
Data Scientist and Data Science: An Overview.

Grady Matthias Oktavian
Data Scientist at Indonesian Unicorn Startup

Outline

1. Sejarah Data Science
2. Traditional vs Modern Data Science
3. Aplikasi Data Science
4. Pekerjaan Data Scientist
5. Masa Depan Karir Data Science
6. Kesimpulan + Q&A

Sejarah Data Science



- **1977** - IASC (Komite Internasional untuk Komputasi Statistika) dibentuk dengan misi: menyambungkan metodologi statistika tradisional dengan ilmu komputasi untuk mengubah data menjadi informasi dan pengetahuan.
- **1999** - Jacob Zahavi, expert data mining, menyebutkan bahwa adanya kebutuhan untuk dikembangkan teknologi-teknologi baru untuk memproses data dalam volume yang besar.
- **2008** - Kata "Data Science" mulai menjadi mainstream.
- **2013** - IBM melaporkan bahwa 90% data di dunia diciptakan hanya dalam 2 tahun terakhir. Revolusi Data Science mulai berlangsung, dan institusi mulai menyadari pentingnya peran ahli data.

Traditional vs Modern Data Science

Aspek	Tradisional	Modern
Goal	Memahami dan melaporkan data yang telah tercatat	Menggunakan data masa lampau untuk memprediksi masa depan
Struktur Data	Data terstruktur dan sudah bersih (biasanya dalam tabel)	Data bisa saja tidak terstruktur dan masih mentah. Jenis-jenis data juga beragam, bisa berupa teks, gambar, dst.
Volume	Lebih kecil, muat di harddisk	Lebih besar, disimpan tidak dalam 1 komputer saja.
Kecepatan	Lebih lambat dan manual	Cepat
Visualisasi	Dalam laporan	Bisa berupa dashboard yang interaktif dan diupdate seiring berjalannya waktu

Apakah tradisional itu buruk?

Tidak.

Justru, metodologi dan prinsip-prinsip dasar statistika harus terus 'dianut' dalam era 'big data'. Modernisasi bukan berarti meninggalkan teori dasar tentang pengolahan data.

Aplikasi Data Science

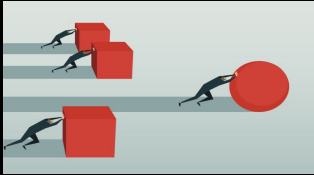
Data bisa berupa apa saja. Kebanyakan data memang tersimpan dalam bentuk 'tabular', namun pada era ini, data bisa berupa gambar, suara, dan teks.



Aplikasi Data Science

Dalam era korporasi, beberapa kasus dimana Data Science dapat diimplementasikan adalah sebagai berikut:

Membandingkan '*Competitive Advantage*'



Segmentasi Customer



Mencari Market/Potensi Bisnis Area Baru



Meminimalisir Fraud



Aplikasi Data Science

Dalam bidang akademisi, Data Science bermanfaat dalam mengolah data yang datang dalam jumlah yang banyak. Contohnya adalah mengolah data RNA-seq (barisan RNA). Beberapa solusi yang dikembangkan adalah "Kallisto", sebuah teknologi sehingga data processing untuk RNA-seq dapat dilakukan dengan 'cepat' menggunakan Mac / Desktop 'biasa'.

<https://github.com/pachterlab/kallisto>



Aplikasi Data Science

Perusahaan seperti 'Netflix' pun bahkan memiliki divisi Riset untuk Machine Learning sendiri.


NETFLIX RESEARCH

HOME RESEARCH AREAS BUSINESS AREAS ARTICLES JOBS

RESEARCH AREAS

Machine Learning

Learning how to entertain the world



About

As the world's leading Internet television network with over 160 million members in over 190 countries, our members enjoy hundreds of millions of hours of content per day, including original series, documentaries and feature films. We invest heavily in machine learning to continually improve our member experience and optimize the Netflix service end-to-end. As researchers, we innovate using machine learning in many areas where we prototype, design, implement, evaluate, and productionize models and algorithms through both offline experiments and online A/B testing.

Recent Job Openings

Senior Research Scientist - Algorithms Engineering
📍 Remote, United States

VIEW ALL JOBS IN THIS AREA ➔

Aplikasi Data Science

The logo for fastText, with 'fast' in red italicized font and 'Text' in blue bold font.

Library for efficient text classification and representation learning

GET STARTED

DOWNLOAD MODELS

Sementara itu, facebook mengembangkan library Python terkait NLP (Natural Language Processing) yang dinamakan 'fasttext'.

Library ini open source dan dapat digunakan siapapun secara gratis.

Intinya...

