Note:

1. This document is to help you as a quick reference for sklearn modules and APIs.
2. In no way it attempts to replace/substitute sklearn documentation.
3. It is not a complete or exhaustive list of sklearn modules and APIs.

List of some important sklearn modules

Module Name	Brief Description
sklearn.feature_extraction	This module implements feature selection algorithms. It currently includes univariate filter selection methods and the recursive feature elimination algorithm.
sklearn.feature_selection	This module implements feature selection algorithms. It currently includes univariate filter selection methods and the recursive feature elimination algorithm.
sklearn.impute	Transformers for missing value imputation
sklearn.linear_model	This module implements a variety of linear models.
sklearn.metrics	This module includes score functions, performance metrics and pairwise metrics and distance computations.
sklearn.model_selection	This module implements various cross validation and HPT techniques.
sklearn.naive_bayes	This module implements Naive Bayes algorithms
sklearn.neighbors	This module implements the k-nearest neighbors algorithm.
sklearn.neural_network	This module includes models based on neural networks.
sklearn.pipeline	This module implements utilities to build a composite estimator, as a chain of transforms and estimators.
sklearn.preprocessing	This module includes scaling, centering, normalization, binarization methods.
sklearn.svm	This module includes Support Vector Machine algorithms.
sklearn.tree	This module includes decision tree-based models for classification and regression.
sklearn.ensemble	This module includes ensemble-based methods for classification, regression and anomaly detection.

Models

Following is the list of most commonly used ML models:

Category	Library module	API
Baseline model	sklearn.dummy	DummyRegressor DummyClassifier
Linear Regression	sklearn.linear_model	LinearRegression
Generic Regression/classification model	sklearn.linear_model	SGDRegressor SGDClassifier
Ridge regression/Classification	sklearn.linear_model	Ridge Ridge Classifier
Lasso regression	sklearn.linear_model	Lasso
RidgeCV	sklearn.linear_model	RidgeCV
LassoCV	sklearn.linear_model	LassoCV
Perceptron classifier	sklearn.linear_model	Perceptron
Logistic Regression	sklearn.linear_model	LogisticRegression
Logistic Regression CV	sklearn.linear_model	LogisticRegressionCV
Naïve Bayes	sklearn.naïve_bayes	BernoulliNB MultinomialNB Categorical NB
Naïve Bayes	sklearn.naïve_bayes	GaussianNB
Nearest Neighbours	sklearn.neighbors	KneighborsClassifier KneighborsRegressor
Nearest Neighbours	sklearn.neighbors	RadiusNeighborsClassifier RadiusNeighborsRegressor
Support Vector Machines	sklearn.svm	LinearSVC LinearSVR
Support Vector Machines	sklearn.svm	SVC SVR
Support Vector Machines	sklearn.svm	NuSVC NuSVR

Decision Trees	sklearn.tree	DecisionTreeRegressor DecisionTreeClassifier
Decision Trees	sklearn.tree	plot_tree
Voting models	sklearn.ensemble	VotingRegressor VotingClassifier
Bagging models	sklearn.ensemble	BaggingRegressor BaggingClassifier
Random Forest	sklearn.ensemble	RandomForestRegressor RandomForestClassifier
Adaptive Boosting	sklearn.ensemble	AdaBoostRegressor AdaBoostClassifier
Gradient Boosting	sklearn.ensemble	GradientBoostingRegressor GradientBoostingClassifier
Clustering	sklearn.cluster	KMeans
Clustering	sklearn.cluster	Agglomerative Clustering
Artificial Neural Networks	sklearn.neural_network	MLPRegressor MLPClassifier
Hyper-parameter tuning	sklearn.model_selection	GridSearchCV
Hyper-parameter tuning	sklearn.model_selection	RandomizedSearchCV
Meta estimators	sklearn.multiclass	OneVsRestClassifier
Meta estimators	sklearn.multiclass	OneVsOneClassifier
Meta estimators	sklearn.multiclass	OutputCodeClassifier
Meta estimators	sklearn.multioutput	MultiOutputClassifier MultiOutputRegressor
Meta estimators	sklearn.multioutput	ClassifierChain RegressorChain

Metrics

Following is the list of most commonly used ML metrics APIs:

