



LEARNING PROGRESS REVIEW

Week 1

Entropy Team

DAFTAR ISI

1.

Introduction to Data Science

Pengenalan bidang *data science*

2.

Data Science Methodology

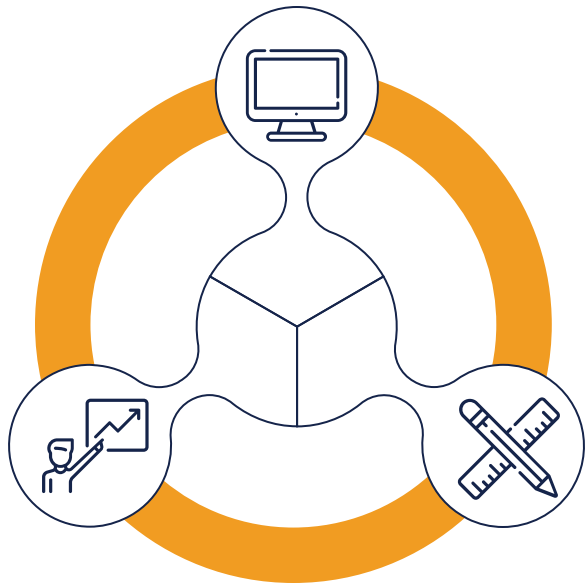
Metodologi yang digunakan dalam *data science*

01

INTRODUCTION TO DATA SCIENCE

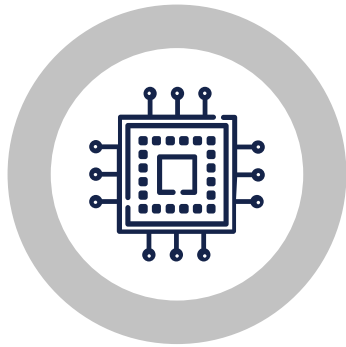
Pengenalan bidang
data science

Data Science



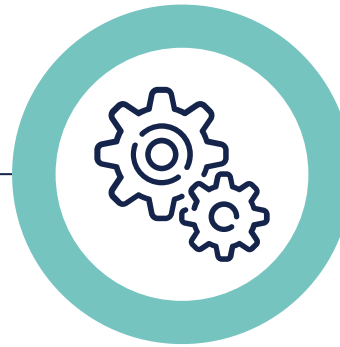
- *Data science* adalah bidang ilmu multidisiplin yang menyatukan **statistik**, **pemrograman**, dan **bisnis**
- *Data science* dapat digunakan untuk menemukan ***insight***, menentukan **keputusan**, dan atau men-**deploy** sebuah model dari kumpulan data

Bagian Data Science



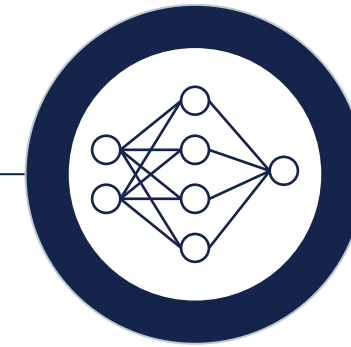
Artificial Intelligence

Mesin dapat **menjalankan tugas** yang biasanya butuh kecerdasan **manusia**



Machine Learning

Bagian dari AI untuk **melatih** mesin agar memahami sebuah **pola** dari data



Deep Learning

Neural network, salah satu **model** dari ML yang terinspirasi oleh **struktur otak** manusia

