

Pertemuan 3

Mata Kuliah Pilihan Data Science (3 SKS)

Data Science Process and Implementation



UIN SUSKA RIAU

Mustakim, S.T., M.Kom

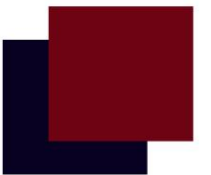


Program Studi Sistem Informasi

<https://mustakim.predatech.org> | mustakim@uin-suska.ac.id | +6285275359942



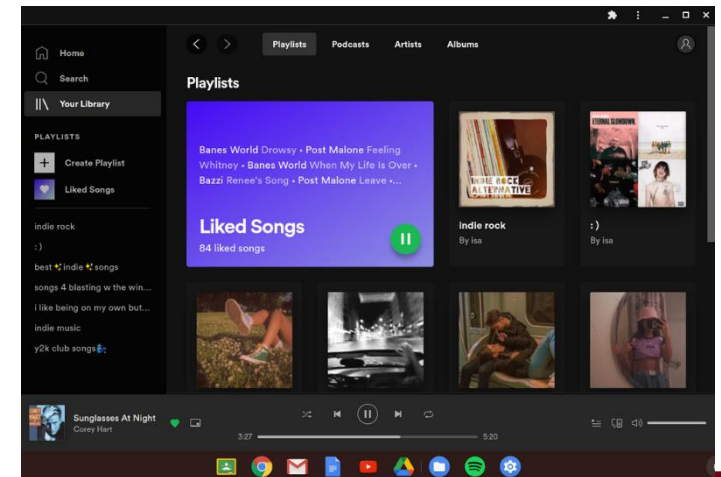
Data Science dan Kebutuhannya



Data Science dan Data Scientist

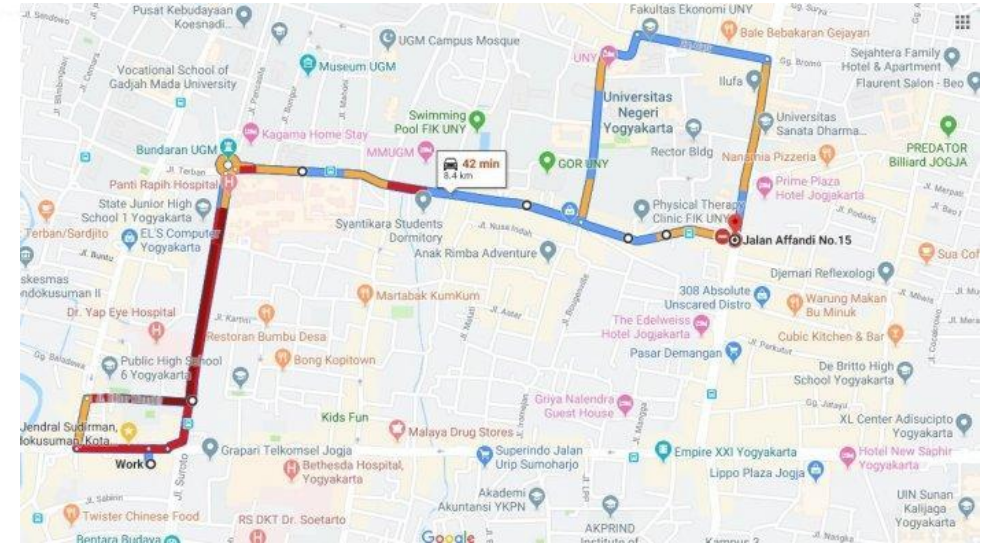
Sejatinya dalam kehidupan sehari-hari kita sudah memanfaatkan buah karya dari para data scientist

- Saat kita browsing di toko online, lalu kita klik salah satu item produk, di bawah browser akan diberikan produk-produk lain yang dibeli bersamaan atau yang mungkin kita sukai juga.
- Sama halnya ketika kita browsing di penyedia streaming lagu dan video. Kita juga akan disuguhi dengan rekomendasi item-item lain untuk didengar atau dilihat



Data Science dan Data Scientist

- Buat kita yang tinggal di kota besar dengan trafik padat, adakah yang belum pernah “ngecek” kemacetan di jalan-jalan kota kita? Ketika kita memilih jalur tercepat dari satu tempat ke tempat lainnya, mesin Google akan memanfaatkan informasi kepadatan lalu-lintas di tiap alternatif jalur untuk memilih yang tercepat. Warna hijau, kuning, oranye dan merah di peta Google telah menjadi informasi penting buat kita!



