



BIG DATA

Nindy Irzavika, S.SI., M.T., Laina Farsiah, S.Si., M.S.,
Nova Tri Romadloni, M.Kom., Rahmatika Pratama Santi, M.T.,
Nabilah, S.Kom., MT., Ani Anisyah, S.Pd., M.T.,
Khairun Amala, S.Si., M.M., Nisa Dwi Angresti, S.SI., M.Kom.,
Dwina Satrinia, S.Kom.,M.T., Dodi Zulherman, M.T.,
Yuwan Jumaryadi, S.Kom., MM., M.Kom.,
Deosa Putra Caniago, S.Kom., M.Kom,
Luki Hernando, M.Kom., Liptia Venica, S.T., M.T.

BIG DATA

Nindy Irzavika, S.SI., M.T., Laina Farsiah, S.Si., M.S.,
Nova Tri Romadloni, M.Kom., Rahmatika Pratama Santi, M.T.,
Nabilah, S.Kom., MT., Ani Anisyah, S.Pd., M.T.,
Khairun Amala, S.Si., M.M., Nisa Dwi Angresti, S.SI., M.Kom.,
Dwina Satrinia, S.Kom.,M.T., Dodi Zulherman, M.T.,
Yuwan Jumaryadi, S.Kom., MM., M.Kom.,
Deosa Putra Caniago, S.Kom., M.Kom.,
Luki Hernando, M.Kom., Liptia Venica, S.T., M.T.



BIG DATA

Copyright© PT Penamudamedia, 2024

Penulis:

Nindy Irzavika, S.Si., M.T., Laina Farsiah, S.Si., M.S., Nova Tri Romadloni, M.Kom., Rahmatika Pratama Santi, M.T., Nabilah, S.Kom., MT., Ani Anisyah, S.Pd., M.T., Khairun Amala, S.Si., M.M., Nisa Dwi Angresti, S.Si., M.Kom., Dwina Satrinia, S.Kom.,M.T., Dodi Zulherman, M.T., Yuwan Jumaryadi, S.Kom., MM., M.Kom., Deosa Putra Caniago, S.Kom., M.Kom., Luki Hernando, M.Kom., Liptia Venica, S.T., M.T.

ISBN:

978-623-8586-16-5

Desain Sampul:

Tim PT Penamuda Media

Tata Letak:

Enbookdesign

Diterbitkan Oleh

PT Penamuda Media

Casa Sidoarium RT 03 Ngentak, Sidoarium Dodeam Sleman Yogyakarta

HP/Whatsapp : +6285700592256

Email : penamudamedia@gmail.com

Web : www.penamuda.com

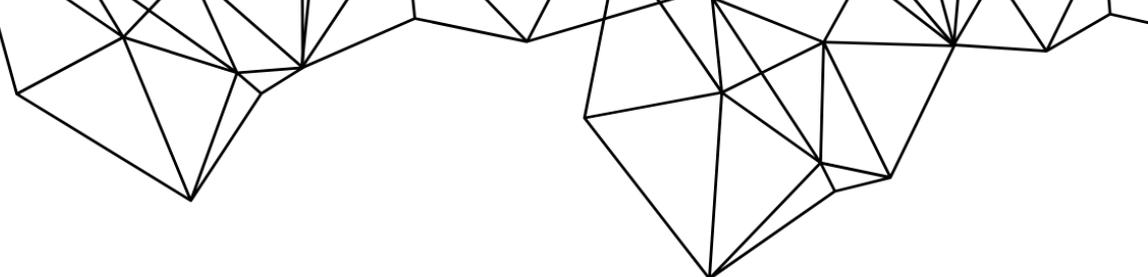
Instagram : @penamudamedia

Cetakan Pertama, April 2024

xii + 181, 15x23 cm

Hak cipta dilindungi oleh undang-undang

Dilarang memperbanyak sebagian atau seluruh isi buku tanpa izin Penerbit



Kata Pengantar

Big Data merupakan sumber informasi yang melimpah dan berpotensi besar, memungkinkan kita untuk mendapatkan wawasan yang bernilai dan membuat keputusan yang lebih baik. Melalui buku ini, kami mengajak pembaca untuk menjelajahi konsep-konsep utama Big Data, seperti analisis data, teknik pengolahan, keamanan data, dan penerapan praktis dalam berbagai bidang.

