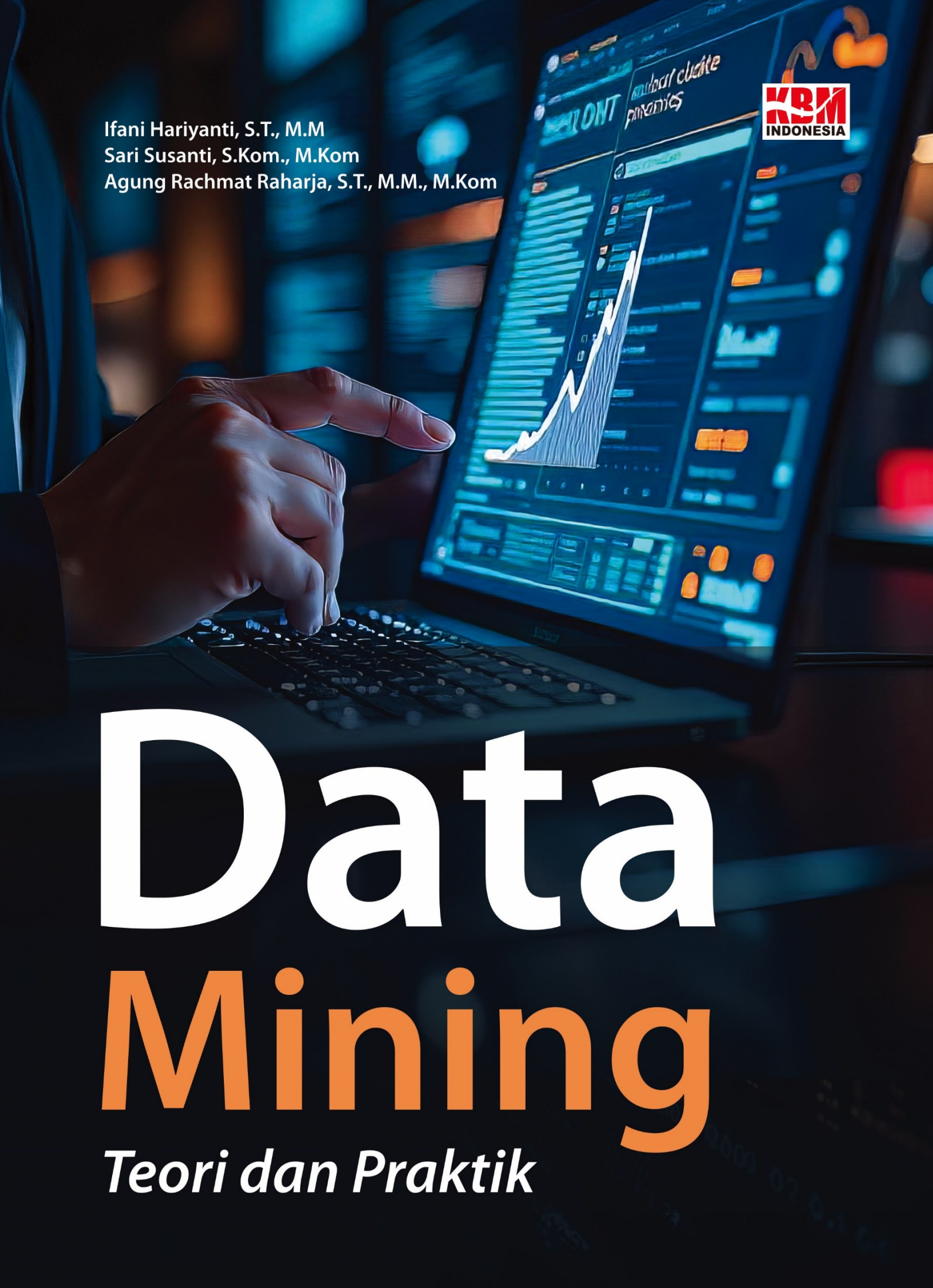

Ifani Hariyanti, S.T., M.M
Sari Susanti, S.Kom., M.Kom
Agung Rachmat Raharja, S.T., M.M., M.Kom

KBM INDONESIA

Data Mining

Teori dan Praktik

DATA MINING

TEORI DAN PRAKTIK

Ifani Hariyanti S.T., M.M.
Sari Susanti, S.Kom., M.Kom.
Agung Rachmat Raharja, S.T., M.M., M.Kom.

