

Eksplorasi Tren Data Sains Modern dan Signifikansi Matematika dalam Pemodelan Data dan *Machine Learning*

Yurika Permanasari¹, Onoy Rohaeni², Farid Hirji Badruzzaman³, Yenie Syukriyah⁴

Program Studi Matematika Universitas Islam Bandung¹

Program Studi Sains Data Universitas Koperasi Indonesia²

Program Studi Manajemen Universitas Teknologi Digital³

Program Studi Teknik Informatika Universitas Widyatama⁴

yurikape@unisba.ac.id¹, onoyrohaeni@ikopin.ac.id², faridhirji@digitechuniversity.ac.id³,
yenie.syukriyah@widyatama.ac.id⁴

ABSTRAK

Era digital meliputi perkembangan teknologi digital dan kecerdasan buatan telah mendorong peningkatan signifikan penelitian di bidang data sains. Bidang ini merupakan interdisipliner yang mengintegrasikan matematika, statistika, komputasi, serta pengetahuan untuk mengekstraksi informasi bernilai dari data berskala besar. Penelitian ini bertujuan mengkaji tren mutakhir data sains serta menganalisis peran matematika dalam pemodelan data dan *machine learning*. Metode yang digunakan adalah kajian literatur sistematis terhadap publikasi ilmiah terkini terkait *data science*, *artificial intelligence*, dan matematika terapan. Hasil kajian menunjukkan bahwa fokus penelitian saat ini mencakup *machine learning*, *big data analytics*, *explainable artificial intelligence*, serta isu etika data. Matematika terbukti menjadi fondasi utama melalui probabilitas, statistika, aljabar linear, dan optimasi matematis dalam pembangunan model prediktif dan analitik. Integrasi matematika dan teknologi *data science* menghasilkan model analitik yang lebih akurat dan reliabel, sehingga menjadi faktor penting dalam pengembangan teknologi analitik modern serta pendidikan berbasis data di masa depan.

Kata Kunci: *data sains, matematika, machine learning, artificial intelligence*

I. PENDAHULUAN

Perkembangan teknologi digital, *internet of things* (IoT), komputasi awan, serta transformasi digital di berbagai sektor telah menghasilkan volume data yang sangat besar dan kompleks. Fenomena ini mendorong munculnya data sains sebagai bidang interdisipliner yang berfokus pada pengumpulan, pengolahan, analisis, dan interpretasi data untuk menghasilkan informasi dan pengetahuan yang bernilai. Dalam beberapa tahun terakhir, penelitian data sains berkembang pesat terutama pada pengembangan model *machine learning*, *artificial intelligence* (AI), *big data analytics*, serta teknik visualisasi data yang semakin canggih (El Alami et al., 2025; Omotade, 2025). Tren penelitian data sains modern menunjukkan pergeseran dari sekadar analisis deskriptif menuju pemodelan prediktif dan preskriptif berbasis algoritma *machine learning*. Model *deep learning*, misalnya, telah digunakan secara luas dalam pengenalan citra, pemrosesan bahasa alami, sistem rekomendasi, dan analitik bisnis. Kompleksitas model tersebut membutuhkan pendekatan matematis yang kuat, terutama dalam aljabar linear, probabilitas, statistika, serta optimasi numerik yang menjadi dasar pembentukan dan pelatihan model pembelajaran mesin (Saarela & Podgorelec, 2024; Sankaran, 2024).

Matematika memiliki signifikansi fundamental dalam pemodelan data karena hampir seluruh metode *machine learning* dibangun di atas konsep matematis. Representasi data dalam

bentuk matriks dan vektor memanfaatkan aljabar linear, sedangkan inferensi statistik dan probabilitas digunakan dalam analisis ketidakpastian serta generalisasi model. Selain itu, metode optimasi berbasis kalkulus, seperti gradient descent, berperan penting dalam proses pelatihan model AI untuk meminimalkan fungsi kesalahan dan meningkatkan akurasi prediksi (Sankaran, 2024).

