



Pengantar Data Sains

#1 Meeting

Overview Data Sains

Ferdian Bangkit Wijaya, S.Stat., M.Si

NIP. 199005202024061001

Dunia Kita Dibanjiri Data



"Without data, you're just another person with an opinion." - W. Edwards Deming

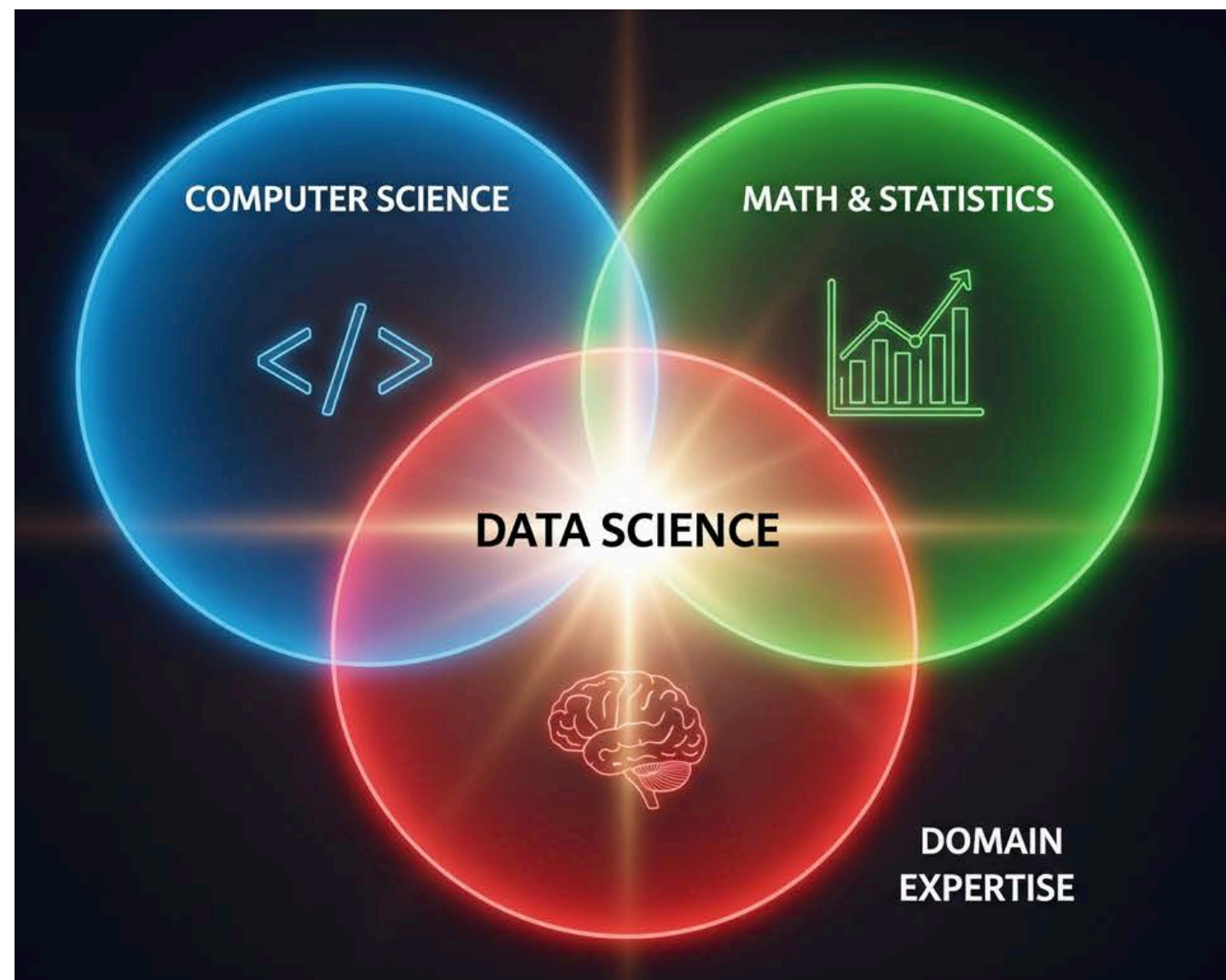
Ledakan Data (Data Deluge):

- 2.5 Quintillion bytes data dibuat setiap hari.
- Pada tahun 2025, diperkirakan total data global akan mencapai 175 zettabytes.
- Sumber data: Media sosial, transaksi e-commerce, sensor IoT (Internet of Things), GPS, data medis, dan banyak lagi.

