

Learning Progress Review #Week1

Introduction to Data Science

Oleh:

MARVEL TEAM

FIKRIE

NATALIA

SATRIA

Apa itu Data Science

Data Science adalah sebuah profesi yang mengolah data dengan metode ilmiah untuk menghasilkan informasi yang akurat dan bernilai tinggi sehingga dapat membantu pengambilan keputusan

Ilmu Data Science sangat dibutuhkan pada dunia bisnis diantaranya adalah dalam bisnis ritel, distribusi, kedokteran dan masih banyak lagi.

Data Science harus menguasai ilmu multi disiplin, yaitu matematika statistik, ilmu komputer, dan pengetahuan terhadap salah satu domain.

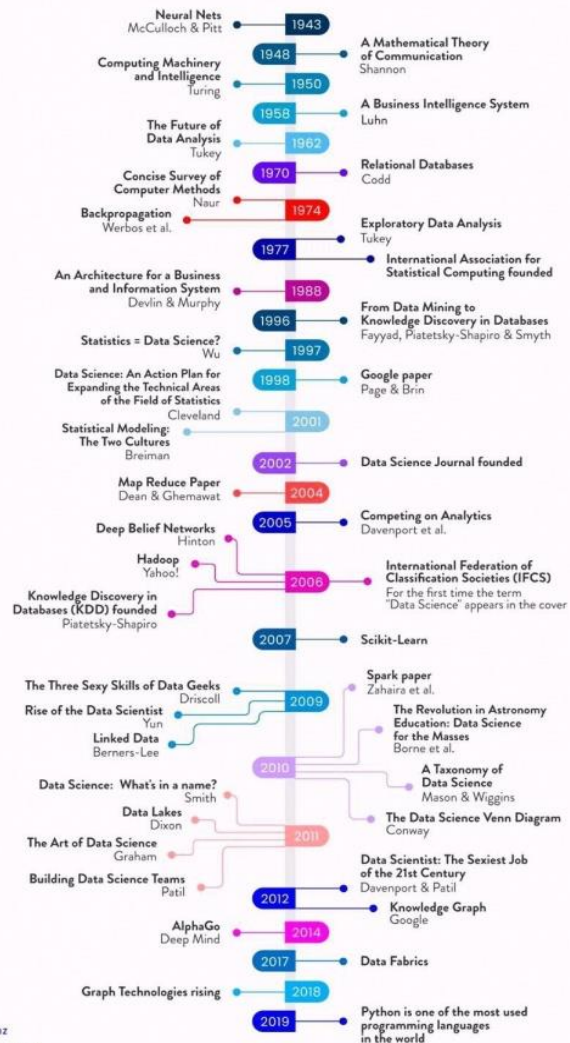
Siapa saja yang bisa menjadi seorang Data Scientist

Untuk menjadi seorang Data Scientist tidak hanya terpaku di beberapa background Pendidikan saja.

Tentu hal tersebut menjadi sesuatu kewajiban bagi seseorang yang bercita-cita menjadi Data Scientist

Hal yang utama bila ingin terjun di dunia Data Science adalah mempelajari 3 keilmuan yang wajib di kuasai, diantaranya yaitu : Statistik/Matematika, ilmu komputer/pemrograman, dan Bisnis

DATA SCIENCE TIMELINE



@faviavaz
@heizelvazquez

Sejarah dan evolusi Data Science

- 1996 : Sebelum data science, popular terlebih dahulu istilah data mining. Dalam sebuah artikel yang disebut from data mining to knowledge Discovery in database pada tahun 1996 yang merujuk pada keseluruhan Proses menemukan informasi yang berguna dari data.
- 2001 : William S. Cleveland membawa data mining ke level yang lebih tinggi. Dia melakukannya dengan menggabungkan computer science dengan Data mining.
- 2003 – 2005 : Pada saat ini, bermula Ketika web 2.0 muncul di mana situs web tidak lagi hanya pamflet digital. Situs web pada tahun ini adalah Myspace, Facebook, Youtube.
- 2010 : Dalam Dat Science, biasanya membutuhkan komputasi parallel. Seperti MapReduce, Hadoop, dan Spark. Jadi, kebangkitan **big data**. Pada tahun 2010 memicu munculnya **ilmu data** untuk mendukung Kebutuhan bisnis untuk mendapatkan wawasan dari kumpulan data Terstruktur mereka yang sangat besar.
- 2011 – Sekarang : Sekarang masyarakat umum menganggap Data Science sebagai penelitian yang berfokus pada **Machine Learning dan Artificial Intelligence** tetapi industry ini memperkerjakan Data Scientist sebagai Analisis,

