# Machine Learning Algorithms: Descriptions and Functions

## Random Forest

Definition: Random Forest is an ensemble learning method that combines multiple decision trees to improve accuracy.

What it does: Aggregates predictions from several trees to reduce overfitting and enhance generalization.

Application: Used in classification, regression, and feature selection across various industries.

Advantage: Handles missing data well and works effectively with large, complex datasets.

## Decision Tree

Definition: Decision Tree is a supervised learning algorithm that splits data into branches based on decision rules.

What it does: Breaks down complex decisions into simpler structures using a tree-like model.

Application: Common in medical diagnoses, credit scoring, and business decision-making.

Limitation: Prone to overfitting, which can reduce generalizability.

## Logistic Regression

Definition: Logistic Regression is used for binary classification, predicting the probability of an event occurring.

What it does: Fits a logistic function to the data, outputting probabilities instead of direct values.

Application: Widely used in healthcare (disease prediction) and marketing (customer churn analysis).

Advantage: Simple to implement and interpretable for binary classification tasks.

## Support Vector Machine (SVM)

Definition: SVM is a supervised learning algorithm for classification and regression tasks.

What it does: Finds the optimal hyperplane that separates data points into distinct classes.

Application: Effective for image recognition, text classification, and bioinformatics.

Advantage: Works well with high-dimensional data and provides flexibility with kernel functions.

## K-Nearest Neighbors (KNN)

Definition: KNN is a simple, non-parametric algorithm used for classification and regression by comparing data points with their nearest neighbors.

What it does: Classifies or predicts based on the majority or average of the nearest neighbors.

Application: Used in recommendation systems, pattern recognition, and anomaly detection.

Limitation: Computationally expensive for large datasets due to the distance calculation for every query.

## Gradient Boosting

Definition: Gradient Boosting is an ensemble learning algorithm that sequentially combines weak classifiers to improve accuracy.

What it does: Optimizes predictions by focusing on the errors of previous models and correcting them in subsequent iterations.

Application: Used in fraud detection, risk modeling, and structured data analysis.

Advantage: Provides high predictive accuracy but requires careful tuning to avoid overfitting.

## XGBoost

Definition: XGBoost (Extreme Gradient Boosting) is a scalable, distributed gradient boosting library designed for high performance and efficiency.

What it does: Combines decision trees sequentially while minimizing errors using gradient descent.

Application: Widely used in Kaggle competitions, predictive modeling, and feature engineering.

Advantage: Offers regularization techniques to prevent overfitting and supports parallel computation for faster training.

## AdaBoost

Definition: AdaBoost (Adaptive Boosting) combines a series of weak classifiers to create a strong classifier by assigning weights to difficult examples.

What it does: Focuses on incorrectly classified samples in each iteration to improve accuracy.

Application: Used in image detection, sentiment analysis, and bioinformatics.

Advantage: Simple to implement and improves model performance by reducing bias and variance.