Bagaimana kita bisa memanfaatkan lautan data ini?

Reinsel, D., Gantz, J., & Rydning, J. (2018). The Digitization of the World – From Edge to Core. IDC White Paper.

Apa Itu Data Sains?



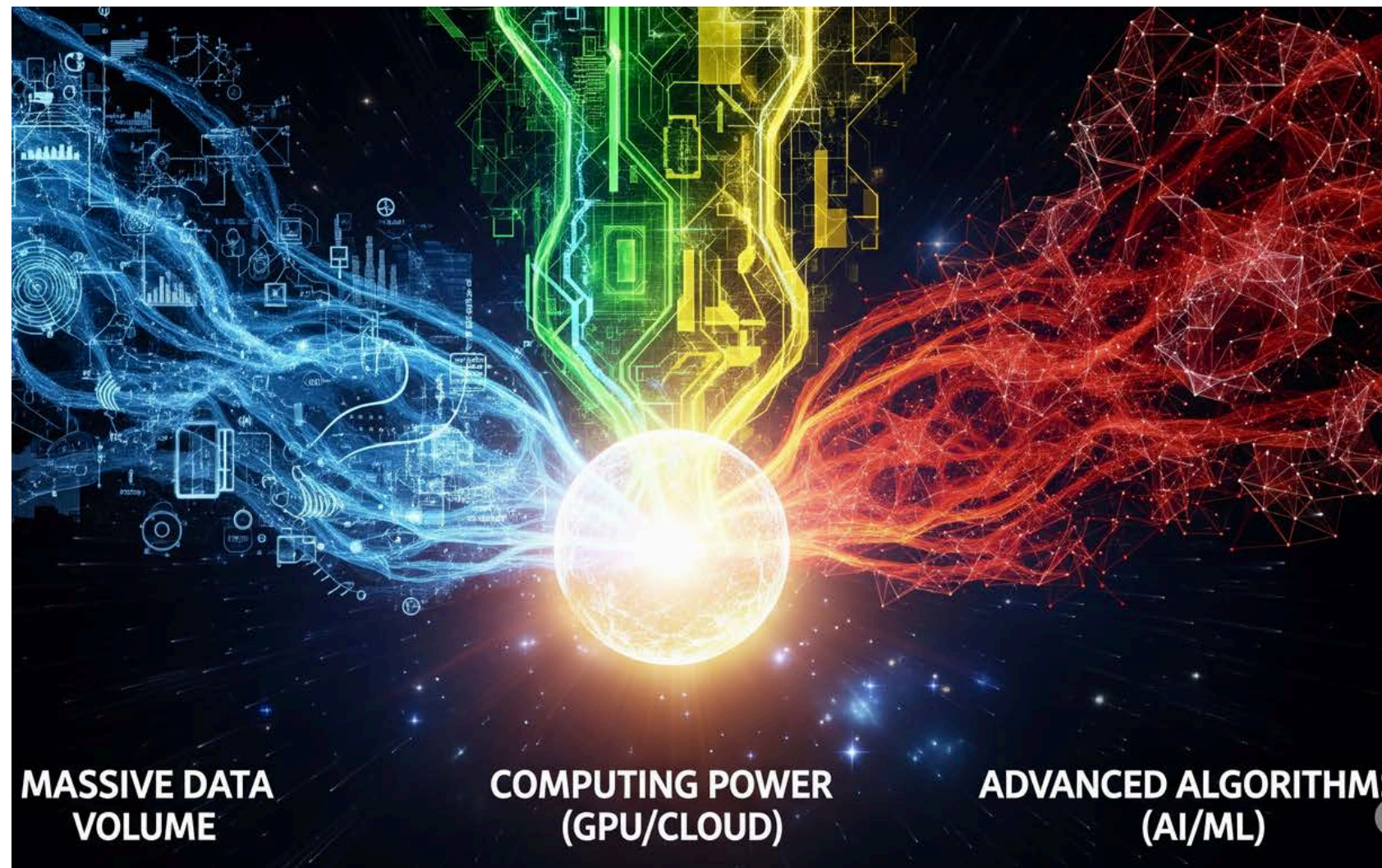
Data Sains adalah bidang interdisipliner yang menggunakan metode, proses, algoritma, dan sistem ilmiah untuk mengekstrak pengetahuan (knowledge) dan wawasan (insights) dari data terstruktur maupun tidak terstruktur.

