Introduction to Data Science & Data Science Methodology

19 Feb 2024

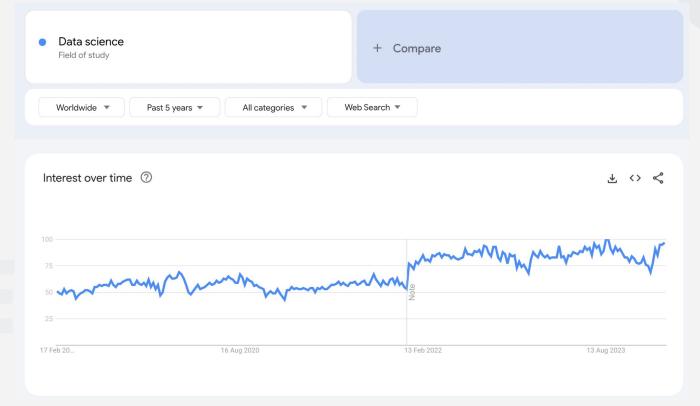


## **Agenda**

- 1. What is Data Science
- 2. Why Data Science
- 3. Data Science in Business
- 4. Data Science Success Story
- 5. What is Data Science Methodology?
- 6. What is Business Understanding?
- 7. What are Analytic Approaches?
- 8. What is Data Requirements?
- 9. What is Data Collection
- 10. What is Data Understanding?
- 11. What is Data preparation?
- 12. What is Modelling?
- 13. What is model evaluation?
- 14. What is deployment?



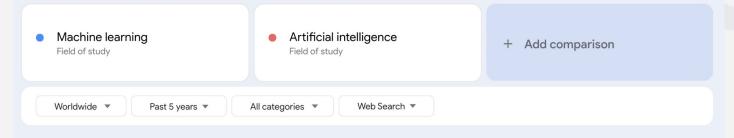
## **Popularitas Data Science**

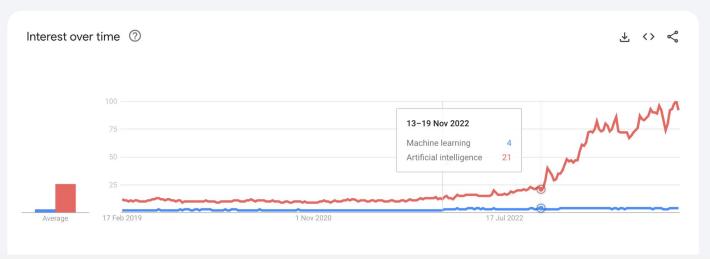


## **Popularitas Data Science**



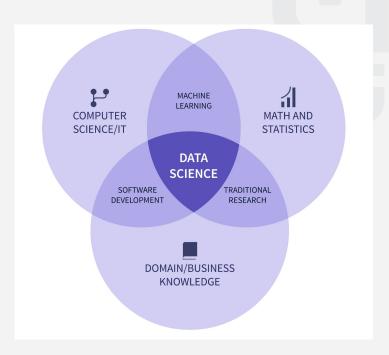
## **Popularitas Data Science**





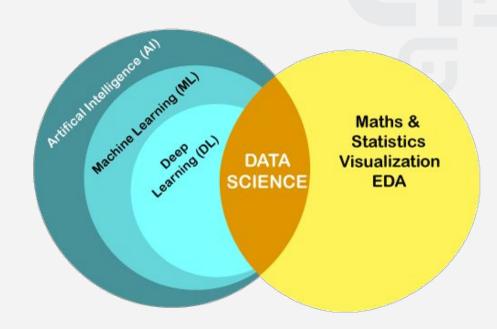
### **Data Science adalah ...**

Sebuah pengetahuan yang
menyatukan statistik dan
matematika, ilmu komputer dan
bisnis atau domain pengetahuan
lainnya untuk mengekstrak
pengetahuan (insight) dari data
terstruktur dan tidak terstruktur.

