



DATA MINING

dan PENERAPAN METODE

Febri Liantoni

DATA MINING

dan PENERAPAN METODE

Data mining merupakan proses untuk mendapatkan informasi yang berguna dari basis data yang besar yang perlu diekstraksi agar menjadi informasi baru. Data mining atau yang biasa disebut juga dengan Knowledge Discovery in Database (KDD) berguna dalam proses pengumpulan dan pengolahan data yang valid dan dinilai potensial serta mudah dipahami polanya dan dapat membantu dalam pengambilan keputusan. Data yang diolah dengan metode yang efektif dan efisien akan menghasilkan informasi yang akurat. Pada buku ini akan dibahas materia terkait pengenalan data mining, forecasting, classification, clustering dan association.



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

ISBN 978-623-487-009-1



9 786234 870091

DATA MINING DAN PENERAPAN METODE

Febri Liantoni



eureka
media aksara

PENERBIT CV. EUREKA MEDIA AKSARA

DATA MINING DAN PENERAPAN METODE

Penulis : Febri Liantoni

Desain Sampul : Eri Setiawan

Tata Letak : Rizki Rose Mardiana

ISBN : 978-623-487-009-1

Diterbitkan oleh : **EUREKA MEDIA AKSARA, JUNI 2022**
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 : 2022

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

Alhamdulillah, segala puji dan syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan buku ini. Buku ini disusun agar dapat membantu para mahasiswa dalam mempelajari data mining beserta penerapannya.

Tidak lupa penulis ucapkan terima kasih kepada Kaprodi, Rekan Dosen dan Mahasiswa Prodi Pendidikan Teknik Informatika dan Komputer, Universitas Sebelas Maret atas dukungan, do'a, saran dan kritik dalam penyelesaian buku ini.

Penulis menyadari sepenuhnya bahwa buku ini masih jauh dari sempurna dikarenakan pengalaman dan pengetahuan yang penulis miliki. Oleh karena itu, penulis mengharapkan segala bentuk saran serta masukan bahkan kritik yang membangun dari berbagai pihak. Akhirnya penulis berharap semoga buku ini dapat memberi manfaat bagi perkembangan dunia pendidikan.

DAFTAR ISI

KATA PENGANTAR.....	iii
DAFTAR ISI.....	iv
DAFTAR GAMBAR.....	vii
DAFTAR TABEL.....	ix
BAB 1 PENGENALAN DATA MINING.....	1
A. Pendahuluan.....	1
B. Definisi Data Mining.....	2
C. Manfaat Data Mining.....	6
D. Tahap-Tahap Data Mining.....	7
1. Eksplorasi atau pemrosesan awal data.....	7
2. Membangun model dan validasi.....	7
3. Penerapan.....	7
E. Teknik Data Mining.....	10
1. <i>Forecasting</i> (Prediksi/ Peramalan).....	10
2. <i>Classification</i> (Klasifikasi).....	10
3. <i>Clustering</i> (Klastering).....	11
4. <i>Association</i> (Asosiasi).....	11
F. Kesimpulan.....	11
BAB 2 FORECASTING (PREDIKSI/ PERAMALAN).....	12
A. Pendahuluan.....	12
B. Forecasting.....	13
1. <i>Linear Regression (LR)</i>	15
2. <i>Deep Learning</i>	17
3. <i>Generalized Linear Model (GLM)</i>	19
4. <i>Moving Average</i>	22
5. <i>Seasonal Adjustment</i>	25
C. Penerapan Studi Kasus.....	28
1. Peramalan Bitcoin Menggunakan <i>Double Exponential Smoothing</i>	28
2. Perbandingan Metode Single Exponential Smoothing dan Metode Holt Untuk Prediksi Kasus COVID-19 di Indonesia.....	29
D. Kesimpulan.....	30

