



LEARNING PROGRESS REVIEW WEEK-1

OPTIMISTIC TEAM

ALDIVA WIBOWO

LUTFIA HUMAIROSI

ASPRIZAL RIZKY

MILLENIA WINADYA P

GILANG RAHMAT R

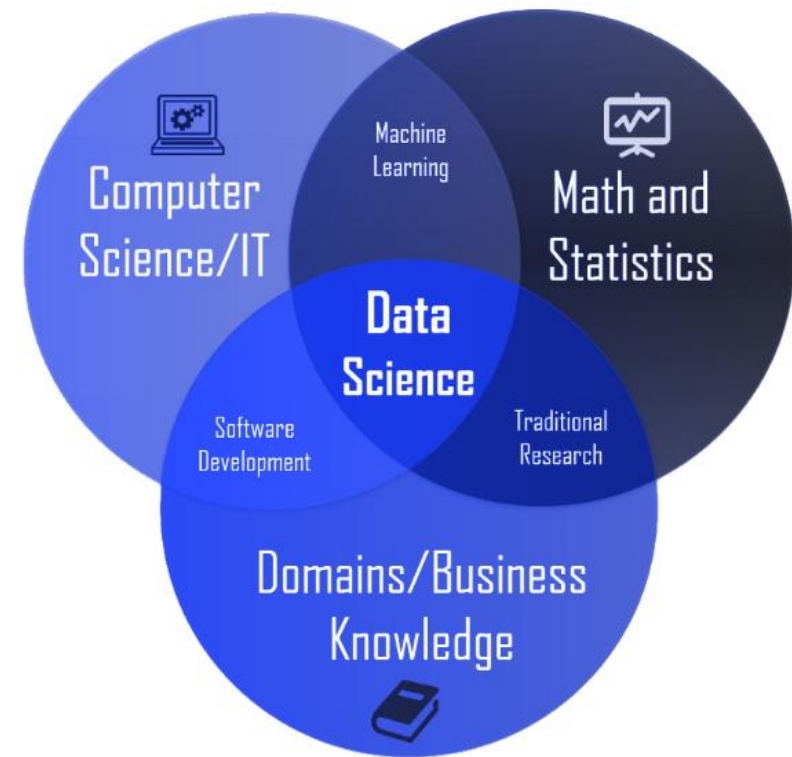
01

INTRODUCTION TO DATA SCIENCE



Apa Itu Data Science?

Data science merupakan bidang studi yang menggabungkan berbagai macam bidang khususnya dari domain expertise, komputer science, pemahaman matematik, dan statistik untuk mendapatkan insight dari sebuah data.



Mengapa Data Science?