Data Science dan Data Scientist

- Apa saja yang sedang “hot” dibicarakan di dunia maya? Berbagai trending di Twitter menjadi salah satu jawabannya. Di situ juga bisa kita dapatkan informasi sentimen atau persepsi, apakah positif atau negatif, terhadap pesan tertentu.

Tren Indonesia

#YangHilangDalamCinta

Streaming siang ini

📌 Dipromosikan oleh Disney+ Hotstar ID

1 · Permainan · Populer

#BlokirKominfo

80,9 rb Tweet

2 · Permainan · Populer

PayPal

157 rb Tweet

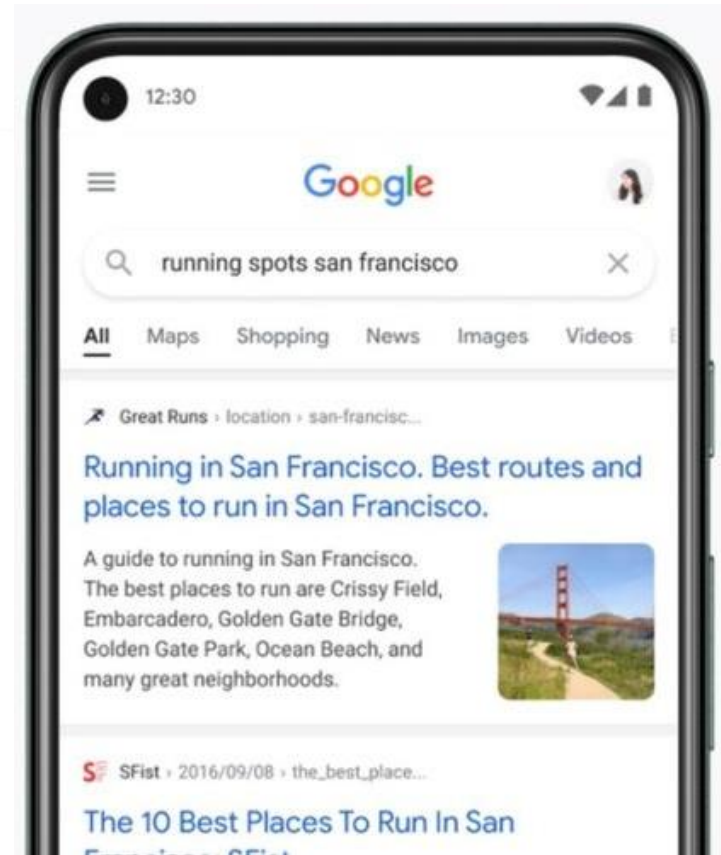
Data Science dan Data Scientis

- Saat kita bepergian, terlebih lagi ke negara 4 musim dimana di suatu wilayah cuacanya dapat berubah dengan cepat ponsel kita menjadi sumber informasi yang penting. Kita bisa cek di sekitaran objek wisata yang akan kita kunjungi, pada hari tanggal dan jam kita berada di sana, cuacanya bagaimana. Apakah akan turun hujan/salju? Angin kencang? Suhu super dingin atau sangat panas? Dari situ, kita bisa menentukan fashion bagaimana yang cocok untuk kita kenakan. Bisa juga kita batal pergi ke objek itu.



Data Science dan Data Scientist

- Pernah membandingkan hasil search di Google dengan keyword tertentu dari satu orang ke orang lain? Bisa beda. Hasil yang diberikan oleh mesin pencari Google akan dibuat sedemikian rupa, dibuat relevan dengan “kebiasaan” pencarian dan browsing kita di Internet.

