



1. General Information

Course Subject	IIMT
Course Number	3642
Course Title	Managing and Mining Big Data
Academic Years	2023-2024
Grading Method	Letter

2. Instructors

Professor Li,Zhepeng
Office: Room 822 /F K.K. Leung Building
Email: zpli@hku.hk
Office: 39102404
Subclasses: 2A

4. Course Description

Course Description	With vast amount of data being available, we are now in the era of "big data". The knowledge and skills on large-scale data management are becoming increasingly important. This course provides a broad introduction to bi, data, including fundamental concepts and basic techniques for data acquisition and aggregation, state-to-the-art algorithms for data analysis, and technologies on modern computing platforms and systems for handling big data.
Prerequisites	IIMT2641: Introduction to business analysis STAT1602

5. Course Objectives

1. To provide students a broad introduction to big data.
2. Understand state-of-the-art algorithms used in big data analytics, and learn how to apply these algorithms to solve big data problems with Python and R.
3. Gain proficiency in applying standard software packages and systems for big data management and analytics.
4. To provide students extensive hands-on programming and analysis experiences on practical problems.

6. Faculty Learning Goals

- Goal 1: Acquisition and internalization of knowledge of the programme discipline
- Goal 2: Application and integration of knowledge
- Goal 3: Inculcating professionalism
- Goal 4: Developing global outlook

6. Faculty Learning Goals
Goal 5: Mastering communication skills
Goal 6: Cultivating leadership

7. Course Learning Outcomes	Aligned Faculty Learning Goals					
Course Teaching and Learning Activities	1	2	3	4	5	6
CLO1. Clearly identify and formulate the relevant problem	✓					✓
CLO2. Select and use effective methods to address the problem		✓				
CLO3. Use software tools to provide solution to the issue at hand		✓		✓		
CLO4. Communicate the solution effectively			✓		✓	✓

8. Course Teaching and Learning Activities		
Course Teaching and Learning Activities #	Expected Study Hours	Study Load (% of study)
T&L1. Interactive lectures	40	25
T&L2. Assignments, tutorials	40	25
T&L3. Self-study, group discussion	20	12.5
T&L4. Project	40	25
T&L5. Quizzes	20	12.5
	Total: 160	Total: 100

9. Assessment Methods			
Assessment Methods	Description	Weight %	Aligned Course Learning Outcomes
A1. Participation	Attendance & discussions	10%	1,2,3,4
A2. Assignments		30%	1,2,3,4
A3. Quizzes		20%	1,2,3,4
A4. Final project		40%	1,2,3,4
A5. Final Exam		0%	

Assessment Rubrics	
A2. Assignments	Effort and accuracy
A+,A,A-	
B+,B,B-	
C+,C,C-	

Assessment Rubrics	
D+,D	
F	
A3. Quizzes	Effort and accuracy
A+,A,A-	
B+,B,B-	
C+,C,C-	
D+,D	
F	
A4. Final project	Effort and accuracy
A+,A,A-	
B+,B,B-	
C+,C,C-	
D+,D	
F	

10. Course Grade Descriptors	
A+,A,A-	<ul style="list-style-type: none"> - Demonstrate a strong understanding of all relevant knowledge - Handling questions professionally - High participation in discussions - Present arguments that have an element of originality - Achieve a standard of excellent performance in the exams with very accurate computation and very good analytical and problem solving skills - Excellent performance in assignments
B+,B,B-	<ul style="list-style-type: none"> - Demonstrate a good understanding of all relevant knowledge - Handling questions in a logically way - Good participation in discussions - Present arguments that go beyond the lecture and textbook - Achieve a standard of good performance in the exams with accurate computation and good analytical and problem solving skills - Good performance in assignments
C+,C,C-	<ul style="list-style-type: none"> - Demonstrate a basic understanding of the concepts involved - Fairly address questions as set - Some participation in discussions - Present arguments in well-structure manner - Meet a standard of acceptable performance in the exams with reasonable computation and acceptable analytical and problem solving skills - Acceptable performance in assignments
D+,D	<ul style="list-style-type: none"> - Demonstrate a minimum understanding of the concepts involved - Barely address questions as set - Minimal or no participation in discussions - Present arguments in marginally acceptable manner - Meet a standard of marginally acceptable performance in the exams with some errors in computation and barely adequate analytical and problem solving skills - Marginally acceptable performance in assignments
F	<ul style="list-style-type: none"> - Demonstrate a poor understanding of the concepts involved - Unable or unwilling to handle questions

10. Course Grade Descriptors

	<ul style="list-style-type: none"> - Minimal or no participation in discussions - Present arguments poorly - Fail to meet a standard of passing the exams with major errors in computation and inadequate analytical and problem solving skills - Poorly performance in assignments
--	---

11. Course Content and Tentative Teaching Schedule

Topic/ Session	Date	Time	Content	Readings	Assignments	Other information
1			Supervised learning: Poisson regression			
2			Unsupervised learning: Principal component analysis			
3			Classification: Logistic regression and linear discriminate analysis			
4			Regularization and model selection: LASSO, Ridge and Subset selection			
5			Hyperparameter tuning: cross-validation			
6			Support vector machine			
7			Supervised modelling process			
8			Tree-based methods: bagging, random forest			
9			Tree-based methods: boosting			
10			Neural network and deep learning: introduction			
11			Neural network and			

11. Course Content and Tentative Teaching Schedule

			deep learning: CNN and RNN			
12			Neural network and deep learning: application (image, natural language processing, reinforcement learning)			
13			Project presentation			

12. Required/Recommended Readings & Online Materials

Reading	<ul style="list-style-type: none">• Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2008, Springer. http://statweb.stanford.edu/~tibs/ElemStatLearn/download.html.• Christopher Bishop. Pattern Recognition and Machine Learning, Springer.• Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press. http://www.deeplearningbook.org/
Textbook	Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, <i>An Introduction to Statistical Learning with Application in R</i> , First Edition, 2013, Springer. http://www-bcf.usc.edu/~gareth/ISL/ .

13. Means / Processes for Student feedback on Course

✓	Conducting mid-term survey in addition to SETL around the end of the semester
	Online response via Moodle site
	Others

14. Course Policy

An orderly learning environment is extremely important for this course. Academic dishonesty includes cheating, plagiarism, unauthorized collaboration, falsifying academic records, and any act designed to avoid participating honestly in the learning process. Any dishonesty will result in an F in the course.

15. Additional Course Information

No late assignment will be accepted.
Lecture notes and self-learning materials will be uploaded on Moodle.
The instructor reserves all the rights to make necessary changes to the syllabus. These changes will be announced as soon as possible.