BAB 3 CLASSIFICATION (KLASIFIKASI).....	31
A. Pendahuluan.....	31
B. <i>Classification</i>	31
1. <i>Decision Tree</i>	32
2. <i>Naive Bayes Classifier</i>	41
3. <i>K-Nearest Neighbor</i>	46
C. Penerapan Studi Kasus	61
1. Klasifikasi Daun dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor	61
2. Perbaikan Kontras Citra Mammogram pada Klasifikasi Kanker Payudara Berdasarkan Fitur Gray-Level Co-Occurrence Matrix.....	63
3. Penerapan Algoritma C4.5 dalam Memprediksi Lama Studi Mahasiswa	64
D. Kesimpulan.....	65
BAB 4 CLUSTERING (KLASTERING).....	67
A. Pendahuluan.....	67
B. <i>Clustering</i>	67
1. K-Means	71
2. K-Medoids	74
3. Self-Organizing Map (SOM)	76
4. Fuzzy C-Means (FCM)	78
5. Hoshen-Kopelman Algorithm.....	79
6. DBSCAN	80
7. OPTICS	80
8. Hierarchical Clustering.....	80
C. Penerapan Studi Kasus	84
1. Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means	84
2. Image Cluster Berdasarkan Warna Untuk Identifikasi Kematangan Buah Tomat Dengan Metode Valley Tracing ..	86
3. Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret).....	88
D. Kesimpulan.....	91

BAB 5 ASSOCIATION (ASOSIASI)	93
A. Pendahuluan	93
B. Association.....	93
1. FP-Growth.....	94
2. A Priori	94
3. Coefficient of Correlation	95
4. Chi-Squared	96
5. TOPSIS.....	96
6. SAW	97
7. Weighted Product (WP).....	99
C. Penerapan Studi Kasus.....	100
1. Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan.....	100
D. Kesimpulan	101
DAFTAR PUSTAKA.....	103
TENTANG PENULIS	106

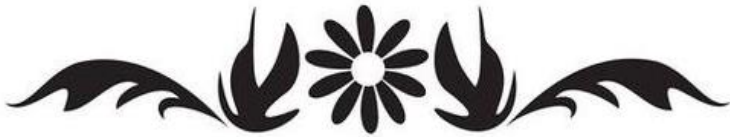
DAFTAR GAMBAR

Gambar 1 Jumlah Pengguna Internet di Indonesia.....	3
Gambar 2 Ilustrasi Data Mining.....	3
Gambar 3 Diagram Hubungan Data Mining	4
Gambar 4 Proses <i>Knowledge Discovery in Database (KDD)</i>	8
Gambar 5 <i>Linear Regresession</i>	15
Gambar 6 Model <i>Deep Learning</i>	17
Gambar 7 <i>Neural Network</i>	18
Gambar 8 <i>Linear Graph</i>	20
Gambar 9 Normal Distribution.....	20
Gambar 10 <i>Exponential</i>	21
Gambar 11 <i>Binary Outcomes</i>	21
Gambar 12 Peramalan Bitcoin dengan Alpha 0.9.....	28
Gambar 13 Plot Data Hasil Peramalan <i>Single Exponential Smoothing</i>	29
Gambar 14 Plot Data Hasil Peramalan Metode <i>Holt</i>	30
Gambar 15 <i>Decision Tree</i>	33
Gambar 16 Node-Node dalam <i>Decision Tree</i>	33
Gambar 17 Alur <i>Naive bayes Classifier</i>	45
Gambar 18 Ilustrasi KNN.....	47
Gambar 19 <i>Logistic Regresion</i>	53
Gambar 20 Perbandingan Jaringan Saraf dengan Jaringan Saraf Tiruan.....	59
Gambar 21 Struktur Neural Network	60
Gambar 22 Struktur Neural Network	61
Gambar 23 Tahapan Algoritma KNN.....	62
Gambar 24 Tahapan Penelitian	63
Gambar 25 Pohon Keputusan	65
Gambar 26 Perbedaan <i>Hierarchical dan Partitional</i>	70
Gambar 27 Dendrogram.....	70
Gambar 28 Ilustrasi Algoritma K-Means.....	73
Gambar 29 Ilustrasi Kelemahan K-Means.....	74
Gambar 30 K-Medoids Clustering	75
Gambar 31 Self-Organizing Map.....	77
Gambar 32 Perbandingan Hard Clustering vs Soft Clustering.....	78
Gambar 33 Hoshen-Kopelman Algorithm.....	80

Gambar 34 Ilustrasi Algoritma Hierarchical Clustering 81
Gambar 35 Ilustrasi Single Linkage 82
Gambar 36 Ilustrasi Centroid Linkage..... 83
Gambar 37 Ilustrasi Complete Linkage 83
Gambar 38 Ilustrasi Average Linkage..... 84
Gambar 39 Diagram Penelitian..... 85
Gambar 40 Langkah Penelitian..... 87
Gambar 41 Langkah Penelitian..... 89
Gambar 42 Alur Proses Pengolahan Data..... 90