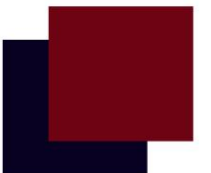




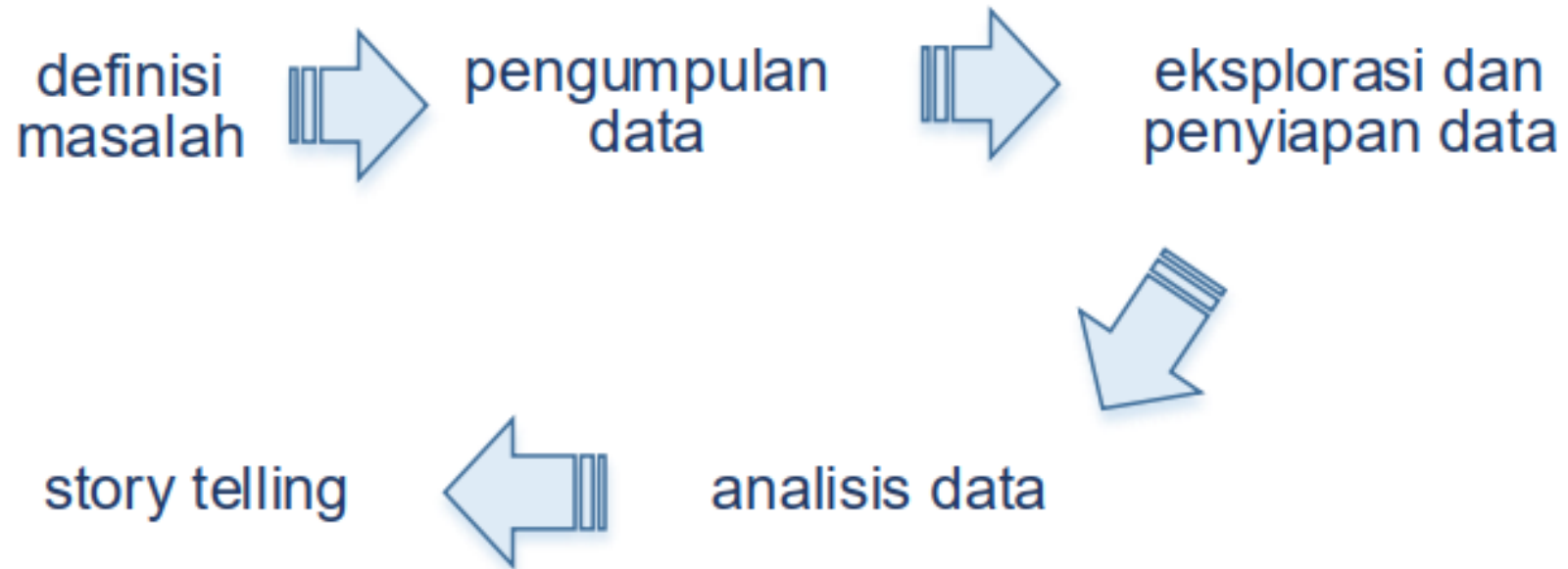
Apa Saja yang Dikerjakan Data Scientist?

Umumnya data scientist dibutuhkan oleh organisasi-organisasi yang telah memiliki sistem sistem teknologi informasi operasional sebagai sumber data

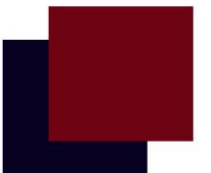
- Organisasi bisnis (misalnya perusahaan e-commerce, bank, transportasi dan pariwisata)
- Organisasi pemerintah yang memberikan layanan kepada masyarakat, misalnya untuk meningkatkan produktivitas pegawai dan memperbaiki layanan.
- Organisasi riset di bidang sains, kebutuhan akan berbeda, misalnya untuk menemukan model dari data yang bermanfaat untuk melakukan prediksi di masa depan.
- Model itu misalnya model prediksi panen tanaman, bencana, kebutuhan energi, kebutuhan transportasi penduduk, kerusakan lingkungan, dsb



Tahapan Data Science



Gambar 1.2. Tahapan data science.

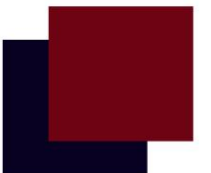


Tahapan Data Science

Pertama, tahap pendefinisian masalah. Data scientist mendapatkan kebutuhan organisasi yang harus dicarikan jawaban atau solusi dari data, misalnya menurunkan biaya produksi dan membuat pelanggan belanja lebih sering

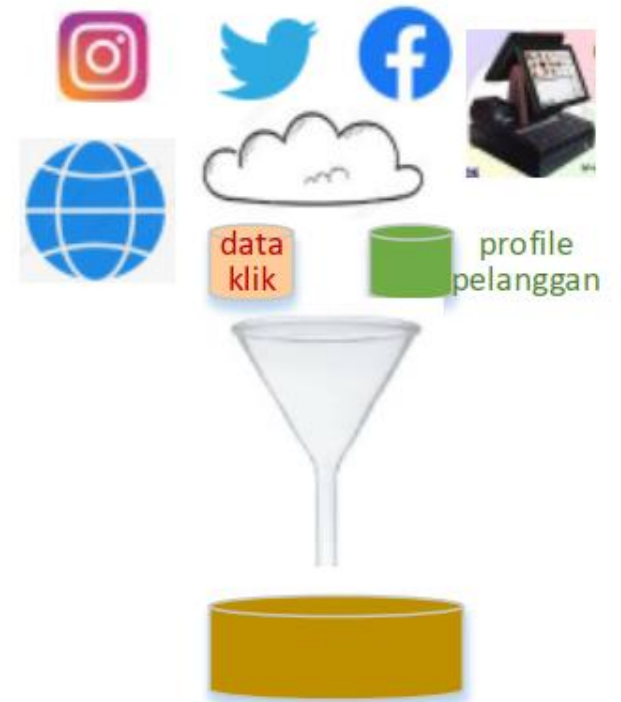


Gambar 1.3. Hal-hal berharga (insights) apa yang dapat digali dari data?



