litnus

BUKU AJAR

PENGANTAR DATA SAINS

Halim Fathoni - Dwi Handoko - Rizka Permata
Atika Arpan - Dian Meilantika

BUKU AJAR

PENGANTAR DATA SAINS

Halim Fathoni - Dwi Handoko - Rizka Permata
Atika Arpan - Dian Meilantika

litnus

PENGANTAR DATA SAINS

Ditulis oleh:

Halim Fathoni
Dwi Handoko
Rizka Permata
Atika Arpan
Dian Meilantika

Diterbitkan, dicetak, dan didistribusikan oleh
PT Literasi Nusantara Abadi Grup
Perumahan Puncak Joyo Agung Residence Blok B11 Merjosari
Kecamatan Lowokwaru Kota Malang 65144
Telp : +6285887254603, +6285841411519
Email: literasinusantaraofficial@gmail.com
Web: www.penerbitlitnus.co.id
Anggota IKAPI No. 340/JTI/2022

Hak Cipta dilindungi oleh undang-undang. Dilarang mengutip atau memperbanyak baik sebagian ataupun keseluruhan isi buku dengan cara apa pun tanpa izin tertulis dari penerbit.

Cetakan I, Desember 2025

Perancang sampul: Rosyiful Aqli
Penata letak: D Gea Nuansa

ISBN : 978-634-234-884-0

viii + 228 hlm. ; 15,5x23 cm.

©Desember 2025

Kata Pengantar

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan **Buku Ajar Pengantar Sains Data** ini dengan baik. Buku ajar ini disusun untuk membantu mahasiswa dan pembaca dalam memahami konsep dasar hingga penerapan praktis sains data secara sistematis, terstruktur, dan mudah dipahami.

Sains data merupakan bidang yang sangat relevan dengan perkembangan teknologi saat ini, di mana pengolahan dan analisis data menjadi kunci penting dalam pengambilan keputusan. Melalui buku ajar ini, penulis berusaha menyajikan materi yang mencakup teori, metode, serta contoh penerapan yang diharapkan dapat mendukung proses pembelajaran, penelitian, maupun praktik di lapangan.

Tujuan penulisan dan adanya buku ini adalah untuk memberikan pemahaman dasar mengenai konsep, metode, serta ruang lingkup sains data, sekaligus menjadi panduan pembelajaran bagi mahasiswa dalam mengikuti perkuliahan maupun praktik di bidang terkait. Selain itu, buku ini diharapkan dapat menjadi referensi bagi dosen, peneliti, dan praktisi dalam mengembangkan serta menerapkan metode analisis data, membantu pembaca memahami penerapan sains data di berbagai bidang seperti bisnis, kesehatan, pemerintahan, maupun teknologi, serta mendorong terciptanya budaya pengambilan keputusan berbasis data (data-driven decision making) baik di lingkungan akademik maupun industri.

Penyusunan buku ajar ini tentu tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan masukan, motivasi, serta dorongan dalam proses penulisan buku ini.

Penulis menyadari bahwa buku ajar ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi penyempurnaan di masa mendatang. Semoga buku ajar ini dapat memberikan manfaat yang luas, khususnya bagi mahasiswa, dosen, praktisi, maupun siapa pun yang tertarik dalam bidang sains data.

Daftar Isi

BAB VII

BAB VIII

BAB IX

BAB X

BAB XI

BAB I

Pendahuluan Data Science

A. Apa Itu Data Science?

Dalam konteks perkembangan digital dewasa ini, istilah data science sering kali muncul sebagai salah satu konsep yang paling sentral sekaligus strategis. Hal ini tidak terlepas dari kenyataan bahwa dunia modern ditandai oleh percepatan eksponensial dalam produksi data yang bersumber dari beragam aktivitas manusia, baik dalam ruang personal, sosial, maupun profesional. Peningkatan masif tersebut terjadi akibat semakin terintegrasinya perangkat digital ke dalam hampir seluruh aspek kehidupan. Kehadiran media sosial, platform e-commerce, layanan keuangan digital, hingga sensor cerdas yang terhubung melalui Internet of Things telah membentuk ekosistem baru yang menuntut kemampuan untuk memahami, mengolah, serta memanfaatkan data secara optimal.

Secara konseptual, data science dapat dipandang sebagai disiplin ilmu multidimensi yang menggabungkan berbagai metode, pendekatan, serta kerangka kerja dari statistika, ilmu komputer, dan pemahaman domain tertentu untuk menghasilkan pengetahuan yang relevan. Pada hakikatnya, data science bukan sekadar praktik analisis data, melainkan sebuah

proses ilmiah yang bertujuan untuk memprediksi, mengoptimalkan, serta mendukung pengambilan keputusan berbasis data melalui integrasi teknik statistik dengan pemrograman komputer (Dhar, 2013). Artinya, data science tidak hanya menekankan proses perhitungan matematis, tetapi juga memfasilitasi proses pengambilan keputusan yang lebih cerdas, efisien, dan akurat dalam berbagai konteks. Penting untuk disadari bahwa data science lahir bukan dalam ruang hampa. Ia merupakan jawaban atas tantangan baru yang tidak mampu sepenuhnya dijawab oleh disiplin yang sudah mapan seperti statistika klasik maupun business intelligence. Dalam statistika tradisional, fokus utama adalah inferensi, yakni menarik kesimpulan dari sampel untuk mewakili populasi. Namun, dalam dunia nyata, data sering kali hadir dalam bentuk yang sangat kompleks: tidak terstruktur, berskala masif, serta berubah dengan cepat. Kondisi ini memerlukan pendekatan baru yang lebih fleksibel, eksperimental, serta iteratif, sehingga muncullah data science sebagai kerangka yang dapat menjembatani kekosongan tersebut.

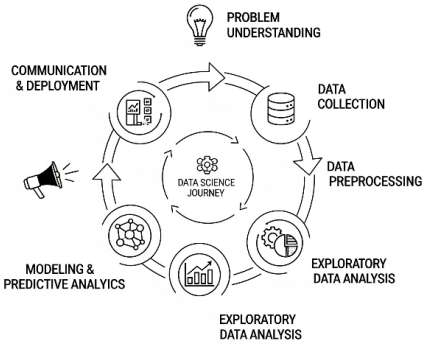
Berbeda dengan business intelligence yang umumnya terbatas pada penyajian laporan historis serta visualisasi deskriptif, data science berorientasi pada eksplorasi, prediksi, dan pemodelan. Ia memungkinkan organisasi tidak hanya memahami apa yang terjadi di masa lalu, tetapi juga memprediksi kemungkinan yang akan terjadi di masa depan. Dengan kata lain, data science lebih bersifat prospektif daripada retrospektif. Hal inilah yang menjadikan disiplin ini sangat penting bagi organisasi modern yang ingin tetap kompetitif dalam lingkungan yang sangat dinamis.

