

Dr. (H.C) Zaid Romegar Mair, S.T., M.Cs., Fitrah Maharani Humaira, S.Si, M.Kom.,  
Vina Lusiana, S.Pd, M.Si., Farida Yunita, S.T., M.T., Rudolf Sinaga, S.Kom, M.S.I.,  
N. Nelis Febriani SM, S.Kom., M.Kom., Muhammad Faqih Dzulqarnain, S.T., M.CS.,  
Castaka Agus Sugianto, M.Kom., MCS, Nuk Ghurroh Setyoningrum, S.Kom, M.Cs.,  
Santi Prayudani, S.Kom., M.Kom., Farid, M.Kom., Samsinar, S.Kom., MMSI.,  
Nur Oktavin Idris, M.Eng., Nova Agustina, S.T., M.Kom., Erna Piantari, S.Kom., M.T.

# DATA SCIENCE



# DATA SCIENCE

Dr. (H.C) Zaid Romegar Mair, S.T., M.Cs

Fitrah Maharani Humaira, S.Si, M.Kom

Vina Lusiana, S.Pd, M.Si

Farida Yunita, S.T., M.T.

Rudolf Sinaga, S.Kom, M.S.I

N. Nelis Febriani SM, S.Kom., M.Kom

Muhammad Faqih Dzulqarnain, S.T., M.CS.

Castaka Agus Sugianto, M.Kom., MCS

Nuk Ghurroh Setyoningrum, S.Kom, M.Cs

Santi Prayudani, S.Kom., M.Kom

Farid, M.Kom

Samsinar, S.Kom., MMSI

Nur Oktavin Idris, M.Eng.

Nova Agustina, S.T., M.Kom.

Erna Piantari, S.Kom., M.T.



# Data Science

Copyright © PT Penamuda Media, 2024

## Penulis:

Dr. (H.C) Zaid Romegar Mair, S.T., M.Cs  
Fitrah Maharani Humaira, S.Si, M.Kom  
Vina Lusiana, S.Pd, M.Si  
Farida Yunita, S.T., M.T.  
Rudolf Sinaga, S.Kom, M.S.I  
N. Nelis Febriani SM, S.Kom., M.Kom  
Muhammad Faqih Dzulqarnain, S.T., M.CS.  
Castaka Agus Sugianto, M.Kom., MCS  
Nuk Ghurroh Setyoningrum, S.Kom, M.Cs  
Santi Prayudani, S.Kom., M.Kom  
Farid, M.Kom  
Samsinar, S.Kom., MMSI  
Nur Oktavin Idris, M.Eng.  
Nova Agustina, S.T., M.Kom.  
Erna Piantari, S.Kom., M.T.

## ISBN:

9786238686230

## Penyunting dan Penata Letak:

Tim PT Penamuda Media

## Desain Sampul:

Tim PT Penamuda Media

## Penerbit:

PT Penamuda Media

## Redaksi:

Casa Sidoarum RT03 Ngentak, Sidoarum Godean Sleman Yogyakarta

Web: [www.penamudamedia.com](http://www.penamudamedia.com)

E-mail: [penamudamedia@gmail.com](mailto:penamudamedia@gmail.com)

Instagram: [@penamudamedia](https://www.instagram.com/penamudamedia)

WhatsApp: [+6285700592256](https://wa.me/6285700592256)

Cetakan Pertama, Juli 2024

x + 238 halaman; 15 x 23 cm

Hak cipta dilindungi undang-undang

Dilarang memperbanyak maupun mengedarkan buku dalam bentuk dan dengan cara apapun tanpa izin tertulis dari penerbit maupun penulis



# Kata pengantar

Dengan senang hati, kami mempersembahkan buku "Data Science" ini kepada para pembaca setia yang tengah menjelajahi dunia ilmu data. Buku ini bertujuan untuk memberikan pemahaman mendalam tentang konsep, metode, dan aplikasi dalam bidang sains data yang semakin vital dalam era digital ini. Semoga buku ini dapat menjadi panduan yang bermanfaat bagi pembaca dalam memahami dan mengaplikasikan pengetahuan data science secara efektif. Selamat membaca!



# DAFTAR ISI

Kata pengantar.....	v
DAFTAR ISI .....	vi
BAB 1 Pengantar Data Science .....	1
A.    Data <i>Science</i> .....	1
B.    Data <i>Scientist</i> .....	3
C.    Sejarah dan Perkembangan Data <i>Science</i> .....	4
BAB 2 Konsep Dasar Statistik untuk Ilmu Data .....	7
A.    Pengertian Data.....	8
B.    Populasi dan Sampel.....	13
C.    Metode Statistika .....	14
BAB 3 Pengumpulan Data: Sumber, Metode dan Etika.....	18
A.    Definisi Pengumpulan Data.....	18
B.    Pentingnya Pengumpulan Data .....	18
C.    Jenis-Jenis Sumber Data.....	20
D.    Teknik Pengumpulan Data Kualitatif.....	23
E.    Teknik Pengumpulan Data Kuantitatif .....	26
F.    Validitas dan Reliabilitas Data .....	30
G.    Pengolahan dan Analisis Data .....	33
H.    Etika dalam Pengumpulan Data .....	36
I.    Studi Kasus Pengumpulan Data .....	39
J.    Tantangan dalam Pengumpulan Data .....	43
K.    Teknologi dalam Pengumpulan Data.....	47