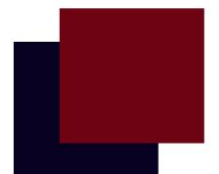
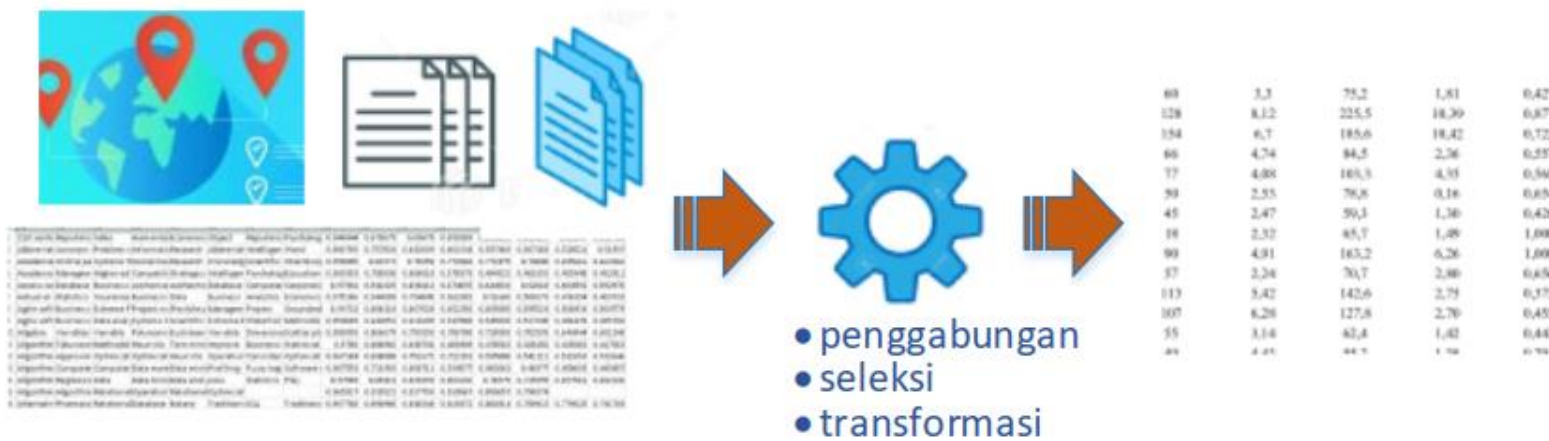
Tahapan Data Science

Kedua, tahap pengumpulan data. Berdasar insights yang akan digali, data scientist perlu merumuskan data apa saja yang dibutuhkan. Data itu dapat saja sudah tersedia semua atau baru sebagian. Jika baru sebagian, misalnya baru tersedia data transaksi sedangkan untuk menggali insights dibutuhkan data profile pelanggan dan Twitter, maka data scientist perlu mencari dan mengumpulkan data, yang dapat berasal dari satu atau lebih sumber



