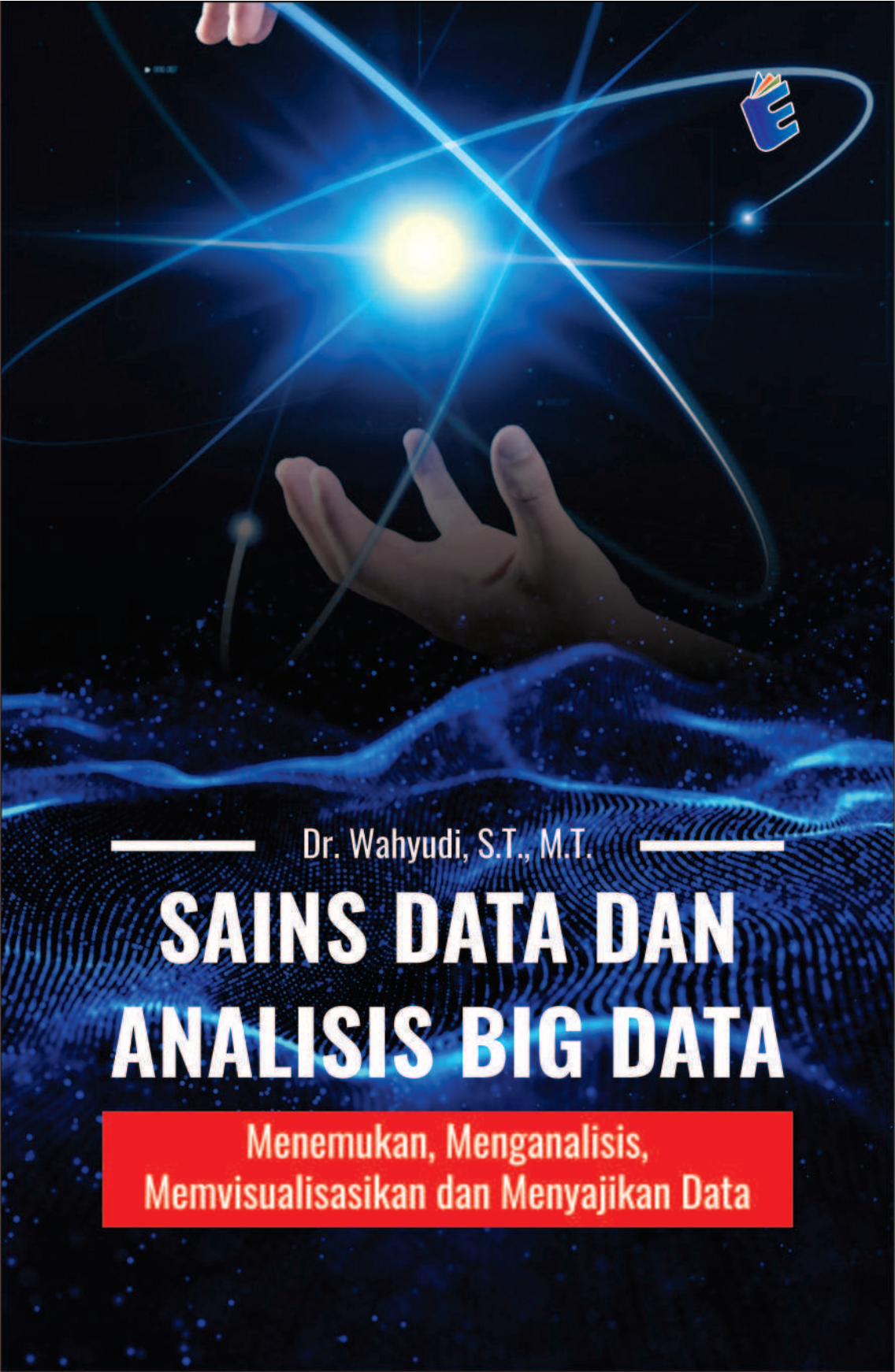

Dr. Wahyudi, S.T., M.T.

