**🔹 Sampling Techniques**

* **Bagging**
  + Avoids or allows overfitting
* **Boosting**
  + Handles heteroskedasticity
  + Uses weak base models
* **Stacking**
  + Combines multiple models
  + *Note: Still unclear*

**🔹 Types of Methods**

* **GLM / Linear (Additive)**
  + Assumes linear behavior
* **Tree-based Methods (e.g., DT, RF, XGBoost, LightGBM, CatBoost)** *(Non-parametric)*
  + Sub-samples the population
  + Assumes unknown distributions
* **Non-linear Methods (e.g., Deep Learning, NN, RNN)** *(Non-linear)*
  + Complex but assumes an underlying distribution
* **Unsupervised Learning**
  + Factor Analysis → related columns
  + Cluster Analysis → related rows
* **Expert Models**
  + **AHP** – Internal expert *on top* of a baseline model
  + **Conjoint Analysis** – Expert user, no need for baseline model
  + **Discrete Choice** – Non-expert user, no need for baseline model
  + **Max Diff** – Rushing user, no need for baseline model
* **Time Series Models**
  + **Time-dependent data**: ARIMA, AutoARIMA, ETS, Facebook Prophet
  + **Repeated measurements with no clear pattern**: TBATS, LSTM

**🔹 Further Enriching Techniques**

* **Feature Selection**
  + Parsimony: “Less is more”
* **L1 / L2 Regularization**
  + Solves multicollinearity
* **Support Vectors**
  + Solve multidimensional outliers by removing them
* **Outlier Flagging**
  + Solve multidimensional outliers by flagging with a binary variable
* **Sectorization of Rows**
  + Solves non-independence of rows (e.g., ratings by sector)
* **PCA / SEM**
  + Addresses multicollinearity and sub-population shadowing in tree-based models

**🔹 SVM (Support Vector Machines)**

* Deals with outliers
* Uses the **Kernel Trick**

**🔹 Tree-based Methods (again for emphasis)**

* Includes Random Forest
* Handles unknown distributions