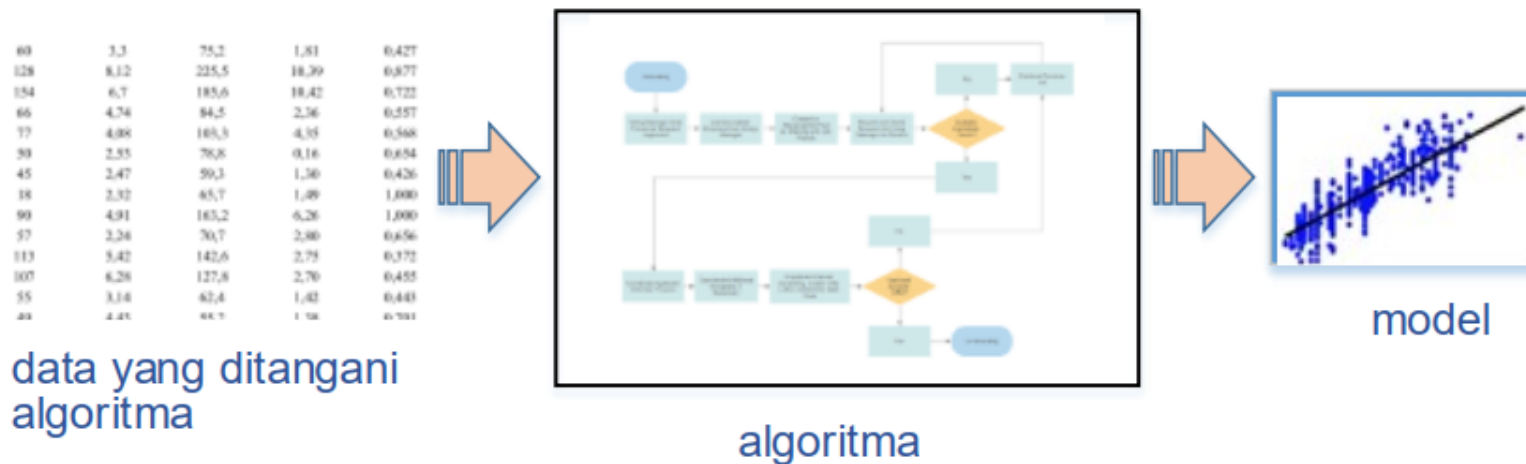
Tahapan Data Science

Ketiga, tahap eksplorasi dan penyiapan data. Setelah data terkumpul, seluruh komponen data perlu dipelajari dengan seksama. Misalnya, jika data berbentuk sebuah tabel, maka makna dan nilai tiap kolom harus dipahami. Untuk memahami data yang cukup kompleks dan berukuran besar, seringkali perlu dibuat visualisasi, kadang juga perlu komputasi statistik untuk mendapatkan ringkasan data (mencari rata-rata, median, minimum, maksimum juga distribusi data).



Tahapan Data Science

Keempat, tahap analisis data. Jika data yang disiapkan sudah bagus, tahap ini dapat dilakukan dengan relatif lebih mudah, asalkan data scientist sudah menguasai teknik/ algoritma, teknologi atau tools yang akan digunakan. Berdasarkan insights yang akan digali, di sini dipilih teknik atau algoritma yang sesuai (dapat berasal dari algoritma Machine Learning yang merupakan subset dari Artificial Intelligent atau Kecerdasan Buatan).

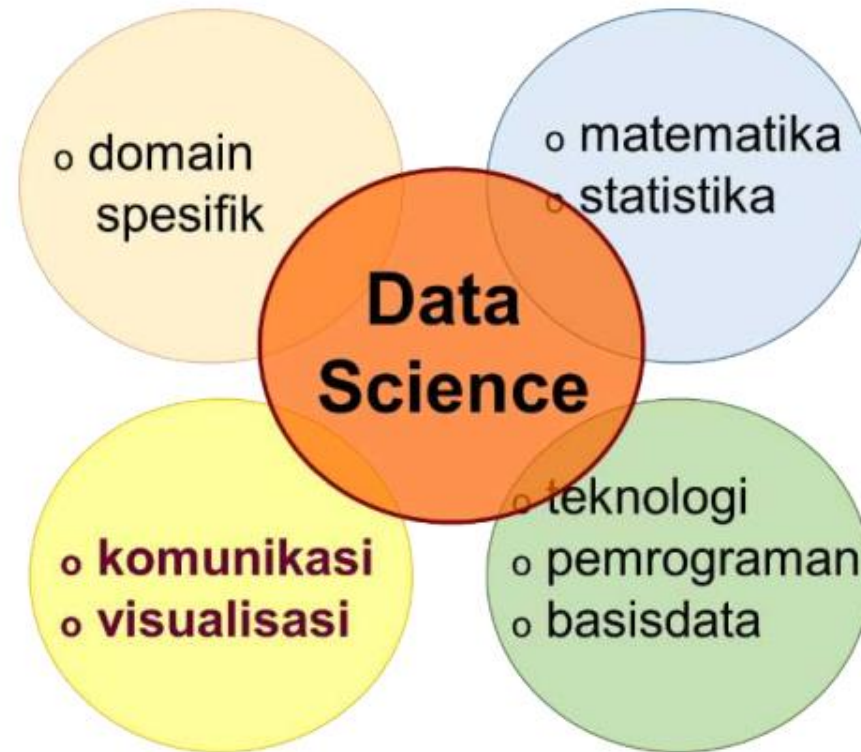


Tahapan Data Science

Kelima, storytelling. Seorang data scientist harus mampu untuk mengkomunikasikan proses dan hasil temuan analisis data dengan sistematis, menarik, tidak ambigu dan mudah dipahami bagi orang-orang (yang berkepentingan dengan proses maupun hasil itu)