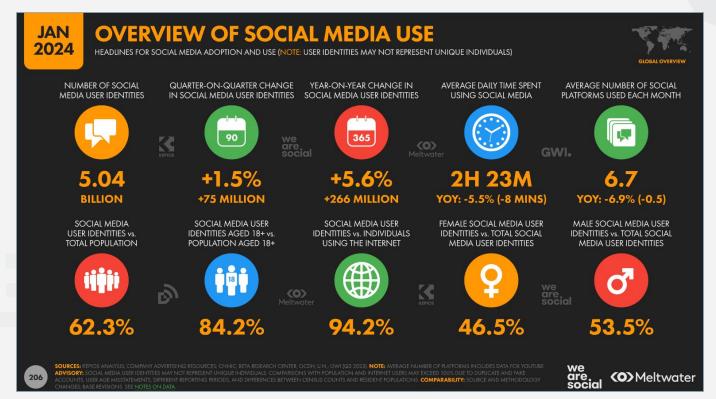


### **Data Science adalah ...**

- Data Science menggunakan Al untuk menganalisis dan menginterpretasikan data yang kompleks, memprediksi hasil, dan membuat keputusan.
- Alur kerja Data Science mungkin melibatkan komponen Al untuk memproses dan menganalisis data dalam skala dan kecepatan yang melebihi kemampuan manusia.



## Mengapa DS: Pertumbuhan Data



## Mengapa DS: Pertumbuhan Data

- Data Everywhere: Setiap detik, jumlah data yang besar dihasilkan dari smartphone,
   media sosial, transaksi online, perangkat IoT, dan lainnya.
- Melebihi Kapasitas Manusia: Volume dan kompleksitas data telah melampaui kemampuan analisis manual. Ilmu data menyediakan alat dan metodologi untuk memahami data ini.
- **Mendukung Pengambilan Keputusan**: Di era informasi, keunggulan kompetitif dan keputusan strategis bergantung pada wawasan yang didorong oleh data.

# DS dalam Bisnis: Pengambilan Keputusan Berbasis Data

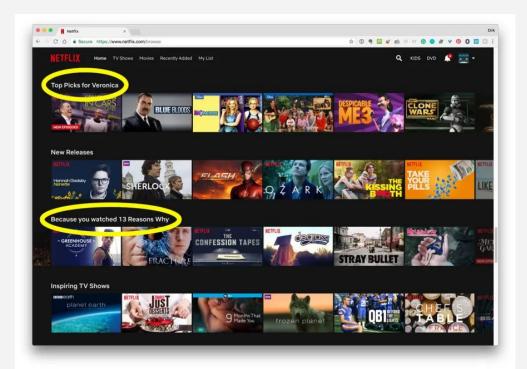
- Strategi yang Terinformasi: Memanfaatkan jumlah data yang besar untuk menginformasikan keputusan strategis bisnis, mengurangi tebakan dan keputusan berbasis intuisi.
- Manajemen Risiko: Memanfaatkan analitik prediktif dan pembelajaran mesin untuk mengidentifikasi dan mengurangi risiko sebelum mereka menjadi masalah kritis.
- Efisiensi Operasional: Menerapkan ilmu data untuk mengoptimalkan operasi bisnis, mulai dari manajemen rantai pasokan hingga layanan pelanggan, yang mengarah pada penghematan biaya yang signifikan dan peningkatan tingkat layanan.



## DS dalam Bisnis: Mendorong Inovasi dan Pertumbuhan Bisnis

- Pengembangan Produk: Memanfaatkan data pelanggan dan tren pasar untuk mendorong inovasi dalam pengembangan produk, menyesuaikan penawaran untuk memenuhi dan mengantisipasi kebutuhan pelanggan.
- Ekspansi Pasar: Mengidentifikasi tren yang muncul dan pasar yang belum terjamah melalui analisis data, mendukung keputusan strategis tentang ke mana harus berekspansi selanjutnya.
- Pengalaman Pelanggan: Meningkatkan perjalanan pelanggan melalui pengalaman yang dipersonalisasi, memanfaatkan ilmu data untuk memahami preferensi dan perilaku, meningkatkan loyalitas dan nilai seumur hidup.

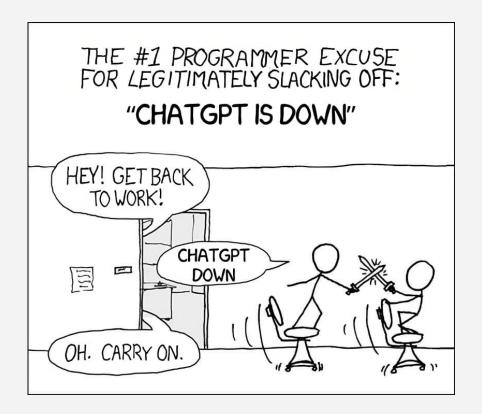
# Transforming Engagement dengan Data Science: Netflix



