

# Perkenalan Data Science

---



SISTEM INFORMASI

# DATA SCIENCE



Data science adalah kegiatan menganalisis data dengan ilmu statistic, membuat, melatih dan evaluasi model machine learning dengan tujuan dapat mengambil business insight dalam data



Data Scientist adalah seorang spesialis yang menggunakan keahlian mereka dalam statistic dan model machine learning untuk membuat prediksi dan menjawab pertanyaan-pertanyaan bisnis



Capture data: harus punya datanya dulu



Manage & clean data: harus mengatur data yang tidak perlu bisa dihapus



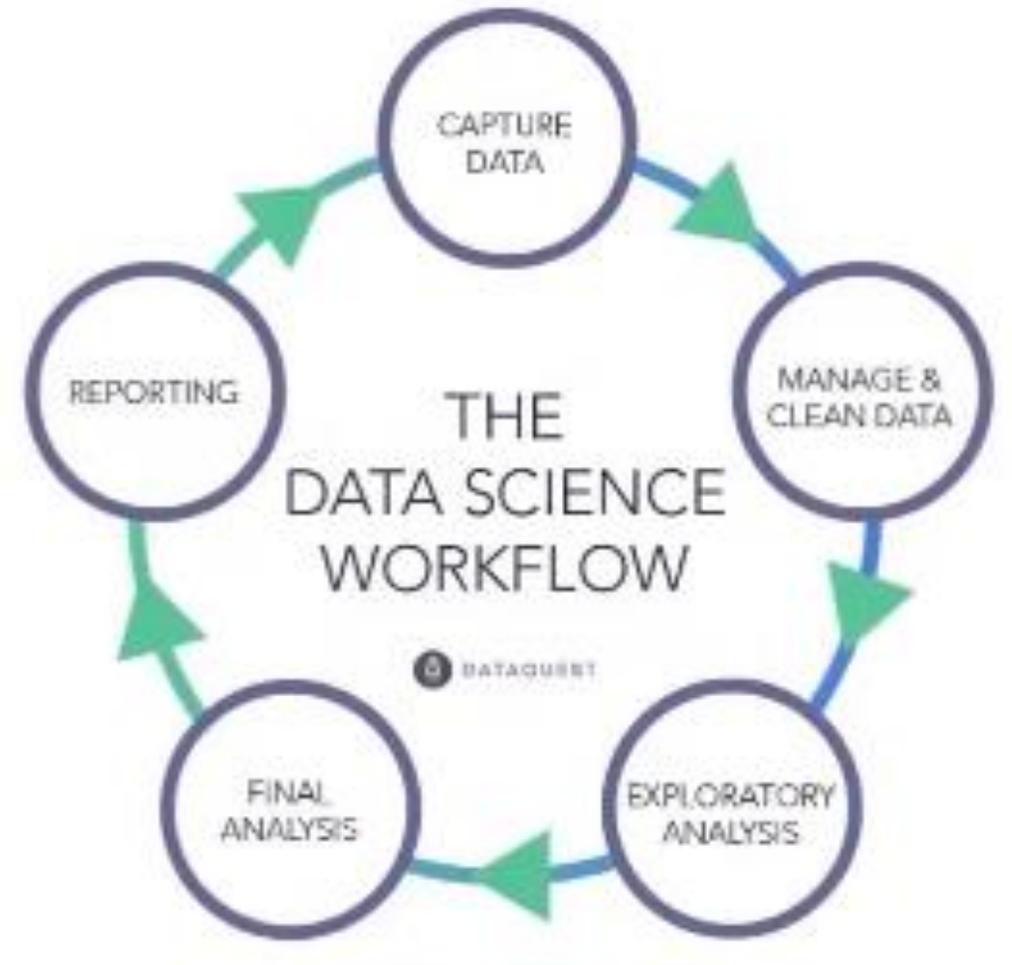
Exploratory Analysis: digali apa yang bisa didapat dari data yang kita punya



