

Nama : Helmi Saifullah Mansur

Kelas : UGM07

Universitas Gadjah Mada

Sesi Industri dan Proyek

## **DATA ENGINEERING**

### **How Data is Engineered in Traveloka**

- Perlu persiapan dalam menghadapi dunia kerja di data science nantinya maka diperlukan adanya portofolio dan project yang mampu mengasah kemampuan terlebih mampu menyelesaikan masalah di industri, selain itu kemampuan problem solving juga sangat dibutuhkan nantinya
- Untuk tools yang ada di data science tidak perlu untuk menguasai semuanya, tetapi harus mengetahui fundamentalnya dan mengetahui alasan dari pemilihan tools tersebut
- Data Engineering di Traveloka meliputi :
  - a) Core Data Engineering
    - Data Infrastructure
    - Data ingestion to data warehouse
    - Data retrieval from data warehouse
    - Governance and security
    - Maintain data team's core tools/library
  - b) Data Warehouse Engineering
    - Data warehouse modelling
    - Maintain scalable ETL platform to handle hundreds of ETL every day/hour
  - c) Machine Learning Platform Engineering
    - Maintain machine learning platform to enable data scientist to productionize ML model
    - Maintain experimentation platform
- Data yang bocor menjadi tanggung jawab semua orang yang ada di organisasi, maka dari itu seorang data scientist harus mempertimbangkan siapa saja yang memiliki akses terhadap data, melakukan pengecekan apakah ada kebocoran data serta melakukan mitigasi agar tidak terjadi kebocoran data. Selanjutnya jangan takut untuk melakukan report dalam artian harus segera ditangani seperti melakukan meeting dengan software engineering dan sebagainya

- ML Model Development Challenges :
  - a) Explore large amount of data
  - b) Provision infrastructure needed to train the model
  - c) Serving the machine learning model
  - d) Model re-training
  - e) Feature engineering