DS2XX3 Data Processing Workshop I

1. COURSE TITLE

Data Processing Workshop I

2. COURSE CODE

DS2XX3

3. PRE-REQUISITE

Nil

4. CO-REQUISITE

Nil

5. NO. OF UNITS

1

6. CONTACT HOURS

42

7. OFFERING UNIT

Data Science Programme, Division of Science and Technology

8. SYLLABUS PREPARED & REVIEWED BY

Prepared by Dr. Rui Meng

Reviewed by Dr. Zhe Xuanyuan

9. AIMS & OBJECTIVES

This workshop aims to lead students to learn independent design, research, and coding on data processing. It will help students understand the concept of applying data processing techniques to solve real problems. By doing course project in groups, students will learn how to cooperate with team members, how to collect data, pre-processing data, store data and perform data analysis.

10. COURSE CONTENT

1. Python Essentials
    a. Introduction to IPython and Jupyter
    b. Python Basics
2. Introduction to NumPy
    a. Data types
    b. Numpy arrays
    c. Indexing arrays
    d. Basic operations and manipulations on arrays

3. Data Manipulation with Pandas
    a. Introduction to Pandas objects
    b. Data indexing and selection
    c. Operating on data in Pandas
    d. Hierarchical indexing
4. Cleaning and Preparing Data
    a. Handling missing values
    b. Combining datasets, aggregation and grouping
    c. Vectorized string operation
5. Data Acquisition
    a. Reading and writing data in various formats
    b. Web scrapping basics
6. Data Storage and Database Access
    a. Database access with DB-API2
    b. SQLAlchemy
7. Data Visualization
    a. Various plots, e.g., scatter plots, line plots, box plots, bar charts,and histograms with matplotlib
    b. Customizing plots, important attributes and arguments
8. Real-World Modelling and Problem Solving

11. COURSE INTENDED LEARNING OUTCOMES (CILOS) WITH MATCHING TO PILOS

Programme Intended Learning Outcomes (PILOs)

Programme Title: Bachelor of Science (Honours) in Data Science	
PILO	**Upon successful completion of the Programme, students should be able to:**
PILO 1	Describe and explain the fundamental knowledge required to support the study and applications of Data Science;
PILO 2	Competently apply a wide range of programming concepts to software development in data collection and analysis;
PILO 3	Formulate novel methods in data information gathering and analysis to solve real world problems;
PILO 4	Collaborate and function effectively in team work with proficient communication and effective interpersonal skills;
PILO 5	Stay abreast of contemporary issues in Data Science and develop life-long effective learning skills to meet the needs of the Data Science discipline.

CILOs-PILOs Mapping Matrix

Course Code & Title: DS2XX3 Data processing workshop I		
CILO	**Upon successful completion of the course, students should be able to:**	**PILO(s) to be addressed**
CILO 1	**Explain** the basic knowledge of different steps in data processing;	PILO 2
CILO 2	**Develop** applications using Python to perform data processing;	PILOs 2,5
CILO 3	**Write** complete documents for a simple project to be presented;	PILO 4
CILO 4	**Collaborate** in a team to complete a simple project.	PILO 4

12. TEACHING & LEARNING ACTIVITIES (TLAS)

CILO No.	TLAs
CILO 1	● **Lecture**: The instructor will explain the course material in detail. ● **Hands-on practices**: Students will practice the knowledge learnt in the lectures.
CILO 2	● **Lecture**: The instructor will explain the course material in detail. ● **Assignment**: Each student is required to independently work on 1 assignment. ● **Hands-on practices**: Students will practice implementing applications for data processing using Python.
CILO 3	● **Project**: For each project, students are required to write software requirement specification and software design specification. Students will present the projects.
CILO 4	● **Project**: Each student is required to design and implement one project. For each project, students are required to write software requirement specification and software design specification. Students will present the projects.

13. ASSESSMENT METHODS (AMS)

Type of Assessment Methods	Weighting	CILOs to be Addressed	Description of Assessment Tasks
Hands-on practices	30%	1-2	Hands-on practices will test the students' ability to use the technologies taught in this course. It will also test and reward students' problem solving capabilities and coding styles.
Quizzes	20%	2-3	Quizzes will assess the students' understanding of the knowledge on python and data processing.
Projects	50%	3-4	Projects will enable students to apply learnt techniques and skills, and demonstrate their understanding of taught topics.

14. TEXTBOOKS / RECOMMENDED READINGS

1. Mastering python for data science. Samir Madhavan. Packt Publishing, 2015.
2. Python Data Science Handbook: Essential Tools for Working with Data. Jake VanderPlas. O'Reilly Media, Inc., 2017.

3. Practical Web Scraping for Data Science. Seppe Vanden Broucke, Bart Baesens. Apress, 2018.
4. Java for Data Science. Richard M. Reese, Jennifer L. Reese. Packt Publishing., 2017.
5. Python Programming for Beginners: An Introduction to the Python Computer Language and Computer Programming, Jason Cannon, 2014.
6. Data Analytics for Beginners: Basic Guide to Master Data Analytics, Paul Kinley, CreateSpace Independent Publishing Platform, 2016.
7. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd Edition, Wes McKinney, O'Reilly Media, 2017.