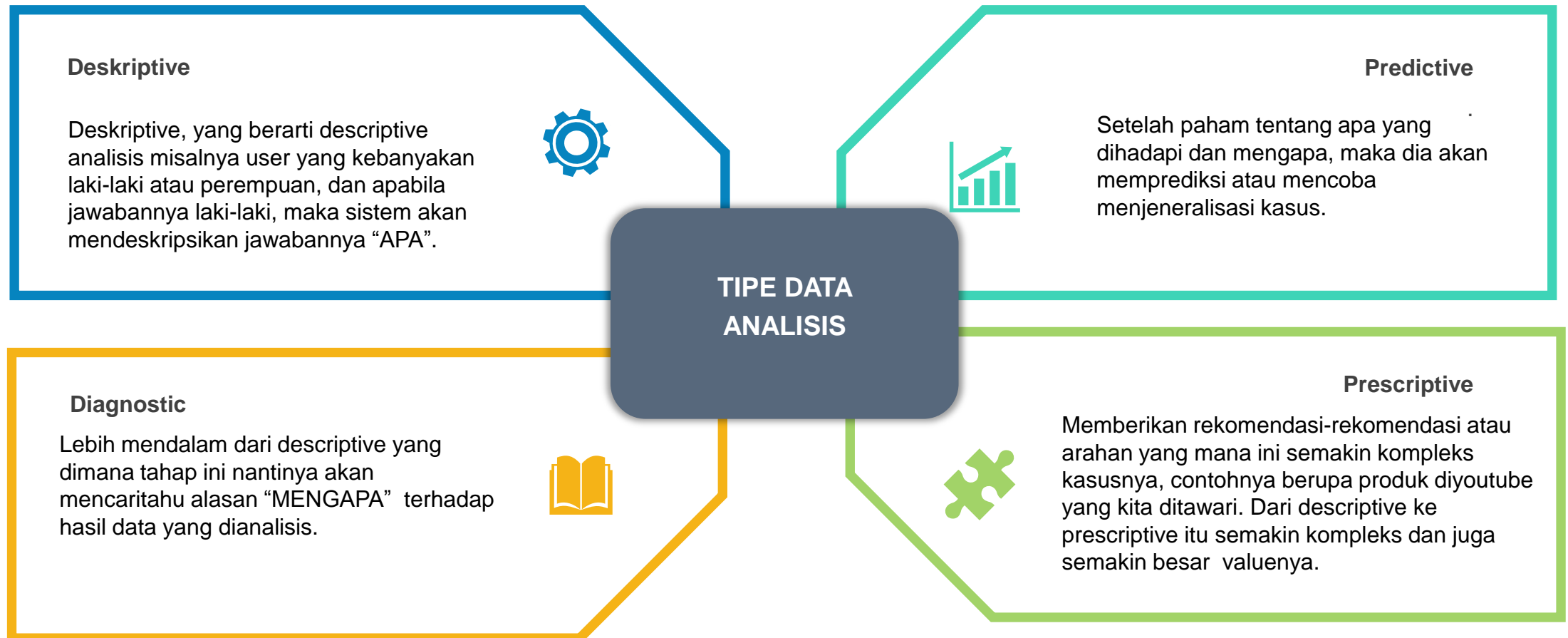


Menurut Forbes, pada tahun 2016 ada sekitar 2.900 postingan untuk mencari seorang data scientist di setiap bulannya di seluruh dunia. Menurut McKinsey Global Institute pada tahun 2019 diperkirakan hampir 200.000 membuka posisi tersebut.

berdasarkan riset dari IBM, kebutuhan sumber daya manusia di bidang *Data Science* mencapai dua juta jiwa pada tahun 2020.

Selain itu, pekerjaan terkait keilmuan ini menempati ranking pertama dari 10 profesi paling masa depan.

Produk Data Science



Karir Di Data Science



Data Scientist

Seorang Data Scientist menganalisa data , Mengembangkan pemodelan, dan Mencari insight untuk menghasilkan produk tertentu.



Data Engineer

Seorang data engginer mengubah data mentah menjadi data siap pakai Menggunakan bahasa pemrograman tingkat lanjut, Memasang machine learning, Mengatur infrastruktur data



Data Analyst

Seorang Data Analyst Mengeksplorasi data, Menganalisa data, dan Menjelaskan serta memvisualisasikan hasilnya.

02

INTRODUCTION TO DATA SCIENCE METHODOLOGY

Data Analytics Lifecycle

Data science pada dasarnya digunakan untuk mendukung pengambilan keputusan (decision support) yang tepat pada suatu bisnis dengan menggunakan hasil pengolahan data.

Data Analytics Lifecycle

Merupakan pola untuk melakukan analisis data dalam menyelesaikan permasalahan bisnis melalui pendekatan yang mendorong nilai bisnis serta inovasi melalui peningkatan berkelanjutan (continues improvements).





Data Analytics Lifecycle

- **Planning Analytics (Business Understanding):** mengidentifikasi tujuan bisnisnya. Intinya adalah merencanakan pemecahan masalah berdasarkan pemahaman bisnis yang sesuai.
- **Descriptive Analytics (Performance Analytics):** mengidentifikasi apa yang terjadi dalam suatu bisnis. Dengan tujuan melihat gambaran keadaan bisnis melalui analisa kinerja berdasarkan data yang tersedia.
- **Diagnostic Analytics (Identify Business Causes):** mengidentifikasi mengapa hal itu terjadi dalam bisnis. Tujuannya untuk melihat apa yang menyebabkan bisnis tersebut terjadi masalah kemudian menentukan cara pemecahan masalah yang dapat diterapkan pada bisnis.
- **Predictive Analytics (Predict Using Data Models):** mengidentifikasi apa yang akan terjadi selanjutnya. Dengan menggunakan data yang telah dikumpulkan kemudian membuat model untuk memprediksi hasil dari pemecahan masalah.
- **Prescriptive Analytics (Solution Optimization):** mengidentifikasi apa yang harus dilakukan untuk memecahkan masalah. Dengan mengoptimalisasi model yang telah dibuat sebelumnya supaya dapat membuat solusi pemecahan masalah yang lebih baik lagi untuk bisnis.





Data Science Methodology

Data Science Methodology

Merupakan framework yang dipakai oleh Data Scientists untuk memproses data dengan metode-metode, proses-proses, dan heuristik-heuristik apa saja yang akan digunakan untuk memperoleh jawaban atau hasil.

Framework dibutuhkan Data Scientists untuk menangani data yang tidak terstruktur serta melakukan prescriptive analytics.

Diterapkan pada data yang memiliki volume sangat besar dan ketika terdapat penggabungan dari ML ke dalam pemodelan preskriptif dan otomatisasi.