L.	Tren Masa Depan dalam Pengumpulan Data .....	50
M.	Kesimpulan.....	53
BAB 4 Pemrosesan Awal Data: Pembersihan dan Transformasi .....		55
A.	Pembersihan Data.....	56
B.	Transformasi Data .....	59
BAB 5 Eksplorasi Data (Visualisasi Dan Analisis).....		64
A.	Pendahuluan .....	64
B.	Bagian II: Visualisasi Data .....	73
C.	Bagian III: Analisis Data .....	78
D.	Kesimpulan.....	81
BAB 6 Model Prediktif: Regresi dan Klasifikasi .....		82
A.	Metode Regresi .....	83
B.	Metode Klasifikasi .....	89
BAB 7 Model Unsupervised learning: Clustering dan Reduksi Dimensi.....		93
A.	<i>Clustering</i> .....	96
B.	Reduksi Dimensi.....	100
BAB 8 Model Evaluasi: Metrik Kinerja dan Validasi.....		104
A.	Pendahuluan .....	104
B.	Akurasi.....	105
C.	Binary classification Problem.....	108
BAB 9 Pembelajaran Mendalam ( <i>Deep learning</i> ): Konsep dan Aplikasi.....		118
A.	Konsep Pembelajaran Mendalam ( <i>Deep learning</i> ) .....	118
C.	<i>Convolutional Neural network (CNN)</i> .....	126
D.	<i>Artificial Neural Network (LSTM)</i> .....	129



E.	<i>Artificial Neural Network (ANN)</i> .....	131
F.	<i>Recurrent Neural Network (RNN)</i> .....	134
BAB 10 <i>Big Data</i> dan Teknologi Terkait.....		136
A.	Pengenalan <i>Big Data</i> .....	137
B.	Arsitektur dan Infrastruktur <i>Big Data</i> .....	140
C.	Analisis dan Pemrosesan Data .....	143
D.	Keamanan dan Privasi dalam <i>Big Data</i> .....	146
E.	Aplikasi dan Implementasi <i>Big Data</i> .....	148
BAB 11 <i>Machine Learning Ops (MLOps)</i> dan Praktik Terkini.....		151
A.	Latar Belakang Sejarah Dan Perkembangan <i>Machine Learning Ops (MLOps)</i> .....	151
B.	Pengertian <i>Machine Learning Ops (MLOps)</i> .....	153
C.	Komponen Utama <i>MLOps</i> .....	155
D.	Praktik Terbaik dalam <i>MLOps</i> .....	159
E.	Lampiran Modul Praktikum <i>MLOps</i> .....	163
BAB 12 Etika Data (Tantangan dan Prinsip) .....		168
BAB 13 <i>Natural Language Processing (NLP)</i> dalam Ilmu Data .....		175
A.	Pengenalan dan Sejarah <i>Natural Language Processing (NLP)</i> .....	176
B.	Teknik dan Algoritma NLP .....	179
C.	Aplikasi NLP dalam Ilmu Data .....	181
D.	Masa Depan NLP dan Tren Terkini .....	184
BAB 14 Pengantar Studi Kasus dan Aplikasi Industri dalam Ilmu Data .....		187
A.	<i>Tools</i> Penerapan Studi Kasus.....	188
B.	Studi Kasus di Bidang Kesehatan.....	189





- C. Aplikasi di Bidang Kesehatan..... 189
- D. Studi Kasus di Bidang Industri..... 194
- E. Aplikasi di Bidang Industri..... 194
- BAB 15 Rekayasa Fitur: Ekstraksi dan Seleksi Fitur..... 198
  - A. Definisi Ekstraksi dan Seleksi Fitur ..... 198
  - B. Teknik Dasar Ekstraksi Fitur..... 199
  - C. Teknik Dasar Seleksi Fitur..... 205
  - D. Tantangan pada Proses Rekayasa Fitur..... 208
- Daftar Pustaka ..... 211
- Tentang Penulis ..... 223





# DATA SCIENCE

Buku Data Science berisi tentang bagaimana data dan analisisnya mampu mengubah cara kita memahami dunia. Dari konsep dasar hingga aplikasi praktisnya, penulis membawa pembaca melalui perjalanan yang mendalam tentang bagaimana data science menjadi tulang punggung dalam pengambilan keputusan yang cerdas di berbagai industri.

Dengan penekanan pada teknik analisis data, visualisasi yang kuat, dan pemahaman mendalam tentang algoritma machine learning, buku ini tidak hanya memperkenalkan konsep-konsep kunci, tetapi juga memberikan wawasan yang berharga tentang bagaimana data dapat diubah menjadi wawasan berharga. Cocok untuk pembaca dengan beragam latar belakang, buku ini memadukan teori dengan studi kasus yang relevan untuk membantu pembaca memahami dan menguasai esensi dari sains data. Menarik, informatif, dan praktis, "Data Science" adalah panduan yang tak ternilai bagi siapa pun yang ingin memahami kekuatan data di era informasi yang terus berkembang.



PT Penerbit Penamuda Media  
Godean, Yogyakarta  
085700592256  
@penamuda\_media  
penamuda.com