Machine Learning

Supervised Learning

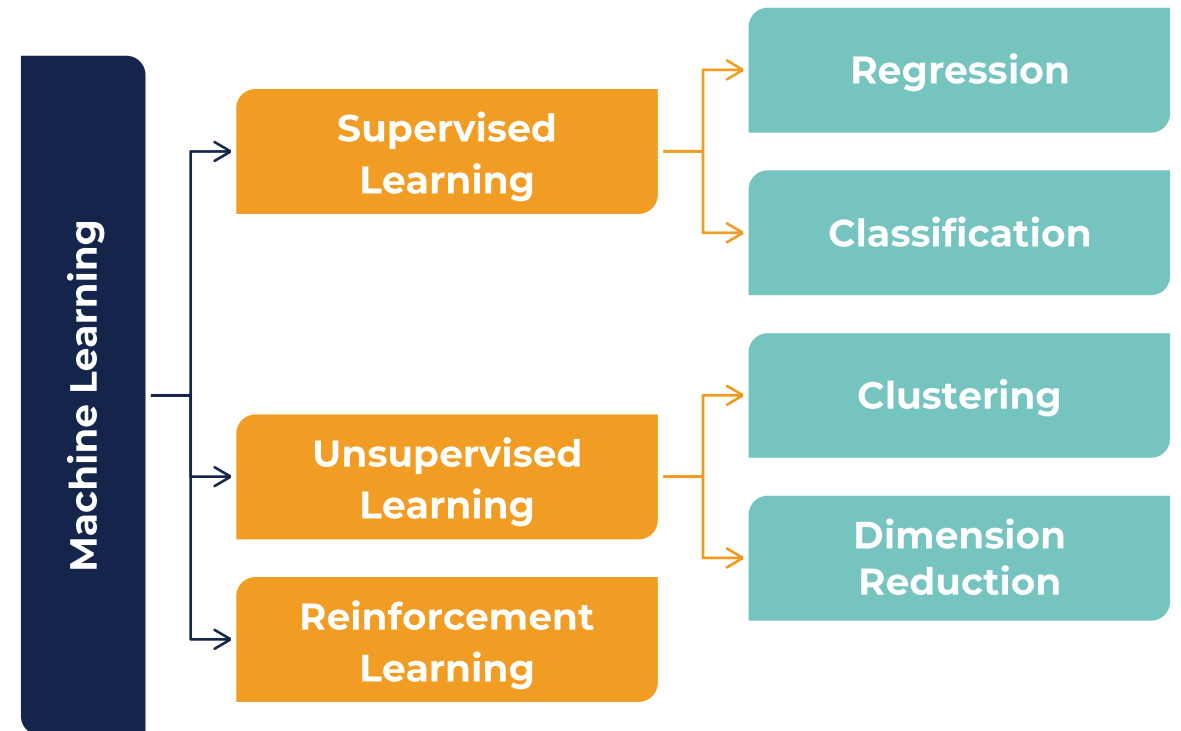
Sistem belajar berdasarkan dataset yang **memiliki label**

Unsupervised Learning

Sistem belajar berdasarkan dataset yang **tidak memiliki label**

Reinforcement Learning

Sistem belajar berdasarkan **reward** dan **punishment** dari pengalaman yang dilakukan secara berulang



Exploratory Data Analysis



- *Exploratory data analysis (EDA)* yaitu proses untuk mengeksplorasi data dan **menemukan *insight***
- *Data scientist* tidak selalu harus *men-deploy ML model*, bisa jadi hanya melakukan EDA

Alur Data

Pengumpulan data
dari berbagai sumber

Raw Data



Persiapan data untuk
analisis dan *modeling*

Data Preprocessing



Penyampaian *insight*
dari data

Data Insight



Data Warehouse

Sentralisasi dan
manajemen data

Data Processing

Analisis data dan atau
pembuatan model

Role dalam Data Science

	DE	DA	DS	MLE
RAW DATA	●	●	●	●
DATA WAREHOUSE	●	●	●	●
DATA PREPROCESSING	●	●	●	●
DATA INSIGHT	●	●	●	●
DATA MODELING	●	●	●	●
MODEL PERFORMANCE	●	●	●	●
MODEL DEPLOYMENT	●	●	●	●

DE : Data Engineer
DA : Data Analyst
DS : Data Scientist
MLE : Machine Learning Engineer

Disclaimer:

- Jobdesc tiap role pada tiap perusahaan bisa saja berbeda
- Jobdesc dari suatu role bisa saja merangkap role lain

Implementasi Data Science



- Mendeteksi dan **mencegah** adanya tindakan **penipuan** dengan melihat aktivitas mencurigakan secara *real time*



- Menilai **kelayakan** kredit seseorang atau bisnis kecil yang mengajukan **pinjaman**



- Menampilkan **konten** yang **dipersonalisasi** berdasarkan data aktivitas pengguna untuk meningkatkan pengalaman pengguna



- Membuat **playlist** lagu pada platform musik berdasarkan data lagu yang dimainkan menurut **genre** atau **emosi** yang dirasakan