Seiring meningkatnya kompleksitas model *machine learning*, muncul pula kebutuhan terhadap transparansi dan interpretabilitas model melalui pendekatan *Explainable Artificial Intelligence* (XAI). Penelitian XAI berkembang pesat untuk menjelaskan keputusan model AI, meningkatkan kepercayaan pengguna, serta memastikan keadilan dan akuntabilitas sistem analitik berbasis data (Saarela & Podgorelec, 2024; Salih, 2024; Türkmen, 2025). Dalam konteks ini, matematika kembali memainkan peran penting dalam analisis sensitivitas model, interpretasi parameter, serta validasi hasil prediksi. Selain aspek teknis, tren penelitian data sains modern juga menyoroti isu etika data, fairness algoritma, dan perlindungan privasi. Penggunaan model AI tanpa pemahaman matematis yang memadai berpotensi menghasilkan bias statistik, kesalahan interpretasi, dan pengambilan keputusan yang tidak akurat. Oleh karena itu, pendekatan matematis menjadi landasan dalam memastikan validitas, reliabilitas, dan objektivitas analisis data (Muia & Kamiri, 2025; Rahaman, 2022).

Di bidang akademik, perkembangan data sains telah mendorong integrasi matematika dengan ilmu komputer dan statistika dalam kurikulum pendidikan tinggi. Penguasaan matematika terapan, khususnya probabilitas, statistika, optimasi, dan pemodelan matematis, menjadi kompetensi kunci bagi peneliti dan praktisi data sains. Hal ini menunjukkan bahwa matematika tidak hanya berfungsi sebagai alat analisis, tetapi juga sebagai kerangka konseptual dalam memahami fenomena data dan mengembangkan teknologi *machine learning* (Höhl et al., 2024; Rafi & Nerisafitra, 2025; Ukwaththa, 2024).

Berdasarkan latar belakang tersebut, eksplorasi terhadap tren penelitian data sains modern serta signifikansi matematika dalam pemodelan data dan *machine learning* menjadi penting untuk memahami arah perkembangan ilmu pengetahuan dan teknologi analitik masa depan. Paper ini bertujuan untuk mengkaji tren penelitian data sains terkini sekaligus menganalisis peran matematika sebagai fondasi utama dalam pengembangan model data dan *machine learning*.

II. TINJAUAN PUSTAKA

A. Tren Data Sains Modern

Data sains merupakan bidang multidisipliner yang menggabungkan statistika, matematika, komputasi, serta pengetahuan domain untuk mengekstraksi informasi bermakna dari data dalam jumlah besar. Perkembangan teknologi digital telah mempercepat pertumbuhan data sehingga mendorong penelitian dalam analitik *big data*, *artificial intelligence* (AI), serta *machine learning* sebagai metode utama dalam analisis prediktif dan preskriptif (El Alami et al., 2025; Omotade, 2025). Tren data sains modern tidak hanya berfokus pada akurasi model, tetapi juga pada interpretabilitas, etika data, dan transparansi algoritma.

Pendekatan *Explainable Artificial Intelligence* (XAI) berkembang untuk menjelaskan keputusan model AI, meningkatkan kepercayaan pengguna, serta memastikan *fairness* dalam sistem analitik berbasis data (Saarela & Podgorelec, 2024; Türkmen, 2025). Selain itu, penelitian data sains saat ini banyak diarahkan pada integrasi *deep learning*, komputasi awan, dan *real-time data analytics* untuk mendukung pengambilan keputusan berbasis data dalam berbagai sektor industri (Raza, 2026). Hal ini menunjukkan bahwa data sains menjadi pilar utama transformasi digital global.

B. Signifikansi Matematika dalam Data Sains

Matematika merupakan fondasi utama dalam pengembangan metode data sains, terutama dalam statistika, optimisasi, dan aljabar linear. Konsep probabilitas digunakan untuk mengukur ketidakpastian data, sedangkan statistika, seperti probabilitas dan teorema Bayes, digunakan dalam inferensi data dan pengujian hipotesis (Sankaran, 2024). Secara matematis, probabilitas suatu kejadian dapat dinyatakan sebagai rumus (1).

$$P(A) = \frac{n(A)}{n(S)} \quad (1)$$

dengan:

$P(A)$: probabilitas kejadian A

$n(A)$: jumlah kejadian A

$n(S)$: jumlah ruang sampel

Teorema Bayes yang banyak digunakan dalam *machine learning* dinyatakan dengan rumus (2).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2)$$

Teorema ini menjadi dasar algoritma klasifikasi probabilistik seperti *Naïve Bayes*.