DAFTAR TABEL

Tabel 1 Perbedaan Peranan Data Mining	5
Tabel 2 Tipe Data pada Data Mining.....	5



DATA MINING DAN PENERAPAN METODE

Febri Liantoni



BAB

1

PENGENALAN DATA MINING

A. Pendahuluan

Seiring kemajuan zaman, teknologi dalam proses pengumpulan dan penyimpanan data 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. Kebutuhan informasi yang tinggi kadang tidak sebanding dengan penyajian informasi yang memadai. Informasi yang disajikan sering kali masih harus digali dari data dalam jumlah besar. Salah satu contoh yaitu data yang tumbuh dalam bidang kesehatan. Data kesehatan menyimpan banyak sekali data-data yang terkait dalam lingkungan kesehatan seperti data pasien, data obat, data penyakit, yang sangat penting untuk dapat diolah supaya lebih bermanfaat. Metode tradisional yang biasa digunakan untuk menganalisis data, tidak dapat menangani data dalam jumlah besar. Dalam menciptakan efisiensi pengumpulan data yang besar tersebut tentunya diperlukan metode baru yang dapat menjawab kebutuhan tersebut. Oleh karena itu data tersebut dapat diolah menjadi pengetahuan menggunakan teknik yang disebut Data Mining. Sebagai bidang ilmu yang relatif baru, data mining menjadi pusat perhatian para akademis maupun praktisi. Beragam penelitian dan pengembangan data mining banyak diaplikasikan pada berbagai bidang.

Data mining merupakan proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu

BAB 2

FORECASTING (PREDIKSI/ PERAMALAN)

A. Pendahuluan

Forecasting atau peramalan adalah metode untuk memperkirakan informasi yang bersifat prediktif untuk menentukan atau mengambil kebijakan pada masa depan dengan menggunakan data historis atau yang telah ada sebagai acuan. Contoh kasus yang diambil sebagai implementasi dari Forecasting ini adalah penerapan *forecasting straight line method* dalam pengadaan unit mendatang. Hasil Penelitian ini dinilai dapat memudahkan pihak perusahaan terkait pengadaan unit mendatang, untuk meminimalisir segala kemungkinan yang tidak diinginkan. Dimana pada akhir-akhir ini banyak problema mengenai salah dalam pengambilan keputusan sehingga mengulangi sebuah kesalahan yang sama pada tahun lalu. Dengan adanya *forecasting* (peramalan) ini bisa sebagai acuan pertimbangan untuk perbaikan di tahun ini dan kedepannya. Penggunaannya bisa digunakan dalam berbagai kasus seperti analisis pasar dana menejemen, *corporate* dan *risk management*, serta *fraud detection* dan *minning unusual patterns*.

Berikut langkah-langkah cara menghitung exponential smoothing sebagai berikut (Oliver, 1984) :

1. Menghitung koefisien α .
2. Menghitung nilai peramalan periode pertama.
3. Menghitung nilai peramalan di keseluruhan periode.
4. Menghitung akurasi peramalan dengan peta kontrol Tracking Signal.
5. Membuat grafik peta kontrol Tracking Signal.

BAB 3

CLASSIFICATION (KLASIFIKASI)

A. Pendahuluan

Data mining adalah salah satu hal yang penting dalam teknologi informasi. Data yang diolah dengan metode yang efektif dan efisien akan menghasilkan informasi yang akurat. Banyak beragam metode yang digunakan dalam pengolahan data yang digunakan sesuai tipe data yang akan diolah. Salah satu metode data mining adalah klasifikasi. Klasifikasi merupakan teknik atau metode dalam data mining yang bertujuan untuk mengelompokan data berdasarkan keterikatan data terhadap data sample. Metode yang akan dibahas kali ini antara lain adalah *Decision Tree*, *Naive Bayes*, *K-Nearest Neighbor (KNN)*, *Linear Discriminant Analysis*, *Logistic Resgion*, *Support Vector Machine*, *Neural Network*.

B. Classification

Klasifikasi adalah salah satu dalam data mining. Klasifikasi merupakan pengelompokan fitur ke dalam kelas yang sesuai. Vektor fitur telah tersedia dan diketahui kelas-kelasnya, lalu vektor fitur pelatihan tersebut digunakan untuk merancang pemilihan. Pengenalan pola ini disebut bimbingan, supervised.