Salah satu keunggulan fundamental dari data science adalah kemampuannya untuk mengintegrasikan aspek teoretis dengan aspek aplikatif secara bersamaan. Dari sisi teori, seorang data scientist dituntut untuk memiliki pemahaman mendalam mengenai teori probabilitas, distribusi data, hingga konsep inferensi statistika. Akan tetapi, teori semata tidaklah cukup. Data yang dihadapi pada era digital memiliki karakteristik unik: besar, heterogen, dan sering kali tidak terstruktur. Oleh karena itu, selain keterampilan teoretis, seorang data scientist juga harus

BAB II

Road Map Proyek Data Science

Dalam konteks transformasi digital yang semakin pesat, data science tidak hanya dipandang sebagai sekumpulan teknik analisis data, melainkan juga sebagai suatu **journey** atau perjalanan sistematis yang melibatkan serangkaian tahapan terstruktur. Setiap tahapan dalam **road map** proyek data science memiliki peran esensial yang saling berhubungan, sehingga membentuk suatu siklus berkesinambungan dari pemahaman masalah hingga implementasi solusi. Tahapan ini sekaligus menjadi kerangka kerja metodologis yang memungkinkan seorang praktisi data, baik akademisi maupun profesional industri, untuk menavigasi kompleksitas data dan menghasilkan pengetahuan yang dapat diimplementasikan secara nyata.



Gambar 2.1 Perjalanan Data Science (Data Science Journey)

A. Problem Understanding (Pemahaman Masalah)

Tahap pertama dalam perjalanan data science adalah pemahaman masalah (problem understanding). Pada tahap ini, seorang data scientist berupaya mengidentifikasi inti persoalan yang dihadapi organisasi atau individu yang menjadi pemilik data. Pemahaman masalah bukan hanya terbatas pada definisi teknis dari suatu persoalan, melainkan juga melibatkan interpretasi konteks bisnis, sosial, maupun ilmiah yang melatarbelakanginya. Dengan kata lain, data science tidak bisa dilepaskan dari realitas domain tempat data itu dihasilkan. Oleh sebab itu, sebelum melangkah pada pengumpulan maupun pengolahan data, data scientist harus terlebih dahulu menjawab pertanyaan fundamental: apa sebenarnya masalah yang ingin diselesaikan, dan mengapa masalah tersebut penting untuk diteliti atau dicari solusinya?

Sebagai contoh, dalam dunia bisnis ritel, permasalahan dapat berwujud rendahnya tingkat retensi pelanggan. Pemahaman masalah dalam kasus ini tidak semata-mata mengamati angka penurunan penjualan, tetapi juga mencakup pertanyaan strategis seperti faktor-faktor yang mendorong pelanggan berhenti berbelanja, peran kompetitor, serta dinamika kebutuhan konsumen. Sementara dalam bidang kesehatan, problem

BAB III

Pembersihan Data *(Data Munging)*

A. Definisi dan Konsep Dasar Data Munging

Dalam kerangka kerja sains data, data munging sering juga disebut sebagai data wrangling merupakan tahapan yang berfungsi untuk mentransformasi data mentah menjadi format yang bersih, konsisten, serta mudah dianalisis. Tahap ini dipandang sebagai fondasi utama, sebab kualitas informasi yang dihasilkan dari analisis statistik maupun algoritme pembelajaran mesin pada akhirnya sangat ditentukan oleh kualitas data input.

Kandel et al. (2011) menegaskan bahwa data munging tidak sekadar aktivitas teknis berupa pembersihan error, melainkan meliputi serangkaian proses yang lebih luas, mulai dari deteksi ketidakteraturan, perbaikan struktur, hingga pengayaan data agar sesuai dengan kebutuhan analitik. Data yang diperoleh dari lapangan atau sistem dunia nyata biasanya tidak sempurna: terdapat duplikasi, format yang tidak konsisten, nilai yang hilang, ataupun variabel yang kurang relevan. Jika kondisi ini dibiarkan, hasil analisis dapat mengalami bias serius.

Sejalan dengan hal tersebut, Provost dan Fawcet (2013) mengingatkan bahwa kualitas model prediktif selalu bergantung pada kualitas data masukan. Dengan kata lain, kegagalan dalam tahap munging dapat menggagalkan keseluruhan siklus proyek data science. Oleh karena itu, pemahaman mendalam mengenai konteks domain data, keterampilan teknis dalam mengelola dataset, serta kepekaan analitis merupakan kombinasi kompetensi yang wajib dimiliki oleh seorang praktisi.

Lebih jauh, data munging tidak dapat dipandang hanya sebagai aktivitas mekanis. Ia merupakan proses interpretatif yang menuntut penggabungan antara intuisi analis dengan prinsip metodologis. Dalam banyak kasus, pembersihan data memerlukan kompromi, misalnya dalam memutuskan apakah suatu nilai yang hilang sebaiknya diimputasi, dihapus, atau dibiarkan sebagaimana adanya. Keputusan tersebut menuntut justifikasi akademis sekaligus pertimbangan etis, sehingga menjadikan data munging sebagai proses yang tidak kalah kompleks dibandingkan tahap analisis lanjutan.

B. Tahapan Utama dalam Data Munging

Data munging, yang juga dikenal sebagai data wrangling, merupakan proses penting dalam analisis data modern. Proses ini bertujuan untuk mengubah data mentah yang biasanya masih berantakan, tidak konsisten, dan sulit digunakan, menjadi dataset yang bersih, terstruktur, dan siap dianalisis. Dalam penelitian maupun implementasi industri, kualitas data sering kali menjadi faktor penentu keberhasilan analisis. Seperti pepatah yang populer di dunia data: **"garbage in, garbage out"**, artinya kualitas hasil analisis sangat bergantung pada kualitas data yang digunakan.

BAB IV

Machine Learning, Deep Learning, dan *Artificial Intelligence*

Artificial Intelligence (AI) dan Data Science adalah dua bidang yang saat ini menjadi motor penggerak revolusi teknologi. AI sendiri lahir dari gagasan dasar yang dikemukakan Alan Turing (1950) dalam makalahnya "Computing Machinery and Intelligence", yang memperkenalkan Turing Test sebagai cara mengukur kecerdasan mesin . Pertanyaan Turing "Can machines think?" menjadi landasan filosofis sekaligus praktis dari riset AI modern.

