

Concept	Type	Description	Example	Pref. use case
Data	Structured	Organised data in tabular format with rows & cols.	House prices datasets with columns for features	Most ML models expect structured data.
	Unstructured	Data without predefined format, such as images/text & audio	Text reviews, images of houses	Image, text & speech recognition tasks.
	Semistructured	Data with some organisational str. like XML or JSON	JSON data of house listings	Data with nested attributes like web or API data.

Concept	Type	Description	Example	Pref. use case
Features	Numerical	Continuous or discrete values that can be measured on a scale	Sq. footage, no. of rooms	Regression model, Tree based model
	Categorical	Distinct groups or categories without inherent order	Neighbourhood types like 'urban' or 'rural'	Tree based models, Neural networks (after encoding)

Binary	Data with only 2 possible values	Ocean proximity (Yes/No)	Logistic regression, SVM, tree based model.
Ordinal	Categorical data with defined order	Quality rating (low, med., high)	Tree based models, regression after encoding

Concept	Type	Description	Example	Preferred use case
Labels	Binary	The target variable has 2 possible values	Will the house sell? (Yes/No)	Binary classification model like logistic regression.
Multi Class		The target variable has more than 2 possible categories	Predicting housing condition (good, fair, poor)	Multi Class classification model.
	Continuous	The target var. is a continuous numerical value.	House price in dollars	Regression models for continuous targets

Concept	Type	Description	Example	Pref use case
Model	Linear models	models that assume linear relationship b/w features & targets.	Linear regression for house price	Preferrd when data has a roughly linear relationship.
	Tree based model	Models that use decision rules - based on feature values	Decision trees, Random Forest	Effective for capturing non linear patterns, handles missing data.
	Neural networks	Models with layers of neurons that learn complex patterns	CNN for image data	Preferrd for tasks with complex patterns like image recognition.
	Ensemble Models	Models that combine multiple algorithms to improve performance	Random Forest, Gradient boosting	Preferrd for high accuracy, reduces overfitting

Concept	Type	Description	Example	Pre use case
Supervised Learning	Classification	Predicts categorical labels based on features	Classifying housing types (luxury, standard)	Tasks requiring distinct label data for distinct categories.
	Regression	Predicts continuous values based on features	Predicting house price	Used in price prediction, demand forecasting
Unsupervised Learning	Clustering	Groups data points into clusters based on similarity	Grouping houses based on neighbourhood types	Data exploration, customer segmentation
	Association	Finds relationships or associations b/w variables	Discovering patterns in purchasing behaviour	Market basket analysis, recommendation system.
	Dimensionality Reduction	Reduces feature space while preserving data str.	Principle component analysis	Feature selection, visualization, improving model efficiency

Concept	Type	Description	Example	Pref Use Case
Cross validation	K-fold	Divides data into K subsets & uses each as a test case once	5 fold CV for housing model	Reliable for model evaluation, prevents overfitting
	Stratified K-fold	Ensures each fold has a similar dist ⁿ of classes	Stratified Split for imbalanced data	Preferred for classification with imbalanced classes
	LOOCV	Each data point is used as a single test example in its own fold	Leave one out CV for small housing data	Small datasets where maximizing data usage is crucial
	Time series	Use past data for training & future data for testing	Predicting housing prices over time	Time dependent data like forecasting & sequential data tasks