

Menurut Data Robot, Data Science adalah bidang studi yang menggabungkan keahlian domain, keterampilan pemrograman, dan pengetahuan matematika dan statistik untuk mengekstrak wawasan yang bermakna dari data. Tujuannya adalah untuk mengekstrak sebuah pengetahuan atau informasi dari data. Pendapat lainnya, data science adalah keilmuan yang menggunakan data dalam jumlah besar dan algoritma yang rumit, sehingga butuh keahlian ilmu komputer yang mendalam. Dengan kata lain, data science adalah ilmu mengekstraksi pola yang berguna dari kumpulan data dengan menggunakan kekuatan komputer.

Hal Varian, seorang ahli ekonom Google dan dosen Ilmu Komputer, Bisnis, dan Ekonomi UC Berkeley, mengatakan bahwa kemampuan mengambil, memahami, memroses, dan menyaring nilai dari suatu data serta memvisualisasikannya adalah keahlian yang semakin penting di dekade yang akan datang. Semakin banyak perusahaan yang menyadari pentingnya data science, AI, dan machine learning. Terlepas dari industri atau ukurannya, organisasi yang ingin tetap kompetitif di era big data perlu mengembangkan dan menerapkan kemampuan data science secara efisien atau berisiko tertinggal.

Big Data adalah kumpulan proses yang terdiri dari beberapa volume data dengan kapasitas yang besar secara terstruktur maupun tidak terstruktur yang biasanya digunakan untuk kegiatan bisnis. Jenis data yang tersedia lebih banyak, juga bervariasi daripada DBMS (Database Management System), volumenya, dan proses kecepatannya. Pada tahun 2001, Doug Laney memperluas gagasan tentang Big Data. Dia menjelaskan bahwa Big Data mencakup peningkatan variasi data yang dihasilkan oleh suatu organisasi atau perusahaan dan kecepatan dimana data dibuat dan juga diperbaharui. Menurutnya, Big Data adalah data yang berisi variasi yang lebih besar dan datang dalam volume yang meningkat dan dengan kecepatan yang semakin tinggi.

Big data tersebut dikenal dengan “Three Vs” atau 3 huruf V, sebagai berikut.

1. Volume

Ukuran data yang dimiliki oleh big data memiliki kapasitas yang lebih besar. Dapat dilakukan proses data dengan ukuran yang besar untuk menjalankannya.

2. Velocity

Kecepatan transfer data juga berpengaruh dalam proses pengiriman data dengan efektif dan stabil. Big data memiliki kecepatan yang memungkinkan untuk dapat diterima secara langsung (Real-time).

3. Variety

Variasi data yang digunakan oleh big data lebih banyak daripada yang digunakan sistem database SQL. Jenis data yang masih bersifat tradisional, lebih terstruktur daripada data yang lebih terstruktur. Contohnya Text, audio, dan video merupakan data yang belum terdefinisi secara langsung juga harus melalui beberapa tahap untuk diproses dalam sebuah database.

Big data memiliki fungsi penting dalam proses pengembangan dan penyempurnaan dari sebuah aplikasi. Berikut ini adalah beberapa penjelasan mengenai fungsi dari big data.

1. Dapat menentukan penyebab suatu masalah atau kegagalan secara real time

Secara real time, menentukan dan menganalisa suatu penyebab dari permasalahan yang terjadi pada sistem. Kemudian dengan penggunaannya juga dapat meminimalisir terjadinya kegagalan dalam proses penyimpanan data.

2. Pengambilan keputusan yang cerdas dan tepat

Big data juga dapat digabungkan dengan sistem dan teknologi yang cerdas seperti IOT (Internet Of Things) dan AI (Artificial Intelligence), memberikan dan menyimpan data informasi yang dibutuhkan dalam pengembangan sebuah produk.

3. Mendeteksi sebuah anomali atau perilaku yang menyimpang dalam struktur bisnis

Mampu untuk mendeteksi secara cepat dan tepat, bentuk atau proses yang menyimpang dan berhenti karena ada kesalahan dari sisi teknis maupun non teknis. Big data

juga dapat merencanakan beberapa opsi untuk mengurangi dan mengatasi anomali dengan lebih cepat untuk membantu aktivitas bisnis perusahaan atau sebuah organisasi.

4. Mengurangi biaya, waktu, dan meningkatkan performa suatu aplikasi

Penggunaan penyimpanan data yang menggunakan sistem big data akan dapat mengurangi biaya yang harus dikeluarkan oleh perusahaan yang menggunakannya. Kemudian, waktu dalam mengelola dan menjalankan sebuah operasi akan menjadi lebih cepat dengan transfer data diatas rata-rata database lain. Peningkatan performa juga menjadi kelebihan tersendiri untuk mendukung pengembangan perangkat lunak.

Terdapat tiga pilar penting yang mendukung data science, yaitu,

1. Bisnis

Seorang data scientist yang mengolah data berdasarkan ilmu data science harus bisa mengolah data menjadi informasi yang bisa dipahami untuk membantu perancangan strategi guna menyelesaikan masalah bisnis.

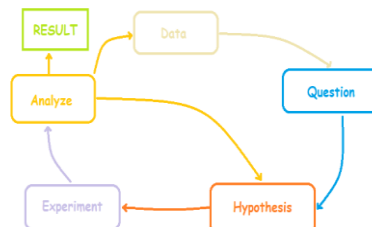
2. Matematika dan Statistika

Data science sangat membutuhkan ilmu matematika karena data harus diolah secara kuantitatif. Banyak permasalahan dalam bisnis yang dapat diselesaikan dengan membuat model analitik dengan dasar matematika.