SAINS DATA DAN ANALISIS BIG DATA

Menemukan, Menganalisis,
Memvisualisasikan dan Menyajikan Data

eureka
media aksara
Anggota IKAPI
No. 225/JTE/2021

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

ISBN 978-623-151-730-2

SAINS DATA DAN ANALISIS BIG DATA:
Menemukan, Menganalisis, Memvisualisasikan dan Menyajikan Data

Dr. Wahyudi, S.T., M.T.

PENERBIT CV. EUREKA MEDIA AKSARA

SAINS DATA DAN ANALISIS BIG DATA:
Menemukan, Menganalisis, Memvisualisasikan dan Menyajikan Data

Penulis : Dr. Wahyudi, S.T., M.T.

Desain Sampul : Ardyan Arya Hayuwaskita

Tata Letak : Uli Mas'uliyah Indarwati

ISBN : 978-623-151-730-2

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

Redaksi:
Jalan Banjaran, Desa Banjaran RT 20 RW 10 Kecamatan Bojongsari
Kabupaten Purbalingga Telp. 0858-5343-1992
Surel : eurekamediaaksara@gmail.com
Cetakan Pertama : 2023

All right reserved

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

KATA PENGANTAR

Alhamdulillahirobbil'alamin. Penulis bersyukur kehadirat Allah SWT berkat rahmat, karunia dan pertolonganNya, penulis dapat menyelesaikan buku berjudul "**Sains Data Dan Analisis Big Data: Menemukan, Menganalisis, Memvisualisasikan dan Menyajikan Data**". Shalawat serta salam semoga senantiasa tercurah atas Nabi Muhammad SAW, para kerabat, serta pengikutnya hingga hari kiamat nanti.

Buku ini hadir untuk menambah literasi tentang teknologi informasi. Buku ini merupakan seri ketiga dari beberapa buku sains data. Buku ini menjelaskan perspektif lain bagaimana cara Menemukan, Menganalisis, Memvisualisasikan dan Menyajikan Data, dan Siklus Analisis Data.

Penulis menyadari bahwa dalam penulisan buku ini masih banyak terdapat kekurangan, untuk itu penulis mengharapkan kritik dan sarannya guna penyempurnaan buku ini di masa mendatang

Padang, September 2023

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

SAINS DATA DAN ANALISIS BIG DATA:
Menemukan, Menganalisis,
Memvisualisasikan
dan Menyajikan Data

BAB 1 PENGANTAR ANALISIS BIG DATA

Banyak yang telah ditulis tentang Big Data dan kebutuhan analitik tingkat lanjut dalam industri, akademisi, dan pemerintah. Ketersediaan sumber data baru dan munculnya peluang analisis yang lebih kompleks telah menciptakan kebutuhan untuk memikirkan kembali arsitektur data yang ada untuk memungkinkan analisis yang memanfaatkan Big Data. Selain itu, terdapat perdebatan yang signifikan dengan sebaik-baiknya. Bab ini menjelaskan beberapa konsep utama untuk memperjelas apa yang dimaksud dengan Big Data, mengapa analitik tingkat lanjut diperlukan, bagaimana Data Science berbeda dari Business Intelligence (BI), dan peran baru apa yang baru apa yang dibutuhkan untuk ekosistem Big Data yang baru.

A. Gambaran Umum Big Data

Data diciptakan secara konstan, dan dengan kecepatan yang terus meningkat. Ponsel, media sosial, teknologi pencitraan untuk menentukan diagnosis medis. Banyak ciptakan data baru, dan harus disimpan di suatu tempat untuk beberapa tujuan. Perangkat dan sensor

BAB 2 SIKLUS ANALISIS DATA

Proyek sains data berbeda dari sebagian besar proyek Intelijen Bisnis tradisional dan banyak proyek analisis data karena proyek sains data lebih bersifat eksploratif. Untuk alasan ini, sangat penting untuk mengaturnya dan memastikan para peserta teliti dan ketat dalam pendekatan, agar tidak menghambat eksplorasi.