Broad category	API
Classification	sklearn.metrics.accuracy_score
Classification	sklearn.metrics.precision_score
Classification	sklearn.metrics.recall_score
Classification	sklearn.metrics.f1_score
Classification	sklearn.metrics.classification_report
Classification	sklearn.metrics.confusion_matrix
Classification	sklearn.metrics.precision_recall_curve
Classification	sklearn.metrics.roc_curve
Classification	sklearn.metrics.ConfusionMatrixDisplay
Classification	sklearn.metrics.PrecisionRecallDisplay
Classification	sklearn.metrics.RocCurveDisplay
Regression	sklearn.metrics.explained_variance
Regression	sklearn.metrics.r2_score
Regression	sklearn.metrics.mean_absolute_error
Regression	sklearn.metrics.mean_squared_error
Regression	sklearn.metrics.mean_squared_log_error
Regression	sklearn.metrics.mean_absolute_percentage_error
Classification	sklearn.metrics.hinge_loss
Classification	sklearn.metrics.log_loss
Classification	sklearn.metrics.balanced_accuracy_score
Classification	sklearn.metrics.roc_auc_score
Classification	sklearn.metrics.top_k_accuracy_score

Data Preprocessing, feature selection and model selection

Following is the list of most commonly used data preprocessing, feature selectiona nd model selection APIs:

Broad category	Sub category	Library module	API
Data pre-processing	Training and test	sklearn.model_selection	train_test_split
Data pre-processing	Feature extraction	sklearn.feature_extraction	DictVectorizer
Data pre-processing	Handling missing values	sklearn.impute	SimpleImputer
Data pre-processing	Handling missing values	sklearn.impute	KNNImputer
Data pre-processing	Feature extraction	sklearn.impute	MissingIndicator
Data pre-processing	Feature scaling	sklearn.preprocessing	StandardScaler
Data pre-processing	Feature scaling	sklearn.preprocessing	MinMaxScaler
Data pre-processing	Feature scaling	sklearn.preprocessing	MaxAbsScaler
Data pre-processing	Feature encoding	sklearn.preprocessing	OneHotEncoder
Data pre-processing	Feature encoding	sklearn.preprocessing	LabelEncoder
Data pre-processing	Feature encoding	sklearn.preprocessing	OrdinalEncoder
Data pre-processing	Feature encoding	sklearn.preprocessing	LabelBinarizer
Data pre-processing	Feature encoding	sklearn.preprocessing	MultiLabelBinarizer
Data pre-processing	Preprocessing	sklearn.preprocessing	add_dummy_feature

Broad category	Sub category	Library module	API
Feature Selection	Filter feature selection	sklearn.feature_selection	VarianceThreshold
Feature Selection	Filter feature selection	sklearn.feature_selection	SelectKBest
Feature Selection	Filter feature selection	sklearn.feature_selection	SelectPercentile
Feature Selection	Filter feature selection	sklearn.feature_selection	GenericUnivariateSelect
Feature Selection	Wrapper feature selection	sklearn.feature_selection	RFE
Feature Selection	Wrapper feature selection	sklearn.feature_selection	RFECV
Feature Selection	Wrapper feature selection	sklearn.feature_selection	SelectFromModel
Feature Selection	Wrapper feature selection	sklearn.feature_selection	SequentialFeatureSelector
Data pre-processing	Dimensionality reduction	sklearn.decomposition	PCA
Data pre-processing	Feature transformation	sklearn.preprocessing	FunctionTransformer
Data pre-processing	Feature transformation	sklearn.preprocessing	PolynomialFeatures
Data pre-processing	Feature transformation	sklearn.preprocessing	KBinsDiscretizer
Data pre-processing	Feature transformation	sklearn.compose	ColumnTransformer