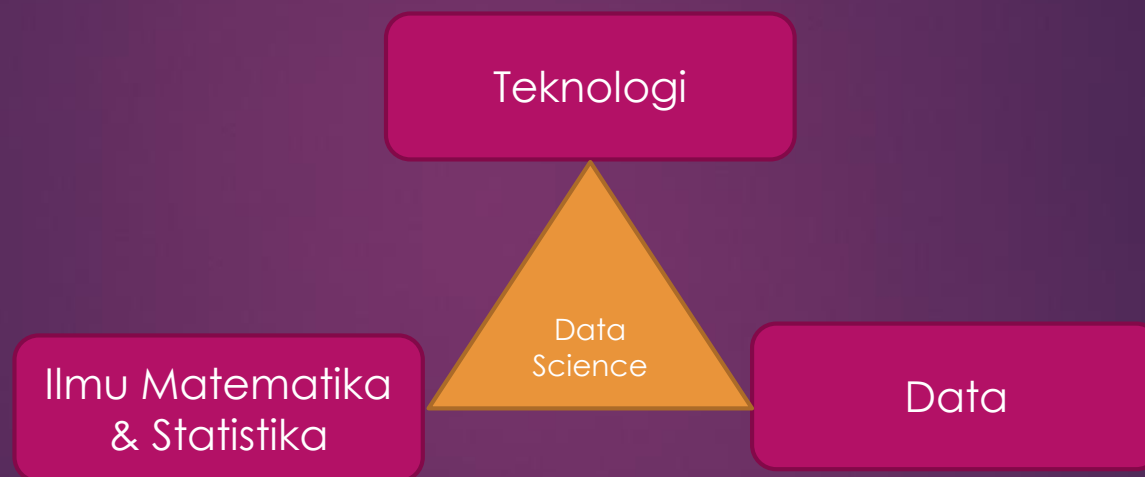
Faktor-faktor yang berkontribusi pada pertumbuhan data

Digitalisasi aktivitas manusia.
Setiap aktivitas manusia saat ini tercatat dalam perangkat elektronik pribadi yang terkoneksi. Kumpulan aktivitas yang terdigitalisasi dan terkoneksi inilah yang menjadi tambang data bagi Data Science.

Kepentingan ekonomi industri.
Memahami keinginan calon konsumen melalui pemahaman pola perilaku aktivitas manusia saat ini terus dikembangkan. Pola perilaku ini tercatat dalam perangkat elektronik pribadi setiap manusia. Hal ini penting dilakukan agar industri secara efektif menghasilkan produk yang tepat sasaran dan laku di pasaran.

Domain Data Science

Kepentingan industri dalam dunia yang terdigitalisasi saat ini mengumpulkan tiga faktor, Yaitu,



Data Science berada pada tiga irisan faktor tersebut. Berpikir kritis apa yang dapat dilakukan dengan melimpahnya data dan membuatnya memiliki nilai yang bermanfaat. Memahami bagaimana data didapatkan, diolah, dianalisa dan disajikan dalam presentasi yang mudah dipahami.

Tim Data Science dan Peran Pentingnya

Dalam bekerjanya sebuah Big Data yang memiliki output yang bernilai dan bermanfaat, setidaknya ada tiga peranan mendasar, yaitu :

Data Engineer

- Memastikan penambangan data.
- Mengklasifikasi data mentah.
- Memastikan infrastuktur penambangan data berjalan baik.

Data Scientist

- Klasifikasi data yang diperlukan.
- Menganalisa data dengan memahami pola perilaku data.
- Memberikan kesimpulan dan mempresentasikannya.

Data Analyst

- Mengeksplorasi hasil temuan dan analisis data.
- Mempresentasi hasil temuan eksplorasi data.
- Mengkomunikasikan dengan pihak terkait.

Data Science dalam Bisnis

Berapa biaya yang dibutuhkan untuk menjual suatu produk dan darimana asal pendapatan (revenue) suatu perusahaan? Apakah 80% pendapatan berasal dari 20% pelanggan?



Semua pertanyaan ini dapat dijawab dengan menggunakan data. Perusahaan menangkap data. Lalu, mengaplikasikan algoritma pada data tersebut dan menganalisisnya.

Pada akhirnya, **data itu tentang biaya (cost)**. Dengan menggunakan Data Science, perusahaan dapat mengembangkan bisnisnya.

Data Science dalam Bisnis

Apa yang dilakukan dan dihasilkan oleh seorang Data Scientist di dalam bisnis suatu perusahaan?



1. Insights

Mengekstraksi *insight* dari Data menggunakan metode A/B Testing atau Observation

2. Basic Machine Learning Models

Membangun Basic Machine Learning Models dengan cara: Classification, Regression, atau Clustering

3. Advance Machine Learning Models

Contoh: Stock Market Analysis and Prediction, Self-Driving Car, Recommender System, dan Image Object Classification

Data Science dalam Bisnis

Ketika mempekerjakan seorang Data Scientist,
recruiter akan bertanya:

Apakah Anda suka main
dengan Data?

Apakah Anda dapat
melakukan analisis data?