## Transforming Engagement dengan Data Science: Netflix

- Tantangan: Melibatkan audiens global dengan konten yang dipersonalisasi dari data film yang terus berkembang.
- Solusi: Sebuah mesin rekomendasi yang didukung oleh machine learning, menganalisis data pengguna, metadata konten, dan konteks tayangan untuk menyesuaikan rekomendasi konten.
- Hasil:
  - Peningkatan engagement dan kepuasan pengguna
  - Penurunan tingkat penghentian langganan
  - Pertumbuhan cepat menjadi lebih dari 200 juta pelanggan, menetapkan standar baru dalam penyampaian konten

## **ChatGPT - Revolutionizing Conversational Al**



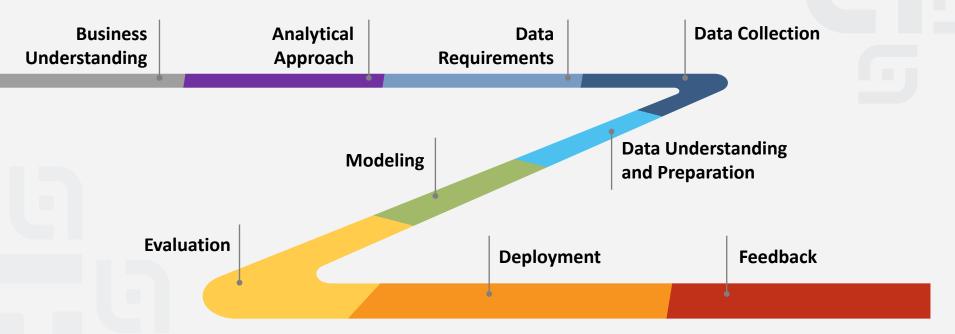
## **ChatGPT - Revolutionizing Conversational Al**

- Pengembangan: Dibuat oleh OpenAI, ChatGPT memanfaatkan kekuatan model GPT (Generative Pre-trained Transformer), yang dilatih pada teks internet yang beragam untuk memahami dan menghasilkan respons teks yang mirip manusia.
- **Kemampuan:** ChatGPT dapat menjawab pertanyaan, menulis esai, menciptakan puisi, menghasilkan kode, dan lainnya, menunjukkan tingkat pemahaman dan kreativitas yang belum pernah ada sebelumnya dalam pemrosesan bahasa alami (NLP).

### Dampak:

- Mengubah interaksi pengguna dengan AI, membuat bot percakapan menjadi lebih serbaguna dan mirip manusia.
- Memungkinkan aplikasi inovatif di berbagai bidang seperti pendidikan, layanan pelanggan, hiburan, dan pemrograman, menunjukkan potensi Al untuk meningkatkan kemampuan manusia.
- Memicu diskusi global tentang penggunaan etis Al, menyoroti pentingnya pengembangan dan penerapan Al yang bertanggung jawab.

## **Data Science Methodology**

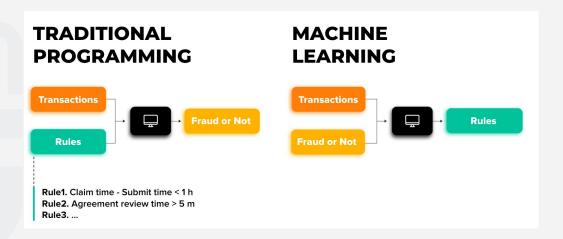


## **DS Methodology (1): Business Understanding**

- Menetapkan tujuan dan masalah yang perlu dipecahkan. Sangat penting untuk memiliki tujuan yang jelas dan memahami konteks bisnis.
  - Metrik kuantitatif, contoh: peningkatan CTR sebesar X%.
  - Detail lingkup kerja (In / Out of scope).
- Contoh Netflix: Netflix bertujuan untuk meningkatkan engagement dan retensi penonton dengan merekomendasikan konten yang sangat relevan untuk setiap pengguna.

## DS Methodology (2): Analytical Approach

- Tentukan bagaimana menganalisis data untuk mengatasi masalah bisnis. Ini melibatkan pemilihan model dan teknik yang tepat.
  - apakah seluruh masalah perlu diselesaikan dengan AI/ML/DL?





## **DS Methodology (2): Analytical Approach**

 Contoh Netflix: Netflix memilih untuk menggunakan model ML untuk menganalisis pola tontonan dan memprediksi preferensi pengguna.



**SIB Cycle 6** | 2024

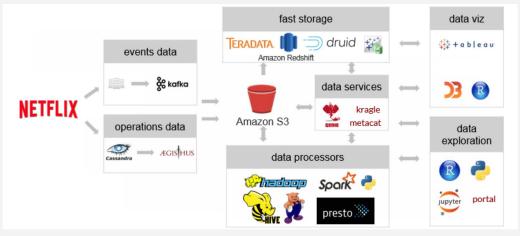
## **DS Methodology (3): Data Requirements**

- Identifikasi data yang diperlukan untuk analisis. Ini termasuk menentukan sumber dan jenis data yang dibutuhkan.
- Contoh Netflix: Data yang diperlukan meliputi riwayat tontonan pengguna, pencarian, rating, dan metadata konten.