Ini bukan hanya tentang statistik atau ilmu komputer, ini adalah gabungan dari ketiganya:

- Ilmu Komputer & TI: Hacking skills, rekayasa perangkat lunak, manajemen basis data.
- Matematika & Statistika: Pembelajaran mesin (machine learning), pemodelan statistik.
- Keahlian Domain/Bisnis: Pengetahuan konteks industri (misal: keuangan, kesehatan, marketing) untuk mengajukan pertanyaan yang tepat dan menginterpretasikan hasil.

Dhar, V. (2013). "Data science and prediction." *Communications of the ACM*, 56(12), 64-73

Mengapa Data Sains Penting?



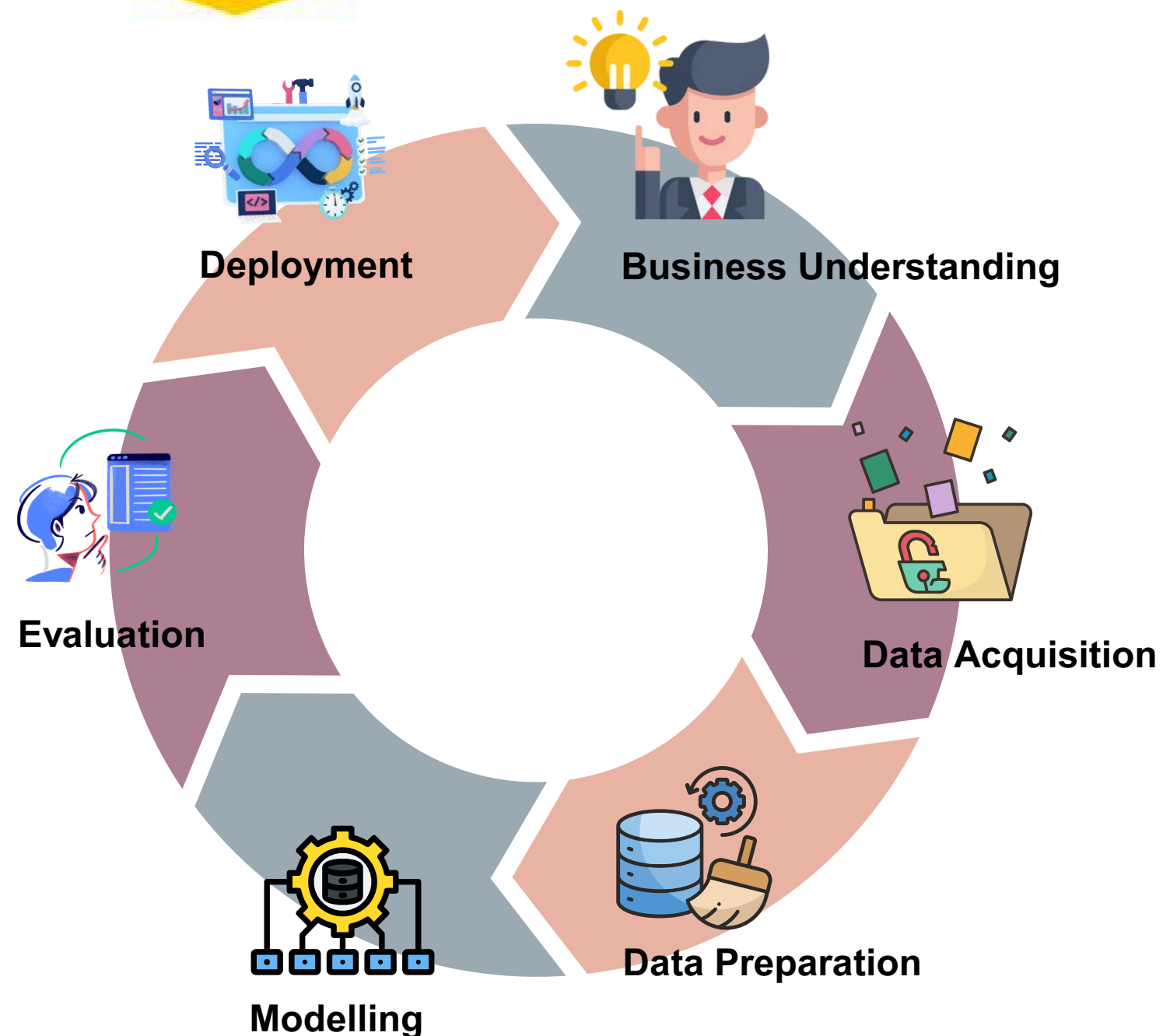
“The Perfect Storm”

