**Rangkuman Fundamental Data Science**

**Amazon** mengemukakan bahwa data science adalah wawasan yang sangat berguna untuk kebutuhan bisnis. Ilmu ini merupakan penggabungan prinsip dan praktik dari bidang matematika, statistik, *artificial intelligence,*dan *computer engineering* untuk menganalisis data dalam jumlah besar.

Menurut Jose **Antonio Ribeiro** pada publikasinya di*Big Data for Executives and Professionals*, data science adalah kegiatan ilmiah yang mencakup pemahaman bisnis, persiapan data, analisis eksplorasi, pemodelan, aplikasi komputasi, evaluasi wawasan, dan tindakan untuk menghasilkan hasil serta  menerapkan pengetahuan statistik, matematika, dan komputasi.

Adapun definisi lain, menurut IBM (*International Business Machine*) menyebutkan bahwa data science adalah menggabungkan matematika dan statistik, pemrograman, analitik, *artificial intelligence*, dan *machine learning*.

**Data Science vs. Data Scientist vs. Data Engineering**

| **Data Science** | **Data Scientist** | **Data Engineering** | **Data Analyst** |
| --- | --- | --- | --- |
| Data Science merupakan disiplin ilmu. | Data scientist adalah praktisi dalam bidang data science. | Data engineer merupakan software engineer yang tugas utamanya adalah menyiapkan data yang dikoleksi dari berbagai sumber informasi dengan membangun sistem data atau infrastruktur agar dapat mudah untuk dianalisa dan mendukung kebutuhan perusahaan. | Data analyst adalah profesi pekerjaan yang bertugas untuk membaca serta menganalisis data yang ada dari perusahaan. |

**Mengapa Data Science Begitu Penting?**

Sederhananya data science membantu organisasi atau perusahaan mendapatkan *insight*(wawasan) tentang bisnis mereka sendiri, tingkat keberhasilan strategi mereka, kinerja anggotanya, dan sebagainya.

**Dampak Data Science di Dunia**

1. **Promosi pelestarian lingkungan**. Tahukah Anda? Selain digunakan untuk membuat sebuah keputusan yang baik untuk organisasi atau perusahaan, data science memiliki peran penting dalam keberhasilan pengelolaan krisis iklim global, *lho*. PBB mengembangkan 17 tujuan pembangunan berkelanjutan atau *Sustainable Development Goals*(SDGs) untuk mengukur dan melacak bagaimana negara menyesuaikan faktor lingkungan, sosial, dan ekonomi untuk memperlambat perubahan iklim.
2. **Kemajuan bidang kesehatan**. Selama pandemi COVID-19, layanan kesehatan dan lembaga pemerintah menggunakan data analyst untuk melacak kasus, memroyeksikan penyebaran infeksi, menginformasikan keputusan kebijakan kesehatan, dan meningkatkan hasil layanan kesehatan.
3. **Pemberdayaan negara berkembang**. Setiap tahun, Bank Dunia menilai negara-negara dalam skala 1-100. Nilai 1 mewakili kurangnya data statistik penting dan kapasitas analisis, dan 100 mewakili kapasitas statistik negara maju. Selama pandemi, 80% kantor statistik nasional di negara berpenghasilan rendah hingga menengah memerlukan dukungan tambahan untuk melakukan pengumpulan dan analisis data penting.
4. **Peningkatan *customer insight.***Salah satu dampak paling menguntungkan yang dimiliki data science terhadap bisnis adalah mengevaluasi, memprediksi, dan memberikan rekomendasi yang akurat dan dipersonalisasi kepada pelanggan.
5. **Mengembangkan pemimpin yang lebih cerdas**. Data scientist dan pemimpin bisnis telah bekerja sama untuk lebih memahami cara sebuah data dapat menjadi jembatan untuk menginformasikan strategi organisasi yang lebih baik, menginformasikan keputusan, dan mencapai hasil yang lebih baik.

**Kalkulus**

Sebagian besar bidang data science tentunya memerlukan pemahaman tentang prinsip dasar kalkulus dan pengaruhnya terhadap model *machine learning.*Namun, perlu digaris bawahi bahwa kalkulus pada data science tidak seperti kalkulus pada kelas matematika di sekolah atau perguruan tinggi. Berikut merupakan beberapa konsep kalkulus yang mungkin bisa digunakan data scientist.