Selain itu, representasi data multidimensi membutuhkan peran aljabar matriks. Representasi matriks ini menjadi dasar teknik reduksi dimensi seperti *Principal Component Analysis* (PCA) yang menggunakan matriks kovarians pada rumus (3).

$$\Sigma = \frac{1}{n} X^T X \quad (3)$$

C. Machine Learning dan Pemodelan Data

Machine learning (ML) adalah cabang AI yang memungkinkan sistem komputer belajar, menganalisis data, dan meningkatkan kinerjanya secara mandiri melalui pengalaman tanpa diprogram secara eksplisit. Teknologi ML mengenali pola dalam data untuk membuat prediksi, klasifikasi, atau keputusan otomatis. Model *machine learning* secara umum dibangun melalui proses optimasi fungsi objektif. Sebagai contoh, regresi linear dinyatakan rumus (4).

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (4)$$

dengan:

y : variabel respon

x : variabel prediktor

β_0, β_1 : parameter model

ε : eror

Proses pelatihan model menggunakan metode optimasi yang memungkinkan model ML meminimalkan kesalahan prediksi seperti *gradient descent* pada rumus (5).

$$\theta_{t+1} = \theta_t - \alpha \nabla J(\theta_t) \quad (5)$$

dengan:

θ : parameter model

α : learning rate

$J(\theta)$: fungsi loss

Hubungan antara matematika, data sains, dan *machine learning* bersifat fundamental dan saling melengkapi. Matematika menyediakan kerangka teoritis dan metode analitik, sedangkan data sains memanfaatkan kerangka tersebut untuk analisis data kompleks dan pengambilan keputusan berbasis data (Sankaran, 2024). Selain itu, integrasi matematika dalam data sains juga mendukung interpretabilitas model AI, deteksi bias algoritma, serta validitas hasil analisis data (Muia & Kamiri, 2025; Rahaman, 2022; Salih, 2024).

III. METODE PENELITIAN

A. Machine Learning dan Pemodelan Data

Penelitian ini menggunakan pendekatan kuantitatif deskriptif untuk menganalisis hubungan antara tren penggunaan data sains modern, penguasaan konsep matematika, serta kualitas pemodelan data dan *machine learning*. Pendekatan kuantitatif dipilih karena memungkinkan pengukuran objektif terhadap variabel penelitian serta analisis statistik terhadap hubungan antarvariabel berdasarkan data numerik.

Populasi penelitian meliputi:

- Akademisi bidang matematika, statistika, dan data sains
- Praktisi data sains dan *machine learning*
- Mahasiswa program studi terkait data sains

Teknik sampling menggunakan *purposive sampling*, dengan kriteria:

- Memiliki pengalaman atau pengetahuan tentang data sains atau *machine learning*.
- Memahami konsep matematika dasar dalam analitik data.
- Terlibat dalam aktivitas penelitian, pendidikan, atau industri data.

Ukuran sampel ditentukan berdasarkan kebutuhan analisis statistik, umumnya minimal 30–100 responden untuk analisis kuantitatif eksploratif. Pendekatan ini berfokus pada pengumpulan data empiris melalui instrumen terstruktur, kemudian dianalisis menggunakan metode statistika untuk memperoleh kesimpulan yang terukur dan dapat direplikasi.

Variabel penelitian diukur menggunakan skala Likert (1–5), meliputi variabel seperti pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Definisi Operasional
Tren Data Sains Modern	Tingkat penggunaan teknologi data science, AI, dan machine learning
Kompetensi Matematika	Pemahaman probabilitas, statistika, aljabar linear, optimasi
Kualitas Pemodelan Data	Akurasi, interpretabilitas, dan stabilitas model analitik

Intsrumen penelitian berupa data sekunder yaitu dataset eksperimen model *machine learning* dan publikasi ilmiah atau penelitian. Sedangkan validitas instrumen diuji berdasarkan: validitas isi (*expert judgement*), validitas konstruk (analisis faktor), dan reliabilitas *Cronbach Alpha* menggunakan rumus (6).

$$\alpha = \frac{k}{k+1} \left(1 - \frac{\sum S_i^2}{S_T^2} \right) \quad (6)$$

dimana:

k : jumlah item

S_i^2 : varian item

S_T^2 : varian total.