Tiga Pendorong Utama (The "Perfect Storm"):

1. Volume Data yang Masif (Big Data): Sensor, media sosial, dan perangkat digital menghasilkan data lebih banyak dari sebelumnya (prinsip Volume, Velocity, Variety).
2. Peningkatan Kekuatan Komputasi: Biaya penyimpanan dan pemrosesan data menurun drastis. Cloud computing membuat komputasi canggih dapat diakses oleh semua orang.
3. Algoritma yang Semakin Canggih: Terobosan dalam Machine Learning dan Deep Learning memungkinkan kita untuk memecahkan masalah yang sebelumnya dianggap mustahil.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety.

Lifecycle Project



Tahapan Umum Model CRISP-DM (Cross-Industry Standard Process for Data Mining):

1. Pemahaman Bisnis/Masalah (Business Understanding): Apa masalah yang ingin kita selesaikan? Apa tujuan akhirnya?
2. Pengumpulan & Pemahaman Data (Data Acquisition & Understanding): Di mana data berada? Bagaimana cara mendapatkannya? Apakah data ini relevan?
3. Persiapan Data (Data Preparation): Membersihkan data (menangani nilai yang hilang, duplikat), transformasi, dan rekayasa fitur (feature engineering). (Seringkali memakan 80% waktu proyek!)
4. Pemodelan (Modeling): Memilih dan menerapkan algoritma machine learning atau statistik.
5. Evaluasi (Evaluation): Apakah model bekerja dengan baik? Apakah menjawab pertanyaan bisnis di awal?
6. Penerapan (Deployment): Mengintegrasikan model ke dalam sistem produksi atau menyajikan wawasan kepada pemangku kepentingan.

