

Virtual Internship Experience

id/x partners

Data Understanding

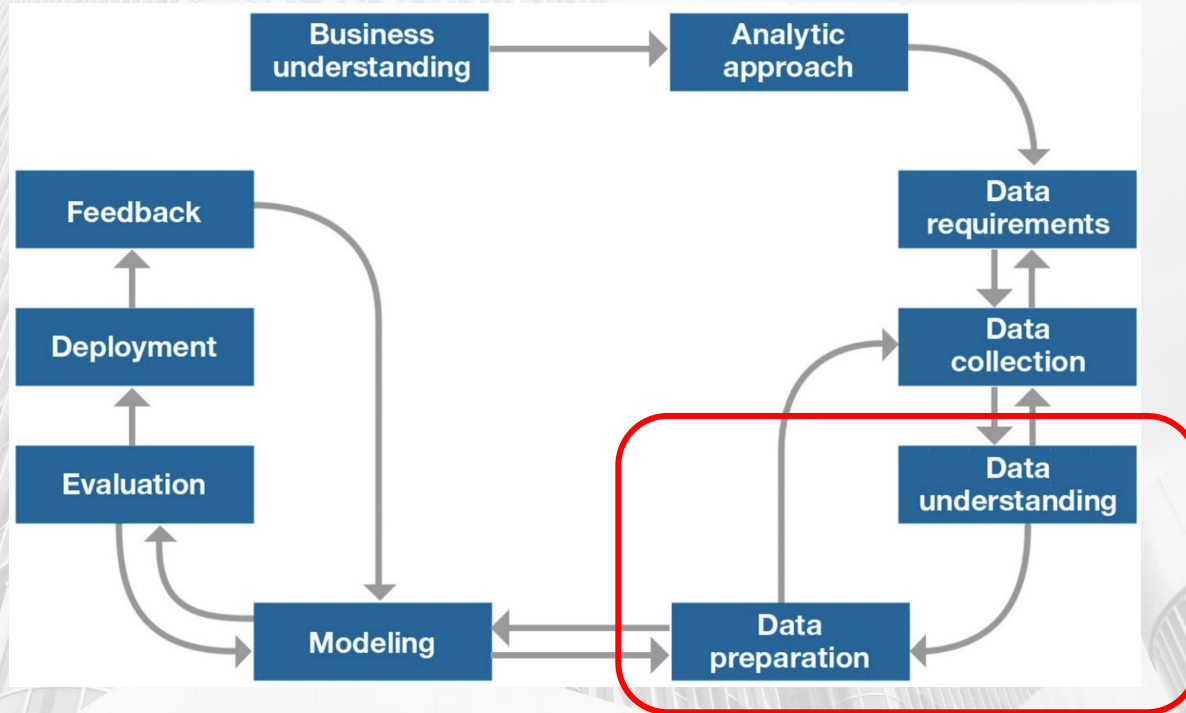
Disclaimer

“Dokumen ini memiliki hak cipta. Barang siapa yang menyebarluaskan atau menduplikasi tanpa izin dari instansi terkait dapat diproses sesuai dengan ketentuan hukum yang berlaku.”

Outline

- 1 **Kesesuaian Data**
- 2 **Eksplorasi Data**
- 3 **Verifikasi Kualitas Data**

Metodologi Data Science



Pertanyaan pemandu dalam Data Understanding:

- Atribut (kolom) mana yang terlihat paling menjanjikan?
- Atribut mana yang terlihat tidak relevan dan tidak perlu diikutsertakan?
- Apakah ada cukup data untuk menarik kesimpulan yang dapat digeneralisasikan atau membuat prediksi yang akurat?
- Apakah ada terlalu banyak atribut untuk metode pemodelan yang dipilih?
- Apakah kamu akan menggabungkan berbagai sumber data? Jika iya, apakah ada hal yang mungkin akan menimbulkan masalah saat penggabungan
- Sudahkan kamu mempertimbangkan bagaimana missing value ditangani di setiap sumber data?