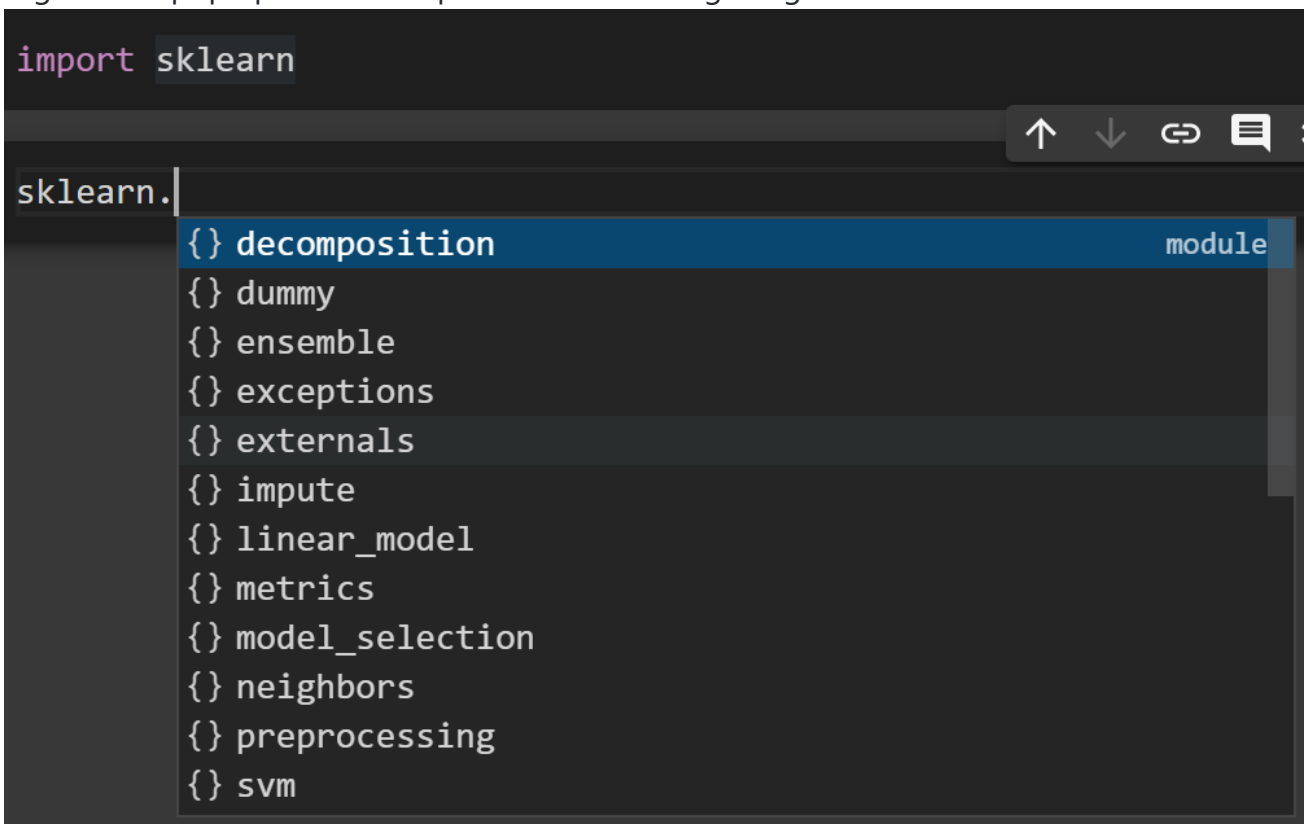
Broad category	Sub category	Library module	API
Data pre-processing	Feature transformation	sklearn.compose	TransformedTargetRegressor
Data pre-processing	Chaining transformers	sklearn.pipeline	Pipeline
Data pre-processing	Chaining transformers	sklearn.pipeline	FeatureUnion
Model Selection	Cross validation	sklearn.model_selection	KFold
Model Selection	Cross validation	sklearn.model_selection	LeaveOneOut
Model Selection	Cross validation	sklearn.model_selection	ShuffleSplit
Model Selection	Cross validation	sklearn.model_selection	cross_val_score
Model Selection	Cross validation	sklearn.model_selection	cross_validate
Model Selection	Cross validation	sklearn.model_selection	learning_curve
Model Selection	Cross validation	sklearn.model_selection	validation_curve
Target identification	Target identification	sklearn.utils.multiclass	type_of_target