Tonggak penting lain terjadi pada Konferensi Dartmouth (1956), di mana John McCarthy, Marvin Minsky, dan Claude Shannon secara resmi memperkenalkan istilah "Artificial Intelligence". Sejak saat itu, AI berkembang sebagai disiplin tersendiri yang mencoba meniru cara berpikir, belajar, dan mengambil keputusan seperti manusia.

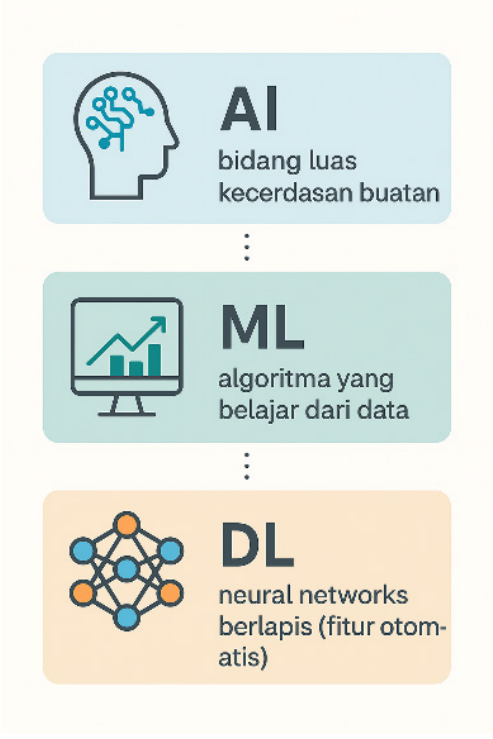
Namun, perjalanan AI tidak selalu mulus. Pada 1970–1990-an, penelitian AI sempat mengalami AI Winter akibat keterbatasan data dan komputasi. Baru pada awal abad ke-21, dengan munculnya big data dan

komputasi GPU, AI kembali bangkit. Machine Learning (ML) dan Deep Learning (DL) menjadi fondasi utama kebangkitan ini.

Menurut The Data Science Handbook (2024), "data is the new oil, and machine learning and deep learning are the engines that refine it into knowledge." Dengan kata lain, Data Science hadir sebagai bidang yang memadukan statistik, ilmu komputer, dan domain knowledge, dengan ML dan DL sebagai jantungnya. Kini, AI tidak hanya menjadi objek penelitian, tetapi juga menjadi fondasi ekonomi digital yang mengubah cara manusia bekerja, berinteraksi, dan mengambil keputusan.

Perbedaan Dasar Artificial Intelligence, Machine Learning, dan Deep Learning

Secara konseptual, AI, ML, dan DL memiliki hubungan hierarkis:



Gambar 4.1. Perbedaan Dasar Artificial Intelligence,

BAB V

Visualisasi dan Metrik Sederhana

Visualisasi data adalah cara mengubah data mentah menjadi representasi grafis (grafik, peta, diagram) agar pola, tren, dan anomali lebih mudah dipahami dan dikomunikasikan. Visualisasi data ini membantu dua aktivitas inti ilmu data: eksplorasi (menemukan pola dan masalah sebelum pemodelan) dan komunikasi (menyampaikan temuan secara jelas kepada pemangku kepentingan non-teknis). Literatur klasik dan modern sepakat bahwa visualisasi yang baik mempercepat analisis, mengurangi beban kognitif pembaca, dan mencegah kekeliruan interpretasi ketika hanya bergantung pada tabel angka.

Di level praktis, visualisasi bekerja berdampingan dengan metrik sederhana *(simple metrics)* seperti rata-rata, median, kuartil, simpangan baku, korelasi untuk memberi ringkasan kuantitatif yang singkat. Ringkasan ini memperkaya grafik (mis. menambahkan *annotation* nilai median pada boxplot atau menampilkan korelasi pada scatter plot) sehingga pembaca mendapatkan konteks angka sekaligus bentuk distribusinya. Prinsip ini diajarkan luas di materi pengantar visualisasi/data science modern.

Mengapa Visualisasi Data Penting?

1. Mempermudah Eksplorasi Data
   Visualisasi memungkinkan analis dan ilmuwan data untuk dengan cepat memahami struktur data, distribusi nilai, dan hubungan antar variabel. Misalnya, jika kita memiliki ribuan baris data penjualan, histogram dapat membantu melihat distribusi penjualan tanpa harus meneliti setiap nilai satu per satu. *Scatterplot* dapat menunjukkan apakah ada hubungan antara harga dan jumlah pembelian.

2. Mendukung Pengambilan Keputusan
   Keputusan yang baik didasarkan pada informasi yang akurat dan mudah dipahami. Dengan visualisasi, pengambil keputusan seperti manajer atau eksekutif dapat dengan cepat menangkap pesan utama dari data. Misalnya, sebuah grafik tren bulanan lebih efektif daripada tabel angka panjang ketika ingin menilai kinerja penjualan.

3. Meningkatkan Komunikasi
   Visualisasi data berfungsi sebagai bahasa universal yang dapat dimengerti oleh berbagai pihak, termasuk mereka yang tidak memiliki latar belakang teknis. Grafik yang dirancang dengan baik dapat menjelaskan hasil analisis kompleks secara sederhana dan intuitif, sehingga mendukung komunikasi lintas departemen.

Jenis Visualisasi Berdasarkan Tujuan

Setiap jenis visualisasi data memiliki kegunaan tertentu. Beberapa jenis yang sering digunakan antara lain:

1. Distribusi: Histogram dan boxplot digunakan untuk melihat sebaran data serta mendeteksi nilai yang berbeda jauh (outlier).
2. Komposisi: Diagram lingkaran (pie chart) dan diagram batang bertumpuk (stacked bar chart) digunakan untuk menunjukkan bagian dari keseluruhan.
3. Perbandingan: Diagram batang (bar chart) dan garis (line chart) digunakan untuk membandingkan kategori atau tren antar waktu.