Final Analysis: Analisis secara keseluruhan



Reporting (kesimpulannya)





# Capture data

- Proses untuk mendapatkan data dari berbagai sumber, beberapa yang sering digunakan:
  1. Data berbasis file (csv, xls)
  2. Data berbasis database (sql)
  3. Data berbasis response (JSON)

# Manage & Clean Data

---

- Proses untuk memilih data yang penting untuk dianalisis kedepannya, biasanya dengan memilih kolom apa saja yang diperlukan dan dihilangkan nilai yang tidak perlu seperti data duplicate, null, NaN (not a number)



# Exploratory Data Analysis (EDA)

---

- Lebih banyak dalam memvisualisasikan data yang diperoleh dan mengira apakah data tersebut bisa menyelesaikan permasalahan yang didapat



# Final Analysis

---

- Final analysis ini membuat visualisasi data hasil dari pengolahan data sehingga dapat menjawab permasalahan yang didapat



# Reporting

- Reporting memberikan data hasil olahan dalam berbagai bentuk sesuai dengan kebutuhan, bisa file based, database, atau dalam JSON



# Perkenalan Data Engineer

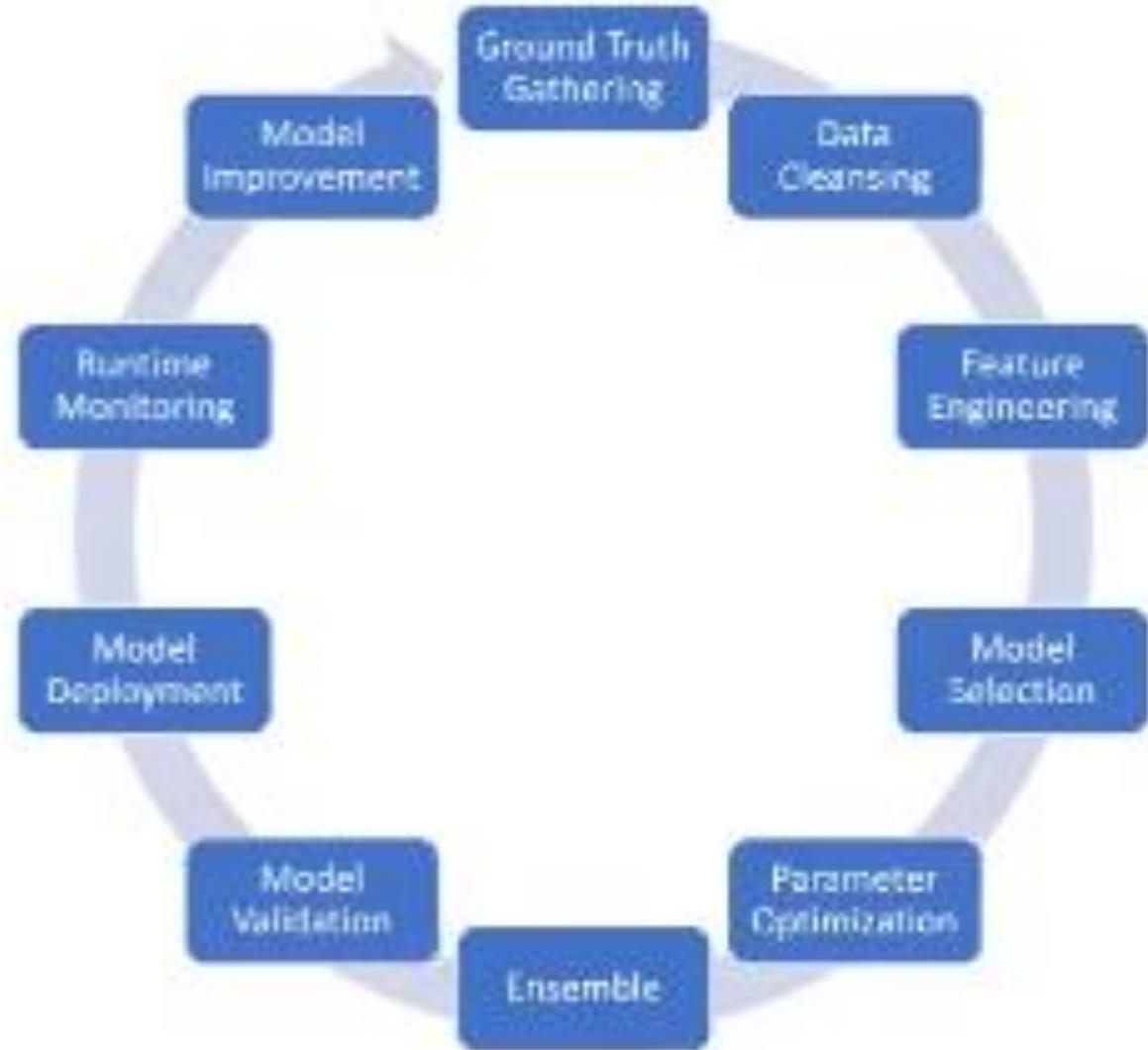


Kegiatan untuk mempersiapkan data (loading data, cleaning data) sehingga dapat diolah secara mudah



Data engineering membangun dan mengoptimalkan system sehingga data scientist dan data analyst dapat melakukan pekerjaan dengan baik

Workflow,  
Workload  
Data  
Engineer





Gatehring and Data Cleaning: tahapan mengumpulka data dari berbagai sumber



Feature Engineering: menentukan target atau data seperti apa yang ingin dihasilkan dari data mentah yang didapatkan dari sumber data



Model selection: menentukan model machine learning yang tepat untuk masalah yang dihadapi



Model learning exploration: melakukan optimasi parameter, membangun model, serta validasi model machine learning yang dibangun



Model Development: membuat sedemikian rupa agar data yang dihasilkan dari machine learning dapat dikonsumsi oleh aplikasi yang lain, serta memonitor jalannya aplikasi machine learning yang dibuat

# Perkenalan Data Analysis

---



Data analysis: kegiatan untuk mengatur dan mengumpulkan data sehingga dapat diperoleh informasi yang bermanfaat



Data analyst: bertugas menyampaikan nilai dari data yang didapat dari machine learning model untuk menjawab atau menentukan keputusan dalam bisnis

