Rakamin Academy

P

Laporan Final Project – Stage 2

(dipresentasikan setiap sesi mentoring)

1. Apakah sudah melakukan pengecekan data bermasalah seperti missing values, invalid values, atau data duplicate dan sudah membersihkannya?

Setelah di cek ternyata data tidak memiliki data duplicate sedangkan jumlah baris yang memiliki missing value terdapat 760 atau sekitar 15%. Tindakan yang dilakukan untuk missing value tersebut, yaitu

- Imputasi numerikal dengan median
- Imputasi dengan designation (liat distribusi antara umur dan designation)
- Imputasi kategorikal bisa modus
- Imputasi kategorikal bisa korelasi dengan variabel lain
- NumberOfTrips nilai null disebabkan karena customer tidak pernah ikut trip
- NumberOfFollowups nilai null disebabkan karena customer belum dilakukan followup kembali setelah pitching
- NumberOfChildrenVisitting nilai null disebabkan customer tidak punya anak yang ikut

Lalu, Kolom DurationOfPitch memiliki outlier 2 data dengan boxplot karena jauh Q3 dan MonthlyIncome memiliki banyak outlier, sehingga tindakan yang dilakukan yaitu menghilangkan outlier dengan Z-score dan IQR dimana data yang digunakan saat ini sebesar 92.29%.

2. Apakah sudah menentukan feature apa saja yang akan digunakan, atau perlu ditambahkan, dan reformatting feature sesuai dengan kebutuhan?

Penting! Saat memanipulasi data, mohon dilakukan dengan alasan yang jelas, dan tidak melakukan penambahan feature baru tapi tidak ada alasan yang mendasari langkah tersebut diperlukan.

Semua feature digunakan untuk modelling (tidak ada yang dihapus) dikarenakan semua feature masih relevan.