The primary types of models in supervised and unsupervised machine learning are categorized based on their function and the type of data they are trained on.

## Supervised Learning Models (Prediction) 🎯📈

Supervised learning models are trained on **labeled data** to predict a specific output.1 They are divided into two main groups:

### 1. Classification Models (Discrete Output)

These models predict a discrete category or class label.2

| Model Type | Concept/Mechanism | Common Use Cases |
| --- | --- | --- |
| **Linear Models** | Uses a linear boundary to separate classes. | Spam filtering, Yes/No predictions. |
| **Logistic Regression** | Estimates the probability of an instance belonging to a particular class (binary or multi-class). |  |
| **Non-Linear/Tree-Based Models** | Uses a series of splits based on features to arrive at a decision. | Credit risk analysis, churn prediction. |
| **Decision Trees** | A single tree structure of decisions. |  |
| **Random Forest** | An ensemble of many uncorrelated decision trees (bagging). |  |
| **Gradient Boosting Machines (GBM)** | An ensemble method that builds trees sequentially, with each new tree trying to correct the errors of the previous one. |  |
| **Instance-Based Models** | Classifies new data based on the majority class of its nearest neighbors. | Simple recommendation systems. |
| **K-Nearest Neighbors (KNN)** |  |  |
| **Support Vector Machines (SVM)** | Finds the optimal hyperplane that maximizes the margin between classes. | Image recognition, text categorization. |
| **Neural Networks (Deep Learning)** | Uses layers of interconnected nodes (perceptrons) to learn complex patterns. | Image/video processing, large-scale classification. |

### 2. Regression Models (Continuous Output)

These models predict a continuous numerical value.

| Model Type | Concept/Mechanism | Common Use Cases |
| --- | --- | --- |
| **Linear Regression** | Finds the best-fit straight line (or hyperplane in multiple dimensions) to model the relationship between variables. | House price prediction, sales forecasting. |
| **Regularized Regression** | Linear models that add a penalty term to prevent overfitting. | Financial modeling, highly variable data. |
| **Ridge Regression** | Adds an L2 penalty (squared magnitude of coefficients). |  |
| **Lasso Regression** | Adds an L1 penalty (absolute magnitude of coefficients), often forcing some coefficients to zero (feature selection). |  |
| **Tree/Ensemble Models** | The regression versions of the classification models, which predict a numerical mean at the leaf nodes. | Predicting maintenance time, stock volume. |
| **Decision Tree Regressor** |  |  |
| **Random Forest Regressor** |  |  |

## Unsupervised Learning Models (Discovery) 🔍

Unsupervised learning models are trained on **unlabeled data** to find hidden structures, patterns, or relationships.3 They are divided into two main groups:

### 1. Clustering Models (Grouping Data)

These models partition the dataset into groups (clusters) such that data points within the same cluster are more similar to each other than those in other clusters.4

| Model Type | Concept/Mechanism | Common Use Cases |
| --- | --- | --- |
| **Centroid-Based** | Partitions data by minimizing the distance between points and the center of their assigned cluster. | Market segmentation, document grouping. |
| **K-Means** | Partitions n observations into k clusters. |  |
| **Density-Based** | Identifies clusters as areas of high density separated by areas of low density. | Identifying anomalies, geographical data analysis. |
| **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)** |  |  |
| **Distribution-Based** | Assumes data points are generated from a mixture of several probability distributions (e.g., Gaussian distributions). | Statistical inference. |
| **Gaussian Mixture Models (GMM)** |  |  |
| **Hierarchical Clustering** | Builds a tree-like hierarchy of clusters (dendrogram). | Taxonomy creation, biological classification. |

### 2. **Dimensionality Reduction Models (Simplifying Data)5**

These models reduce the number of features (dimensions) in a dataset while retaining most of the important information.

| Model Type | Concept/Mechanism | Common Use Cases |
| --- | --- | --- |
| **Linear Techniques** | Projects the high-dimensional data onto a lower-dimensional subspace while preserving variance. | Feature compression, data visualization. |
| **Principal Component Analysis (PCA)** | Finds orthogonal components (principal components) that capture the maximum variance. |  |
| **Non-Linear Techniques (Manifold Learning)** | Seeks to map data onto a low-dimensional "manifold" that is naturally embedded within the high-dimensional space. | Bioinformatics, visualizing complex relationships. |
| **t-distributed Stochastic Neighbor Embedding (t-SNE)** | Optimally suited for data visualization. |  |
| **Autoencoders (Deep Learning)** | A neural network designed to learn a compressed representation (encoding) of the input data. | Denoising images, feature extraction. |

### 3. Association Rule Mining

These models discover interesting relationships and co-occurrence patterns among variables in large datasets.

| Model Type | Concept/Mechanism | Common Use Cases |
| --- | --- | --- |
| **Apriori Algorithm** | Identifies frequent item sets in a transaction dataset. | **Market Basket Analysis** (e.g., "Customers who buy milk and bread also tend to buy butter"). |