|  |  |  |  |
| --- | --- | --- | --- |
| Step Number | Step | Description | Relevant scikit-learn Modules & Functions (examples) |
| 1 | Problem Definition, Scoping & Framing | Define the ML task, set clear objectives, determine constraints, and establish a structured approach for solving the problem. | No specific module. Use dataset sources. |
| 2 | Data Exploration & Understanding | Perform statistical analysis, compute descriptive statistics, inspect dataset properties, visualize key insights. | pandas.DataFrame.describe()  pandas.DataFrame.info()  pandas.DataFrame.value\_counts()  pandas.DataFrame.isnull().sum()  seaborn.heatmap()  seaborn.pairplot(),  matplotlib.pyplot.hist() |
| 3 | Data Preparation & Feature Engineering | Clean, transform, and optimize data; handle missing values; encode categories; scale features; engineer new features. | sklearn.preprocessing.StandardScaler()  MinMaxScaler()  SimpleImputer()  OneHotEncoder()  LabelEncoder()  sklearn.feature\_selection.SelectKBest()  sklearn.decomposition.PCA() |
| 4 | Model Selection & Evaluation | Train and compare different ML models, evaluate performance using cross-validation and test sets. | sklearn.model\_selection.train\_test\_split()  cross\_val\_score()  sklearn.linear\_model.LogisticRegression()  SGDClassifier()  Ridge()  Lasso()  sklearn.tree.DecisionTreeClassifier()  sklearn.svm.SVC()  sklearn.neighbors.KNeighborsClassifier()  sklearn.ensemble.RandomForestClassifier()  xgboost.XGBClassifier()  sklearn.metrics.accuracy\_score() |
| 5 | Performance Tuning & Optimization | Apply hyperparameter tuning, regularization, and ensemble techniques to improve performance. | sklearn.model\_selection.GridSearchCV()  sklearn.pipeline.Pipeline()  sklearn.ensemble.AdaBoostClassifier()  xgboost.XGBRegressor()  RidgeCV()  LassoCV() |
| 6 | Results Interpretation & Deployment | Summarize findings, visualize insights, save the model, and integrate it into production. | sklearn.metrics.classification\_report()  confusion\_matrix()  joblib.dump()sklearn.pipeline.Pipeline |