* Penurunan gradient (*Gradient descent*). Pengoptimalan algoritma yang melatih model *machine learning* dari waktu ke waktu dan menjadi lebih akurat.
* Kalkulus multivariabel (*Multivariate Calculus*). *Machine learning*menggunakan kalkulus multivariabel untuk membangun model prediktif.

**Statistika**

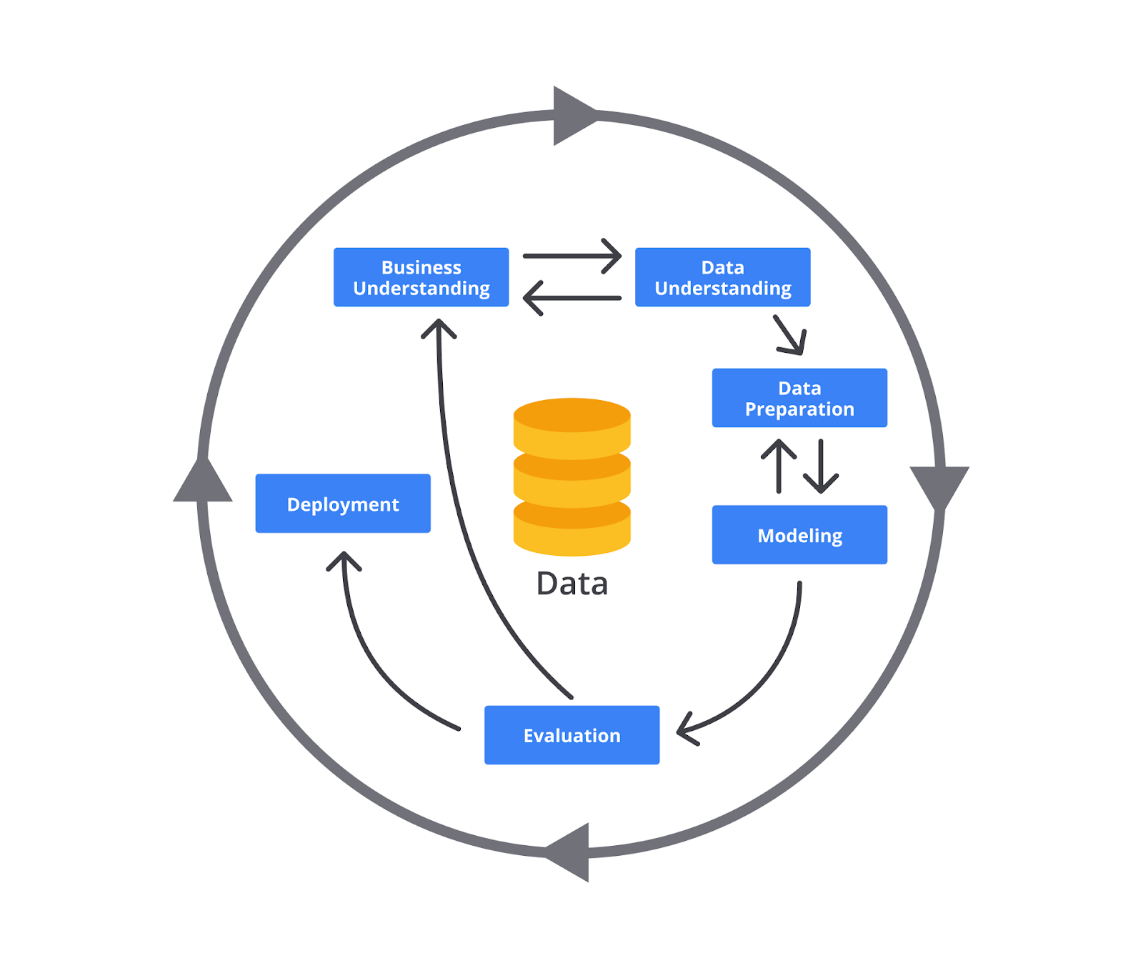
Sejauh ini, statistik adalah matematika terpenting yang perlu Anda ketahui untuk data science. Statistik adalah cabang matematika yang mengumpulkan data dan menganalisis kumpulan data besar untuk menginterpretasikan wawasan yang bermakna. Data scientist menggunakan statistik, sebagai berikut.