Banyak masalah yang tampak besar dan menakutkan pada awalnya dapat dipecah menjadi bagian-bagian yang lebih kecil atau fase-fase yang dapat ditindak lanjuti dengan mudah. Proses yang baik memastikan metode yang komprehensif dan komprehensif dan diulang untuk melakukan analisis. Selain itu, membantu memfokuskan waktu dan energi di awal proses untuk mendapatkan pemahaman tentang masalah bisnis yang harus dipecahkan.

Kesalahan umum yang sering terjadi dalam proyek data science adalah terburu-buru dalam pengumpulan dan analisis data, yang perlu meluangkan waktu yang cukup untuk merencanakan dan menentukan jumlah pekerjaan yang dilakukan, memahami kebutuhan, atau bahkan membingkai masalah bisnis dengan benar. Akibatnya, para

DAFTAR PUSTAKA

C. B. B. D. Manyika, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute, 2011.

D. R. John Gantz, "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and BiggestGrowth in the Far East," IDC, 2013.

http://www.willisresilience.com/emc-datalab [Online].

C. Duhigg, The Power of Habit: Why We Do What We Do in Life and Business, New York: Random House, 2012.

K. Hill, "How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did," Forbes, February 2012. http://hadoop.apache.org [Online].

T. H. Davenport and D. J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," Harvard Business Review, October 2012.

J. Manyika, M. Chiu, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute, 2011.

"Scientific Method" [Online]. Available: http://en.wikipedia.org/wiki/Scientific_method.

"CRISP-DM" [Online]. Available: http://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining.

T. H. Davenport, J. G. Harris, and R. Morison, Analytics at Work: Smarter Decisions, Better Results, 2010, Harvard Business Review Press.

D. W. Hubbard, How to Measure Anything: Finding the Value of Intangibles in Business, 2010, Hoboken, NJ: John Wiley & Sons.

J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein and C. Welton, MAD Skills: New Analysis Practices for Big Data, Watertown, MA 2009.

"List of APIs" [Online]. Available: http://www.programmableweb.com/apis.

B. Shneiderman [Online]. Available: http://www.ifp.illinois.edu/nabhcs/abstracts/shneiderman.html.

"Hadoop" [Online]. Available: http://hadoop.apache.org.

"Alpine Miner" [Online]. Available: http://alpinenow.com.

"OpenRefine" [Online]. Available: http://openrefine.org.

"Data Wrangler" [Online]. Available: http://vis.stanford.edu/wrangler/.

"CRAN" [Online]. Available: http://cran.us.r-project.org.

"SQL" [Online]. Available: http://en.wikipedia.org/wiki/SQL.

"SAS/ACCESS" [Online]. Available: http://www.sas.com/en_us/software/ data-management/access.htm.

"SAS Enterprise Miner" [Online]. Available: http://www.sas.com/en_us/software/analytics/enterprise-miner.html.

"SPSS Modeler" [Online]. Available: http://www-03.ibm.com/software/products/en/category/business-analytics.

"Matlab" [Online]. Available: http://www.mathworks.com/products/matlab/.

"Statistica" [Online]. Available: https://www.statsoft.com.

"Mathematica" [Online]. Available: http://www.wolfram.com/mathematica/.

"Octave" [Online]. Available: https://www.gnu.org/software/octave/.

"WEKA" [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/.

"MADlib" [Online]. Available: http://madlib.net.

K. L. Higbee, Your Memory—How It Works and How to Improve It, New York: Marlowe & Company, 1996.

S. Todd, "Data Science and Big Data Curriculum" [Online]. Available:http://stevetodd.typepad.com/my_weblog/data-science-and-big-data-curriculum/.

T. H Davenport and D. J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," Harvard Business Review, October 2012.