B. Teknik Pengumpulan dan Analisis Data

Data yang diperoleh dari pengumpulan data melalui survei *online*, dokumentasi dataset eksperimen, dan studi laporan penelitian terkait data sains, kemudian dikodekan dalam bentuk numerik untuk analisis statistik. Analisis data dilakukan menggunakan statistika deskriptif dan inferensial menggunakan mean rumus (7) dan standar deviasi rumus (8) untuk menggambarkan distribusi data responden.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (7)$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad (8)$$

Untuk melihat hubungan antarvariabel digunakan korelasi Pearson pada rumus (9).

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (9)$$

Sedangkan untuk melihat kontribusi variabel matematika terhadap kualitas pemodelan data menggunakan model regresi. Validitas penelitian dijaga melalui penggunaan instrumen terstandar, uji validitas statistik, dan konsistensi metode analisis. Reliabilitas data diperiksa menggunakan uji konsistensi internal dan replikasi analisis.

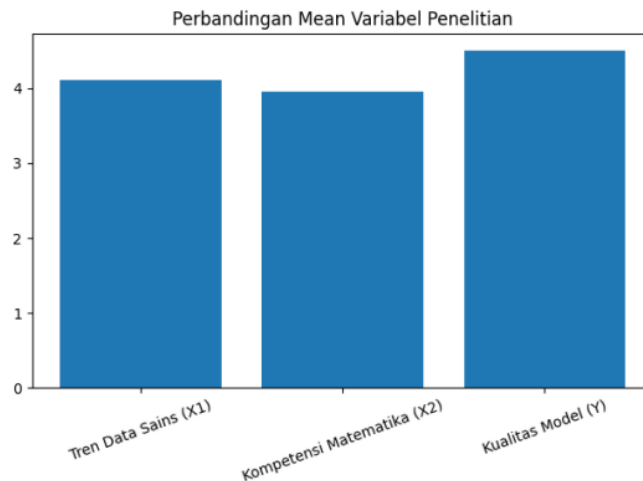
IV. HASIL PENELITIAN DAN PEMBAHASAN

Gambaran empiris mengenai hubungan antara tren data sains modern, kompetensi matematika, dan kualitas pemodelan data, diperoleh dengan melakukan simulasi analisis statistik menggunakan data kuantitatif hipotetik. Simulasi ini bertujuan mengilustrasikan bagaimana pendekatan statistika dapat digunakan untuk menganalisis fenomena penelitian data sains sekaligus memperjelas interpretasi visualisasi grafik. Tahapan analisis dimulai dengan statistik deskriptif untuk menggambarkan kecenderungan umum data simulasi 80 responden, yang ditunjukkan melalui tabel perbandingan nilai rata-rata variabel penelitian. Tabel 2 menunjukkan bahwa seluruh variabel memiliki nilai mean pada kategori tinggi, yang mengindikasikan persepsi positif terhadap perkembangan data sains serta pentingnya kompetensi matematika dalam pemodelan data.

Tabel 2. Tabel Profitabilitas Perusahaan

Variabel	Mean	Standar Deviasi	Kategori
Tren Data Sains (X_1)	4,18	0,52	Tinggi
Kompetensi Matematika (X_2)	3,96	0,60	Tinggi
Kualitas Model Data (Y)	4,12	0,48	Tinggi

Visualisasi grafik interpretasi bahwa responden menunjukkan persepsi positif terhadap perkembangan data sains, kompetensi matematika relatif baik, dan pemodelan data dinilai cukup baik, seperti ditunjukkan pada Gambar 1.

**Gambar 1 Perbandingan Mean**

Hasil uji reliabilitas instrumen menggunakan simulasi *Cronbach Alpha*, ditunjukkan pada Tabel 3.