**Apakah Anda menyukai
model bisnis perusahaan?**



**Pertanyaan ini yang TERPENTING. Hal
yang terpenting dari seorang Data
Scientist adalah PASSIONATE
terhadap bisnis perusahaannya.**

Kasus Sukses dari Penggunaan Data Science



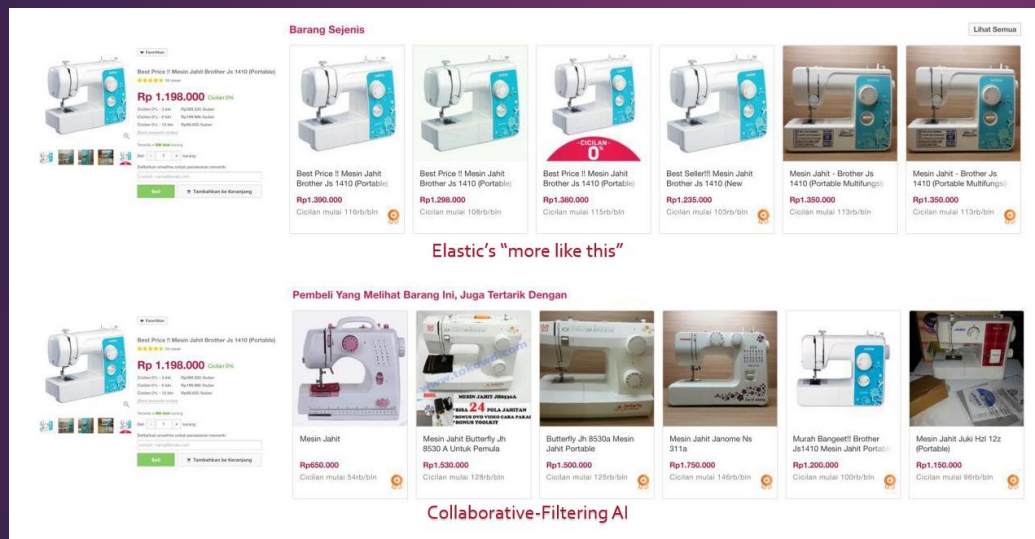
Permainan ini diciptakan oleh Niantic dan didasarkan pada permainan yang mirip bernama Ingress. Pokemon Go adalah sebuah permainan yang tercipta dari hubungan antara big data dan augmented reality.

Kasus Sukses dari Penggunaan Data Science



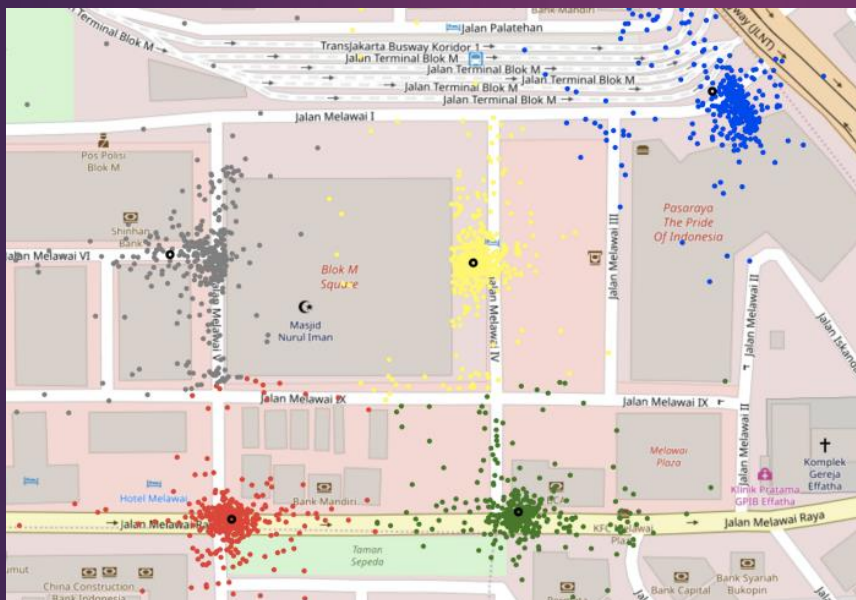
Dengan 30 juta play per hari, Netflix mempunyai akses terhadap seluruh data preferensi penikmat tv atau film sehingga memiliki insight tentang genre apa yang disukai penonton, artis/aktor mana yang lebih disukai penonton, sutradara mana yang memiliki karya film atau serial TV dengan rating tinggi. Berdasarkan insight tersebutlah, Netflix dapat memprediksi House of Card menjadi serial TV yang hits.

Kasus Sukses dari Penggunaan Data Science



Bukalapak, sebuah platform e-commerce menciptakan Reco AI untuk meningkatkan penjualan produk para seller. Penciptaan Reco AI dimulai dari salah satu tipe Recommendation AI yang paling umum yaitu Collaborative Filtering. Kemudian, Reco AI itu divalidasi menggunakan pendekatan A/B Testing. Dengan adanya Reco AI, tercatat pendapatan para penjual di toko Bukalapak naik hingga Rp1 triliun.

Kasus Sukses dari Penggunaan Data Science



Dahulu Gojek belum mempunyai fitur titik lokasi penjemputan. Dengan menggunakan Data Science, saat ini, Gojek sudah memiliki fitur titik lokasi penjemputan. Gojek bereksperimen dengan menggunakan clustering algorithms DBSCAN dan Kmeans.

Meskipun Kmeans tidak memperhitungkan outliers, Gojek memilih metode Kmeans untuk mengembangkan fitur titik lokasi penjemputan karena lebih efektif dan benar dalam mengidentifikasi kluster.

