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.**##### Find how machine learning can be applied to predict strokes using patient data**	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