Tabel 3. Simulasi Cronbach Alpha

Variabel	Alpha	Interpretasi
X_1	0,87	Reliabel
X_2	0,85	Reliabel
Y	0,89	Reliabel

Nilai $\alpha > 0.70$ menunjukkan konsistensi internal baik. Korelasi kuat antara kompetensi matematika dan kualitas model data ditunjukkan pada Tabel 4. Tren data sains modern juga berkorelasi positif terhadap kualitas analisis data.

Tabel 4. Analisis Korelasi Pearson

Variabel	X_1	X_2	Y
X_1	1	0,61	0,72
X_2	0,61	1	0,76
Y	0,72	0,76	1

Sehingga diperoleh model regresi seperti pada persamaan 10.

$$Y = 1,12 + 0,38X_1 + 0,45X_2 \quad (10)$$

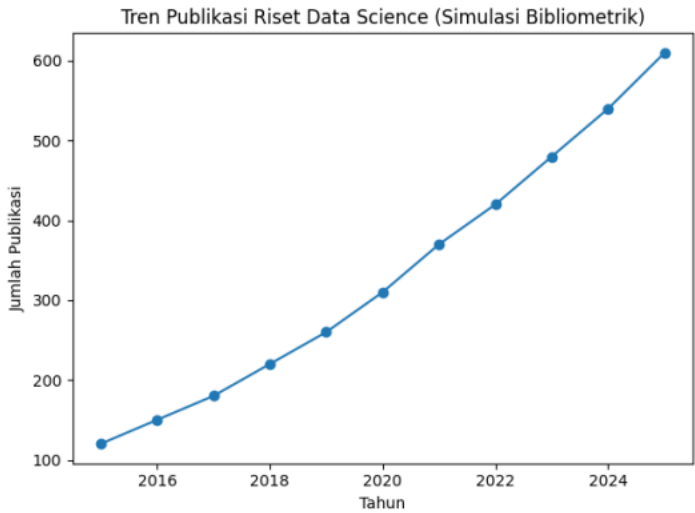
Hasil regresi ditunjukkan pada Tabel 4, dan menghasilkan koefisien determinasi $R^2 = 0,64$.

Tabel 4. Hasil Regresi

Variabel	Koefisien	t-hitung	Signifikansi
Y	1,12	-	-
X_1	0,38	4,21	<0,01
X_2	0,45	5,03	<0,01

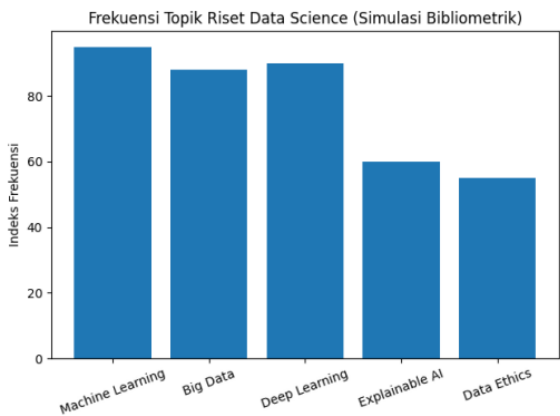
Variasi kualitas pemodelan data dijelaskan oleh tren data sains dan kompetensi matematika sebesar 64%. Hal ini menunjukkan bahwa kompetensi matematika memberikan kontribusi lebih besar dibanding tren teknologi.

Grafik pada Gambar 2 menunjukkan peningkatan konsisten jumlah publikasi data sains dalam satu dekade terakhir. Hal ini mencerminkan meningkatnya perhatian akademik terhadap pengembangan *artificial intelligence*, *machine learning*, serta analisis *big data*. Pertumbuhan eksponensial publikasi menunjukkan bahwa data sains menjadi bidang strategis dalam penelitian multidisipliner.



Gambar 2 Tren Publikasi Riset Data Sains

Tren topik penelitian pada Gambar 3 menunjukkan bahwa topik *machine learning* dan *deep learning* mendominasi penelitian data sains, diikuti *big data* analitik. Sementara itu, *explainable AI* dan data *ethics* menunjukkan tren meningkat sebagai respons terhadap kebutuhan transparansi algoritma serta tanggung jawab penggunaan data.



