Welcome
Feature Engineering 101

DATA ENGINEERING
MABA

Agenda

- Feature Engineering
- Why do it?
- Examples of Feature Engineering
- Exercises and Examples
- Practical of Feature Engineering

Feature Engineering

Creation of new data features for statistical modeling or machine learning through the transformation of existing features.

Why do Feature Engineering?

1. Improve performance of statistical model or machine learning algorithm

2. Transform data to more adequately capture the domain or business question

3. Transform the data into a form that an algorithm can utilize

4. Surface new outcomes from from the data

Examples of Feature Engineering

- Transform Features (e.g. Normalization)
- Combining Features (e.g. Ratios)
- Categorical Coding

Transformation

Open Positions	AvgLeave Balance	scale_positions	leave_balance
17	22	2.95	0.54
12	32	1.71	1.98
7	15	0.47	(0.47)
4	6	(0.28)	(1.76)

Normalization or scaling a variable to make it comparable with other variables

Combination

FTEs	AvgLeave Balance	leave_per_fte
108	22	0.20
55	32	0.58
33	15	0.45
43	6	0.14

Create a new feature by combining existing features

Categorical Coding

Department	IT	Finance	HR
IT	1	0	0
Finance	0	1	0
HR	0	0	1
Strategy	0	0	0

Binary coding categories for modeling purposes

Benefits of Feature Engineering

- More insight (more factors to measure)
- Better models (e.g. higher score/R2)

Thank You
Feature Engineering 101

DATA ENGINEERING
MABA