



NTNU

Norwegian University of Science and Technology

# Basic Data Science Tutorial: Installation Instructions

DFRWS-EU, Oslo, Norway April 24th 2019



NTNU

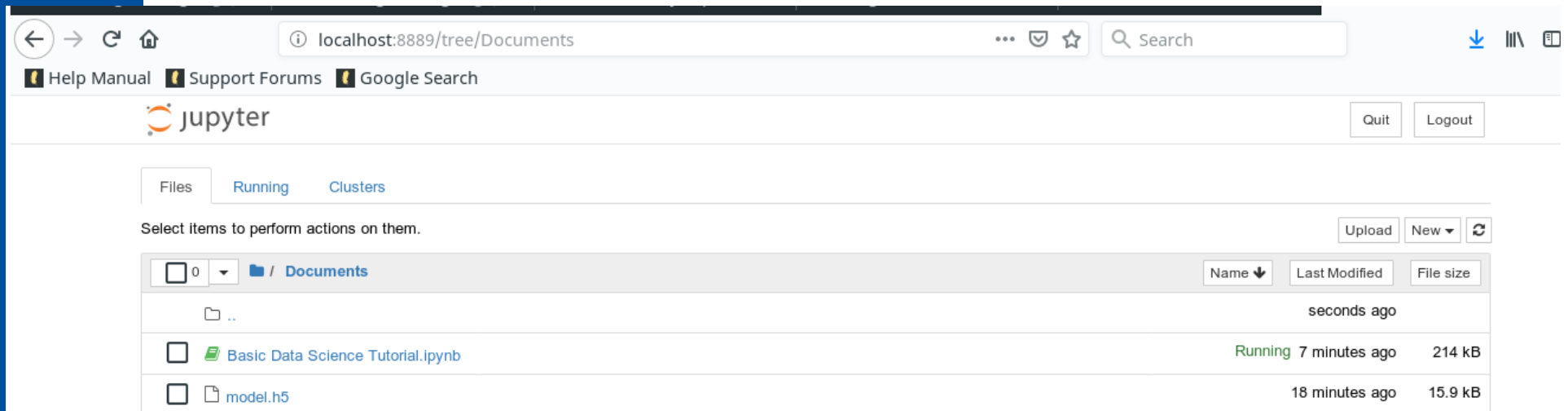
# Overview

- We go over the instructions of installing python, pip, and the necessary python packages to participate in the tutorial, and how to open up the Jupyter notebook in Linux.
- This can also be performed in Windows, but Linux is simpler.

# Installation instructions (Linux command line)

- Make sure you have python 3 installed.
  - `sudo apt-get install python3.7`
- Make sure you have the python package installer.
  - `sudo apt-get install python3-pip`
- Install following packages.
  - `pip3 install jupyter`
  - `pip3 install matplotlib`
  - `pip3 install pandas`
  - `pip3 install scipy`
  - `pip3 install seaborn`
  - `pip3 install scikit-learn`
  - `pip3 install keras`
  - `pip3 install tensorflow`
- Install necessary Jupyter requirements with:
  - `sudo apt install jupyter-core`
- When you have the Jupyter notebook, navigate to the location of the Jupyter notebook (tutorial) and run the following.
  - `~/local/bin/jupyter-notebook`