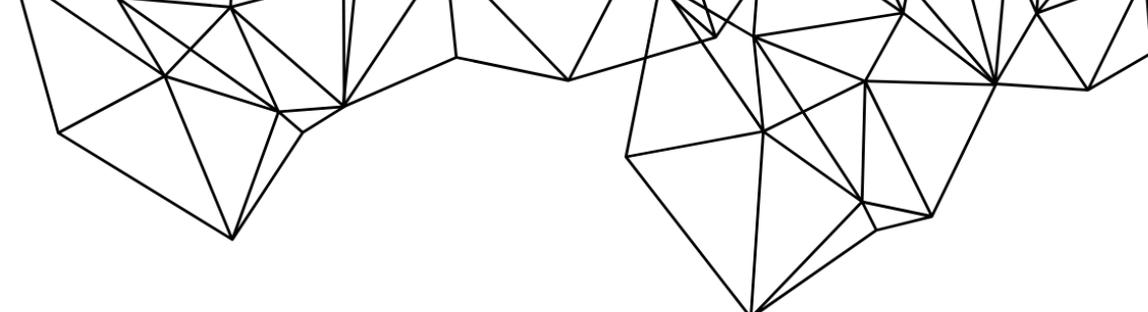
Dengan pemahaman yang mendalam tentang Big Data, pembaca diharapkan dapat mengoptimalkan potensi data yang ada, menemukan pola-pola yang relevan, dan mengambil langkah-langkah strategis berdasarkan analisis yang akurat.

Kami berharap buku ini dapat menjadi panduan yang bermanfaat bagi para profesional, peneliti, dan siapa pun yang tertarik dalam memahami dan memanfaatkan Big Data. Terima kasih kepada semua pihak yang telah mendukung dan berkontribusi dalam pembuatan buku ini.

Semoga buku ini memberikan wawasan yang berharga dan memperkaya pengetahuan pembaca dalam menghadapi tantangan dan peluang yang terkait dengan Big Data di era digital ini.

Hormat kami

Penulis



Daftar Isi

Kata Pengantar v

Daftar Isi..... vi

Bab 1 – Pengantar Big Data.....1

A. Definisi Big Data..... 1

B. Jenis Data pada Big Data..... 3

C. Tantangan dan Peluang Big Data..... 5

D. Implementasi Big Data.....7

E. Dampak Big Data8

Bab 2 – Karakteristik Big Data.....9

A. Volume..... 10

B. Velocity11

C. Variety.....11

D. Veracity.....12

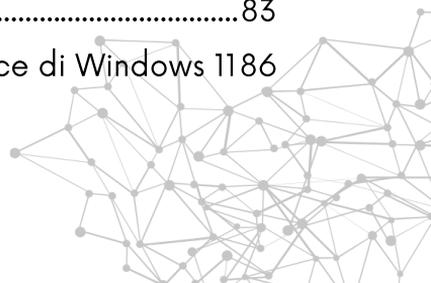
E. Value13

F. Variability.....14

G. Complexity.....14

H. Accessibility.....	15
I. Authenticity.....	16
J. Agility.....	17
K. Analyzability.....	17
L. Availability.....	18
M. Actionability.....	19
N. Alignment.....	20
O. Privacy and Ethics.....	20
Bab 3 – Data Analitik	22
A. Definisi Data Analitik.....	22
B. Teknologi Pendukung Data Analitik.....	25
C. Metode dan Algoritma Data Analitik.....	27
D. Tantangan dan Solusi.....	29
E. Studi Kasus Penjualan E-Commerce	30
Bab 4 – Siklus Hidup <i>Big Data Analytic</i>.....	33
A. Business Case Evaluation.....	35
B. Data Identification.....	35
C. Data Acquisition and Filtering	36
D. Data Extraction	37
E. Data Validation and Cleaning.....	38
F. Data Anggregation and Representation.....	39
G. Data Analysis.....	40

H. Data Visualization.....	41
I. Utilization of Analysis Results.....	42
Bab 5 – Visualisasi Big Data.....	43
A. Pengantar Visualisasi Big Data.....	43
B. Prinsip Dasar Visualisasi Big Data.....	44
C. Jenis dan Kategori Visualisasi Big Data.....	45
D. Alat dan Teknologi Visualisasi.....	51
E. Langkah – langkah dalam Membuat Visualisasi.....	51
Bab 6 – Management Big Data – NOSQL.....	53
A. NoSQL.....	55
B. Teori CAP.....	57
C. Model Basis Data No SQL.....	60
Bab 7 – Exploratory Data Analysis (EDA).....	67
A. Pengertian dan Tujuan EDA.....	67
B. Teknik dan <i>Tools</i> EDA.....	69
C. Langkah-Langkah dalam EDA.....	71
Bab 8 – Hadoop dan MapReduce.....	78
A. Sejarah perkembangan Hadoop.....	82
B. Pengenalan MapReduce.....	83
C. Tutorial Installasi Hadoop MapReduce di Windows 1186	



Bab 9 – Ekosistem Hadoop.....	89
A. Hadoop Distributed File System (HDFS).....	90
B. HBase	91
C. MapReduce.....	92
D. Yarn.....	93
E. Hive	94
F. Pig.....	95
G. Mahout.....	96
H. Sqoop	96
I. Spark.....	97
J. Oozie.....	98
K. ZooKeeper	99
L. Apache Ranger.....	99
Bab 10 – Apache Spark.....	101
A. Pengenalan Apache Spark	102
B. Arsitektur Apache Spark.....	107
C. Konfigurasi Apache Spark.....	110
Bab 11 – Machine Learning & Big Data	116
A. Langkah Tepat Untuk Transisi Efektif ke Machine learning.....	117
B. Apa itu Big Data?	118