BAB VI

Overview Machine Learning vs AI

Kegagalan Awal Gerakan AI dan "Musim Dingin AI" Konsep "machine learning" sebagian besar lahir dari kegagalan awal gerakan AI pada tahun 1960-an, 1970-an, dan 1980-an. Pada masa itu, orang-orang sangat fokus pada gagasan bahwa komputer bisa dibuat untuk berpikir, dan secara luas diharapkan bahwa mesin-mesin berpikir akan segera terwujud dalam beberapa tahun. Bahkan, ada anekdot bahwa Marvin Minsky, salah satu pendiri AI, pernah menugaskan seorang mahasiswa pascasarjana untuk mengerjakan visi komputer selama musim panas. Fokusnya adalah memperlakukan otak manusia sebagai "mesin logika besar" dan membuat komputer meniru pemrosesan logis yang dilakukan manusia.

Namun, AI gagal mencapai ekspektasi yang telah dihasilkan (setidaknya relatif terhadap *hype*-nya), yang menyebabkan "musim dingin AI" di mana pendanaan rendah dan pesimisme tinggi. Inilah sebagian alasan mengapa istilah "kecerdasan buatan" (artificial intelligence) menurun popularitasnya dan digantikan oleh "machine learning".

Pergeseran Fokus dari Mimikri Manusia ke Pencocokan Pola (ML) Pergeseran fokus terjadi dari meniru kecerdasan manusia ke menggunakan komputer untuk melakukan tugas-tugas yang secara historis harus

dilakukan oleh manusia, tanpa berpura-pura melakukannya dengan cara yang mirip manusia. Contohnya termasuk mengenali apakah ada burung dalam foto, menentukan apakah *email* adalah spam, atau mengidentifikasi "peristiwa menarik" dalam deret waktu. Machine learning (ML) dibangun di atas penggunaan komputer sebagai pengganti penilaian manusia dalam situasi spesifik dan terbatas. Teknik-teknik yang dikembangkan ini dapat diterapkan pada banyak bidang, bahkan yang penilaian manusia tidak pernah diterapkan dalam praktik, sehingga ML telah matang menjadi *toolset* standar bagi setiap ilmuwan data.

Pergeseran Jenis Alat yang Digunakan .

- AI secara tradisional menggunakan pendekatan berbasis aturan yang memanfaatkan inferensi logis untuk mencapai kesimpulan, berdasarkan ide bahwa pikiran manusia pada dasarnya bernalar dari prinsip-prinsip dasar saat memecahkan masalah.
- ML kurang berorientasi pada logika dan lebih banyak tentang pencocokan pola. ML menggunakan data pelatihan untuk mempelajari heuristik, mengukur seberapa efektifnya, dan kemudian menerapkan heuristik tersebut untuk membuat prediksi probabilistik tentang situasi di masa depan.

Kebangkitan Kembali AI dengan *Deep Learning* dan LLM Lebih baru ini, kemajuan dalam *deep learning* telah menghidupkan kembali antusiasme terhadap AI. Dengan munculnya *Large Language Models* (LLM) seperti ChatGPT, alat-alat ini cukup fleksibel dan umum sehingga istilah "kecerdasan buatan" menjadi lebih tepat. Namun, alat-alat ini sama sekali tidak menyerupai mesin logika di masa lalu. Sebaliknya, mereka didasarkan pada pendekatan heuristik dan pencocokan pola ML – hanya saja dibawa ke tingkat yang jauh lebih tinggi.

Hubungan Saat Ini antara ML dan AI Saat ini, ML dan AI dianggap sebagai konsep ortogonal.

BAB VII

Machine Learning Clasification

Sumber-sumber ini memberikan penjelasan komprehensif tentang klasifikasi pembelajaran mesin (machine learning classification), mulai dari definisi dasar, kasus penggunaan, tantangan praktis, jenis-jenis pengklasifikasi spesifik, hingga metode evaluasinya.

Berikut adalah diskusi mendalam tentang Klasifikasi Pembelajaran Mesin:

Apa Itu Pengklasifikasi (Classifier)?

Pengklasifikasi pembelajaran mesin adalah objek komputasi yang memiliki dua tahap utama:

- **Pelatihan (Training)**: Pada tahap ini, pengklasifikasi menerima data pelatihan yang terdiri dari banyak titik data dan label yang benar yang terkait dengannya. Tujuannya adalah untuk **mempelajari pola** tentang bagaimana titik-titik data memetakan ke label tersebut.
- **Prediksi (Prediction)**: Setelah dilatih, pengklasifikasi bertindak sebagai fungsi yang mengambil titik data tambahan dan mengeluarkan **klasifikasi yang diprediksi** untuknya. Prediksi ini terkadang berupa

label spesifik, tetapi lebih sering berupa angka bernilai kontinu (atau beberapa angka) yang dapat dilihat sebagai **skor kepercayaan** untuk label tertentu.

Pengklasifikasi sering diromantisasi sebagai "kotak hitam ajaib" yang dapat mempelajari segalanya dan menyelesaikan masalah bisnis, padahal kenyataannya lebih mendasar. Dibutuhkan banyak pekerjaan untuk mempersiapkan data, mengarahkan kotak hitam pada pertanyaan yang tepat, dan memahami hasilnya.

Kasus Penggunaan Pengklasifikasi

Ada dua kasus penggunaan utama untuk pengklasifikasi:

- **Klasifikasi Langsung (Obvious Use Case)**: Ini terjadi ketika ada hal-hal yang perlu diklasifikasikan secara langsung. Contohnya termasuk komputer yang memutuskan iklan mana yang akan ditampilkan kepada pengguna, atau menandai potensi kasus penipuan kartu kredit untuk diperiksa oleh manusia.
- **Memberikan Wawasan (Insights about the Underlying Data)**: Ini adalah penggunaan yang lebih umum. Klien seringkali tidak hanya tertarik pada prediksi (misalnya, mesin akan gagal), tetapi lebih pada **pola-pola dalam data yang memprediksi kegagalan**. Pola-pola ini dapat membantu mendiagnosis dan memperbaiki masalah. Dalam kasus seperti ini, ilmuwan data perlu membedah pengklasifikasi untuk mengekstrak wawasan bisnis. Tantangannya adalah pengklasifikasi yang paling akurat terkadang paling sulit untuk dipahami secara dunia nyata.

Masalah Praktis dalam Klasifikasi

Beberapa masalah praktis muncul dalam implementasi klasifikasi ML:

- **Ketersediaan Data Pelatihan Berlabel Benar**: Klasifikasi ML didasarkan pada gagasan memiliki data pelatihan berlabel benar dalam jumlah yang cukup. Namun, ini seringkali merupakan kemewahan yang tidak tersedia di dunia nyata.