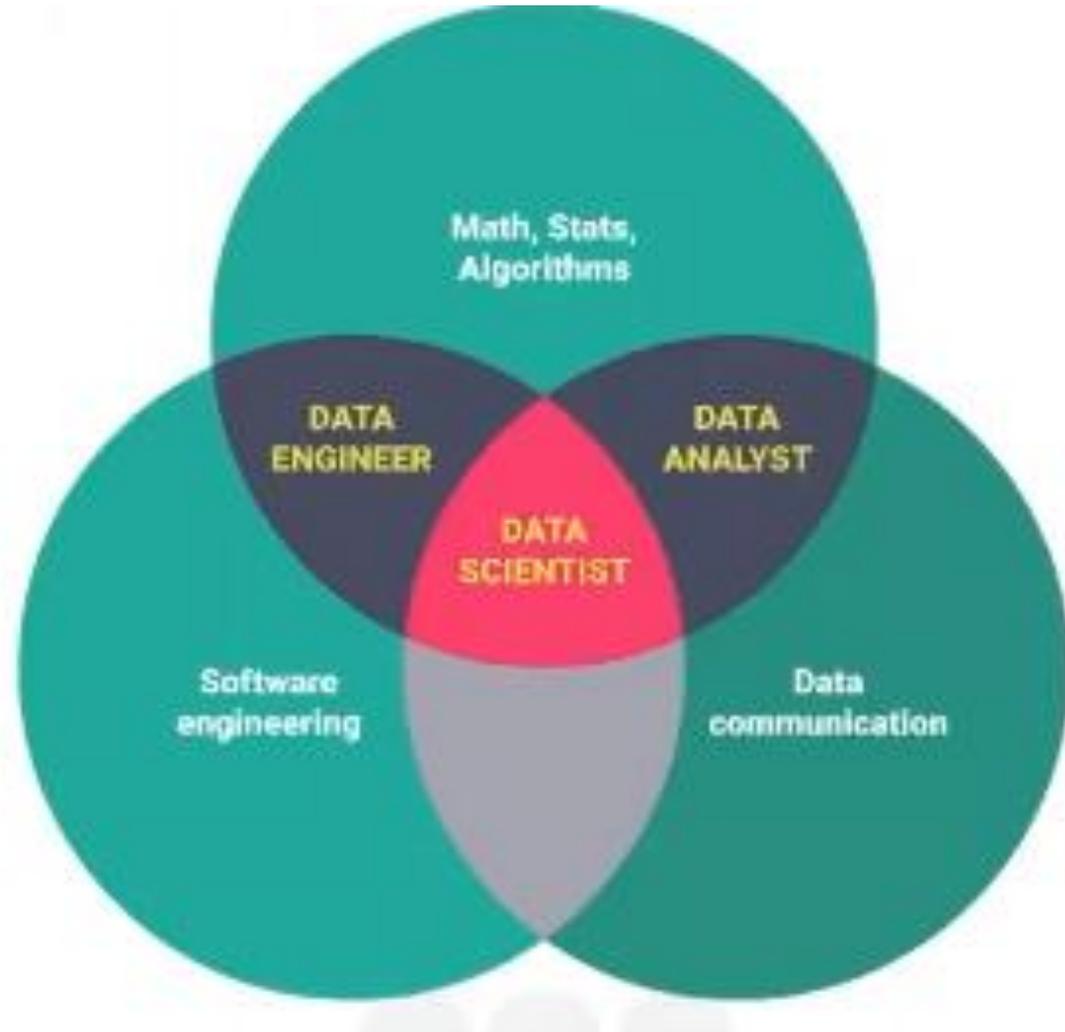
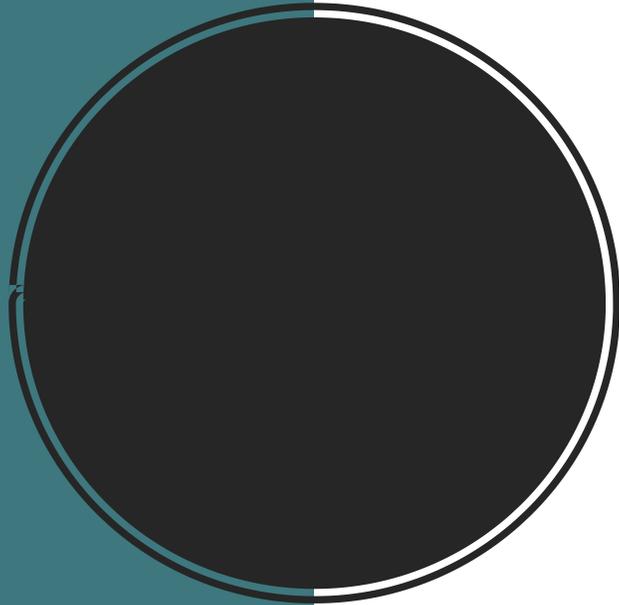


Workflow, Workload Data Analyst

- 
- Pada umumnya data analyst lebih banyak terlibat dalam komunikasi data dengan pimpinan atau client yang memerlukan hasil olahan data dari machine learning, kemudian memvisualisasikan data tersebut dan menjelaskan insight daripada data yang didapatkan

# Perbedaan Workflow Data Scientist, Engineer, Analyst

---



**TERIMA KASIH**