CSCI316 Big Data Mining Implementation and Techniques

Laboratory 1

Objective

- To set up the programming environment
- To practice basic Python programming for data analytics

1. Setup the programming environment

Install all the required software (see the guideline document in the "Resources" section on Moodle).

We use Jupyter Notebook as the programming tool for Python.

(a) Start the Anaconda Prompt (search "anaconda" in the Windows 10 task bar).
(b) Type "jupyter notebook" in Anaconda Prompt and Jupyter notebook will be running in a Web browser.
(c) Create a new Python 3 notebook and then begin your implementation work.
(Note. If you use MacOS or Linus, just type "jupter notebook" in the Terminal.)

You can develop your Python code in a notebook in Jupyter Notebook (see Step 2 below).

2. Programming for data analytics

Python is a general-purpose programming language. This subject uses Python for data analytics. Some basic Python programming skills (including some necessary data structures and APIs) are demonstrated in the first part of Lecture 2.

To prepare yourself for subsequent lab tasks, walk through reproduce in Jupyter the Python codes in the lecture slides.

Do NOT attempt to copy and paste the codes from the slides—*that is error-prone*! Besides reproducing the code in Jupyter Notebook, it is also important to analyse the syntax of the Python programming language.

A note about Jupyter Notebook

You can present texts in the **Markdown format**. In the laboratory and assignment submissions, you WILL need to include annotations and explanations of your implementation and methodology. The deliverables of your individual and group assignments WILL consist of a **notebook source file** (with the .ipybn extension) and a **PDF document** which is generated from the source file. You can simply use the Web browser's PDF printing or "print as PDF" function to create the PDF.

Group Forming

Besides working on the tasks of this laboratory, please also form groups for the two group-based assignments. Group size is limited to 2-3 members (not smaller than 2 not larger than 3). Members in the same group must be either all full-time students or all part-time students.