BAB VIII

Visualisasi Lasso dan *Feature Selection, Big Data*, dan Database

A. Visualisasi Lasso Dan Feature Selection

Dalam era digital yang semakin berkembang, data telah menjadi salah satu aset paling berharga bagi organisasi, perusahaan, maupun lembaga penelitian. Volume data yang terus bertambah, baik yang terstruktur maupun tidak terstruktur, menuntut adanya metode analisis yang mampu memberikan wawasan bermakna serta mendukung pengambilan keputusan yang tepat (Collins et al., 2021). Salah satu cabang penting dalam analisis data adalah regresi, sebuah metode statistika yang digunakan untuk memahami hubungan antara variabel input (predictors atau independent variables) dengan variabel output (response atau dependent variable) (Ir & Tarumingkeng, 2025b).

Namun, ketika jumlah variabel prediktor semakin banyak, tantangan baru muncul. Tidak semua variabel memberikan kontribusi signifikan dalam memprediksi output. Beberapa variabel mungkin memiliki pengaruh

yang kecil, bahkan menambah noise dalam model sehingga menurunkan performa prediksi. Di sinilah peran feature selection (pemilihan fitur) menjadi sangat penting, karena dapat membantu menyaring variabel-variabel yang paling relevan dan mengabaikan yang tidak berguna.

Salah satu metode populer yang digunakan untuk tujuan ini adalah LASSO Regression (Least Absolute Shrinkage and Selection Operator). Metode ini bukan hanya sekadar teknik regresi biasa, tetapi juga berfungsi sebagai mekanisme seleksi variabel dengan cara melakukan regularisasi terhadap koefisien regresi (Nurfitri Imro'ah, 2020).

Urgensi dan Relevansi LASSO dalam Data Sains

Dalam praktik nyata, data ilmiah dan bisnis modern seringkali memiliki ribuan hingga jutaan fitur (Wibowo, 2025). Misalnya:

- Bidang Kesehatan: Prediksi penyakit menggunakan ribuan marker genetik.
- Bidang Keuangan: Analisis risiko kredit dengan banyak indikator ekonomi.
- Bidang Pemasaran: Menentukan faktor yang memengaruhi keputusan pembelian dari berbagai variabel perilaku pelanggan.
- Bidang IoT dan Big Data: Prediksi konsumsi energi dari ratusan sensor dalam waktu nyata.

Dalam kondisi ini, metode regresi linear klasik sering kali tidak memadai karena:

- Rentan terhadap overfitting jika jumlah variabel terlalu besar.
- Sulit diinterpretasikan jika semua variabel dimasukkan.
- Tidak efisien secara komputasi ketika skala data sangat tinggi.

BAB IX

Time Series Analysis, Probability , Statistics

A. *Time Series Analysis*

Analisis deret waktu (Time Series Analysis) merupakan salah satu cabang penting dalam ilmu data dan statistika yang berfokus pada data yang dikumpulkan atau direkam secara berurutan berdasarkan waktu (Laksma Pradana, 2025). Tidak seperti data cross-sectional yang menggambarkan kondisi pada satu titik waktu, data deret waktu menekankan pada perubahan, pola, serta dinamika data sepanjang periode tertentu.

Data deret waktu dapat ditemukan di berbagai bidang kehidupan sehari-hari, misalnya:

- Ekonomi dan Bisnis: harga saham, inflasi, nilai tukar mata uang, penjualan bulanan.
- Sains dan Teknik: suhu harian, curah hujan, tekanan udara, penggunaan energi listrik.
- Kesehatan: jumlah pasien harian, perkembangan penyebaran penyakit, detak jantung.

- Teknologi: trafik jaringan internet, jumlah pengguna aplikasi dari waktu ke waktu.

Pentingnya Analisis Deret Waktu

Analisis deret waktu memiliki peran yang sangat strategis dalam pengambilan keputusan berbasis data (Yusapra Salim et al., 2024). Beberapa manfaat utamanya antara lain:

- Identifikasi Pola: membantu menemukan pola musiman (seasonality), tren jangka panjang (trend), atau fluktuasi acak pada data.
- Peramalan (Forecasting): memprediksi nilai data di masa depan berdasarkan pola masa lalu. Misalnya, memprediksi permintaan produk untuk menentukan strategi produksi.
- Deteksi Anomali: mengidentifikasi kejadian tidak biasa atau penyimpangan data yang signifikan, seperti mendeteksi serangan siber dari pola trafik jaringan.
- Pengendalian dan Perencanaan: memberikan dasar bagi organisasi dalam perencanaan strategis, pengelolaan risiko, dan evaluasi performa.

Relevansi dalam Data Sains

Dalam konteks Data Sains, analisis deret waktu sangat relevan karena banyak data modern yang bersifat temporal (Arwansyah, 2024). Dengan kemajuan teknologi machine learning dan deep learning, metode analisis deret waktu tidak hanya terbatas pada model klasik seperti ARIMA, tetapi juga berkembang ke model modern seperti LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), dan Temporal Convolutional Networks.

Oleh karena itu, memahami dasar-dasar Time Series Analysis menjadi bekal penting bagi seorang data scientist untuk dapat mengolah, memahami, dan memanfaatkan data berbasis waktu secara efektif dalam mendukung pengambilan keputusan yang lebih baik.

BAB X

Maximum-Likehood Estimation And Optimization

Estimasi kemungkinan maksimum atau Maximum-Likelihood Estimation (MLE) adalah metode statistik untuk mengestimasi parameter dari suatu model statistik. Ide dasarnya adalah menemukan nilai parameter yang paling mungkin (maksimum) menghasilkan data yang telah diobservasi. Metode ini telah menjadi salah satu pilar utama dalam bidang ekonometri, statistik, dan pembelajaran mesin.

> *"Maximum-likelihood estimation (MLE) is a method for estimating the parameters of a statistical model given data. The method finds the parameters that maximize the likelihood function, which is the joint probability density function of the observed data as a function of the parameters."(Greene, 2012)*

A. Prinsip Dasar

Prinsip dasar MLE adalah memilih parameter yang membuat data sampel kita paling mungkin terjadi. Misalkan kita memiliki data sampel $X_1, X_2, \ldots, X_n$ yang diasumsikan independen dan identik terdistribusi (i.i.d) dari

suatu distribusi probabilitas dengan parameter θ. Fungsi likelihood, L(θ|x), adalah produk dari fungsi densitas probabilitas (PDF) atau fungsi massa probabilitas (PMF) dari setiap titik data:

$$L(\theta|x)=f(X_1|\theta)\times f(X_2|\theta)\times\cdots\times f(X_n|\theta)$$

Mengoptimalkan fungsi log-likelihood setara dengan mengoptimalkan fungsi likelihood aslinya karena fungsi logaritma adalah fungsi yang monotonik. Keuntungan menggunakan log-likelihood adalah mengubah produk menjadi penjumlahan, yang menyederhanakan proses diferensiasi dan perhitungan, serta membantu menghindari masalah numerik seperti underflow ketika produk dari banyak probabilitas kecil menjadi sangat dekat dengan nol (Casella, G., & Berger, 2020).

