🔍 Topic/Tool	📌 Purpose / Goal	🛠️ Scikit-learn Module / Function
Machine Learning Types	Understand types of problems (classification, regression, clustering)	-
Supervised Learning	Learn from labeled data	LogisticRegression, DecisionTreeClassifier, etc.
Unsupervised Learning	Discover patterns without labels	KMeans, PCA, etc.
Train-Test Split	Prevent overfitting by testing on unseen data	train_test_split()
Model Training (Fit)	Train model on features and labels	.fit(X_train, y_train)
Prediction	Use trained model to make predictions	.predict(X_test)
Evaluation Metrics	Measure model performance	accuracy_score, f1_score, classification_report
Confusion Matrix	Understand types of prediction errors	confusion_matrix()
ROC Curve & AUC	Visual + numeric metric for binary classifier performance	roc_curve, roc_auc_score
Missing Values Handling	Replace or remove missing data	SimpleImputer
Feature Scaling	Normalize feature values for better performance	StandardScaler, MinMaxScaler
Categorical Encoding	Convert text categories to numbers	LabelEncoder, OneHotEncoder
Pipeline	Combine preprocessing + modeling into 1 step	Pipeline, make_pipeline
ColumnTransformer	Apply different preprocessing to different columns	ColumnTransformer
Cross-validation	Check model stability by training/testing on different folds	cross_val_score, GridSearchCV
Hyperparameter Tuning	Find the best settings for a model	GridSearchCV, RandomizedSearchCV
Ensemble Learning	Combine multiple models to improve accuracy	RandomForestClassifier, VotingClassifier, Bagging
Bias-Variance Tradeoff	Understand model underfitting vs overfitting	Plot learning curves / use cross-validation