

Machine Learning Model Types and Difference

1. Clustering vs Classification

	Clustering	Classification
Definition	It is an unsupervised learning technique used to group similar data points together based on their features without prior knowledge of the group labels.	It is a supervised learning technique where the goal is to assign data points to predefined categories or labels.
Labels	In clustering no labeled data is used during the training/testing process of the model. The model identifies the inherent grouping of the data.	In classification labeled data is required for training/testing the model for mapping from input features to output labels.
Goal	To discover the inherent structure and patterns in the data.	To predict the category to which a new data point belongs based on learned patterns from labeled data.
Complexity	Clustering is less complex compared to Classification.	Classification is more complex compared to Clustering.
Techniques	Common techniques for clustering includes K-means, Hierarchical Clustering, Principal Component Analysis(PCA), etc.	Common techniques for classification includes Logistic Regression, Decision Tree, Support Vector Machine, etc.
Example	A company wants to segment its customers based on purchasing behavior to tailor marketing strategies. Using an unsupervised learning algorithm like K-means clustering, they group customers based on similarities in purchase history, frequency, and amount spent. This results in discovering natural customer segments without predefined labels.	An email service provider wants to classify incoming emails as "spam" or "not spam". Using a supervised learning algorithm like support vector machines (SVM), they train the model on a labeled dataset of emails. The model learns to recognize patterns associated with spam and non-spam emails, enabling it to classify new emails accurately.

2. Regression vs Classification

	Regression	Classification
Definition	It is a supervised learning technique used to predict a continuous output variable based on input features.	It is a supervised learning technique used to predict a categorical output variable.
Output Type	The output is continuous/real value.	The output is a discrete class label (Either binary or multi-label).
Techniques	Common techniques for regression include Linear Regression, Polynomial Regression, Random Forest Regressor, etc.	Common techniques for classification includes Logistic Regression, Decision Tree, Support Vector Machine, etc.
Evaluation Metrics	Evaluated using metrics like mean absolute error (MAE), root mean squared error (RMSE), and R-squared, etc.	Evaluated using metrics like accuracy, precision, F1 score, etc.
Example	A real estate firm wants to predict the selling price of houses based on features like size, location, number of bedrooms, and age of the house. They use a regression model like linear regression to predict the continuous variable of house prices.	A healthcare provider wants to predict whether a patient has diabetes based on their medical data. Using a supervised learning algorithm like logistic regression, they train the model on a labeled dataset of patient records with features such as age, BMI, blood pressure, and glucose levels. The model learns to distinguish between diabetic and non-diabetic patients, allowing accurate predictions for new patients.