02

DATA SCIENCE METHODOLOGY

Metodologi yang digunakan
dalam *data science*

Metodologi Data Science

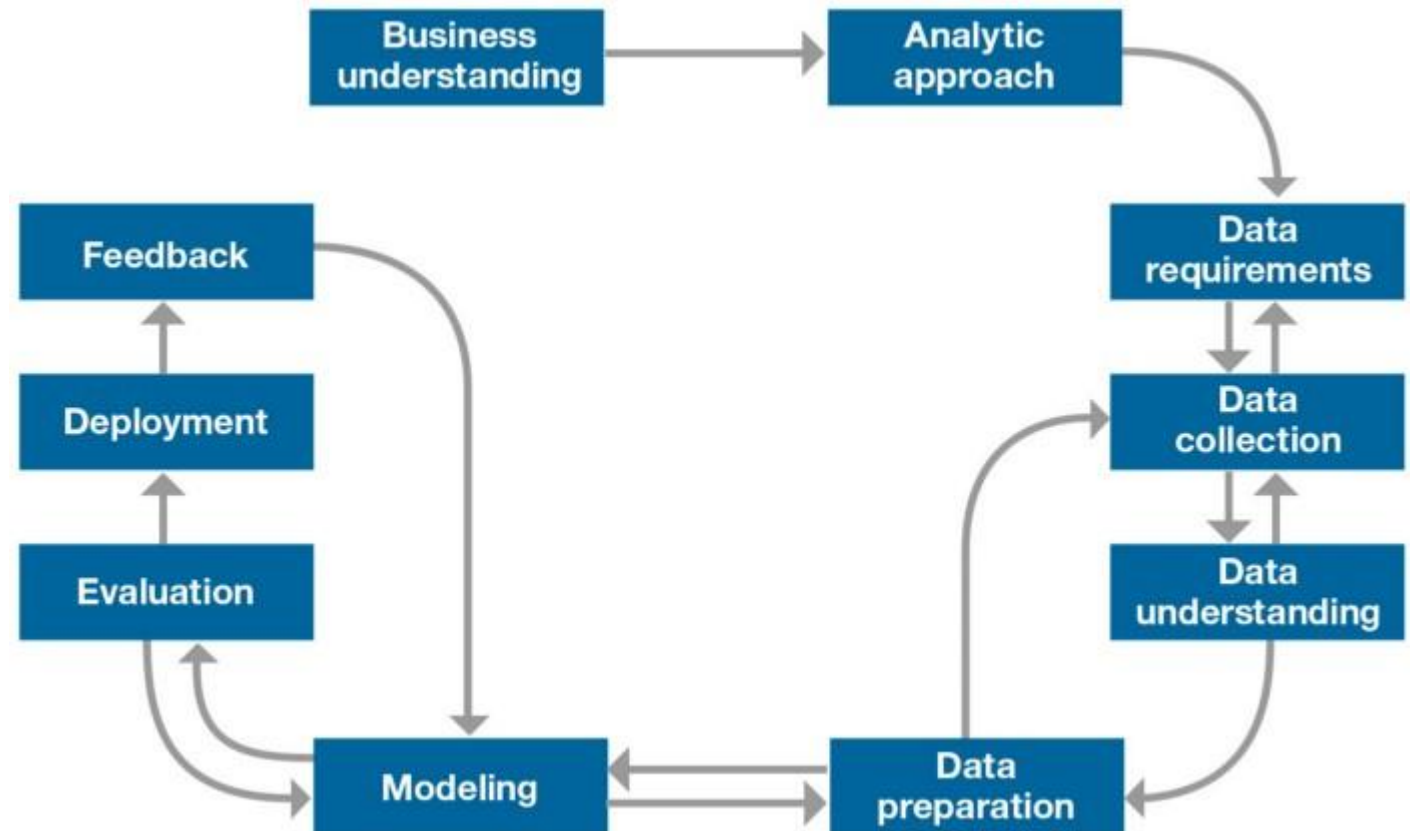


- **Langkah** yang digunakan dalam proyek *data science* agar mendapatkan hasil yang optimal, yaitu dapat **menjawab pertanyaan** dari suatu masalah yang ingin diselesaikan
- Metodologi ini **tidak bergantung** pada teknologi atau *tools* tertentu

Tahapan Metodologi

Terdapat 10 tahapan metodologi dengan metode CRISP-DM*

**Cross-industry standard process for data mining*



Sumber: <https://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science>

Business Understanding

What is the problem that you are trying to solve?

Masalah apa yang sedang coba diselesaikan



- Merupakan langkah awal yang perlu dilakukan dalam metodologi *data science*.
- *Business understanding* dilakukan dengan tujuan untuk **memahami** permasalahan bisnis yang ingin ditangani
- Mendefinisikan **permasalahan, tujuan** yang akan dicapai, serta kebutuhan **solusi** yang diperlukan dari sudut pandang **bisnis**

Analytic Approach

How can you use data to answer the question?