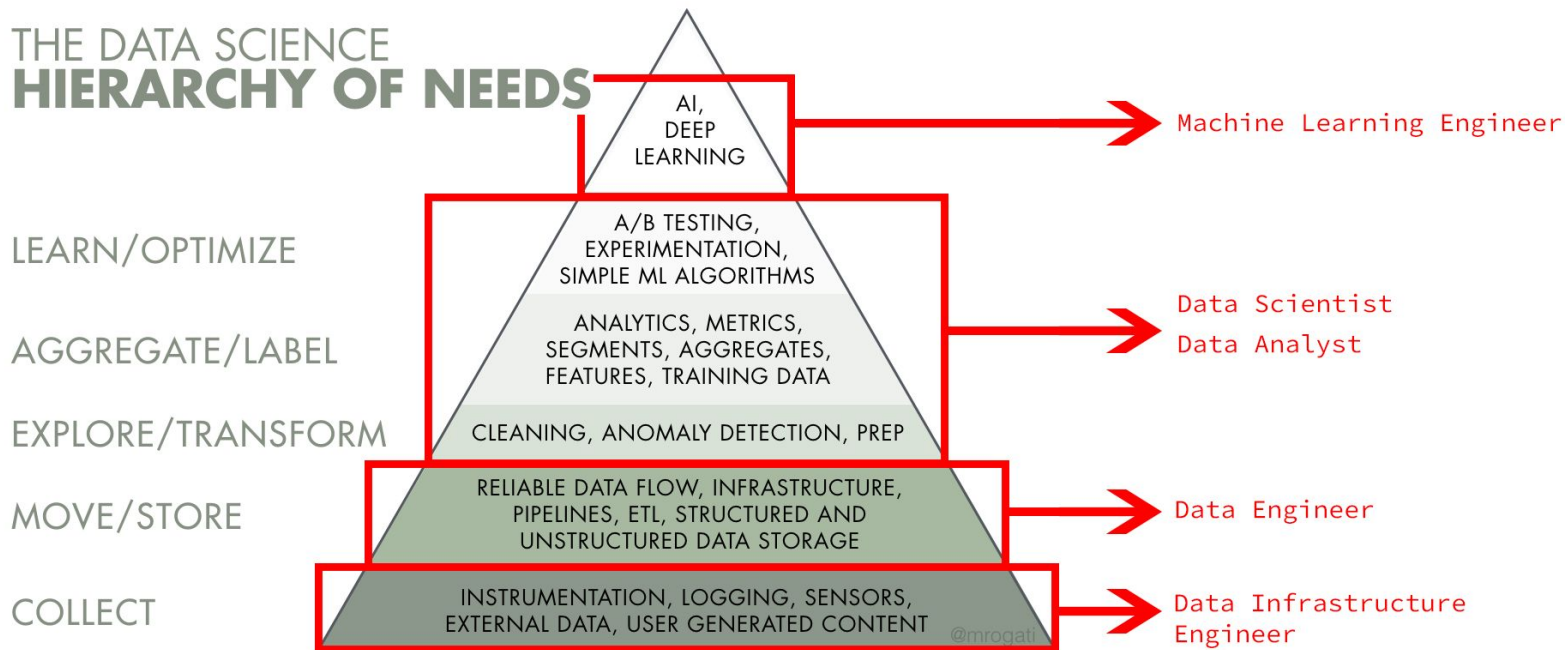
Dimana ada data, di situ ada kesempatan bagi Data Scientist untuk menggali data untuk menemukan insights, dan bahkan membuat algoritma/model untuk mendukung business area terkait.

Jadi, apa sih yang dikerjakan Data Scientist?

Memberikan solusi terhadap sebuah business problem yang melibatkan data. Secara general, itu yang dikerjakan.

Data Science Hierarchy of Needs

THE DATA SCIENCE HIERARCHY OF NEEDS



Examples of Data Science 'Projects':

- Recommendation Engine**
- Fraud Detection**
- Spam Detection**
- Forecasting (Weather, Stock, etc)**
- Image Classification**
- Sentiment Analysis**

Di masa depan...

Role 'Data Scientist' akan semakin terspesialisasi. There will be room for everyone.

Seperti apa contohnya?

1. **Data Engineering / Data Warehousing**

Data collection, transformation, ETL, dll.

2. **Data Mining and Statistical Analysis**

Fokus pada eksplorasi data dan visualisasi.

3. **Cloud Architect/Engineer**

Implementasi Cloud Infrastructure.



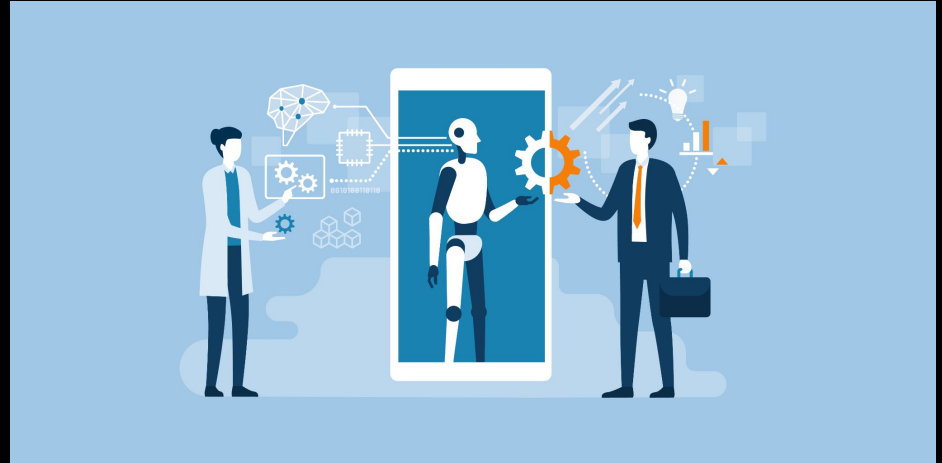
4. Business Intelligence and Strategy

Fokus pada merencanakan penyelesaian masalah business, dan juga berkomunikasi dengan stakeholders.

5. Machine Learning / AI Engineer

Fokus pada pembuatan pipeline deep learning supaya 'training model' dapat dilakukan seefektif mungkin.

6. Market or Field Related Data Scientist



Tips for University Students

1. Jangan takut mencoba! Jangan takut/males 'Googling', riset sendiri, dll.
2. Harus terus mau belajar dan keep up-to-date dengan perkembangan teknologi zaman now (apalagi anak IT...)
3. Pahami fundamental nya, jangan terlena dan bingung dengan 'buzzwords'.

Kesimpulan

1. Data Science adalah cabang keilmuan yang melibatkan statistika dan ilmu komputer untuk menyelesaikan masalah berdasarkan data.
2. Ke depannya, data akan semakin banyak, peran data scientist semakin penting di berbagai bidang, dan role data scientist semakin terspesialisasi.

Q & A