

Tugas Personal ke – 1

1. Penjelasan Tentang Data Science

- a. Data Science adalah ilmu tentang bagaimana cara mengekstrak pengetahuan secara umum dari data, bukan hanya mengolah angka. Untuk workflow dalam data science, ada beberapa tahapan, :

- i. Pengumpulan data (Data Collection)

Pada tahapan ini kita akan menjalani proses pengumpulan data dari berbagai sumber, seperti database, file excel/csv, data public seperti Kaggle.

- ii. Pembersihan Data (Data Cleaning)

Tahapan ini dilakukan untuk proses membersihkan data dari, duplikasi, missing values, menstandarisasi format data, dan memperbaiki data yang outlier atau diluar konteks table.

- iii. Eksplorasi Data (Data Understanding/exploration)

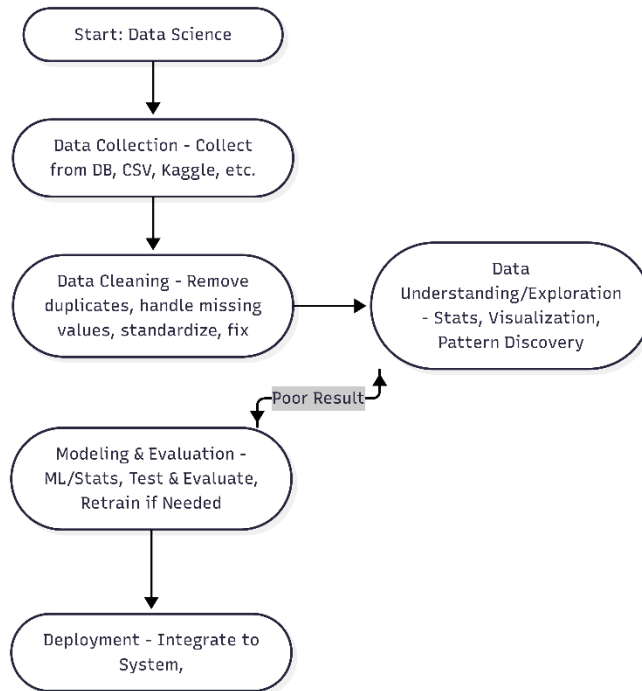
Eksplorasi data untuk memahami karakteristik data, lalu menggunakan statistic deskriptif dan visualisasi. Tujuannya untuk menemukan pola dan hubungan antara data

- iv. Modeling & Evaluation

Menerapkan algoritma Machine Learning atau statistic, : klasifikasi, regresi, dan klustering. Lalu akan di evaluasi dengan train training, jika hasilnya buruk, akan kembali ke tahap eksplorasi.

- v. Deployment

Model sudah diterapkan kedalam system. Missal integrase ke CRM. Lalu membuat dashboard/report untuk stakeholder agar bisa memahami hasilnya.



b. Contoh Penerapan Data Science di dunia Industri

Analisis Project sebuah perusahaan dengan data yang di input kedalam ERP

Sebuah perusahaan ingin menganalisis Revenue vs Cost

i. Data Collection

Data yang dibutuhkan untuk analisis tersebut adalah biaya marketing, biaya produksi, biaya distribusi, pendapatan perbulan. Data ini di dapat dari inputan user kedalam ERP

ii. Data Cleaning

Menghapus data duplikat, data yang belum approve, mengatasi revenue yang belum di input karena delay.

iii. Data Understanding

Cek terhadap outlier data, scatter plot Revenue vs Cost apakah ada hubungannya.

iv. Modeling & Evaluation

Train dan tes model dengan data historis, dan evaluasi performa model dengan (MAE, *mean absolute error*).

v. Deployment

Buat dashboard interaktif (*Tableau, PowerBI*), manajemen tim bisa melihat dan juga prediksi revenue dan profit perusahaan. Lalu bisa dimonitoring juga apakah pola Revenue vs Cost ini berubah.

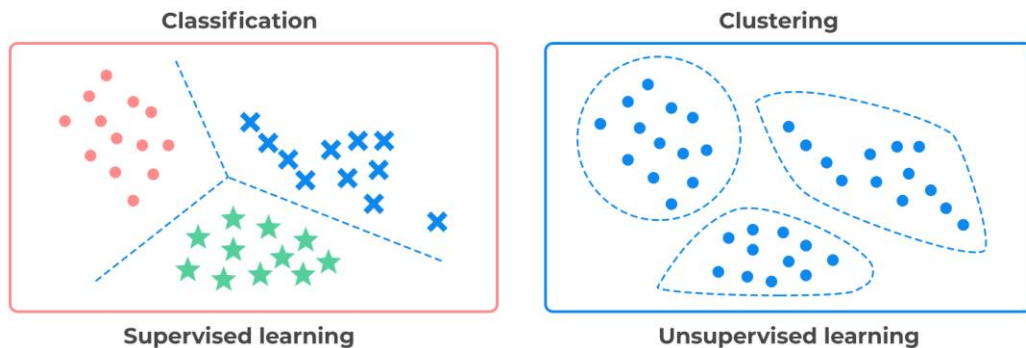
c. Perbedaan antara Supervised dan UnSupervised

Untuk perbedaan ini akan disajikan dalam bentuk table, agar mempermudah untuk melihat perbedaannya

Supervised Learning	UnSupervised Learning
Model dilatih dengan data yang memiliki label. Jadi ada input dan output yang memiliki target	Model dilatih tanpa label. Jadi hanya input aja tanpa target
Memprediksi output yang dihasilkan berdasarkan inputan	Menemukan pola, pengelompokkan, atau struktur data tersembunyi
Output berupa prediksi angka	Output berupa cluster, atau representasi data baru



Supervised vs. Unsupervised Learning



Contoh dataset Kaggle untuk supervised dan unsupervised learning :

Titanic Dataset => terkait data penumpang selamat atau tidak selamat

Mall Customers Dataset => segmentasi pelanggan