Subject Overview

CSCI316: Big Data Mining Techniques and Implementation

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Teaching Team

Lectures:

- Dr. Su Guoxin
  - Email: guoxin@uow.edu.au

Tutorials/Workshops:

- Mr. Sionggo Japit
  - Email: sjapit@uow.edu.au

Assessments

- 2 Individual Assignments 2 * 15% = 30%
- 1 Group Assignment (two questions) 2 * 10% = 20%
- Final Exam 50%

❖ All submissions (except final exam) are uploaded via Moodle.

Subject Learning Outcomes

1. Demonstrate an understanding of Big Data project lifecycle and related concepts.
2. Demonstrate an understanding of Big Data processing models and methodologies.
3. Understand and implement data pre-processing and post-processing techniques for Big Data applications.
4. Understand and implement data mining algorithms for Big Data applications.
5. Understand and implement real-time processing and stream mining methods for Big Data applications
6. Demonstrate the ability to develop Big Data application by using popular programming libraries and software platforms.

Learning Outcomes

- Knowledge
  - Understanding techniques of building big data applications
  - Dive into selected data mining / machine learning algorithms
- Practical Skills
  - Implementation of Big Data applications.
  - Hands-on experience with some popular Big Data and machine learning tools (e.g., Apache Spark and TensorFlow)
- Teamwork
  - Build up team spirit in solving challenging problems.

Subject Materials

- Subject Materials are available on Moodle:
  - ❖ Subject Outline (including reference books)
  - ❖ Lecture Slides
  - ❖ Laboratories
  - ❖ Assignments
- Major reference text:

[1] Han, J., Pei, J. and Kamber, M., 2012. Data mining: concepts and techniques (3ed). Elsevier

[2] Aurelien Geron. 2022. Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow (3ed)

[3] Chambers, B. and Zaharia, M., 2017. Spark: The Definitive Guide. O'Reilly.

Lecture Topics

Lecture	Contents	Reference/ reading
1	Introduction to big data, Programming basics, Data collection	[1] Chapter 1
2	Data Pre-processing	[1] Chapter 2 & 3
3	Big Data Project Life-cycle	[2] Chapter 2
4	Classification by Splitting Data Sets	[1] Chapter 8
5	Probabilistic Classification and Model Evaluation	[1] Chapter 8
6	Handling Massive Data Sets	[3] Chapter 1 & 24
7	Training Artificial Neural Networks	[2] Chapter 10

Note. There are **more than 7 lecture notes** on Moodle, but the content are grouped into **7 lecture themes** above.

Programming & Software

- **Python 3** (plus libraries)
  - Scientific-computing & data analytics libraries: **Numpy**, **Pandas**, **Matplotlib** and others (https://www.scipy.org/)
  - Machine linear library: **Scikit-Learn** (https://scikit-learn.org/stable/)
- Big data processing platform **Apache Spark** (https://spark.apache.org)
  - **PySpark** (Spark's Python APIs)
- Deep learning library **TensorFlow** (https://www.tensorflow.org/)
- All software are available or downloaded in Google Colab. They can can also be installed in your personal computers.

What contents to expect in this subject?

- Big data includes a very large and fast-growing area which encompasses a large collection of techniques
  - ranging from maths and algorithms to programming libraries and APIs, and from prototype development to tools and frameworks
  - including data storing, streaming, querying, analytics and artificial intelligence (machine learning, deep learning, image processing, LLM)
- The subject contents include two perspectives.
  - On the **practical** perspectives, you will learn the whole life cycle of big data project, and the tools and libraries to build big data and machine learning applications.
  - On the **fundamental/theoretical** perspective, you will learn some selected models/algorithms in depth. You will test your knowledge by writing code in a low level.
- *Why learn both levels?*
  - A natural learning process, a solid understanding
  - Cannot use a tool in a good way if you lack knowledge of it
  - Foster an ability to learn and use other models, algorithms, methods, libraries, tools, etc.

Questions

UNIVERSITY
OF WOLLONGONG
AUSTRALIA