Accessing sklearn modules and APIs documentation in Google Colab

1. In a Google colab document, import the module of interest. e.g. `sklearn`, as following:

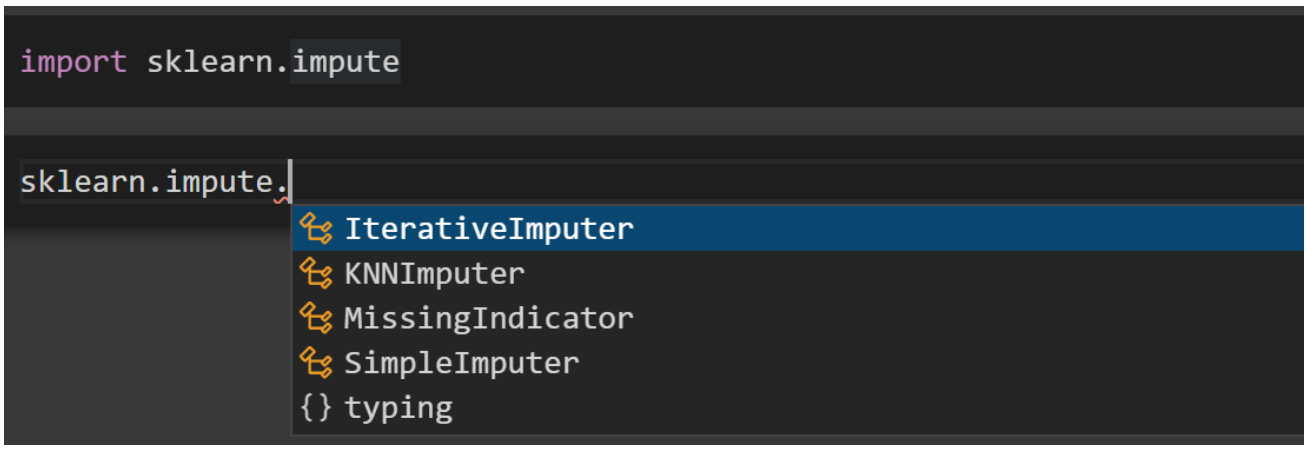
2. ```
   import sklearn
   ```
   ```

3. Type `sklearn.` (notice the dot `.` after module name) and press `Ctrl + Space Bar` together. A pop up will show up as in the following image:



as you can see in the above image, a scrollable list of all sub modules is presented.

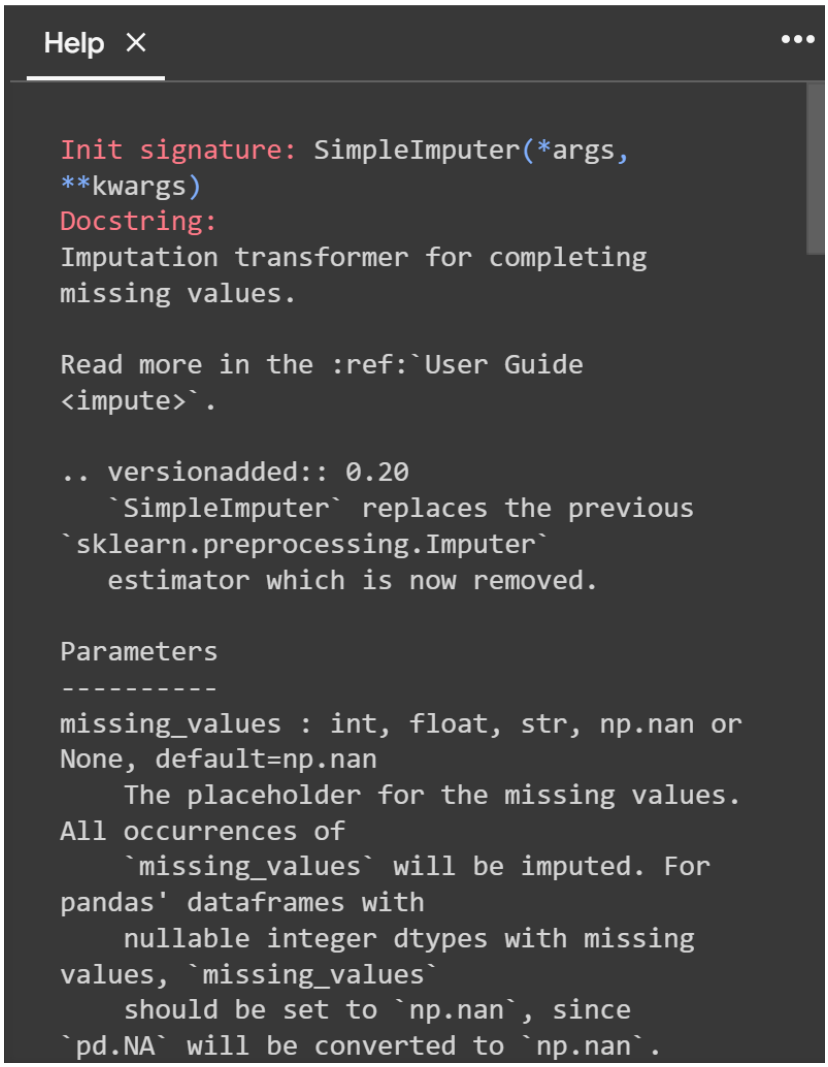
4. Let's say you want to see APIs in the `impute` module. Then follow the same procedure as above, you will see a pop up like following:



5. Now, say you want to know more about `SimpleImputer` and want to see what it does, its signature, info on its parameters and possibly some examples, then use `?` operator as following

```
from sklearn.impute import SimpleImputer
?SimpleImputer
```

6. A scrollable panel will open like following:



7. Instead of `?` operator you can use `help()` like following:

```
from sklearn.impute import SimpleImputer
help(SimpleImputer)
```

8. Instead of showing same documentation of `SimpleImputer` API in a separate panel, it will be shown as output of the code cell, as shown below: