# Scikit-learn

<https://scikit-learn.org/stable/>

Scikit-learn is regarded as a simple and efficient tool that one can use for predictive data analysis. Scikit-learn provides one with the opportunity to engage in machine learning within the programming language of python.

Classification: Identifying which category an object belongs to.

Regression: Predicting a continuous-valued attribute associated with an object.

# W3schools machine learning tutorial

Machine learning is making the computer learn from studying data and statistics.

It is a step into the direction of artificial intelligence.

It is a program that analyses data and learns to predict the outcome.

Text

Description automatically generated

# Datacamp tutorial

<https://www.datacamp.com/tutorial/machine-learning-python>

Text, timeline

Description automatically generated

## Exploratory data analysis

Takes place after reading in the dataset

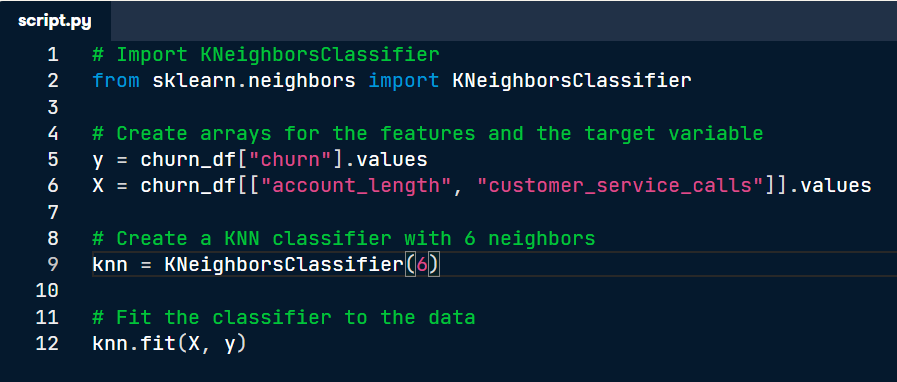
<https://campus.datacamp.com/courses/supervised-learning-with-scikit-learn/classification-1?ex=1>

Text

Description automatically generated

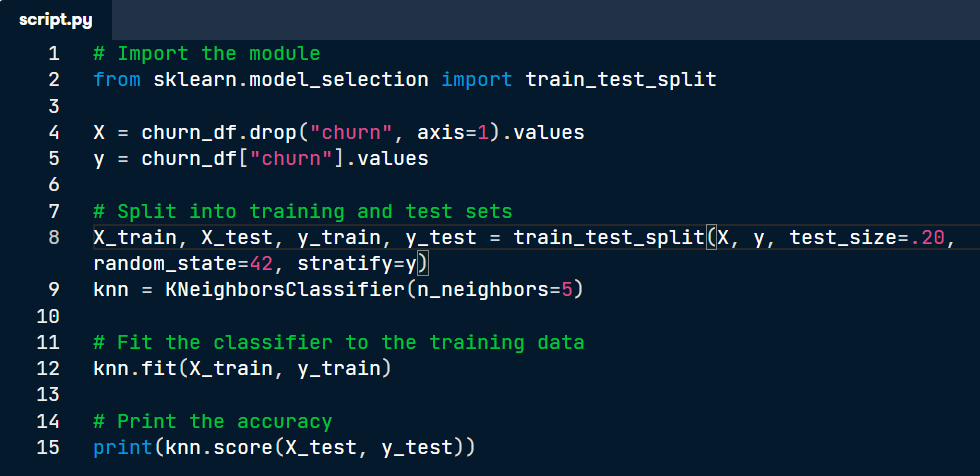
Text

Description automatically generated



A screenshot of a computer

Description automatically generated with medium confidence



# Data camp course machine learning with scikit-learn

## Classification

Text

Description automatically generated with medium confidence

### KNN Classifier

A screenshot of a computer

Description automatically generated

### KNN Making a prediction on a non-labelled piece of data

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

### Measuring model performance

Text

Description automatically generated

Text

Description automatically generated

### Train/test split

Text

Description automatically generated

## Regression

Text

Description automatically generated with medium confidence

Text

Description automatically generated

### Reshaping

Text

Description automatically generated

Loss function

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

The above code is based on the boston house dataset

### Fit & predict for regression

Text

Description automatically generated

### 5 fold cross validation

Text

Description automatically generated

### Regularized regression

A picture containing text

Description automatically generated

Often hearing this term – fit on the training predict on the test

Text

Description automatically generated

## Part 3 Fine-tuning your model

### Confusion Matrix

Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

Table

Description automatically generated

### ***\*Logistic regression is used with classification***

### Logistic regression

Graphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated

### Logistic regression – ROC curve

A picture containing text

Description automatically generated

Text

Description automatically generated

Text

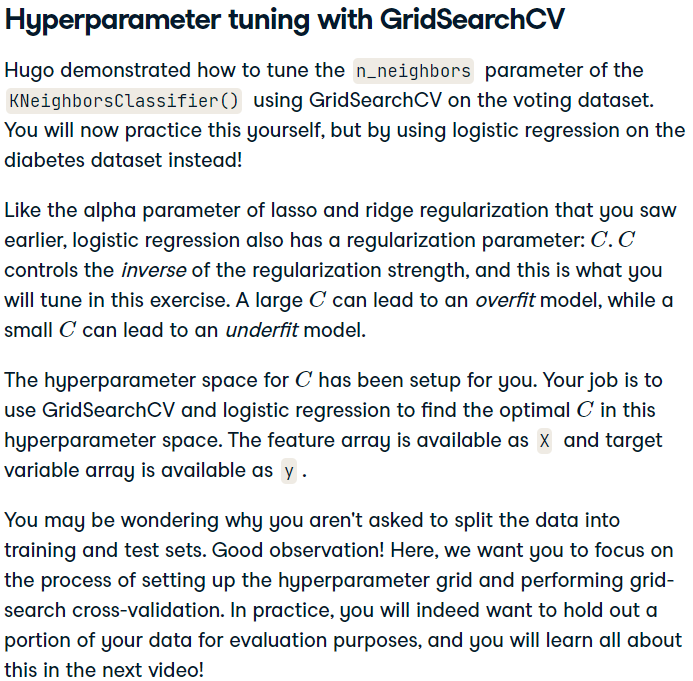
Description automatically generated

### AUC computation

Text

Description automatically generated

### Hyperparamter tuning with gridsearchCV



A screenshot of a computer

Description automatically generated

### Hyperparameter tuning with RandomizedSearchCV

Text

Description automatically generated

### Hold-out set in practice I: Classification

Text

Description automatically generated

### Hold-out set in practice II: Regression

Text

Description automatically generated

Text

Description automatically generated

## Preprocessing Part 4 Pipelines

### Exploring categorical features

A picture containing graphical user interface

Description automatically generated

### Creating dummy variables

Text

Description automatically generated

Text

Description automatically generated

### Regression with categorical features

Text

Description automatically generated

A screenshot of a computer

Description automatically generated

### Dropping missing data

Text

Description automatically generated

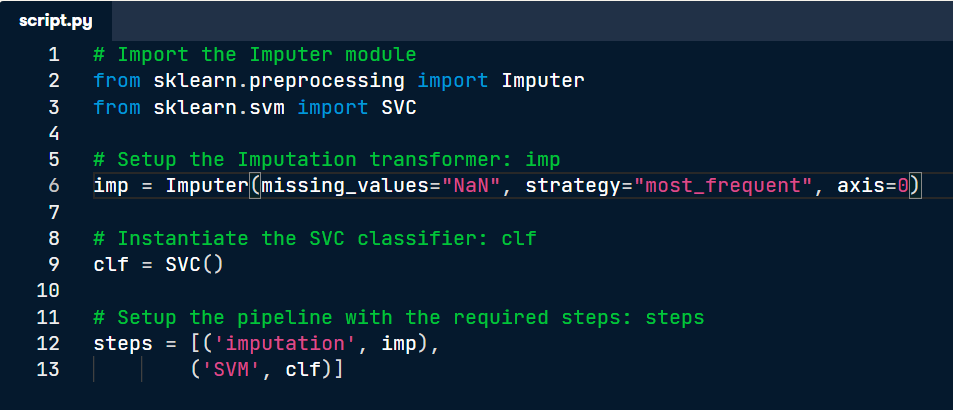
Text

Description automatically generated

### Imputing data

Text, letter

Description automatically generated



Text

Description automatically generated

Text

Description automatically generated

### Centering and scaling

Text

Description automatically generated

A screenshot of a computer

Description automatically generated

### Centering and scaling in a pipeline

Text

