Unsupervised ml : anomaly (outliers)detection - clustering (centroids) - how far from the centroid should be considered outliers??
Solutions: Parametr tuning or al interpretation
The selection of parameter k - elbow point: the number of clusters that make the error start decreasing significantly

1. K- means
2. Genetic algorithms - fitness function

Supervised ML: with known actions (predictors) - configuration of ml algorithms - depending on the dependent variable we have classifiers and regressors - deployment

Validation: 1. holdout (train-split split)
2. K-fold cross validation

3. Stratified cross validation
4. Leave one out cross validation (loocv)
5. Time series validation

Classification : boolean outcome variable
Regression: numeric ~
Discretisation: converting regression problem to classification one (problem: oversimplifying the problem and harder to explain in reality later)

How to engineer a ml pipeline
Predictors: Data and feature engineering
Outcome variable
Training
Configuration (**suitable ml algorithm** and based assumptions)
E.g 1) naive-bayes: independence assumption: features must be independent
Of each other: have to consider feature selection ;
2) linear regression: normality test first (shapiro-wilk test)

Hyperparameter (instruments) tuning:
1. random search and grid search
2. Muti-objective genetic algorithm

Deployment : retraining , latency, data privacy, network connectivity , cost

Real-time inference
Batch inference
Model deployment onto edge