# You should see something like this



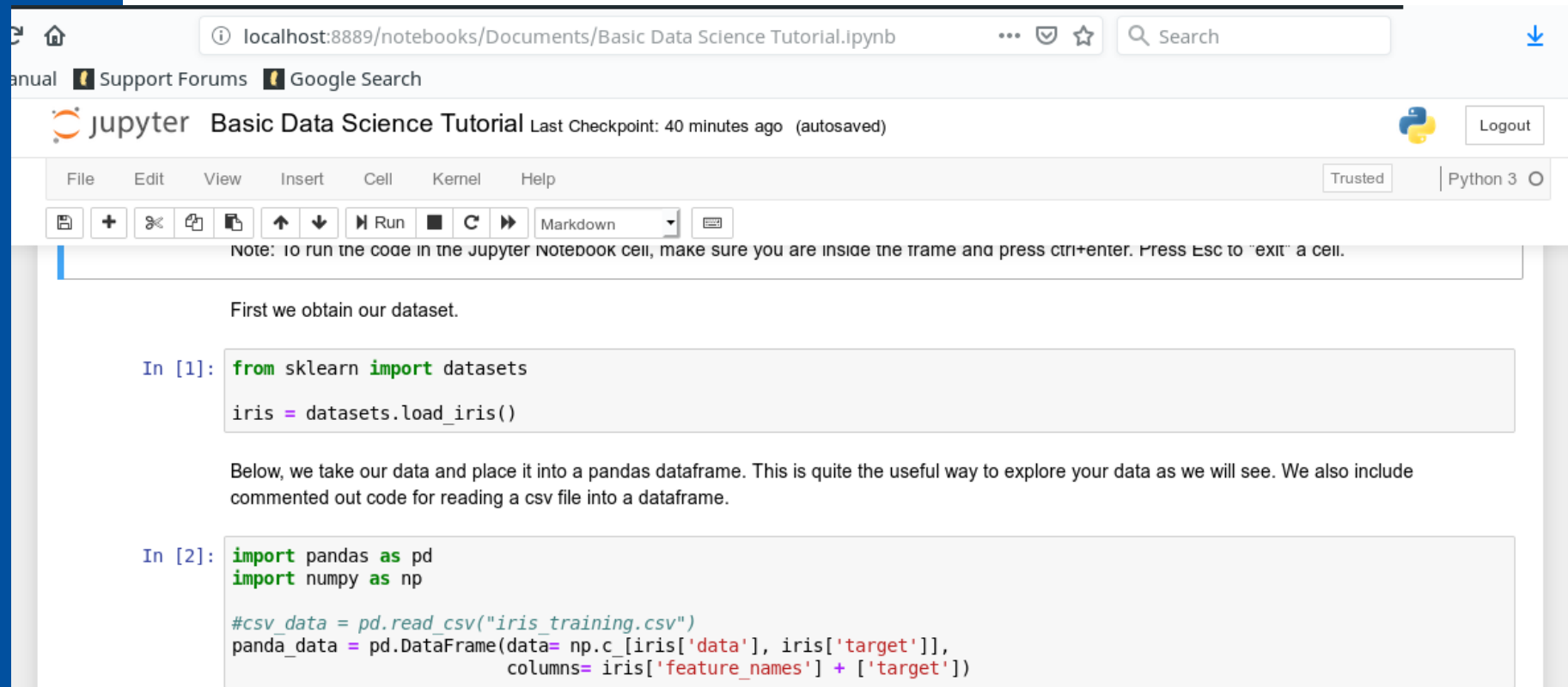
The screenshot shows the JupyterLab interface in a web browser. The address bar indicates the URL is localhost:8889/tree/Documents. The interface includes a navigation bar with 'Help Manual', 'Support Forums', and 'Google Search'. Below the navigation bar, there are tabs for 'Files', 'Running', and 'Clusters'. The 'Files' tab is active, showing a file browser view of the 'Documents' directory. The browser displays a table of files with columns for 'Name', 'Last Modified', and 'File size'. The files listed are: '..' (seconds ago), 'Basic Data Science Tutorial.ipynb' (Running 7 minutes ago, 214 kB), and 'model.h5' (18 minutes ago, 15.9 kB).

Name	Last Modified	File size
..	seconds ago	
Basic Data Science Tutorial.ipynb	Running 7 minutes ago	214 kB
model.h5	18 minutes ago	15.9 kB

- Click on the tutorial (file with the .ipynb extension).

# The Jupyter Notebook is now open

- To run all code in a “cell”, click or press enter on a cell and press “ctrl+enter”.
- To exit a cell, press “esc”.



The screenshot shows a web browser window displaying a Jupyter Notebook. The address bar shows the URL: localhost:8889/notebooks/Documents/Basic Data Science Tutorial.ipynb. The notebook title is "Basic Data Science Tutorial" with a last checkpoint of 40 minutes ago. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Help) and a toolbar with icons for file operations, navigation, and execution. A note at the top of the notebook area states: "Note: To run the code in the Jupyter Notebook cell, make sure you are inside the frame and press ctrl+enter. Press Esc to 'exit' a cell." Below this, there are two code cells. The first cell contains the code to load the Iris dataset using sklearn. The second cell contains code to load the data into a pandas DataFrame, with some code commented out for reading a CSV file.

```
Note: To run the code in the Jupyter Notebook cell, make sure you are inside the frame and press ctrl+enter. Press Esc to "exit" a cell.
```

First we obtain our dataset.

```
In [1]: from sklearn import datasets
iris = datasets.load_iris()
```

