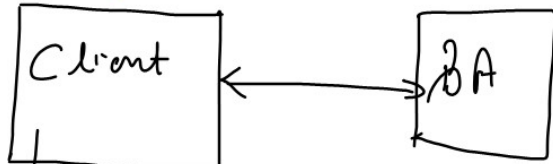


ML Projects Pipeline:

Telecom:



Objective

Solve Problem (Customers are churning)
Business Goal (Customers segmentation)

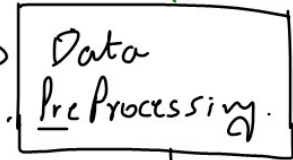
Correlation

$-K < 1$ if A, B ↓
 0.8

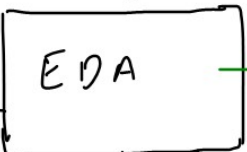


Where is the data?

Primary → Ownership
Secondary → Duration

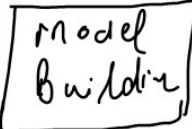


Extract Transform Load.
ETL
Data cleaning
feature engineering
optimizing feature

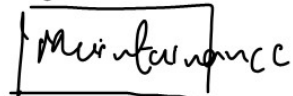


Exploratory Data Analysis

Univariate ✓
Bivariate ✓
Multivariate



Visualization



→ Unsupervised

→ Explore
→ cluster/segm
→ Analysis

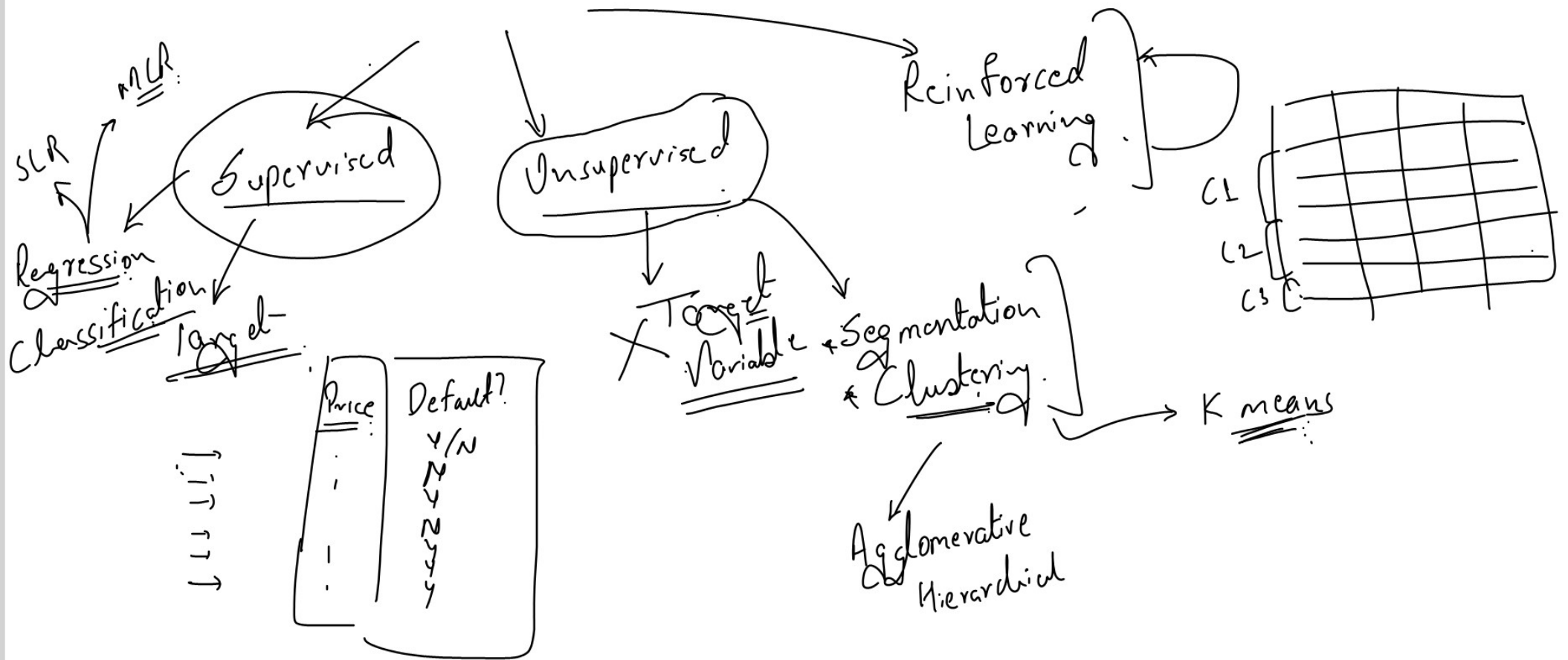
→ Supervised

→ Prediction
→ Target Variable

→ Reinforcement

→ Do → learn
→ Optimised





A hand-drawn diagram illustrating a data structure. It features a 4x4 grid of cells. The columns are labeled n_1 , n_2 , n_3 , and n_4 from left to right. To the right of the grid is a vertical column of 10 circles, each containing a number from 1 to 10. This column is labeled 'Array' and 'prop' at the top. Arrows point down from the top of the grid and the array column.

Correlation Coefficient:

$$-1 < r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} < 1$$

Linear Regression

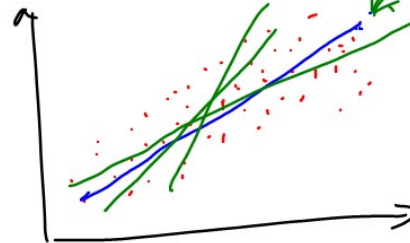
~~2~~ → multiple

Depression und

$\frac{G}{L} \text{ ind}$

x_1, x_2, x_3

$$y = m_1x_1 + m_2x_2 + m_3x_3 + \dots + m_nx_n + c$$



Best fit line

$$\boxed{y = 1.7x_1 + 2.3x_2 - 4.2x_3 + 12.652}$$

Null Values

Imputation

- Mean
- Median
- Mode

Predictive Modeling :

200	16.1	45	
100	16.1	45	