Data Science for Computer and Systems Sciences (DAMI)

Department of Computer and Systems Sciences, DSV, Stockholm University

Lab Template Files

This .zip folder contains all the code templates needed for the labs in the course Data Mining (DAMI), including the Jupyter notebooks explained during the videos and others notebooks provided as practice labs. These **labs are not graded and do not need to be submitted anywhere**.

The course contains **six labs** that will guide you through the data mining process starting from data exploration, then moving on to clustering and classification, model evaluation, and finally model deployment. At the end of the labs and homework assignments you will have the tools to design and run an interactive web application that uses a trained machine learning (ML) model to make predictions for a specific classification task. Each student will train a custom ML model, run the interactive website to simulate input data, and check visually how well the model performs.

Lab 0 is an introduction for those students that do not have previous experience in programming or have not used the terminal/console to install custom packages in Python. Labs 1 to 5 elaborate on specific Data Science techniques, including methods for data exploration, pre-processing, unsupervised/supervised learning, model evaluation, and deployment of ML models in production.

The course labs provide implementation of the main data science techniques explained during the course. They are intended to give the basis that allow students solve the **three homework assignments** (these **are graded** and more information is found on iLearn). The labs are developed incrementally on the following questions:

Lab 0:	How to start programming in Python?
Lab 1:	How to understand and process my original dataset to create a machine learning model?
Lab 2/3:	Which type of models can I train based on the characteristics of my data?
Lab 4:	Having multiple models, how can I evaluate which one might perform better on unseen data?
Lab 5:	How can I reuse my trained model in a production-ready environment?

Videos: Each lab session has a set of video recordings that explain the topic step-by-step. The links to the videos are published on the **iLearn** page of the DAMI course.

Main Lab Template: All the Jupyter notebooks described in the videos are provided so that students can follow along.

Practice Labs: All labs propose an extra task as an opportunity to delve deeper into concepts that will be used in the homeworks assignments. These are not mandatory and not graded.

Instructions to run Jupyter notebooks

As explained in the videos for Lab0, the execution of Jupyter notebooks (extension *.ipynb*) requires the installation of an additional package using the command `pip install jupyter` in the console. After you have installed it, you can choose the editor that will be used to run the notebooks. Next, we describe two possible ways to execute the lab files provided in this folder:

1. Native Jupyter Server

This option comes by default after installing the `jupyter` package, for external reference you can read [here](#).

- Open a terminal/console pointing at the folder where you extracted all the files (e.g. `C:/Users/student/DAMI-Labs/`)
- Execute the command `jupyter notebook`
- A new window will open in the browser with all the files. If not, copy the prompted localhost URL (e.g. `http://localhost:XXXX/?token=XXX`) and paste it in your browser.
- Click any notebook to run it or edit it.

2. VSCode (RECOMMENDED)

This option requires you to install the development environment VSCode, which can be downloaded [here](#)

- Install VSCode and the extension called *Python*
- Open VSCode and click on *File > Open Folder*, selecting the folder where you extracted all the lab files
- You should be able to see the files in the left tab of the screen
- Click any notebook to run or edit it

Pointers to the Lab files

Lab 0

The files for the first part of the Lab0 can be found in the folder **./dami_dsv/introduction/**. It includes the slides and fundamentals of Python programming, package installation and jupyter notebooks. The Jupyter notebook used in the last video of Lab0 is **./Lab0-IntroToPython.ipynb**

Lab 1

The jupyter notebooks used as introduction to Pandas and Plotting can be found in the folder **./dami_dsv/data_preprocessing/**. These labs are optional for students without previous experience using the packages *pandas* and *matplotlib*.

The main topics of the Lab1 are developed in the Jupyter notebooks **./Lab1_a_EpidemiologicalModel.ipynb** and **./Lab1_b_Preprocessing.ipynb**

Additional practice tasks for Lab1 are provided in the file **./practice_labs/Lab1_practice.ipynb**

The topics of the Lab1 are linked with **Homework Assignment 1**.

Lab 2

The main topics of the Lab2 are developed using the Jupyter notebook **./Lab2-UnsupervisedLearning.ipynb**

Additional practice tasks for Lab2 are provided in the file **./practice_labs/Lab2_practice.ipynb**

The topics of the Lab2 are linked with **Homework Assignment 2**.

Lab 3

The main topics of the Lab3 are developed using the Jupyter notebook **./Lab3-SupervisedLearning.ipynb**

Additional practice tasks for Lab3 are provided in the file **./practice_labs/Lab3_practice.ipynb**

Lab 4

The main topics of the Lab4 are developed using the Jupyter notebook **./Lab4-ModelEvaluation.ipynb**

Additional practice tasks for Lab4 are provided in the file: **./practice_labs/Lab4_practice.ipynb**

Lab 5

The main topics of the Lab4 are developed using the Jupyter notebook **./Lab5-ModelDeployment.ipynb**. The files to run the interactive web platform are located in the folder **./dami_dsv/model_deployment/**

The topics of the Lab3, Lab4, and Lab5 are linked with **Homework Assignment 3**.

Additional Bibliography

If you don't have prior experience in Python programming you can refer to:

- *Python Data Analytics* by Fabio Nelli. **[Download from SU Library]**

To know more about Data Science with Python packages you can refer to:

- *Python Data Science Handbook* by Jake VanderPlas. **[Free Web Version]**