Tujuan MLE adalah menemukan θ^ML yang memaksimalkan fungsi log-likelihood:

$$\theta\hat{}ML=argmax\theta\ell(\theta|x)$$

Maximum-Likelihood Estimation (MLE) adalah cara yang sangat umum untuk membingkai sejumlah besar masalah dalam ilmu data. Intinya, proses ini melibatkan hal-hal berikut:

- Anda memiliki distribusi probabilitas yang dicirikan oleh beberapa parameter, yang disebut θ. Misalnya, dalam distribusi normal, θ terdiri dari dua angka: rata-rata dan standar deviasi.
- Anda mengasumsikan bahwa proses dunia nyata dijelaskan oleh distribusi probabilitas dari keluarga ini, tetapi Anda tidak membuat asumsi tentang θ.
- Anda memiliki kumpulan data yang disebut X, yang diambil dari proses dunia nyata.
- Anda kemudian mencari nilai θ yang memaksimalkan probabilitas P(X|θ).

Sebagian besar model klasifikasi dan regresi *machine learning* berada di bawah payung ini. Meskipun bentuk fungsionalnya sangat bervariasi,

BAB XI

Pemodelan Stokastik

Pemodelan stokastik (**stochastic modeling**) adalah pendekatan dalam ilmu probabilitas yang digunakan untuk mempelajari proses yang tidak sepenuhnya deterministik, melainkan dipengaruhi oleh unsur ketidakpastian dan keacakan. Berbeda dengan model deterministik yang selalu menghasilkan keluaran sama jika kondisi awalnya sama, pemodelan stokastik justru menekankan pada variasi kemungkinan hasil yang dapat terjadi. Secara sederhana, pemodelan stokastik tidak hanya mempelajari satu variabel acak, tetapi juga suatu proses yang berubah seiring waktu dan bergerak mengikuti aturan probabilistik tertentu.

Menurut (Cady, 2025) dalam bukunya The Data Science Handbook, Second Edition (2025):

> *"Stochastic modeling refers to a collection of advanced probability tools for studying not a single random variable, but, instead, a process that changes over time in a way that is partly random."*

Artinya, pemodelan stokastik berfokus pada analisis proses yang tidak hanya melibatkan satu variabel acak, melainkan suatu rangkaian peristiwa

yang berkembang secara bertahap, di mana sebagian dari perubahannya ditentukan oleh probabilitas.

Mengapa Pemodelan Stokastik Penting?

Pemodelan stokastik sangat penting karena memungkinkan kita untuk membuat keputusan yang lebih baik dalam menghadapi ketidakpastian. Dengan memahami rentang kemungkinan hasil dan probabilitasnya, kita dapat:

- Mengelola Risiko: Dalam keuangan, model stokastik digunakan untuk menilai risiko portofolio investasi. Dalam asuransi, mereka membantu perusahaan menghitung premi yang adil berdasarkan kemungkinan klaim di masa depan.
- Merencanakan Strategi: Dalam logistik dan manajemen rantai pasokan, model ini membantu mengoptimalkan inventaris dan jadwal pengiriman dengan mempertimbangkan fluktuasi permintaan yang acak.
- Membuat Kebijakan Publik: Dalam epidemiologi, model stokastik digunakan untuk memprediksi penyebaran wabah penyakit, membantu pemerintah merencanakan respons yang efektif.

"Proses stokastik adalah generalisasi dari variabel acak, di mana kita memiliki himpunan variabel acak yang terindeks oleh waktu." (Sheldon M. Ross, 2014). Pemodelan stokastik memiliki relevansi yang sangat luas, dari analisis pergerakan harga saham, dinamika pertumbuhan populasi, hingga pemodelan antrean di pusat layanan.

A. Konsep Dasar Dan Jenis Model Stokastik

Inti dari pemodelan stokastik adalah proses stokastik. Proses stokastik adalah kumpulan variabel acak yang diindeks oleh waktu atau ruang. Formalnya, sebuah proses stokastik didefinisikan sebagai $\{X_t, t \in T\}$, di mana X_t adalah variabel acak pada waktu t, dan T adalah himpunan indeks (seringkali waktu). Konsep ini memungkinkan kita untuk memodelkan

Daftar Pustaka

Aldisa, R. T., Maulana, P., & Abdullah, M. A. (2022). Penerapan Big Data Analytic Terhadap Strategi Pemasaran Job Portal di Indonesia dengan Karakteristik Big Data 5V. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *3*(3), 267. https://doi.org/10.30865/json.v3i3.3905

Analisis, P., Dan, K., Pada, R., & Dasar, S. (2021). *MANGGALI statistika dasar . Sedangkan pada mahasiswa . keseluruhan mahasiswa telah memiliki laptop*. *1*, 177–184.

Anggraini, D., & MA, D. (2025). Pengertian Statistik Dan Manfaat Statistik Dalam Kehidupan Sehari-Hari Understanding Statistics and the Benefits of Statistics in Everyday Life. *Jurnal Intelek Insan Cendekia (JIIC)*, *2*(5), 8767–8774.

Arwansyah. (2024). Model Prediksi Deret Waktu Menggunakan Deep Convolutional LSTM. *Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, *2024*(2), 21–25.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/

Brilliant, M., Handoko, D., & Sriyanto, S. (2017). Implementation of Data Mining using Association Rules for Transactional Data Analysis. *3rd International Conferences on Information Technology and Business (ICITB)*, 177–180.

Chyan, P., Gustiana, Z., Arni, S., Yasir, A., Husain, H., Dermawan, B. A., Oktarino, A., Indrayana, I. P. T., Siregar, A. M., Gormantara, A., Mumpuni, I. D., Prihatmono, M. W., Andisana, I. P. G. S., Possumah, L. M. A., Atho'illah, I., Aisyah, S., Santi, S., Setyoningrum, N. G., Farizy, S., & Afifah, V. (2024). Pengantar Data Science: Mengambil

Keputusan Berdasarkan Data. In *PT. Mifandi Mandiri Digital Redaksi*.