Bagaimana menggunakan data untuk menjawab pertanyaan yang ada



- Tahap mendefinisikan masalah dalam **konteks statistik** untuk memperoleh hasil yang diinginkan secara efektif
- Pendekatan analitik yang dipilih **bergantung permasalahan** yang akan diselesaikan dalam konteks *business requirement* yang telah didapat dalam tahap *business understanding*
- Metode analisa meliputi **descriptive analytics**, **diagnostic analytics**, **predictive analytics**, **prescriptive analytics**

Data Requirement

What data do you need to answer the question?
Data yang dibutuhkan untuk dapat menjawab pertanyaan



- Menentukan **data** apa saja yang **dibutuhkan** untuk menjawab permasalahan yang telah dirumuskan
- Penggunaan data-data yang tidak sesuai akan berakibat hasil yang didapat kurang memuaskan

Data Collection

Where is data coming from & how will you get it?

Dari mana data berasal dan bagaimana cara mendapatkannya



- Menentukan **sumber** data yang tersedia, bisa dari data **internal** dan **eksternal**
- Menentukan bagaimana **cara mengumpulkan** data yang diperlukan
- Apabila data yang dibutuhkan **tidak tersedia**, maka perlu dilakukan perbaikan kembali pada tahap *data requirement* seperti menambahkan atau menyederhanakan data

Data Understanding

Is the collected data representative of the problem to solved?

Apakah data merepresentasikan masalah yang ingin diselesaikan



- Mempelajari dan melihat **kualitas data**
- Melakukan **pengecekan data** untuk mengetahui apakah masih terdapat *missing values, imbalanced data, outlier*, salah format dan lain-lain yang **memerlukan perbaikan**
- Dapat menggunakan **statistik deskriptif** dan teknik **visualisasi** untuk membantu memahami isi data, menilai kualitas data, dan menemukan **insight awal** dari data tersebut

Data Preparation

What additional work is required to work with the data?

Proses tambahan apa yang diperlukan untuk mengolah data



- Proses **membangun dataset** yang akan digunakan dalam tahap pemodelan
- Terdiri dari **pembersihan** data, **penggabungan** data, dan **pengubahan** data menjadi variabel yang lebih berguna
- *Data preparation* merupakan tahapan yang paling **memakan** banyak **waktu**. Sebagian *data scientist* mengatakan tahap ini bisa menghabiskan sekitar **70%-80%** dari keseluruhan proses dalam proyek *data science*

Modeling

How can the data be visualized to get the answer?

Bagaimana data dapat divisualisasikan untuk memperoleh hasil?

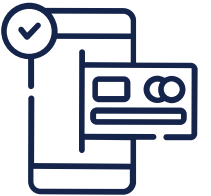


- Pemodelan data berfokus pada **mengembangkan model**, baik itu model deskriptif atau prediktif yang tergantung pada *analytical approach*
- Model **deskriptif** adalah model yang menggambarkan korelasi atau implikasi antardata
- Model **prediktif** adalah jenis model yang menghasilkan prediksi ketika mendapat data baru

Deployment

Can you put the model into practice?

Dapatkah model tersebut dipraktikkan atau diterapkan?



- Ketika **model** yang telah dikembangkan sudah dapat **diterima** dan **disetujui** oleh pemilik proyek/ *stakeholder*, maka model tersebut siap untuk digunakan

Feedback

Can you get constructive feedback to answer the question?

Bisakah mendapat umpan balik yang membangun



- Model yang telah di-*deploy* akan mendapatkan *feedback*
- *Feedback* tersebut dapat digunakan untuk **analisa**, **perbaikan**, dan **pengembangan** model untuk meningkatkan performa dan kegunaannya

THANKS

Entropy Team



Adhang Muntaha Muhammad

<https://www.linkedin.com/in/adhangmuntaha/>



Aziz Fauzi

<https://www.linkedin.com/in/aziz-fauzi-a6904711b/>



Iwan Wahyu

<https://www.linkedin.com/in/iwan-wahyu-setyawan-506809183>



Marcellina Alvita F

<https://www.linkedin.com/in/marcellina-alvita-faustina-63a284226>



Ramadhan Luthfan

<https://www.linkedin.com/in/luthfan-mahathir-91369b18b>

CREDITS: This presentation template was originally created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**