Seperti yang telah dinyatakan sejumlah klasifikasi teknik telah diusulkan dalam literatur. Terutama proses klasifikasi dibagi menjadi beberapa kategori yang berbeda, yang dinamakan sebagai keputusan berbasis pengklasifikasi.

Klasifikasi adalah tugas pembelajaran sebuah fungsi target f yang memetakan setiap himpunan atribut x ke salah satu label kelas

BAB

4

CLUSTERING (KLAUSTERING)

A. Pendahuluan

Salah satu metode dalam data mining yaitu pengelompokan (*Clustering*) dimana metode tersebut mengidentifikasi objek yang memiliki kesamaan karakteristik tertentu, dan kemudian menggunakan karakteristik tersebut sebagai vektor karakteristik atau centroid. *Clustering* adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode Data Mining, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain.

B. *Clustering*

Clustering (pengelompokan) adalah proses pengelompokan atau penggolongan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/*cluster*. Cara kerja dari clustering sendiri yaitu membagi data ke dalam grup-grup yang mempunyai objek yang karakteristiknya sama.

Clustering dapat dianggap sebagai hal yang paling penting dalam masalah *unsupervised learning*. *Unsupervised learning* sendiri adalah salah satu tipe algoritma *machine learning* yang digunakan untuk menarik kesimpulan dari dataset. Oleh karena itu setiap masalah semacam ini, ia berurusan dengan mencari struktur dalam kumpulan yang tidak diketahui datanya. Hal ini dapat didefinisikan bahwa *clustering* merupakan "proses pengorganisasian

BAB

5

ASSOCIATION (ASOSIASI)

A. Pendahuluan

Asosiasi data mining merupakan sebuah algoritma yang memberikan informasi tentang hubungan antar data didalam sebuah database. Suatu item dapat diprediksi dengan menggunakan kemunculan item lain. Asosiasi memberikan faktor yang mempengaruhi kompleksitas dengan aturan, nilai minimum, ukuran, dimensi hingga rata-rata transaksi sebuah item (Hussain dkk., 2019). Asosiasi menemukan pola berdasarkan hubungan antar variabel dalam transaksi yang sama. Teknik ini juga disebut sebagai teknik relasi karna menggunakan hubungan antar item dan menemukan frekuensi kemunculan item yang berbeda dengan frekuensi tertinggi dalam sebuah database.

B. Association

Asosiasi menggunakan pernyataan jika-maka secara berurutan untuk menunjukkan kemungkinan hubungan antar data item atau variabel dalam kumpulan data besar diberbagai jenis database. Asosiasi banyak digunakan untuk membantu menemukan korelasi penjualan dalam data transaksional atau data medis. Berdasarkan histori data penjualan, teknik asosiasi sering digunakan oleh pengecer untuk mendapatkan hubungan jika pelanggan membeli keripik, maka pembeli juga akan membeli air mineral. Untuk itu teknik ini juga disebut sebagai teknik analisis keranjang pasar.

DAFTAR PUSTAKA

- Al Ihsan, N. H. A. S., Dzakiyah, H. H., & Liantoni, F. (2020). Perbandingan Metode Single Exponential Smoothing dan Metode Holt untuk Prediksi Kasus COVID-19 di Indonesia. *Ultimatics: Jurnal Teknik Informatika*, 12(2), 89–94. <https://doi.org/10.31937/ti.v12i2.1689>
- Anggraeni, R. M. (2014). Perbandingan Algoritma Apriori dan Algoritma FP-Growth untuk Rekomendasi Pada Transaksi Peminjaman Buku di Perpustakaan Universitas Dian Nuswantoro. *Teknik Informatika*, 1–6.
- Asmara, A. A. G. A. P. R. (2016). Implementasi Metode Simple Additive Weighting (SAW) Dalam Memprediksi Calon Mahasiswa Dropout STMIK STIKOM Bali. *Sistem Informatika*, 11(1), 11–18.
- Cindy Mutia Annur. (2022). *Ada 204,7 Juta Pengguna Internet di Indonesia Awal 2022*. databoks.
- Elwood S. Buffa, & Sarin, R. K. (1996). *Manajemen Operasi dan Produksi Modern*. Binarupa Aksara.
- Fausett. (1993). *Fundamental of Neural Network: Architectures, Algorithm, And Application*. Prentice Hall.
- Fauzi, A., & Priati. (2017). *Data Mining dengan Teknik Clustering Menggunakan Algoritma K - Means pad Data Transaksi Superstore*.
- Han, J., & Kamber, M. (2006). *Data Mining: Concept and Techniques*. Morgan Kaufmann Publisher.
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. England: Pearson Education.
- Heizer, J., & Render, B. (2009). *Operation Management*. Pearson

Education.