* Kumpulkan (*collect*), tinjau (*review*), analisis (*analyze*), wawasan dari data (*insight from data*).
* Identifikasi dan analisis pola data menjadi *insight*(wawasan) bisnis yang dapat ditindaklanjuti.
* Jawab pertanyaan dengan membuat eksperimen, menganalisis, dan menafsirkan kumpulan data.
* Memahami *machine learning* dan model prediksi (*predictive model*).

**Aljabar Linier**

Aljabar Linier adalah salah satu keterampilan matematika yang banyak digunakan dalam beberapa bagian, diantaranya pada sebagian besar *machine learning* dapat diekspresikan dalam bentuk matriks, kemudian pada deep learning ketika membuat artificial neural network (jaringan syaraf tiruan). Data Scientist memanipulasi dan menganalisis data mentah melalui matriks, baris, dan kolom angka atau titik data.

**Metodologi Data Science**

[](https://www.dicoding.com/academies/615/tutorials/33118)

CRISP-DM memiliki enam tahapan yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation,*dan *Deployment*.

1. ***Business Understanding***. Tahap ini merupakan pemahaman mendalam tentang kebutuhan pelanggan. Kegiatan yang dilakukan pada tahap ini adalah menentukan tujuan bisnis, menilai situasi ketersediaan sumber daya, tentukan tujuan pengumpulan data, dan menghasilkan rencana proyek.
2. ***Data Understanding***. Tahap pemahaman data yaitu mendorong fokus untuk mengidentifikasi, mengumpulkan, dan menganalisis kumpulan data yang dapat membantu Anda mencapai tujuan proyek. Kegiatan pada tahap ini adalah mengumpulkan data awal, menjelaskan data, jelajahi data, dan verifikasi kualitas data.
3. ***Data Preparation***. Fase ini sering disebut “data munging”, yaitu menyiapkan kumpulan data akhir untuk pemodelan. Kegiatan pada fase ini diantaranya memilih data yaitu menetukan set data yang akan digunakan, lalu bersihkan data, bangun data atau dengan kata lain mendapatkan atribut baru yang akan membantu, lalu integrasikan data, dan yang terakhir format data.
4. ***Modeling***. Secara garis besar pada tahap ini Anda akan membuat dan menilai berbagai model berdasarkan beberapa teknik pemodelan yang berbeda. Pada tahap ini terdapat empat tugas, yaitu memilih teknik pemodelan, menghasilkan desain pengujian, membangun model, dan yang terakhir menilai model.
5. ***Evaluation***. Fase evaluasi ini melihat lebih luas model yang paling sesuai dengan bisnis dan yang harus dilakukan selanjutnya. Terdapat tiga kegiatan yang mewakili fase evaluasi, yaitu evaluasi hasil, proses peninjauan, dan tentukan langkah selanjutnya.
6. ***Deployment***. Merupakan tahap terakhir dan tahap yang paling dihargai dari proses CRISP-DM. Perencanaan untuk *deployment* dimulai dari fase *business understanding*dan harus menggabungkan tidak hanya untuk menghasilkan nilai model, tetapi juga cara mengonversi skor keputusan, dan cara untuk menggabungkan keputusan tersebut.

**Keterampilan Data Scientist**

Menjadi seorang data scientist tentu bukan hanya keterampilan teknis yang diperlukan tetapi perlu adanya keseimbangan antara softskill dan hardskill untuk menunjang karier Anda di masa depan. Pada pembahasan kali ini, Anda akan mengeksplorasi keterampilan yang dibutuhkan agar menjadi seorang data scientist yang Andal

**Keterampilan Interpersonal**

1. Komunikasi Efektif
2. Pengetahuan
3. Memiliki Pemahaman Bisnis
4. Kerja Sama
5. Penyelesaian Masalah (*Problem Solving*)

**Keterampilan Teknis**

1. Statistik dan Probabilitas
2. Mengolah Data (Data wrangling dan database management)
3. Pemrograman
4. Machine Learning
5. Analisis Data (Data Analyst)
6. Cloud Computing