Keahlian dan Skill Data Scientist

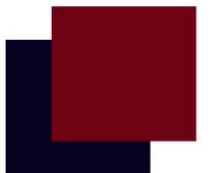
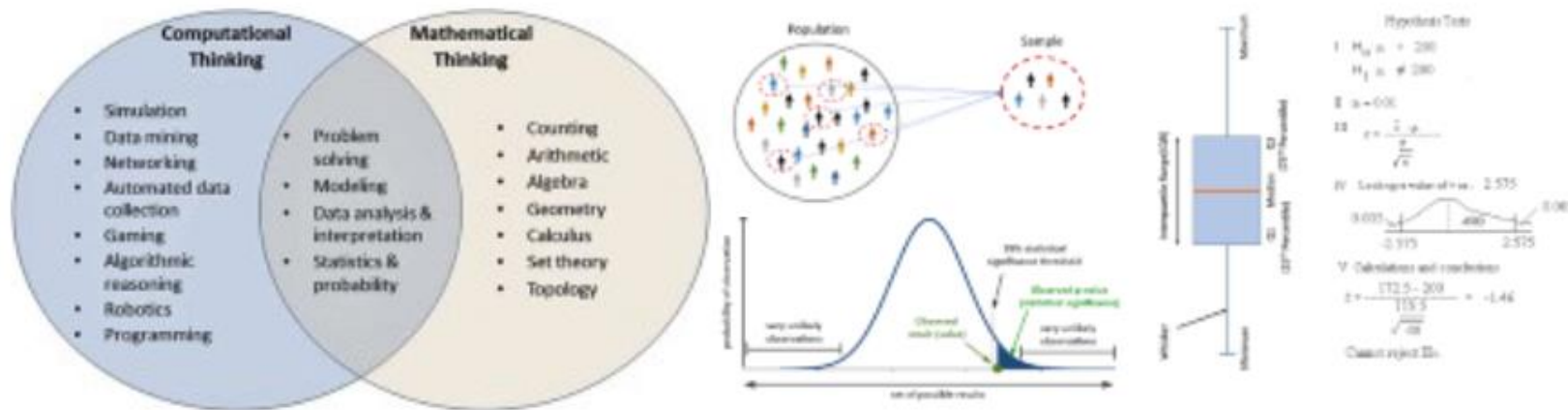


Gambar 1.8. Keahlian dan skill multi-disiplin data scientist.



Matematika dan Statistik

Matematika dalam konteks luas, termasuk kemampuan berpikir secara logis, sistematis, dan matematika diskret. Tidak hanya ilmu matematika seperti aritmatika, aljabar, kalkulus, himpunan, geometri, dsb. Jika insights yang akan digali dari data berupa model, misalnya model yang dapat digunakan untuk melakukan prediksi di masa depan, maka Machine Learning perlu digunakan.



Keahlian dan Skill Data Scientist

Teknologi, Pemrograman dan Basisdata

Data yang akan dianalisis pastilah tersimpan di suatu (atau beberapa) tempat penyimpanan data. Sistem yang menyimpan dan mengelola data dinamakan sistem basisdata. Sistem ini dapat mengelola data berformat terstruktur (bertipe tabular), semi terstruktur (misalnya data dengan format HTML, CSV, JSON dan XML, juga data spasial atau data geografis), maupun tidak terstruktur (misalnya dokumen, email, foto dan video).



Keahlian dan Skill Data Scientist

Komunikasi, Visualisasi dan Softskill Lainnya

Data Scientist harus menguasai berbagai tools untuk visualisasi data (misalnya Excel, Tableau dll) atau mampu memprogram untuk menghasilkan bentuk visual khusus yang menarik (misalnya distribusi data pada peta).

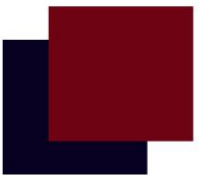


Kebutuhan Data Science

- McKinsey & Company, penyedia layanan konsultasi manajemen dan strategi bisnis, melaporkan bahwa teknologi Artificial Intelligent (AI) yang termasuk Machine Learning makin banyak dibutuhkan karena memberikan keuntungan-keuntungan di bidang bisnis (McKinsey, 2018).
- World Economic Forum (WEC) melaporkan kebutuhan data scientist yang meningkat pada berbagai bidang, misalnya pada industri yang berbasis teknologi informasi, media dan hiburan, layanan finansial dan investasi, layanan profesional, pemerintah, dll. (WEC, 2019).
- Asia-Pacific Economic Cooperation (APEC) pada laporan tahun 2017 menuliskan: *Data Science and Analytics (DSA) skills are in high demand, but supply is critically low with employers facing severe shortages* (APEC, 2017).
- LinkedIn, organisasi yang mengelola jaringan para profesional di Internet yang terbesar, pada laporan tahun 2020 menempatkan data scientist di *top 10 emerging jobs* di berbagai negara, seperti Amerika Serikat (Linkedin-US, 2020), Kanada (Linkedin-CA, 2020), Australia (Linkedin-Aus, 2020). Demikian juga di kawasan ASEAN, seperti Singapore (Linkedin-Sing, 2020) dan Malaysia (Linkedin-Malay, 2020).



