

# BIG DATA ANALYTICS

## KONSEP, IMPLEMENTASI, DAN APLIKASI TERKINI

ZEN MUNAWAR | AGUS MÜLIANTARA | REMUZ MB KMURAWAK | AMNA  
FELIX REBA | ALVIAN M SROYER | DEDE SUKMAWAN | A RAHMAN  
AL HUIJAH ASIANINGRUM | GINA PURNAMA INSANY | SAMUEL A MANDOWEN | KAMDAN  
WILDAN TOYIB | TAGHFIRUL AZHIMA YOGA S | IVANA LUCIA KHARISMA | ISHAK SEMUEL BENO

EDITOR :  
NOVIANTI INDAH PUTRI



# *Big Data Analytics*

Konsep, Implementasi, dan Aplikasi Terkini

## **UU No 28 tahun 2014 tentang Hak Cipta**

### **Fungsi dan sifat hak cipta Pasal 4**

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

### **Pembatasan Pelindungan Pasal 26**

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

- i. Penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
- ii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
- iii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
- iv. Penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

### **Sanksi Pelanggaran Pasal 113**

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 Rp100.000.000 (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 Rp500.000.000,00 (lima ratus juta rupiah).

*Big Data Analytics*  
Konsep, Implementasi, dan Aplikasi Terkini

Zen Munawar  
Agus Muliantara  
Remuz MB Kmurawak  
Amna  
Felix Reba  
Alvian M Sroyer  
Dede Sukmawan  
A Rahman  
Al Hujjah Asianingrum  
Gina Purnama Insany  
Samuel A Mandowen  
Kamdan  
Wildan Toyib  
Taghfirul Azhima Yoga S  
Ivana Lucia Kharisma  
Ishak Samuel Beno

**Penerbit:**



**Anggota IKAPI**  
No. 446/JBA/2022

# *Big Data Analytics: Konsep, Implementasi, dan Aplikasi Terkini*

**Penulis :**

Zen Munawar  
Agus Muliantara  
Remuz MB Kmurawak  
Amna  
Felix Reba  
Alvian M Sroyer  
Dede Sukmawan  
A Rahman  
Al Hujjah Asianingrum  
Gina Purnama Insany  
Samuel A Mandowen  
Kamdan  
Wildan Toyib  
Taghfirul Azhima Yoga S  
Ivana Lucia Kharisma  
Ishak Samuel Beno

**ISBN :** 978-623-93084-9-0 (PDF)

**Editor :** Novianti Indah Putri, S.T., M.Kom.

**Tata Letak :** Yuliana Ayu

**Desain Sampul :** Ayu Lestari

**Penerbit :** Kaizen Media Publishing

**Redaksi :**

Jl. Antapani X, No. 3, Ankid, Antapani, Bandung 40291

Telp/Faks: (022) 20526377

Website: [www.kaizenpublisher.co.id](http://www.kaizenpublisher.co.id) | E-mail: [admin@kaizenpublisher.co.id](mailto:admin@kaizenpublisher.co.id)

**Cetakan Pertama :**

15 Maret 2023

**Ukuran :**

vii, 270, Uk: 15,5 x 23 cm

---

Hak Cipta 2023, Kaizen Media Publishing dan Penulis

Isi diluar tanggung jawab percetakan

**Copyright** © 2023 by Kaizen Media Publishing

All Right Reserved

Hak cipta dilindungi undang-undang

Dilarang keras menerjemahkan, memfotokopi, atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari Penerbit.

---

# KATA PENGANTAR

Puji syukur kami panjatkan kehadiran Allah Yang Maha Esa, karena berkat rahmat dan karunia-Nya sehingga buku *Big Data Analytics: Konsep, Implementasi, dan Aplikasi Terkini* dapat dipublikasikan dan dapat sampai dihadapan pembaca. Buku ini diharapkan dapat hadir memberi kontribusi positif dalam ilmu pengetahuan khususnya terkait kemajuan bidang ilmu kebidanan. Sistematika buku ini mengacu pada pendekatan konsep teoritis dan contoh penerapan.

Buku ini terdiri atas 16 bab yang dibahas secara rinci diantaranya: Pengantar *Big Data*, *Big Data Analytic (BDA)*, *Data Analytics Lifecycle*, BDA dengan *Data Mining, Classification and Prediction, Cluster Analysis, Associations Rules*, Infrastruktur *Big Data*, Visualisasi *Big Data*, *Big Data Tools*, *Security dan Privacy*, *Big Data dan IoT*, *Big Data dan Cloud Computing*, Peluang dan Hambatan, Studi Kasus *Big Data, Blockchain*.

Kami menyadari bahwa tulisan ini jauh dari kesempurnaan dan masih terdapat banyak kekurangan, sejatinya kesempurnaan itu hanya milik Allah Yang Maha Kuasa. Oleh sebab itu, kami tentu menerima masukan dan saran dari pembaca demi penyempurnaan lebih lanjut.