- Hussain, S., Atallah, R., Kamsin, A., & Hazarika, J. (2019). *Classification, Clustering and Association Rule Mining in Educational Datasets Using Data Mining Tools: A Case Study* (hal. 196–211). https://doi.org/10.1007/978-3-319-91192-2_21
- Kohonen, T. (1989). *Self-organizing feature maps*. Springer Berlin Heidelberg.
- Larose. (2005). *Discovering knowledge in data*. Canada: A John Willey & Sons, Inc.
- Liantoni, F. (2011). *Klasifikasi Daun Dengan Centroid Linked Clustering Berdasarkan Fitur Bentuk Tepi Daun*. ITS Library. <http://digilib.its.ac.id/ITS-Undergraduate-3100010041146/15590>
- Liantoni, F. (2015). Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor. *ULTIMATICS, Jurnal Teknik Informatika*, 7(2), 98–104.
- Liantoni, F., & Agusti, A. (2020). Forecasting Bitcoin using Double Exponential Smoothing Method Based on Mean Absolute Percentage Error. *JOIV: International Journal on Informatics Visualization*, 4(2), 91–95. <https://doi.org/10.30630/joiv.4.2.335>
- Liantoni, F., Indra Perwira, R., Muharom, S., Agung Firmansyah, R., & Fahrudi, A. (2019). Leaf classification with improved image feature based on the seven moment invariant. *Journal of Physics: Conference Series*, 1175, 012034. <https://doi.org/10.1088/1742-6596/1175/1/012034>
- Liantoni, F., & Santoso, A. (2020). Perbaikan Kontras Citra Mammogram Pada Klasifikasi Kanker Payudara Berdasarkan Fitur Gray-Level Co-Occurrence Matrix. *SINTECH JOURNAL*, 3(1).

- Makridakis, S., Wheelwright, S. C., & McGree, V. E. (1992). *Metode dan aplikasi peramalan*. Erlangga.
- Oliver, R. M. (1984). Exponential smoothing – I. *Operations Research Letters*, 3(3), 111-117. [https://doi.org/10.1016/0167-6377\(84\)90001-4](https://doi.org/10.1016/0167-6377(84)90001-4)
- Perangin-angin, M. I., Lubis, A. H., & Ikhwan, A. (2017). Association Rules Analysis on FP-Growth Method in Predicting Sales. *International Journal of Recent Trends in Engineering and Research*, 3(10), 58-65. <https://doi.org/10.23883/ijrter.2017.3453.dhcoa>
- Supriyanto, A., Maryono, D., & Liantoni, F. (2021). Predicted Student Study Period with C4.5 Data Mining Algorithm. *IJIE (Indonesian Journal of Informatics Education)*, 4(2), 94-100. <https://doi.org/10.20961/IJIE.V4I2.46265>
- Wahyuni, W. A., & Saepudin, S. (2021). Penerapan Data Mining Clustering Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci. *Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra*.
- Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1-37. <https://doi.org/10.1007/s10115-007-0114-2>
- Yan, X. (2013). Weighted K-Nearest Neighbor Classification Algorithm Based on Genetic Algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 11(10). <https://doi.org/10.11591/telkomnika.v11i10.2534>

TENTANG PENULIS

FEBRI LIANTONI

Penulis kelahiran Magetan ini adalah dosen program studi Pendidikan Teknik Informatika dan Komputer di Universitas Sebelas Maret, sejak tahun 2019. Penulis menyelesaikan pendidikan formal Sarjana Teknik Informatika di Politeknik Elektronika Negeri Surabaya dan Magister Teknik Informatika di Institut Teknologi Sepuluh Nopember. Penulis pernah menjadi dosen program studi Teknik Informatika di Institut Teknologi Adhi Tama Surabaya pada tahun 2015-2019. Penulis memiliki ketertarikan pada fotografi, software engineering, data mining, artificial intelligence dan image processing.