Data Science dan Analisis Data e-Commerce



Sistem Rekomendasi dan Collaborative Filtering

Pada saat berbelanja online, platform e-commerce sering memberikan rekomendasi barang untuk masing-masing konsumen

Terlaris Untukmu



makanan pakan
ikan cupang kecil h...

Rp 7.000



Hand Towel
Microfiber Chenille ...

Rp 7.450



Lihat Semua



POLYBAG UKURAN 40x40
ISI 25 LEMBAR

Polybag Ukuran
40x40 Eceran/Lem...

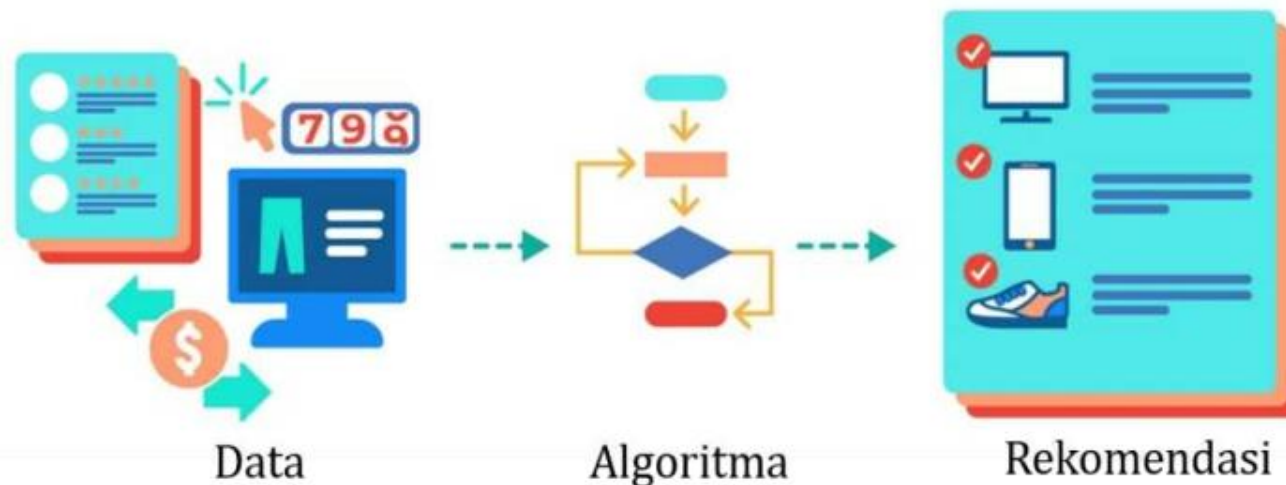
Rp 20.000



Sistem Rekomendasi dan Collaborative Filtering

Berdasarkan [Sivaplan, 1999], sistem rekomendasi pada platform e-commerce secara umum menggunakan data berikut:

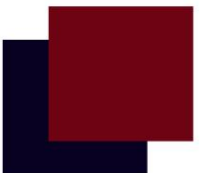
- Rating suatu barang yang dibeli
- Pola perilaku, misalnya, durasi browsing dan jumlah klik
- Transaksi, misalnya, tanggal pembelian, kuantitas, dan harga
- Produksi, misalnya, merk barang, artis favorit, dan topik favorit