Collins, S. P., Storrow, A., Liu, D., Jenkins, C. A., Miller, K. F., Kampe, C., & Butler, J. (2021). *No Title 済無No Title No Title No Title*.

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning*. Wiley-Interscience.

Davenport, T. H., & Patil, D. J. (2012). *Data Scientist: The Sexiest Job of the 21st Century*. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64–73. https://doi.org/10.1145/2500499

Dkk., D. A. P. S. (2024). Jurnal Sistem dan Teknologi Informasi. *Dampak Negatif Perkembangan Teknologi Informasi Terhadap Pergaulan Bebas Pada Remaja Usia 15–20 Tahun*, *1*(2), 1–5.

Fadillaha, Y. Al, Akbarb, A. R., & Gusmanelic. (2024). Strategi Desain Pembelajaran Adaptif Untuk Meningkatkan Pengalaman Belajar di Era Digital. *Jurnal Pendidikan Sains Dan Teknologi Terapan*, *01*(04), 354–362.

Faiqoh, H. (2023). *Penerapan Ensemble Feature Selection untuk Mengurangi Dimensionalitas dalam Prediksi Data Time Series*. 1–84.

F. Cady, *The Data Science Handbook*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2025.

Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 20160112. https://doi.org/10.1098/rsta.2016.0112

Gunawan, A., Ningsih, S., & Lantana, D. A. (2023). Pengantar Basis Data. In *Gastronomía ecuatoriana y turismo local.* (Vol. 15, Issue 2).

Ir, P., & Tarumingkeng, R. C. (2025a). *Big Data dan AI : Sinergi dalam Era Digital*.

Ir, P., & Tarumingkeng, R. C. (2025b). *BIG DATA dan ANALITIK : Menyongsong Era Big Data*.

Jamco, J. C., Kondolembang, F., & Noya Van Delsen, M. S. (2023). Penanganan Multikolinearitas Pada Regresi Linier Berganda Menggunakan Regresi Lasso (Studi Kasus: Distribusi Presentase Produk Domestik Regional Bruto Di Provinsi Maluku Tahun 1999-2021). *PARAMETER: Jurnal Matematika, Statistika Dan Terapannya*, *2*(02), 145–154. https://doi.org/10.30598/parameterv2i02pp145-154

KandelSean, HeerJeffrey, PlaisantCatherine, KennedyJessie, HamFrank, van, Henry, R., WeaverChris, LeeBongshin, BrodbeckDominique, & BuonoPaolo. (2011). Research directions in data wrangling. *Information Visualization*. https://doi.org/10.1177/1473871611415994

K. Pykes, "What is Data Visualization? A Complete Guide to Tools, Techniques, and Best Practices," *DataCamp Blog*, Sept. 19, 2024.

Laksma Pradana, B. (2025). Time Series Forecasting of LQ45 Stock Index Using ARIMA: Insights and Implications. *Journal of Management, Accounting and Business Research (JMABR)*, *1*(1), 27–40. https://doi.org/10.51170/jmabr.v4i.1.160

Luluilmaknun, U. (2023). Kolaborasi dalam Pembelajaran: Belajar dari dan dengan Teman Sebaya. In *Membangun Landasan Matematika*.

Mahalani, A. J., & Rifai, N. A. K. (2022). Least Absolute Shrinkage and Selection Operator (LASSO) untuk Mengatasi Multikolinearitas pada Model Regresi Linear Berganda. *Bandung Conference Series: Statistics*, *2*(2), 119–125. https://doi.org/10.29313/bcss.v2i2.3438

Martias, L. D. (2021). STATISTIKA DESKRIPTIF–ukuran penyebaran data: simpangan rata rata, standar deviasi, jangkauan kuartil dan persentil. *Fihris: Jurnal Ilmu Perpustakaan Dan Informasi*, *16*(1), 40.

Mozin, S. Y., Abdullah, S., & Sawali, N. (2025). Pemanfaatan Teknologi Cerdas Untuk Pelayanan Publik: Study Tentang e-Government Dan

Smart City Berbasis ICT Big Data Dan AI. *JPS: Journal of Publicness Studies*, *2*(2), 117–130.

Muda Harahap, L., Gloria Pakpahan, T., Aulia Wijaya, R., & Zacky Nasution, A. (2024). Publikasi Ilmu Tanaman dan Agribisnis (BOTANI) Dampak Transformasi Digital pada Agribisnis: Tantangan dan Peluang bagi Petani di Indonesia. *Botani*, *1*(2), 99–108.

Nur, A., & Hura, B. K. (2024). Revolusi Logistik di Era Digital: Evaluasi Penggunaan Big Data di Industri Logistik. *Journal Of Informatics And Busisnes* , *2*(3), 443–453.

Nurfitri Imro'ah, I. S. N. N. D. (2020). Analisis Regresi Dengan Metode Least Absolute Shrinkage and Selection Operator (Lasso) Dalam Mengatasi Multikolinearitas. *Bimaster : Buletin Ilmiah Matematika, Statistika Dan Terapannya*, *9*(1), 31–38. https://doi.org/10.26418/bbimst.v9i1.38029

Putri, D. A., & Absharina, E. D. (2025). Eksplorasi Penerapan Teknologi Big Data Dalam Mendorong Inovasi Kesehatan Di Era Digital. *Simtek : Jurnal Sistem Informasi Dan Teknik Komputer*, *10*(1), 19–22. https://doi.org/10.51876/simtek.v10i1.1382

Provost, F., & Fawcet, T. (2013). *Data Science for Business*.

Rahman, O. P. (2025). Perkembangan Teknologi Komputer dan Implikasinya terhadap Masyarakat. *Jurnal BISPENTEK Nurul Hasanah*, *1*(1), 19–20.