Eksplorasi Data



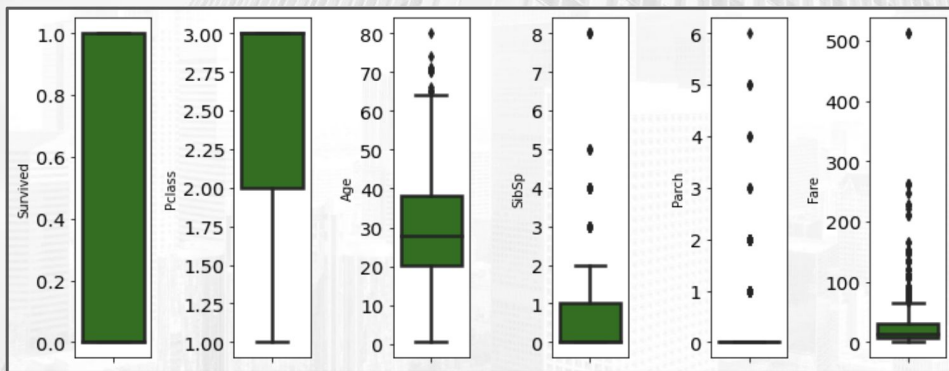
id/x partners

Exploratory Data Analysis (EDA) merupakan proses analisis untuk memahami karakteristik data dan hal-hal apa saja yang harus dilakukan agar data tersebut bisa dipakai untuk modeling.

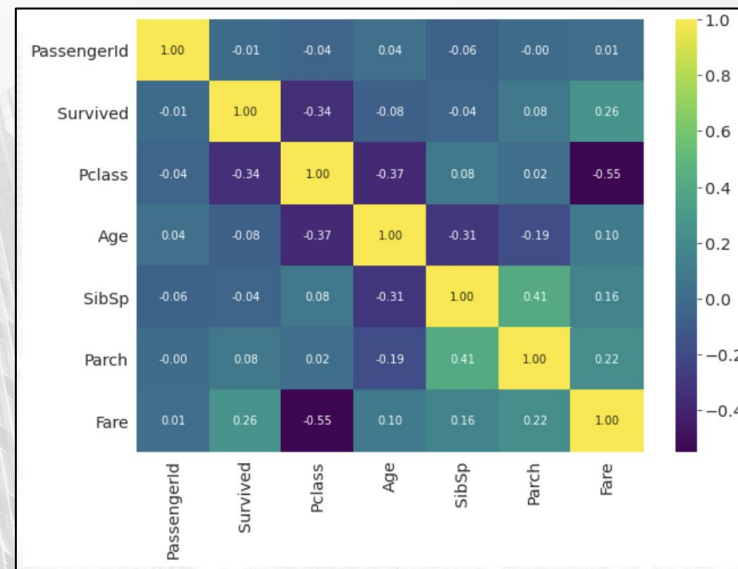
Pertanyaan yang bisa ditanyakan saat melakukan EDA:

1. Bagaimana sebaran nilai dalam feature dan label?
2. Apakah kira-kira feature yang dimiliki cukup baik untuk memprediksi target?
3. Persiapan apa saja yang harus dilakukan sebelum data tersebut dimasukkan menjadi data training model?

Proses EDA biasanya sering dibantu dengan menggunakan *chart/visualisasi* untuk mempermudah *discovery/penemuan* hal menarik dari data yang dieksplorasi.



(a)



(b)

Verifikasi Kualitas Data

Data yang dimiliki pada umumnya jauh dari kata sempurna, artinya memiliki beberapa *inconsistency*, nilai yang hilang, *error*, dsb. Beberapa hal yang bisa diperhatikan saat melakukan pemeriksaan kualitas data:

1. Nilai yang hilang / *missing data*
2. Kesalahan penulisan
3. Kesalahan pengukuran
4. Label yang tidak konsisten
5. Metadata yang buruk



Thank You!



id/x partners