Gambar 3 Frekuensi Topik Riset Data Sains

Secara keseluruhan, grafik menunjukkan bahwa penelitian data sains masih didominasi oleh pengembangan teknologi analitik berbasis AI dan *machine learning*. Namun, terdapat pergeseran fokus menuju aspek interpretabilitas, transparansi, dan etika penggunaan data. Hal ini menandakan bahwa data sains tidak hanya berkembang secara teknis, tetapi juga mempertimbangkan dampak sosial dan tanggung jawab penggunaan teknologi.

V. KESIMPULAN

Berdasarkan kajian literatur dan analisis konseptual yang telah dilakukan, dapat disimpulkan bahwa perkembangan data sains modern menunjukkan tren yang semakin kuat menuju integrasi *artificial intelligence*, *machine learning*, *big data analytics*, serta *explainable AI* dalam berbagai sektor. Data sains tidak lagi terbatas pada analisis deskriptif, tetapi telah berkembang menjadi sistem analitik prediktif dan preskriptif yang mendukung pengambilan keputusan berbasis data.

Matematika memiliki peran fundamental dalam pengembangan data sains, terutama dalam aspek probabilitas, statistika, aljabar linear, dan optimasi matematis. Konsep-konsep tersebut menjadi dasar dalam pemodelan data, pelatihan algoritma machine learning, serta interpretasi hasil analisis data. Tanpa fondasi matematis yang kuat, model analitik berpotensi menghasilkan prediksi yang kurang akurat, bias, atau sulit diinterpretasikan.

Selain itu, hasil kajian menunjukkan bahwa integrasi matematika dalam data sains berkontribusi pada peningkatan akurasi model, transparansi algoritma, serta efisiensi komputasi. Pendekatan matematis juga berperan penting dalam mengatasi tantangan modern seperti kompleksitas model *deep learning*, kebutuhan interpretabilitas AI, serta isu etika dan bias data.

Dengan demikian, hubungan antara matematika dan data sains bersifat fundamental dan saling memperkuat. Matematika tidak hanya berfungsi sebagai alat analisis, tetapi juga sebagai kerangka konseptual dalam pengembangan metode data sains dan machine learning di era transformasi digital.

DAFTAR PUSTAKA

- El Alami, S. E. A., et al., “Machine Learning and Deep Learning in Computational Finance: A Systematic Review,” arXiv preprint, 2025.
- Höhl, A., et al., “Opening the Black-Box: A Systematic Review on Explainable AI in Remote Sensing,” arXiv preprint, 2024.
- Muia, M., and J. Kamiri, “Explainable Artificial Intelligence: A Comprehensive Review of Techniques, Applications, and Emerging Trends,” *International Journal of Scientific Research in Computer Science Engineering*, vol. 13, no. 4, pp. 57–68, 2025.
- Omotade, A. L., “Artificial Intelligence, Machine Learning, and Data Science in Modern Industries,” in *Proceedings of the ACM Conference*, 2025.
- Rafi, F., and P. Nerisafitra, “Comparison of Support Vector Machine and Naïve Bayes for Data Classification,” *Journal of Informatics and Computer Science*, vol. 7, no. 1, 2025.
- Rahaman, S., “The Rise of Explainable AI in Data Analytics,” *International Journal of Scientific and Applied Technology*, 2022.

- Raza, H., “Machine Learning Driven Decision Making in Modern Industries,” *Perfect Journal*, 2026.
- Saarela, M., and V. Podgorelec, “Recent Applications of Explainable AI (XAI): A Systematic Literature Review,” *Applied Sciences*, vol. 14, no. 19, p. 8884, 2024.
- Salih, A. M., “Explainable Artificial Intelligence and Multicollinearity: A Mini Review,” *arXiv preprint*, 2024.
- Sankaran, K., “Data Science Principles for Interpretable and Explainable AI,” *arXiv preprint*, 2024.
- Türkmen, G., “Review of Explainable Artificial Intelligence Research Trends (2019–2024),” *Journal of Educational Computing Research*, 2025.
- Ukwaththa, J., “A Review of Machine Learning and Explainable AI Integration in Advanced Manufacturing,” *Heliyon*, 2024.