Chapman, P., et al. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc

Peran-Peran di Dunia Data

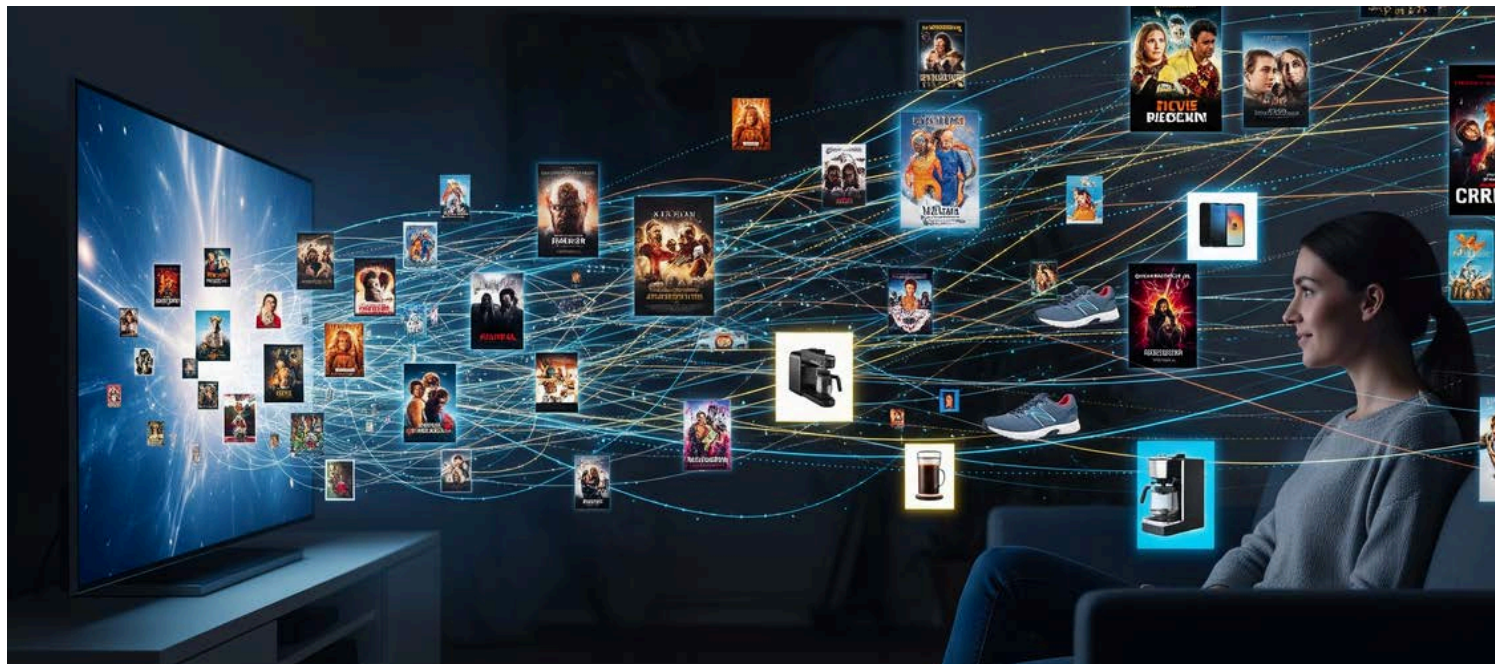


Peran Utama:

- **Data Analyst:** Fokus pada masa lalu. Menganalisis data historis untuk menjawab pertanyaan "apa yang terjadi?" dan "mengapa?". Menggunakan SQL, Excel, dan alat visualisasi (Tableau, Power BI).
- **Data Scientist:** Fokus pada masa depan. Menggunakan statistik dan machine learning untuk membuat prediksi ("apa yang akan terjadi?"). Menggunakan Python/R, membangun model prediktif.
- **Data Engineer:** Fokus pada infrastruktur. Membangun dan memelihara pipa data (data pipelines) dan arsitektur untuk memastikan data mengalir dengan andal dan efisien. Menggunakan SQL, Python, Spark, sistem cloud.



Contoh Aplikasi di Dunia Nyata



E-commerce & Hiburan: Mesin Rekomendasi

- Netflix: Menganalisis riwayat tontonan Anda (dan jutaan pengguna lain) untuk merekomendasikan film atau serial yang kemungkinan besar akan Anda sukai.
- Hasil: 80% dari konten yang ditonton di Netflix berasal dari rekomendasi.
- Amazon/Tokopedia: Menyarankan produk berdasarkan apa yang telah Anda beli atau lihat ("Pelanggan yang membeli ini juga membeli...").
- Teknik: Collaborative Filtering dan Content-Based Filtering.

Koren, Y., Bell, R., & Volinsky, C. (2009). "Matrix factorization techniques for recommender systems." *Computer*, 42(8), 30-37.

Contoh Aplikasi di Dunia Nyata



Kesehatan: Diagnostik Medis Berbasis AI

- Analisis Gambar Medis: Algoritma Deep Learning dilatih pada ribuan gambar rontgen atau MRI untuk mendeteksi tanda-tanda penyakit seperti kanker atau retinopati diabetik.
- Studi Kasus: Model AI Google Health terbukti dapat mendeteksi kanker payudara dari mamogram dengan tingkat akurasi yang sebanding (atau bahkan lebih baik) dari ahli radiologi manusia.
- Dampak: Membantu dokter membuat diagnosis lebih cepat, lebih akurat, dan mengurangi beban kerja.

McKinney, S. M., et al. (2020). "International evaluation of an AI system for breast cancer screening." *Nature*, 577(7788), 89-94

Etika Data Sains



- Kekuatan besar datang dengan tanggung jawab besar.
- Isu Etis yang Krusial:
- Bias Algoritmik: Model dapat mempelajari dan bahkan memperkuat bias yang ada dalam data historis (misalnya, bias rasial dalam sistem peradilan pidana atau bias gender dalam rekrutmen).
- Privasi Data: Bagaimana data pribadi dikumpulkan, disimpan, dan digunakan? (Contoh: Skandal Cambridge Analytica).
- Transparansi & Penjelasan (Explainability): Bisakah kita menjelaskan mengapa sebuah model membuat keputusan tertentu? (Penting untuk model black box seperti deep learning).

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing

Rangkuman

- Data Sains adalah bidang interdisipliner yang mengubah data mentah menjadi wawasan berharga.
- Kebangkitannya didorong oleh ledakan data, kekuatan komputasi, dan algoritma canggih.
- Ini adalah proses siklus yang sistematis, bukan hanya tentang coding.
- Aplikasinya sangat luas, dari merekomendasikan film hingga menyelamatkan nyawa.
- Pertimbangan etis adalah bagian yang tidak terpisahkan.





SEE YOU NEXT WEEK !

Ferdian Bangkit Wijaya, S.Stat., M.Si

NIP. 199005202024061001

ferdian.bangkit@untirta.ac.id