Rahmawati, F., & Suratman, R. Y. (2022). Performa Regresi Ridge dan Regresi Lasso pada Data dengan Multikolinearitas. *Leibniz: Jurnal Matematika*, *2*(2), 1–10. https://doi.org/10.59632/leibniz.v2i2.176

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., … Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *Npj Digital Medicine*, *1*(1), 1–10. https://doi.org/10.1038/s41746-018-0029-1

Reyhan, M., Ahmad, D. R., Ramadhan, N. A., Hidayat N, R., & Kusumasari, I. R. (2024). Penggunaan Data Analisis dan Big Data dalam Strategi Pengambilan Keputusan Keuangan. *Jurnal Akuntansi, Manajemen, Dan Perencanaan Kebijakan*, *2*(2), 9. https://doi.org/10.47134/jampk.v2i2.540

Rismaninda Putri Dwi Prasetya, Azizah, R. N., Halwa, J. B. W., Nugroho, R. H., & Kusumasari, I. R. (2024). Implementasi Penggunaan Data Analytics untuk Mengoptimalkan Pengambilan Keputusan Bisnis di Era Digital. *Jurnal Bisnis Dan Komunikasi Digital*, *2*(2), 12. https://doi.org/10.47134/jbkd.v2i2.3459

Rohadi, D. (2024). PENGARUH ENSEMBLE FEATURE SELECTION PADA PREDIKSI DATA TIME SERIES MENGGUNAKAN GATED RECURRENT UNIT (GRU) DAN BIDIRECTIONAL LONG SHORT-TERM MEMORY (Bi-LSTM) (Studi Kasus: Walmart). *Repository.Uinjkt.Ac.Id*, *16*(1), 1–23.

Rudini, R. (2017). Peranan Statistika Dalam Penelitian Sosial Kuantitatif. *Jurnal SAINTEKOM*, *6*(2), 53. https://doi.org/10.33020/saintekom.v6i2.13

Safar Dwi Kurniawan, M. K., & Rosalina Yani Widiastuti, S.Kom., M. (1967). *Big Data: Mengenal Big Data dan Implementasinya di Berbagai Bidang*.

Salsabila, M. (2022). Pendekatan visual analytics dalam pemodelan prediksi cacat perangkat lunak menggunakan kombinasi pca dan smote. In *Repository.Uinjkt.Ac.Id*.

Sari, D., Islam Negeri Raden Intan Lampung Jalan Letnan Kolonel Jl Endro Suratmin, U. H., Sukarame, K., & Bandar Lampung, K. (2025). Transformasi Akuntansi Manajemen Lingkungan di Era Digital: Peluang dan Tantangan pada Sektor Energi Terbarukan. *Jurnal Ilmiah Ekonomi, Manajemen, Bisnis Dan Akuntansi*, *2*(1), 347–360.

Sari, E. P., Mustamin, S. B., Atnang, M., Sahriani, & Fajar, N. (2024). Studi Literatur Deep Learning dan Machine Learning untuk Analisis dan Prediksi Pasar Saham: Metodologi, Representasi Data, dan Studi

Kasus. *Jurnal Teknologi Dan Sains Modern*, *1*(1), 19–28. https://doi.org/10.69930/jtsm.v1i1.59

Siddik, A., & Yansyah, E. A. (2025). Penerapan Statistik dalam Penelitian Ilmiah: Metode dan Tantangan Application of Statistics in Scientific Research : Methods and Challenges. *JIIC: Journal Intelek Insan Cendekia*, *2*(5), 8759–8762.

Tarumingkeng, R. C. (2025). *Machine Learning dalam Data Science: Teknik dan Kasus*. 62.

Utami, M. R. (2025). Analisis Literature Prediksi Tren Penjualan E-Commerce Berbasis Data Time-Series: Metode Statistik & Machine Learning. *Jurnal Ilmiah Penelitian Mahasiswa*, *3*(2), 331–339.

VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Online Learning. https://www.oreilly.com/library/view/python-data-science/9781491912126/

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*, 1–23. https://doi.org/10.18637/jss.v059.i10

Wibowo, A. (2025). Pengantar AI, Big Data dan Ilmu Data. In *Penerbit Yayasan Prima Agus Teknik*.

Yanke, A., Zendrato, N. E., & Soleh, A. M. (2022). "Handling Multicollinearity Problems in Indonesia's EconomicGrowth Regression Modeling Based on Endogenous EconomicGrowth Theory"Penanganan Masalah Multikolinieritas pada Pemodelan PertumbuhanEkonomi Indonesia Berdasarkan Teori Pertumbuhan EkonomiEndogenous. *Indonesian Journal of Statistics and Its Applications*, *6*(2), 228–244.

Yusapra Salim, A., Yuliani, M., Andayani Komara, M., & Sri Wahyuni, R. (2024). Analisis Deret Waktu Data Perencanaan Tenaga Kerja pada Perusahaan Manufaktur Menggunakan Model ARIMA. *Jurnal Teknologika*, *14*(2), 481–492. https://doi.org/10.51132/teknologika.v14i2.420

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, S., & Stoica, I. (2016). Apache Spark: A unified engine for big data processing. *Communications of the ACM*, *59*(11), 56–65. https://doi.org/10.1145/2934664

BUKU AJAR

PENGANTAR DATA SAINS

Dalam konteks perkembangan digital dewasa ini, istilah data science sering kali muncul sebagai salah satu konsep yang paling sentral sekaligus strategis. Hal ini tidak terlepas dari kenyataan bahwa dunia modern ditandai oleh percepatan eksponensial dalam produksi data yang bersumber dari beragam aktivitas manusia, baik dalam ruang personal, sosial, maupun profesional.

Secara konseptual, data science dapat dipandang sebagai disiplin ilmu multidimensi yang menggabungkan berbagai metode, pendekatan, serta kerangka kerja dari statistika, ilmu komputer, dan pemahaman domain tertentu untuk menghasilkan pengetahuan yang relevan. Sains data merupakan bidang yang sangat relevan dengan perkembangan teknologi saat ini, di mana pengolahan dan analisis data menjadi kunci penting dalam pengambilan keputusan. Melalui buku ajar ini, penulis berusaha menyajikan materi yang mencakup teori, metode, serta contoh penerapan yang diharapkan dapat mendukung proses pembelajaran, penelitian, maupun praktik di lapangan.

Tujuan penulisan dan adanya buku ini adalah untuk memberikan pemahaman dasar mengenai konsep, metode, serta ruang lingkup sains data, sekaligus menjadi panduan pembelajaran bagi mahasiswa dalam mengikuti perkuliahan maupun praktik di bidang terkait. Selain itu, buku ini diharapkan dapat menjadi referensi bagi dosen, peneliti, dan praktisi dalam mengembangkan serta menerapkan metode analisis data, membantu pembaca memahami penerapan sains data di berbagai bidang seperti bisnis, kesehatan, pemerintahan, maupun teknologi, serta mendorong terciptanya budaya pengambilan keputusan berbasis data (data-driven decision making) baik di lingkungan akademik maupun industri.

litnus

literasinusantaraofficial@gmail.com
www.penerbitlitnus.co.id
Literasi Nusantara
literasinusantara_
085755971589

Teknik +17

ISBN 978-634-234-884-0
9 786342 348840