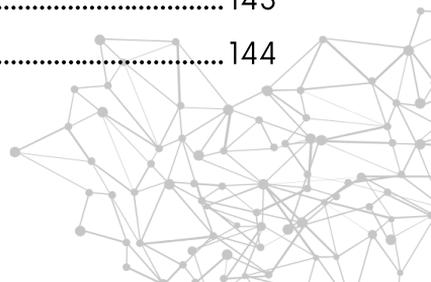
C. Apa itu Machine Learning?	120
D. Hubungan antara Big Data dan Machine Learning...	120
E. Peluang dan Tantangan Data Preprocessing.....	123
F. Redundansi Data.....	123
G. Data Noise.....	123
H. Heterogenitas Data	124
I. Pelabelan Data.....	124
J. Imbalanced Data.....	124
K. Feature Selection	125
L. Metode Machine Learning & Big Data.....	125
M. Kerangka Machine Learning pada Big Data.....	126
N. Contoh Aplikasi Machine Learning Pada Big Data....	127

Bab 12 - Etika Big Data.....130

A. Masalah Privasi dalam Big Data:	130
B. Mitigasi Keadilan dan Bias.....	132
C. Transparansi dan Akuntabilitas.....	133
D. Tantangan Etis dalam Tata Kelola Data	135
E. Tanggung Jawab Sosial dan Big Data.....	137

Bab 13 - Penerapan Big Data pada berbagai bidang 139

A. Bisnis dan pemasaran	141
B. Pemerintahan.....	143
C. Keuangan dan Perbankan.....	144



D. Perawatan dan Kesehatan	145
E. Transportasi.....	147
Bab 14 – Penerapan Hadoop dan Machine Learning pada Big Data.....	149
A. Pencarian Pengetahuan dari Pola <i>Cluster</i> Menggunakan Algoritma K-means <i>Clustering</i> Secara Pararel dengan Hadoop MapReduce	150
B. K-means Paralel dengan Hadoop MapReduce untuk Optimalisasi Pemilihan Inisial <i>Centroid</i>	159
Daftar Pustaka	161
Tentang Penulis	175



Pengantar Big Data

Nindy Irzavika, S.Si., M.T.

A. Definisi Big Data

Data dihasilkan setiap jam, setiap menit, bahkan setiap detik, sehingga jumlah data terus meningkat. Perangkat seluler, media sosial, rekam medis menciptakan data baru dan tersimpan untuk beberapa tujuan. Ledakan jumlah data ini menjadi tantangan dan peluang bagi pelaku bisnis, pemerintah, dan peneliti untuk mengubah data tersebut menjadi bernilai. Data menjadi asset yang sangat bernilai yang bisa dimiliki karena data menghasilkan informasi, pengetahuan, dan membantu stakeholder dalam pengambilan keputusan (Irzavika and Supangkat, 2018).

Big data merupakan terminologi dari cara menganalisis, mengekstraksi informasi secara sistematis, atau menangani kumpulan data yang terlalu besar dan rumit diselesaikan oleh alat pemrosesan data tradisional (Mariani, Tweneboah and Beccar-Varela, 2022). Big data adalah aset informasi yang dicirikan dengan jumlah, kecepatan, dan keanekaragaman yang sangat tinggi sehingga membutuh-

BIG DATA

Dalam buku ini, pembaca akan dihadapkan pada berbagai aspek penting seperti pengumpulan data, penyimpanan data, analisis prediktif, dan visualisasi data. Penekanan diberikan pada peran teknologi seperti machine learning, data mining, dan cloud computing dalam mengolah big data untuk mendapatkan informasi yang bermanfaat.

Penulis juga membahas tantangan yang terkait dengan big data, termasuk privasi data, keamanan informasi, dan etika penggunaan data. Dengan contoh kasus dan studi empiris, pembaca diberikan pemahaman yang mendalam tentang potensi dan risiko yang terkait dengan era big data.

Secara keseluruhan, "Buku Big Data" menjadi panduan yang sangat berharga bagi para profesional TI, analis data, dan peneliti yang tertarik untuk memahami lebih dalam tentang konsep big data dan bagaimana memanfaatkannya secara efektif. Buku ini tidak hanya memberikan wawasan mendalam tentang big data, tetapi juga memberikan strategi praktis untuk mengelola, menganalisis, dan mengambil keputusan berdasarkan informasi yang tersimpan dalam jumlah data yang besar dan kompleks.

ISBN 978-623-8586-16-5



9 786238 586165



PT Penerbit Penamuda Media
Codean, Yogyakarta
085700592256
@penamuda_media
penamuda.com