ASIAN INSTITUTE OF TECHNOLOGY
SCHOOL OF ENGINEERING AND TECHNOLOGY

AT82.01 Computer Programming for Data Science and Artificial Intelligence, 2(0-90)
Semester: August

Objective: The course objective is to provide students hands-on programming skills and best practices related to Data Science. It is a tutorial course in which students will develop programming skills in loading, cleansing, transforming, modeling, and visualizing data.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Manipulate data sets programmatically
2. Perform exploratory data analysis programmatically
3. Build data-driven predictive models programmatically

Prerequisites: None

Course Outline:

I. Fundamentals
   1. Python programming
   2. The Python toolset

II. Working with data
   1. Numerical computation using numpy
   2. Data manipulation using pandas
   3. Exploratory data analysis

III. Data visualization
   1. Matplotlib
   2. Pandas

IV. Statistical analysis
   1. Hypothesis testing using scipy and statsmodels

V. Machine learning tools
   1. Scikit-learn
   2. PyTorch

VI. Machine learning from scratch
   1. Regression
   2. Classification
   3. Neural Network

Laboratory Session(s): Each topic is a series of tutorial sessions.

Learning Resources:

Textbooks: No specific textbook. Lab manuals and online resources will be used.

Reference Books:

Downey, A. (2014), *Think Stats*, 2nd edition, O'Reilly.

Geron, A. (2017), *Hands-On Machine learning with Scikit-Learn & TensorFlow*, O'Reilly.

McKinney, W. (2013), *Python for Data Analysis*, O'Reilly.

VanderPlas, J. (2016), *Python Data Science Handbook: Essential Tools for Working with Data*, O'Reilly.

Journals and Magazines:

IEEE Transactions on Knowledge and Data Engineering, IEEE
ACM/IMS Transactions on Data Science, ACM
Journal of Machine Learning Research (JMLR), Microtome
Neural Networks, *Elsevier*

Others:

Python tutorials available online: https://docs.python.org/3/tutorial/
Jupyter notebook tutorials available online: https://ipython.org/documentation.html
Numpy tutorials available online:https://numpy.org/doc/stable/
Pandas tutorials available online: https://pandas.pydata.org/docs/
Nltk tutorials available online: https://www.nltk.org
Matplotlib tutorials available online: https://matplotlib.org/contents.html
Visdom tutorials available online: https://github.com/facebookresearch/visdom
Scikit-learn tutorials available online: https://scikit-learn.org/stable/user_guide.html
Pytorch tutorials available online: https://pytorch.org/tutorials/

Teaching and Learning Methods:

1. **Use of online tutorials:** Students will make use of online tutorials for self-learning.
2. **Tutorial sessions**: Students will be required to perform a series of exercises and demonstrate their completion.
3. **Homework**: Several homework exercises requiring students to apply the knowledge acquired in the tutorial sessions will be assigned and graded.

Time Distribution and Study Load:

- Tutorial sessions: 45 hours.
- Self study: 100 hours.
- Homework: 35 hours.

Evaluation Scheme:

1. Project: 50% (3 mini-projects with 10% each and 1 mega project with 20%)
2. Lab exercise: 5%
3. Quiz: 10%
4. Final exam: 20%
5. Midterm exam: 15%

A grade of “A” indicates successful completion of all procedures and excellent and insightful understanding of the techniques introduced in the laboratory; “B” indicates mostly successful completion and a good understanding of the techniques; “C” indicates barely acceptable completion and understanding; and “D” indicates inability to complete many procedures and/or poor understanding of the techniques.

Instructor(s): Dr. Chaklam Silpasuwanchai