KBM INDONESIA

PENERBIT KBM INDONESIA

Adalah penerbit dengan misi memudahkan proses penerbitan buku buku penulis di tanah air Indonesia. Serta menjadi media sharing proses penerbitan buku.

DATA MINING
Teori dan Praktik

Penulis
Ifani Hariyanti S.T., M.M.
Sari Susanti, S.Kom., M.Kom.
Agung Rachmat Raharja, S.T., M.M., M.Kom.
Desain Sampul
Aswan Kreatif
Tata Letak
Sofita HM
Editor
Dr. Muhamad Husein Maruapey, Drs., M.Sc.
Background isi buku di ambil dari https://www.freepik.com/

Official
Depok, Sleman-Jogjakarta (Kantor)
Penerbit Karya Bakti Makmur (KBM) Indonesia
Anggota IKAPI/No. IKAPI 279/JTI/2021
081357517526 (Tlpn/WA)

Website
https://penerbitkbm.com
www.penerbitbukumurah.com
Email
naskah@penerbitkbm.com
Distributor
https://penerbitkbm.com/toko-buku/
Youtube
Penerbit KBM Sastrabook
Instagram
@penerbit.kbmindonesia
@penerbitbukujogja

ISBN: 978-634-202-440-9
Cetakan ke-1, Juni 2025
21 x 29 cm, iv + 161 halaman

Sanksi Pelanggaran Pasal 113
Undang-Undang No. 28 Tahun 2014 Tentang Hak Cipta

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp 100. 000. 000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp 500. 000. 000,00 (lima ratus juta rupiah).
3. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf e, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp 1. 000. 000. 000,00 (satu miliar rupiah).
4. Setiap Orang yang memenuhi unsur sebagaimana dimaksud pada ayat (3) yang dilakukan dalam bentuk pembajakan, dipidana dengan pidana penjara paling lama 10 (sepuluh) tahun dan/atau pidana denda paling banyak Rp 4. 000. 000. 000,00 (empat miliar rupiah).

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, karena atas limpahan rahmat dan karunia-Nya, penulisan buku yang berjudul "**Data Mining: Teori dan Praktik**" ini dapat diselesaikan dengan baik. Buku ini disusun sebagai upaya untuk memberikan pemahaman yang komprehensif kepada mahasiswa, dosen, peneliti, dan praktisi mengenai konsep dasar, metode, serta penerapan data mining dalam berbagai bidang.

Seiring dengan pertumbuhan eksponensial data di era digital, kemampuan untuk mengekstraksi informasi dan pengetahuan tersembunyi dari data menjadi semakin penting. Oleh karena itu, data mining hadir sebagai solusi penting dalam mendukung pengambilan keputusan berbasis data (data-driven decision making). Buku ini dirancang untuk menggabungkan teori fundamental dengan pendekatan praktis melalui contoh kasus, algoritma, dan ilustrasi proses mining secara bertahap.

Penulis berharap buku ini dapat menjadi referensi yang berguna dalam kegiatan pembelajaran di perguruan tinggi maupun dalam pengembangan solusi data mining di dunia industri. Materi yang disajikan mencakup tahapan proses data mining, teknik-teknik populer seperti klasifikasi, klastering, dan asosiasi, serta penerapan dengan perangkat lunak pendukung yang umum digunakan.

Ucapan terima kasih penulis sampaikan kepada semua pihak yang telah memberikan dukungan dan motivasi selama proses penulisan buku ini. Kritik dan saran yang membangun sangat penulis harapkan demi penyempurnaan edisi berikutnya.

Akhir kata, semoga buku ini dapat memberikan manfaat yang maksimal dan menjadi kontribusi positif dalam pengembangan ilmu pengetahuan dan teknologi informasi di Indonesia.

Bandung, Juli 2025

Penulis

DAFTAR ISI

REFERENSI

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95.

Waskom, M. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021.

McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 445, 51–56.

van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Tan, P. N., Steinbach, M., & Kumar, V. (2018). *Introduction to Data Mining* (2nd ed.). Pearson.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. Springer.

Liu, B. (2007). *Web data mining: Exploring hyperlinks, contents, and usage data*. Springer.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification* (2nd ed.). Wiley.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, 487–499.

Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, 12(3), 372–390.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI*, 14(2), 1137–1145.
Thomas, M., & Freund, Y. (1999). *Boosting: Foundations and algorithms*. MIT Press.
Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113.
Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31(3), 249–268.
Berrar, D. (2018). *Cross-validation*. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 542–545). Elsevier.
Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
Provost, F., & Fawcett, T. (2013). *Data science for business*. O'Reilly Media.
Biecek, P., & Burzykowski, T. (2021). *Explanatory model analysis*. Chapman and Hall/CRC.
Aggarwal, C. C. (2015). *Data mining: The textbook*. Springer.
Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370–418.
Zhang, H. (2004). The optimality of naive Bayes. *AA*, 1(2), 3.
Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media.
Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.
Larson, R., & Farber, B. (2015). *Elementary statistics: Picturing the world* (6th ed.). Pearson.
NIST/SEMATECH. (2013). *e-Handbook of Statistical Methods*. National Institute of Standards and Technology.
Kelleher, J. D., Mac Carthy, M., & Korvir, B. (2015). *Fundamentals of machine learning for predictive data analytics: Algorithms, worked examples, and case studies*. MIT Press.
IBM. (2020). *SPSS Statistics Documentation*. Retrieved from https://www.ibm.com/docs/en/spss-statistics
Kaggle Inc. (2023). *Kaggle Datasets*. Retrieved from https://www.kaggle.com/datasets
UC Irvine Machine Learning Repository. (2023). Retrieved from https://archive.ics.uci.edu/ml/index.php
Chollet, F. (2018). *Deep learning with Python*. Manning Publications.
Rouse, M. (2021). *Data preprocessing*. TechTarget. Retrieved from https://www.techtarget.com
Microsoft. (2023). *Azure Machine Learning Documentation*. Retrieved from https://docs.microsoft.com/en-us/azure/machine-learning/
IBM Cloud Education. (2022). *What is data mining?* Retrieved from https://www.ibm.com/cloud/learn/data-mining
DataRobot. (2021). *What is data mining?* Retrieved from https://www.datarobot.com/wiki/data-mining/
Tableau. (2022). *Data visualization overview*. Retrieved from https://www.tableau.com/learn/articles/data-visualization
Oracle. (2023). *Introduction to data mining*. Retrieved from https://docs.oracle.com
SAS Institute. (2022). *SAS Visual Data Mining and Machine Learning*. Retrieved from https://www.sas.com

BIODATA PENULIS

Ifani Hariyanti S.T., M.M., lahir di Bandung pada 18 November 1987. Sejak 2020, saya menikmati peran sebagai pengajar di ARS University, dan juga membantu mengelola keuangan universitas. Saya senang dengan hal baru dan senang berpetualang, memungkinkan saya untuk terus tumbuh dalam pekerjaan dan hobi saya.

Sari Susanti, S.Kom., M.Kom., lahir di Bandung 23 Maret 1995. Merupakan seorang Dosen tetap dan Peneliti di Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Adhirajasa Reswara Sanjaya atau yang lebih dikenal (ARS University). Beliau aktif melakukan penelitian dengan kajian di bidang data mining, analisis dan perancangan sistem informasi. Selain itu beberapa kali pernah mendapatkan hibah penelitian yang didanai oleh Kemdikbudristek atau yang sekarang dikenal dengan Kemdiktisaintek.

Agung Rachmat Raharja, S.T., M.M., M.Kom., Lahir di Bandung 23 April 1987. Menempuh pendidikan SDN 1 Sejahtera II, SMP 32 Bandung, SMA PGRI, setelah lulus S1 melanjutkan S2 Magister Manajemen di ARS University dan menempuh pendidikan S2 Magister Komputer di Universitas Langlang Buana. Aktif dalam mengajar dan sekarang menjadi dosen di Universitas Swasta di Bandung