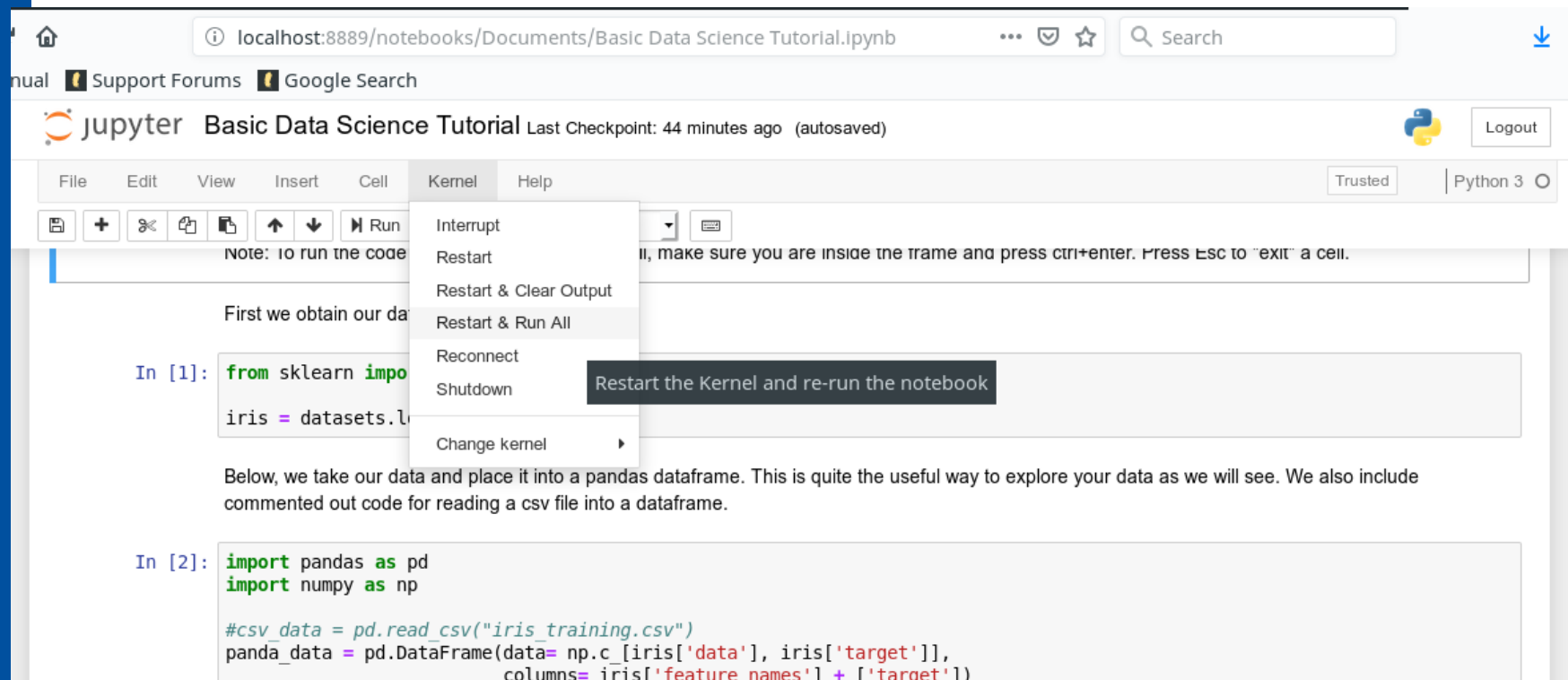
Below, we take our data and place it into a pandas dataframe. This is quite the useful way to explore your data as we will see. We also include commented out code for reading a csv file into a dataframe.

```
In [2]: import pandas as pd
import numpy as np

#csv_data = pd.read_csv("iris_training.csv")
panda_data = pd.DataFrame(data= np.c_[iris['data'], iris['target']],
                          columns= iris['feature_names'] + ['target'])
```

# Final Tests

- To test to see if everything is working as intended, press the Kernel tab at the top and select “Restart and Run All”. The cell with code at the bottom should have output.



The screenshot shows a Jupyter Notebook interface in a browser. The URL is localhost:8889/notebooks/Documents/Basic Data Science Tutorial.ipynb. The notebook title is "Basic Data Science Tutorial" and it was last checkpointed 44 minutes ago. The "Kernel" menu is open, showing options: Interrupt, Restart, Restart & Clear Output, Restart & Run All, Reconnect, Shutdown, and Change kernel. A callout box highlights "Restart & Run All" with the text "Restart the Kernel and re-run the notebook".

Note: To run the code

First we obtain our data

```
In [1]: from sklearn import datasets
iris = datasets.load_iris()
```

Below, we take our data and place it into a pandas dataframe. This is quite the useful way to explore your data as we will see. We also include commented out code for reading a csv file into a dataframe.

```
In [2]: import pandas as pd
import numpy as np

#csv_data = pd.read_csv("iris_training.csv")
panda_data = pd.DataFrame(data= np.c_[iris['data'], iris['target']],
                           columns= iris['feature names'] + ['target'])
```

# At the beginning of the tutorial...

- Open the Kernel tab again, and press “Restart and Clear Output”.
  - We will work through it cell by cell.