Sistem Rekomendasi dan Collaborative Filtering

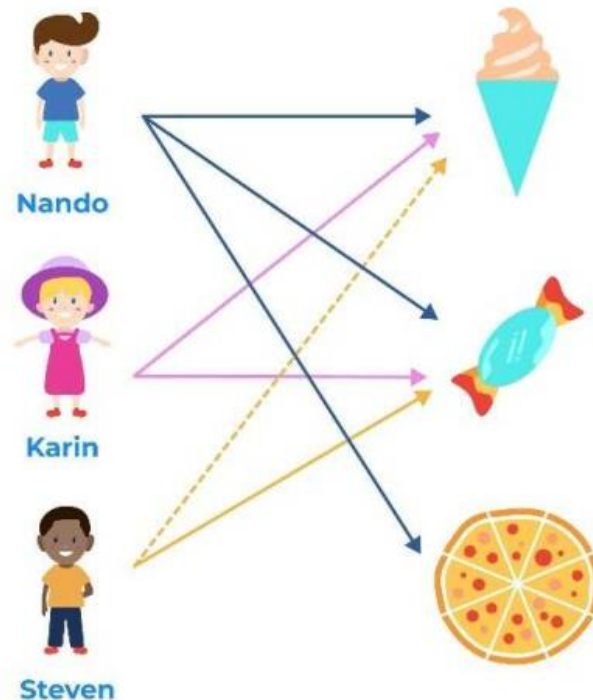
Beberapa jenis sistem rekomendasi. Misalnya, personalized, non-personalized, dan people-to-people correlation.

1. Rekomendasi personalized, rekomendasi diberikan secara otomatis, berdasarkan kesukaan konsumen, misalnya, warna favorit, genre musik, dan genre film [Schafer, 2001].
2. Rekomendasi non-personalized diberikan hanya berdasarkan rating produk yang diberikan oleh konsumen lain [Schafer, 2001].
3. Rekomendasi People-to-people correlation memberikan rekomendasi berdasarkan produk atau rating yang diberikan oleh konsumen lain, yang dianggap mirip dengan konsumen penerima rekomendasi [Sarwar, 2000].



Sistem Rekomendasi dan Collaborative Filtering

Gambar 3.5 memberikan gambaran tentang konsep *collaborative filtering item-based*. Pada gambar tersebut, dapat dilihat bahwa Nando suka permen, es krim, dan pizza. Sedangkan, Karin suka permen dan es krim. Steven menyukai permen. Karena Nando dan Karin, yang menyukai es permen, juga suka es krim, maka seharusnya Steven juga menyukai es krim.



Sistem Rekomendasi dan Collaborative Filtering

Cara kerja sistem rekomendasi dengan algoritma *collaborative filtering item-based* adalah sebagai berikut [Sarwar, 2001]:

1. Mengubah data *rating* konsumen menjadi matriks *rating*. Baris matriks ini mewakili identitas konsumen, sedangkan kolomnya mewakili produk yang diberi *rating*.
2. Membuat *item-to-item similarity matrix*, yaitu matriks yang berisikan nilai kemiripan antara produk satu dengan lainnya. Nilai kemiripan ini, misalnya, didapatkan dengan menghitung korelasi Pearson, jarak Euclidean, atau *cosine similarity measure* antar produk. Hal ini dilakukan untuk mengukur kemiripan produk.
3. Menghitung prediksi *rating* konsumen untuk produk terkait, berdasarkan *similarity matrix* yang sudah didapat sebelumnya.
4. Membuat rekomendasi produk, berdasarkan prediksi *rating* produk, yang telah dihitung sebelumnya.

Sistem Rekomendasi dan Collaborative Filtering

Konsumen	Produk	Rating
Hulk	Armor	5
Hulk	Drone	3
Thor	Helmet	4
Thor	Drone	4
Iron Man	Armor	3
Iron Man	Helmet	5
Black Widow	Armor	4
Black Widow	Helmet	4
Black Widow	Drone	1



	Armor	Helmet	Drone
Hulk	5		3
Thor		4	4
Iron Man	3	5	
Black Widow	4	4	1

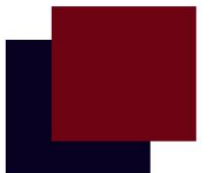


	Armor	Helmet	Drone
Hulk	5		3
Thor		4	4
Iron Man	3	5	
Black Widow	4	4	1

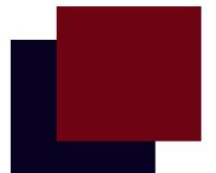
Bangun item-to-item
similarity matrix



Prediksi rating



- Program Data Science. 2020. Pengantar Data Science dan Aplikasinya bagi Pemula. Informatika Universitas Katolik Parahyangan (UNPAR). UNPAR Press
- (APEC, 2017) Asia-Pacific Economic Cooperation (APEC) – Human Resource Development Working Group, Data Science and Analytics Skills Shortage: Equipping the APEC Workforce with the Competencies Demanded by Employers, July 2017.
- (Coursera, 2020) Coursera, Global Skills Index, 2020 (EMC, 2015) EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley Publ., USA, 2015
- Terima Kasih Kepada: Veronica S. Moertini dan Mariskha Tri Adithia





Terima Kasih

