN Veteran Jakarta. Menyelesaikan pendidikan S1 Manajemen Informatika di UPN Veteran Jakarta pada tahun 2000, kemudian melanjutkan pendidikan ke jenjang (S2) Magister Ilmu Komputer di Departemen Ilmu Komputer – FMIPA IPB pada tahun 2008. Penelitian yang pernah dilakukan mayoritas berkaitan dengan bidang *Data Mining*

dan *Machine Learning*, baik dilakukan secara mandiri maupun berkolaborasi dengan mahasiswa dan rekan sejawat. Pada tahun 2020 bergabung bersama beberapa rekan sejawat menulis sebuah bab terkait *Database Management System (DBMS)* dalam sebuah *book chapter* dengan judul “Pengantar Teknologi Informasi”.

Data Mining

dan Aplikasinya



Data mining merupakan suatu alat yang memungkinkan para pengguna untuk mengakses secara cepat data dengan jumlah yang besar juga sebagai suatu proses ekstraksi atau penggalian data dan informasi yang besar, yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Data mining menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui pada data yang telah dikumpulkan. Data mining memungkinkan pemakai menemukan pengetahuan dalam data dari database yang tidak mungkin diketahui keberadaannya oleh pemakai. Teknik data mining digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Maka, buku ini menyajikan segala komponen yang dibutuhkan oleh para pengelola data dalam menjalankan pengelolaannya untuk menciptakan kualitas, yang dapat menjadikan pengolahan data secara efektif. Oleh sebab itu buku ini hadir dihadapan sidang pembaca sebagai bagian dari upaya diskusi sekaligus dalam rangka melengkapi khazanah keilmuan dibidang informatika, sehingga buku ini sangat cocok untuk dijadikan bahan acuan bagi kalangan intelektual dilingkungan perguruan tinggi ataupun praktisi yang berkecimpung langsung dibidang informatika.