Tujuan dari Data Science Methodology adalah untuk membagikan metode yang dapat digunakan untuk dalam data science serta untuk memastikan bahwa data dapat digunakan untuk mengatasi masalah yang relevan dan layak dimanipulasi untuk menjawab pertanyaan yang ada.





Data Science Methodology

Tahapan dalam Data Science Methodology:

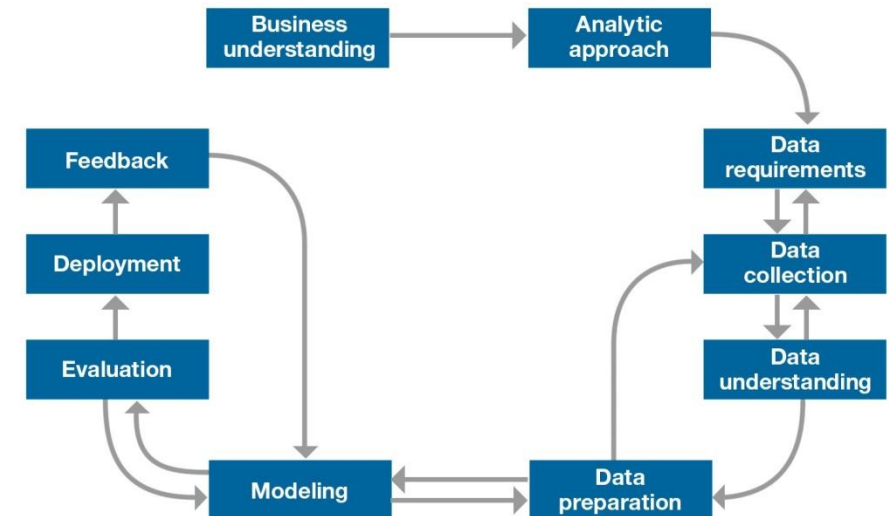
1. Pendekatan masalah, meliputi

- Business Understanding

Pertanyaan yang diajukan pada tahap ini adalah “Apa permasalahan yang ingin kamu coba pecahkan?”. Ini dilakukan untuk mendapatkan gambaran permasalahan yang harus dipecahkan/diselesaikan. Yang harus dilakukan pada tahap ini yaitu menentukan tujuan (goals) dan menganalisa cara untuk mencapai tujuan tersebut (objectives) dimana diikuti dengan pemahaman tentang business metrics yang tersedia.

- Analytic Approach

Pertanyaan yang diajukan pada tahap ini adalah “Bagaimana kamu dapat menggunakan data untuk memecahkan permasalahan?”. Ini dilakukan untuk mencari tahu bagaimana cara menyelesaikan masalah tersebut. Pendekatan yang dipilih nantinya harus sesuai dengan konteks kebutuhan bisnis.





Data Science Methodology

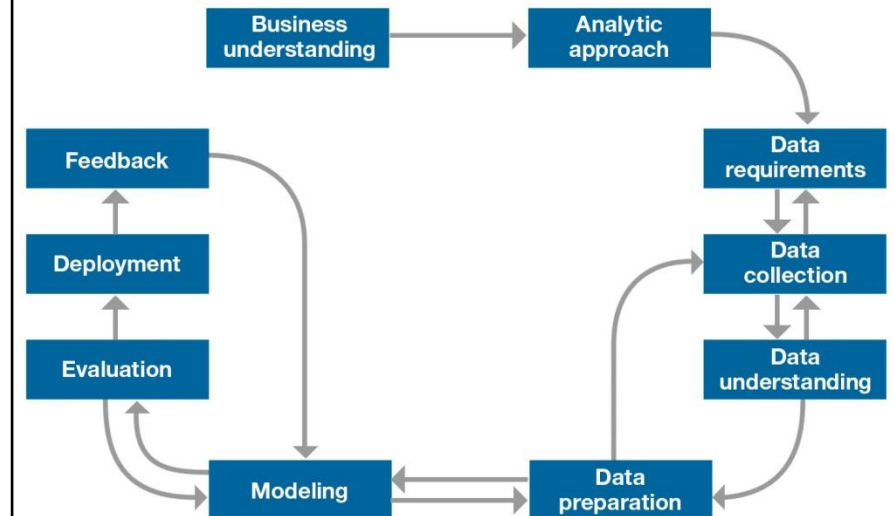
2. Bekerja dengan data, meliputi

- Data Requirements

Pertanyaan: “Data apa yang dibutuhkan untuk memecahkan masalah?”. Dengan menentukan data apa saja yang dibutuhkan, bagaimana format datanya, banyaknya data yang harus digunakan, dan darimana data akan diambil.