3. Teknologi

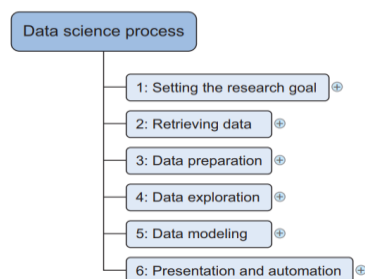
Data science tidak bisa lepas dari teknologi dan kreativitas serta kecerdasan dalam menggunakan keahlian teknis untuk menyelesaikan suatu permasalahan. Seorang data scientist harus mampu berpikir layaknya algoritma dalam memecahkan permasalahan yang paling sulit sekalipun.

Data Scientist adalah seseorang yang dapat mengambil sekumpulan data, mengembangkan kasus penggunaan untuk data tersebut, membuat hipotesis tentang cara menggunakannya, melakukan eksperimen menggunakan hipotesis yang dikembangkan, menganalisis hasil, dan menemukan solusi.



Gambar 1. Siklus analisis-hipotesis-eksperimental

Proses data science secara umum terdiri dari langkah, yang dapat dilihat dalam peta konsep berikut.



Gambar 2. Proses data science

1. Setting the research goal (Atur tujuan penelitian)

Data science sebagian besar diterapkan dalam konteks organisasi. Sebelum melakukan proyek data science, perlu menyiapkan diagram proyek. Ini berisi informasi seperti apa yang akan diteliti, bagaimana manfaat itu untuk perusahaan, data dan sumber daya apa yang dibutuhkan, jadwal, dan output.

2. Retrieving data (Mengambil data)

Langkah kedua yaitu mengumpulkan data. Pada langkah ini, memastikan bahwa dapat menggunakan data dalam proyek, yang berarti memeriksa keberadaan, kualitas, dan akses ke data. Data juga dapat dikirimkan oleh perusahaan pihak ketiga dan dalam berbagai bentuk mulai dari spreadsheet Excel ke berbagai jenis database.

3. Data preparation (Penyiapan data)

Pengumpulan data adalah proses yang rawan kesalahan, kualitas data dapat ditingkatkan dan dipersiapkan untuk digunakan pada langkah-langkah selanjutnya. Terdiri dari 3 langkah, yaitu pembersihan data dengan menghilangkan nilai salah dan inkonsistensi lintas sumber data, integrasi data dengan menggabungkan informasi dari beberapa sumber data, dan transformasi data memastikan bahwa data berada dalam format yang sesuai untuk digunakan dalam model.

4. Data exploration (Mengeksplor data)

Eksplorasi data berkaitan dengan membangun pemahaman yang lebih dalam tentang data. Memahami bagaimana variabel berinteraksi satu sama lain, distribusi data, dan apakah ada outlier. Digunakan statistik deskriptif, teknik visual, dan pemodelan sederhana. Langkah ini sering disingkat EDA, Exploratory Data Analysis.

5. Data modelling or model building (Memodelkan data)

Dalam langkah ini, menggunakan model, pengetahuan, dan wawasan tentang data yang ditemukan pada langkah-langkah sebelumnya untuk menjawab pertanyaan penelitian. Memilih teknik dari bidang statistika, machine learning, riset operasi, dan lainnya. Membangun model adalah proses berulang yang melibatkan pemilihan variabel untuk model, mengeksekusi model, dan diagnosa model.

6. Presentation and automation (Pemaparan dan automasi)

Presentasi hasil ke bisnis. Hasil ini dapat mengambil banyak bentuk, mulai dari presentasi hingga laporan penelitian. Terkadang perlu mengotomatisasi pelaksanaan proses karena bisnis akan menggunakan wawasan yang diperoleh dalam proyek lain atau memungkinkan proses operasional untuk menggunakan hasil dari model.

Salah satu alasan mengapa data science begitu populer adalah banyaknya kemungkinan aplikasi yang muncul. Berikut aplikasinya dalam berbagai bidang.

1. Pemasaran dan Penjualan

Kasus penggunaan khas untuk data science dalam pemasaran adalah rekomendasi produk. Ada algoritma di balik rekomendasi tersebut yang menganggap bahwa pelanggan akan menyukai produk tersebut berdasarkan apa yang sebenarnya dibeli oleh pelanggan lain yang juga melihat produk tersebut.

2. Keuangan

Pada bidang ini, data science digunakan untuk analisis risiko kredit. Terdapat model statistik canggih yang terus diperbarui dan memberikan perkiraan probabilitas default yang baik, membuat keseluruhan proses jauh lebih cepat dan lebih diandalkan.

3. Kesehatan

Kesehatan adalah salah satu industri yang paling menjanjikan dalam hal data science. Ada banyak data yang dihasilkan oleh perangkat yang dikenakan yang terhubung seperti *smartwatches*, termasuk kalori yang dihabiskan, mil berjalan, dan detak jantung. Salah satu aplikasi yang mungkin adalah menjelaskan beberapa penyakit, dan bahkan meningkatkan untuk pergi ke dokter jika menunjukkan perilaku yang mungkin mengindikasikan masalah kesehatan.

Sumber :

Introducing Data Science—Big Data, Machine Learning and More, Using Python Tools

<https://towardsdatascience.com/data-science-101-99e34bea86c>

<https://medium.com/@onejohi/an-introduction-to-data-science-32403b22f5c1>

<https://qlints.com/id/lowongan/data-science-adalah/#.YwT7IXZBy5d>

<https://www.datarobot.com/wiki/data-science/>

<https://penerbitbuku.co.id/pengertian-big-data/>