

# MACHINE LEARNING ALGORITHMS

---

**Regression | Classification | Ensemble | Clustering | Dimensionality Reduction | Scikit Learn**

---

## 1. Supervised :

1.1 - Regression - Simplelinear | Multiplelinear | Polynomial | Ridge & Lasso | Decision Tree Regression | Random Forest Regression | Support Vector Regression (SVR).

1.2 - Classification - Logistic Regression | Naive Bayes | Decision Trees | Random Forests | Support Vector Machines (SVMs) .

1.3 - Algorithms for both Classification and Regression - Decision Trees, Random Forests, Support Vector Machines (SVMs), and K-Nearest Neighbors (KNN).

1.4 - Ensemble learning -

Ensemble learning combines multiple simple models to create a stronger, smarter model. There are mainly two types of ensemble learning:

- Bagging that combines multiple models trained independently.
- Boosting that builds models sequentially, each correcting the errors of the previous one.
- Bagging - Random Forest, Random Subspace Method .
- Boosting - Gradient Boosting Machines (GBM), Extreme Gradient Boosting (XGBoost), AdaBoost (Adaptive Boosting), CatBoost .

## 2. Unsupervised :

2.1 - Clustering - Hard clustering , Soft clustering.

2.1.1 - Centroid based methods :

- K-Means clustering
- Elbow Method for optimal value of k in KMeans
- K-Means++ clustering
- K-Mode clustering
- Fuzzy C-Means (FCM) Clustering

2.1.2 - Density based methods :

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- OPTICS (Ordering Points To Identify the Clustering Structure)

2.1.3 - Connectivity based methods :

- Hierarchical clustering
- Agglomerative Clustering
- Divisive clustering
- Affinity propagation

2.1.4 - Distribution based methods :

- Gaussian mixture models
- Expectation-Maximization Algorithm
- Dirichlet process mixture models (DPMMs)

2.2 - Dimensionality reduction :

- Principal Component Analysis (PCA)
- t-distributed Stochastic Neighbor Embedding (t-SNE)
- Non-negative Matrix Factorization (NMF)
- Independent Component Analysis (ICA)
- Isomap
- Locally Linear Embedding (LLE)

2.3 - Association Rule Mining :

Find patterns between items in large datasets typically in market basket analysis.

- Apriori algorithm
- Implementing apriori algorithm
- FP-Growth (Frequent Pattern-Growth)
- ECLAT (Equivalence Class Clustering and bottom-up Lattice Traversal)

### 3. Reinforcement :

3.1 - Model based methods -

- Markov decision processes (MDPs)
- Bellman equation

- Value iteration algorithm
- Monte Carlo Tree Search

### 3.2 - Model free methods -

- Q-Learning
- SARSA
- Monte Carlo Methods
- Reinforce Algorithm
- Actor-Critic Algorithm
- Asynchronous Advantage Actor-Critic (A3C)

## 4. **Semi supervised** : Semi-Supervised Learning

- Semi Supervised Classification
- Self-Training in Semi-Supervised Learning
- Few-shot learning in Machine Learning

## 5. **Self supervised** : Self-Supervised Learning

## 6. **Forecasting Models :**

- ARIMA (Auto-Regressive Integrated Moving Average)
  - SARIMA (Seasonal ARIMA)
  - Exponential Smoothing (Holt-Winters)
- 
- 

- Regression model evaluation, Loss functions : MAE, MSE, RMSE, R<sup>2</sup>, Huber
- Classification model evaluation : Confusion matrix, ACC, PRE, Recall, F-1 score, Auc-roc
- Classification Loss functions - entropy, Binary cross entropy, categorical cross entropy, sparse categorical cross - entropy, KL Divergence loss, Hinge loss
- Loss function, cost function, Learning rate, Step value, Gradient descent, mini batch GD, Stochastic GD, Vanishing gradient descent.
- Bias, Variance, Bias-variance trade off , under-fitting, over fitting, Regularization , Lasso - L2, Ridge - L1
- Preprocessing - Min-max scaler, Standardscaler, Normalization, Standardization.

- Cross validation - Holdout, LOOCV, Stratified, K-fold
  - Hyperparameter tuning - gridsearch cv, Randomized search cv, Bayesian
-