- Data Collection

Pertanyaan: “Data berasal darimana dan bagaimana cara mendapatkannya?”. Tahap ini menggunakan statistik deskriptif dan visualisasi untuk menilai konten, kualitas, serta insights awal tentang data.





Data Science Methodology

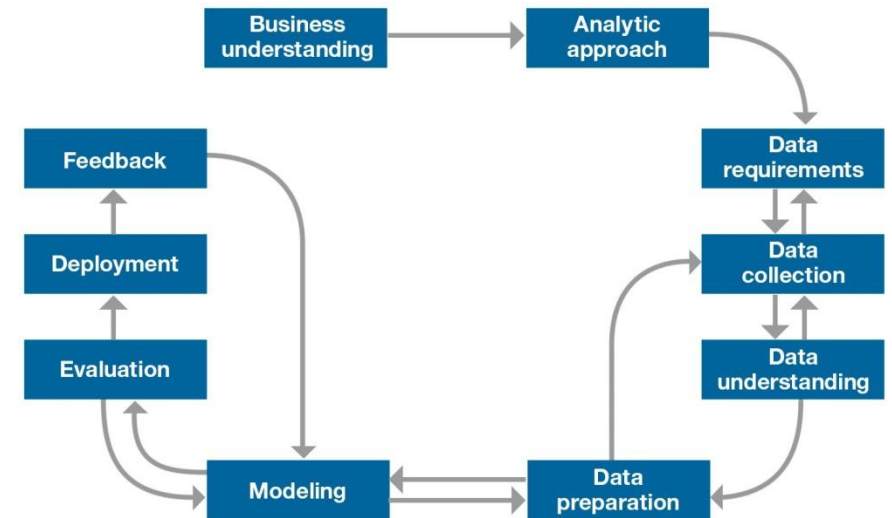
3. Memperoleh solusi, meliputi

- **Modelling**

Pertanyaan: “Cara apa yang dapat digunakan untuk memvisualisasikan data agar masalah dapat terselesaikan?”. Tahap ini melakukan pemodelan data baik dengan pendekatan descriptive analytics (membuat visualisasi dashboard untuk menjelaskan jawaban/hasil) maupun dengan predictive analytics (memerlukan pemodelan dengan bantuan ML).

- **Evaluation**

Pertanyaan: “Apakah model yang dihasilkan benar-benar dapat menyelesaikan permasalahan ataukah masih perlu dilakukan penyesuaian?”. Tahap ini dilakukan secara iteratif selama pengembangan model dan sebelum model digunakan. Penyesuaian model dilakukan dengan evaluation metrics (accuracy, precision, RMSE, dsb) tergantung dari masalah yang ada.





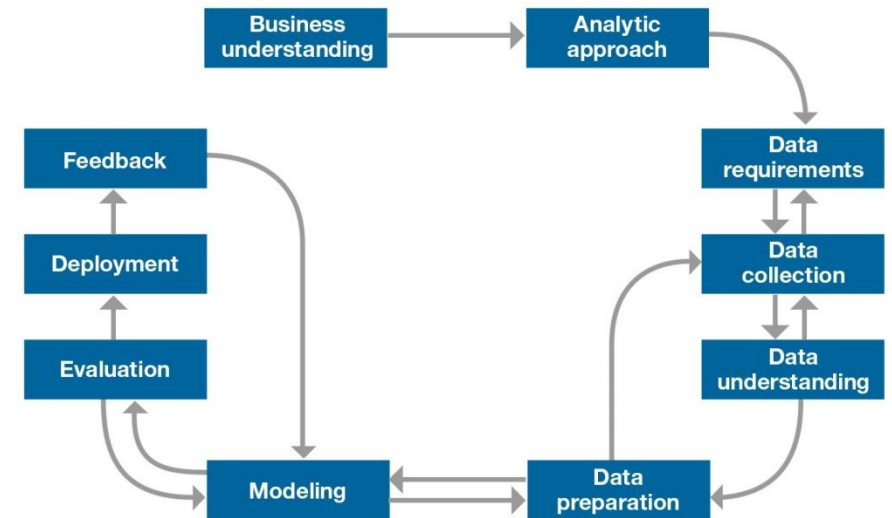
Data Science Methodology


- Deployment

Pertanyaan: “Bagaimana cara mengimplementasikan model pemecahan masalah tersebut?”. Tahap ini dimaksudkan untuk meletakkan hasil model yang terbaik pada sebuah platform supaya bisa dilakukan otomatisasi.

- Feedback

Pertanyaan: “Apakah terdapat feedback yang konstruktif terhadap terselesaikannya suatu masalah?”. Masukan yang baik dapat digunakan untuk mengembangkan model selanjutnya.





TERIMA KASIH