## **DS Methodology (4): Data Collection**

- Kumpulkan data yang telah diidentifikasi dari berbagai sumber. Langkah ini memastikan semua data yang diperlukan tersedia untuk analisis.
- Contoh Netflix: Netflix mengumpulkan data dari platformnya, melacak
   bagaimana pengguna berinteraksi dengan layanan di berbagai perangkat.

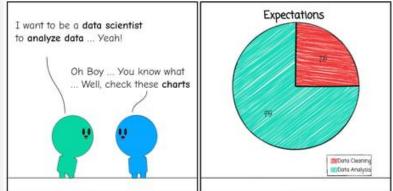


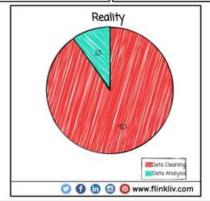
DS Methodology (5): Data Understanding and

**Preparation** 

 Explore dan pre-proses data. Tahap ini termasuk melakukan pembersihan data (data cleaning), melakukan EDA (exploratory data analysis) untuk mengetahui pola (pattern) dan anomali.

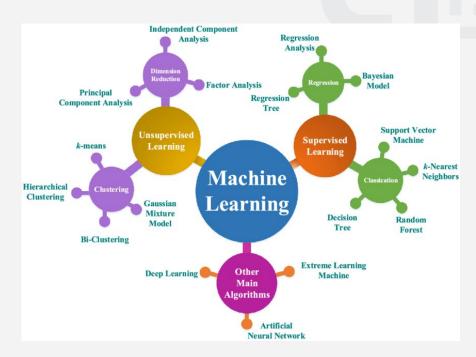
Contoh Netflix: Netflix menganalisis
 pola tontonan untuk mengidentifikasi
 preferensi umum dan membersihkan
 data dari outlier untuk memastikan
 akurasi.





## **DS Methodology (6): Modeling**

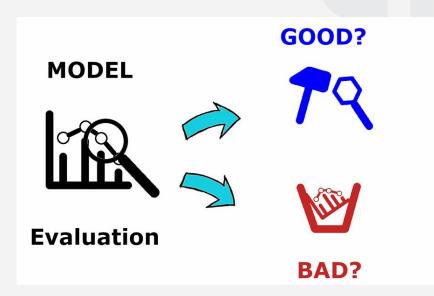
- Kembangkan model yang bisa memprediksi, mengklasifikasi, atau mengelompokkan berdasarkan data yang diberikan. Pemilihan model berdasarkan masalah dan data yang ada.
- Contoh Netflix: Mengembangkan algoritma untuk rekomendasi personal berdasarkan perilaku pengguna dan kesamaan konten.



## **DS Methodology (7): Evaluation**

- Cek kinerja model dengan
  menggunakan metrik yang sesuai.

  Langkah ini memvalidasi seberapa baik
  model memenuhi tujuan bisnis.
- Contoh Netflix: Netflix mengevaluasi model menggunakan metrik seperti akurasi, recall, dan engagement rate.



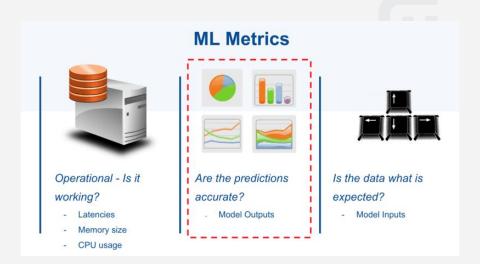
## **DS Methodology (8): Deployment**

- Implementasi model di aplikasi di mana model tersebut bisa memberikan dampak, melibatkan integrasi model ke dalam sistem yang sudah ada.
- Contoh Netflix: Mengintegrasikan sistem rekomendasi ke dalam platform Netflix, secara dinamis mempersonalisasi konten untuk setiap pengguna.



## DS Methodology (9): Feedback

- Pantau kinerja model dan kumpulkan feedback pengguna untuk melakukan perbaikan berkelanjutan. Proses iteratif ini memastikan model tetap efektif seiring waktu.
- Contoh Netflix: Netflix terus
   memperbarui algoritmanya
   berdasarkan data baru dan feedback
   pengguna untuk menyempurnakan
   rekomendasinya.





## **Terima Kasih**



