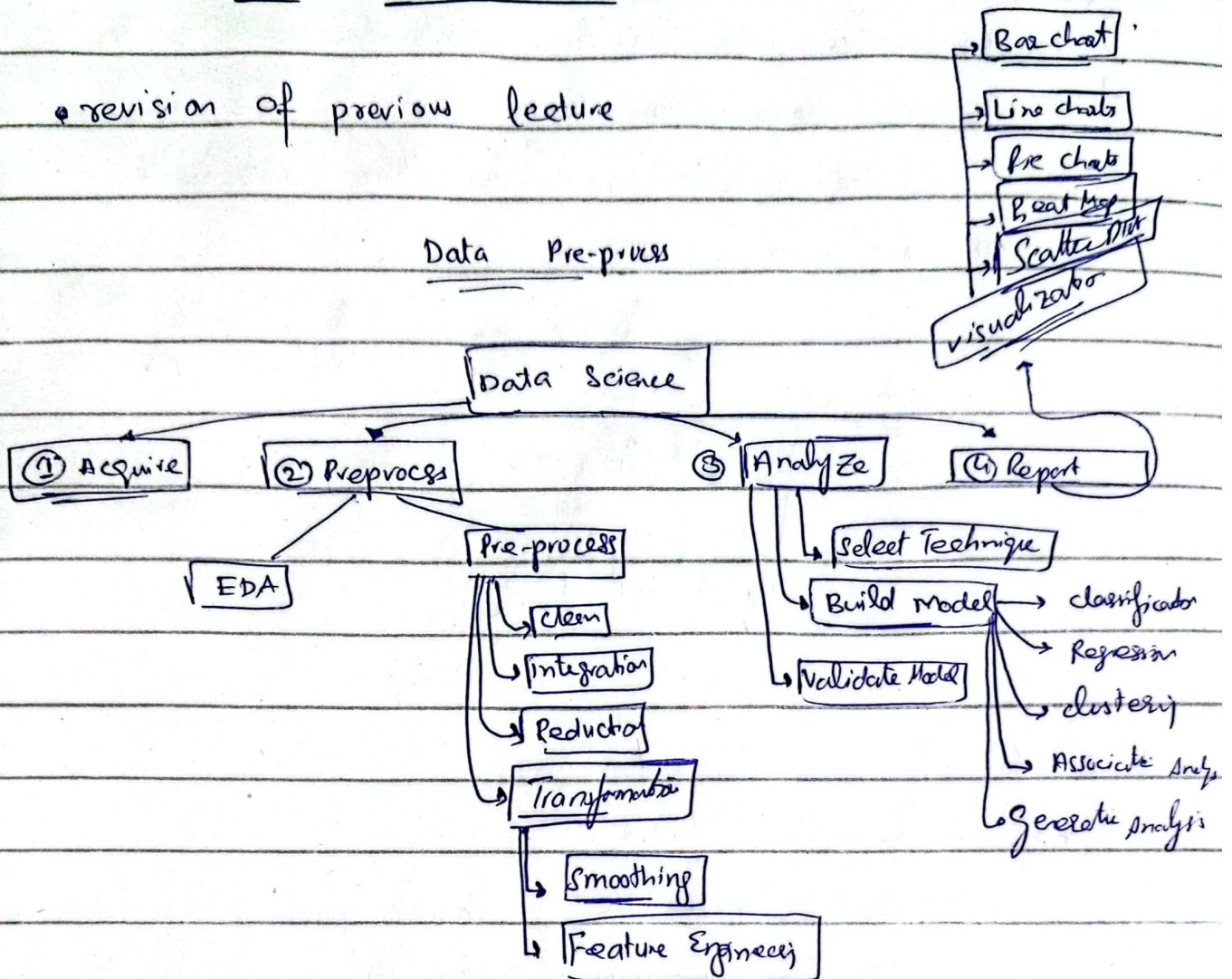


# Lect #16 Data - Preprocess

• revision of previous lecture

## Data Pre-process



## Transformer

→ ① Smoothing

(manage data in such way that model can learn)

→ ② Feature Engineering & construction

(construct/combining columns such way that give better insight)  
(bmi)  $\swarrow$  combin of weight, height + 2

• new meaningful extraction of attribute

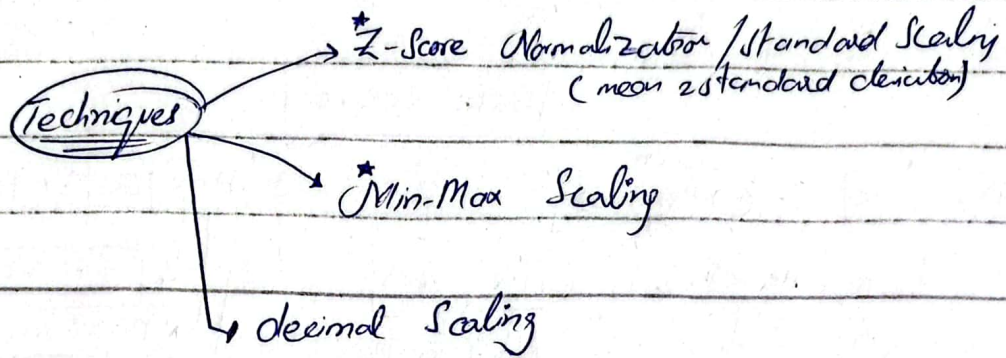
discretization

↳ numerical → Ordinal



### ③ Normalization

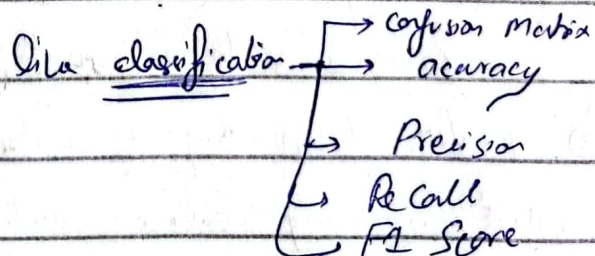
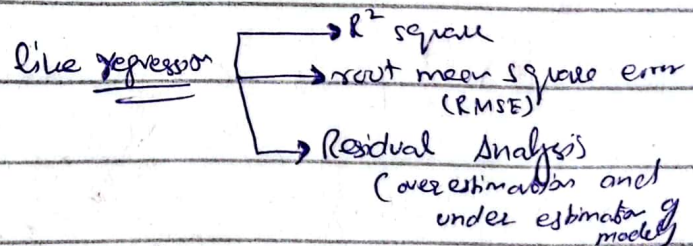
- when values are too high and too low
  - so we keep their range same.
- ⇒ No feature should be dominating over other based on value.



### ④ Analyze

- Choosing Technique
- Build a model
- Model validation (choosing best model)

Based on problems we are trying to solve



Technique → Model → Evaluation