Description automatically generated

### Brining it all together pipeline for classification

Text, letter

Description automatically generated

Text

Description automatically generated

**Text

Description automatically generated**

### Brining it all together pipeline for regression

Text

Description automatically generated

Text

Description automatically generated

# **Common Process**

## Imports

## Instantiate the classifier

## Spilt data into training and test sets

## Fit the model on training

## Predict on the test set

# Kevin Jolly - Machine Learning with Scikit-learn Quick Start Guide

Graphical user interface, text, application, chat or text message

Description automatically generated

Kevin Jolly (2018) *Machine Learning with Scikit-learn Quick Start Guide : Classification, Regression, and Clustering Techniques in Python*. Birmingham, UK: Packt Publishing. Available at: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=e000xww&AN=1936459&site=ehost-live&scope=site (Accessed: 23 July 2022).

Graphical user interface, text, application, email

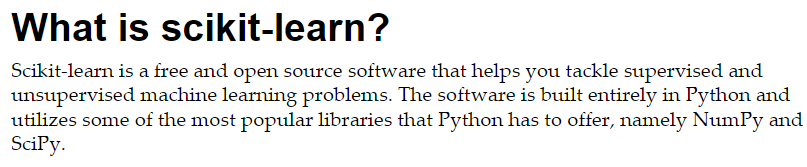
Description automatically generated

Pg. 6

## Supervised learning

Text

Description automatically generated



Pg.7

Text, letter

Description automatically generated

Pg.14

Text

Description automatically generated with medium confidence

Pg.14

## Implementing knn algorithm

Text

Description automatically generated

Text

Description automatically generated

Pg.23

Text

Description automatically generated

Pg.23

Text

Description automatically generated

Pg.24

Text

Description automatically generated

Pg.24

## Explanation overfitting / underfitting

Text, letter

Description automatically generated

Pg.25

Text

Description automatically generatedpg.25

## Scaling for performance

Text

Description automatically generated

Pg.27

A screenshot of a computer

Description automatically generated with medium confidence

Pg.28

Text

Description automatically generated

## Predicting Numeric Outcomes with Linear Regression

Text

Description automatically generated

Diagram

Description automatically generated

Pg.52

# Learning Scikit-learn : Machine Learning in Python: Experience the Benefits of Machine Learning Techniques by Applying Them to Real-world Problems Using Python and the Open Source Scikit-learn Library

Garreta, R. and Moncecchi, G. (2013) *Learning Scikit-learn : Machine Learning in Python: Experience the Benefits of Machine Learning Techniques by Applying Them to Real-world Problems Using Python and the Open Source Scikit-learn Library*. Birmingham, UK: Packt Publishing (Community Experience Distilled). Available at: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=e000xww&AN=673033&site=ehost-live&scope=site (Accessed: 26 July 2022).

Graphical user interface, text, application

Description automatically generated

# Mastering Machine Learning with Scikit-learn - Second Edition

Graphical user interface, text, application

Description automatically generated

Hackeling, G. (2017) *Mastering Machine Learning with Scikit-learn - Second Edition*. Birmingham, UK: Packt Publishing. Available at: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=e000xww&AN=1562686&site=ehost-live&scope=site (Accessed: 26 July 2022).