Akhirnya kami mengucapkan terima kasih yang tak terhingga kepada semua pihak yang telah mendukung dalam proses penyusunan dan perbitan buku ini secara khusus kepada Kaizen Media Publishing. Semoga buku ini dapat bermanfaat bagi pembaca sekalian.

Bandung, Februari 2023

Editor



# DAFTAR ISI

KATA PENGANTAR.....	i
DAFTAR ISI.....	iii
<b>BAB 1 : Pengantar Big Data .....</b>	<b>1</b>
1.1    Pendahuluan .....	1
1.2    Sejarah <i>Big Data</i> .....	2
1.3    Tujuan dan Nilai <i>Big Data</i> .....	5
1.4    Manfaat <i>Big Data</i> .....	8
<b>BAB 2 : Big Data Analytic (BDA) .....</b>	<b>17</b>
2.1    Pengertian <i>Big Data Analytics</i> (BDA).....	17
2.2    Sektor pengguna <i>Big Data Analytics</i> .....	17
2.3    Kegunaan <i>Big Data Analytics</i> untuk Perkembangan Bisnis .....	19
2.4    Langkah-langkah Data Analytics .....	19
2.5    Tipe Data Analytics.....	21
2.6    Teknik <i>Data analytics</i> .....	22
2.7    Peralatan <i>Data analytics</i> .....	29
2.8    Kemampuan yang dibutuhkan untuk menjadi <i>data analyst</i> .....	30
<b>BAB 3 : Data Analytics Lifecycle .....</b>	<b>35</b>
3.1    Pengantar <i>Data Analytics Lifecycle</i> .....	35
3.2    Fase 1 <i>Discovery</i> .....	37
3.3    Fase 2 <i>Data Preparation</i> .....	40
3.4    Fase 3 <i>Model Planning</i> .....	45
3.5    Fase 4 <i>Model Building</i> .....	48
3.6    Fase 5 <i>Communicate Results</i> .....	50
3.7    Fase 6: <i>Operationalize</i> .....	52
<b>BAB 4 : Big Data Analytic dengan Data Mining .....</b>	<b>57</b>
4.1    Pengertian Data .....	57
4.2    Pengertian <i>Big Data</i> .....	57
4.3 <i>Big Data Analytic</i> .....	59
4.4    Teknik <i>Big Data Analytics</i> .....	60
4.5 <i>Data Mining</i> .....	62
4.6 <i>Big Data</i> dan <i>Data Mining</i> .....	63
<b>BAB 5 : Klasifikasi dan Prediksi .....</b>	<b>69</b>
5.1    Pendahuluan .....	69

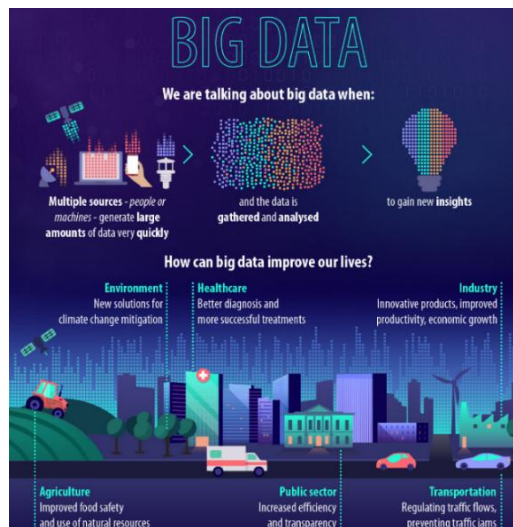


5.2	Apa itu Klasifikasi dan Prediksi? .....	70
5.3	Pengukuran Kesalahan Klasifikasi dan Prediksi .....	75
5.3.1	Ukuran Akurasi Klasifikasi .....	75
5.3.2	Pengukuran Kesalahan Prediksi .....	76
5.4	Masalah Terkait Klasifikasi dan Prediksi .....	78
5.4.1	Menyiapkan Data untuk Klasifikasi dan Prediksi .....	78
5.4.2	Membandingkan Metode Klasifikasi dan Prediksi .....	79
5.4.3	Mengevaluasi Keakuratan Klasifikasi atau Prediktor .....	80
<b>BAB 6</b>	<b>: Analisis Kluster .....</b>	<b>83</b>
6.1	Pendahuluan .....	83
6.2	Definis Analisis Kluster .....	83
6.3	Jenis Data dalam Analisis Kluster .....	84
6.4	Metode Partisi .....	88
6.5	Metode Hierarkis .....	89
6.6	Metode Berbasis Kepadatan .....	90
6.7	Metode Berbasis Grid .....	91
6.8	Metode Pengelompokan Berbasis Model .....	92
6.9	Analisis Kluster Berbasis Kendala .....	94
6.10	Analisis <i>outlier</i> .....	95
<b>BAB 7</b>	<b>: Associations Rules .....</b>	<b>99</b>
7.1	Pengertian <i>Associations Rules</i> .....	99
7.2	Defisini <i>Frequent Item Set</i> dan <i>Parameter Associations Rules</i> .....	100
7.3	Algoritma <i>Associations Rules</i> berdasarkan dimensinya .....	101
7.3.1	Algoritma satu dimensi (Single Dimensional) .....	101
7.3.2	Algoritma dua atau lebih dari dua dimensi (multi dimensional) .....	108
<b>BAB 8</b>	<b>: Infrastruktur Big Data .....</b>	<b>113</b>
8.1	Infrastruktur <i>Big Data</i> .....	113
8.2	Komponen Utama .....	116
8.3	Infrastruktur Teknologi <i>Big Data</i> .....	121
<b>BAB 9</b>	<b>: Visualisasi Big Data .....</b>	<b>125</b>
9.1	Pengantar Visualisasi <i>Big Data</i> .....	125
9.2	Jenis-Jenis Visualisasi <i>Big Data</i> .....	126
9.3	Membuat visualisasi <i>big data</i> yang efektif .....	129
9.4	Aplikasi Visualisasi Data .....	131
9.5	Keuntungan dan Keterbatasan Visualisasi <i>Big Data</i> .....	133

# BAB 1 : PENGANTAR *BIG DATA*

## 1.1 Pendahuluan

**B**ig Data dapat didefinisikan sebagai data dalam jumlah besar, baik terstruktur maupun tidak terstruktur, biasanya disimpan di *cloud* atau di pusat data, yang kemudian dimanfaatkan oleh perusahaan, organisasi, startup, bahkan pemerintah untuk tujuan yang berbeda. *Big Data* adalah istilah untuk kumpulan kumpulan data yang begitu besar dan kompleks sehingga menjadi sulit untuk diproses menggunakan teknik dan alat penambangan data konvensional. Seringkali, banyak analis data profesional menyiratkan proses ekstraksi, transformasi, dan pemuatan (*extraction, transformation, and load*) atau ETL untuk kumpulan data besar sebagai konotasi *Big Data*. Deskripsi populer *Big Data* didasarkan pada tiga atribut utama data: volume, kecepatan, dan variasi (*volume, velocity, and variety*) atau 3V.



Gambar 1. 1 *Big Data*

Sumber: (Society, 2021)

Jumlah data yang besar dan analitis menyebabkan lahirnya istilah *big data* (Munawar & Putri, 2020). Ini bukan istilah yang tepat dan tidak membawa arti tertentu selain gagasan tentang ukurannya. Kata "besar" terlalu umum; pertanyaan seberapa "besar" itu besar dan seberapa "kecil" itu kecil (Smith, 2013) adalah relatif terhadap waktu, ruang, dan keadaan. Dari perspektif evolusioner, ukuran "*Big Data*" selalu berkembang. Jika kita menggunakan kapasitas lalu lintas Internet global saat ini (Cisco, 2023) sebagai tongkat pengukur, arti volume *Big Data* akan berada di antara kisaran terabyte (TB atau 10<sup>12</sup> atau 240) dan *zettabyte* (ZB atau 10<sup>21</sup> atau 270). Berdasarkan tingkat pertumbuhan lalu lintas data historis, Cisco mengklaim bahwa manusia telah memasuki era ZB pada tahun 2015 (Cisco, 2023).

Tujuan utama bab ini adalah untuk memberikan pandangan historis *Big Data* dan untuk menyatakan bahwa ini bukan hanya 3V, melainkan 32V atau 9V. Atribut *Big Data* tambahan ini mencerminkan motivasi sebenarnya di balik analitik *Big Data* (BDA). Dapat dipercaya bahwa fitur yang diperluas ini mengklarifikasi beberapa pertanyaan dasar tentang esensi BDA: masalah apa yang dapat diatasi oleh *Big Data*, dan masalah apa yang tidak boleh disamakan dengan BDA. Perlunya keseimbangan antara teknologi baru untuk memproses data pribadi dengan perlindungan privasi data (Putri, Rustiyana, Herdiana, & Munawar, 2021).

## 1.2 Sejarah *Big Data*

Saat ini dunia komputasi telah berubah dari terpusat ke sistem terdistribusi lalu ke komputasi awan (Putri, Musadad, Munawar, & Komalasari, 2021), dimana *big data* disimpan. Beberapa kajian telah dilakukan mengenai pandangan sejarah dan perkembangan di kawasan BDA. Gil Press (Press, 2013) memberikan sejarah singkat *Big Data* mulai dari tahun 1944, yang didasarkan pada karya Rider (Rider, 1944). Dia meliputi 68 tahun sejarah evolusi *Big Data* antara 1944 dan 2012 dan mengilustrasikan 32 peristiwa terkait *Big Data* dalam sejarah ilmu data terkini. Seperti yang ditunjukkan Press dalam artikelnya, garis tipis antara pertumbuhan data dan *Big Data*