Subject Description Form

Subject Code	AMA563
Subject Title	Principles of Data Science
Credit Value	3
Level	5
Pre-requisite/ Co-requisite/ Exclusion	nil
Objectives	To provide students with a basic introduction to data science along with real examples and case studies from both academic and industrial sources. To introduce basic concepts that are closely related to the computing foundation of data science and analytics. To introduce some software and languages for data analysis.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: (a) understand the foundations of data science; (b) master basic techniques in data analytics; (c) describe some applications of data science; (d) master some software related to data science.
Subject Synopsis/ Indicative Syllabus	Data Analytics: (18 hours) Basic probability theory, random variables, higher order moments. Estimation theory: method of moments, maximum likelihood, sufficient statistics, exponential family, bias, variance, mean squared error, minimum variance unbiased estimator. Hypothesis testing: Neyman-Pearson lemma, significance and power, likelihood ratio test, and information criteria. Case studies: (11 hours) Contemporary applications. Numerical methods: (10 hours) Preliminaries of selected programming language, e.g. R or Python. Data type; basic operations on linear algebra; conditions and loops; functions and recursion; parallel computing; documentation; objects; scoping; file operations; data visualization; statistical packages.
Teaching/Learning Methodology	The subject will mainly be delivered through lectures and tutorials. The lectures will be conducted to introduce the theoretical background, and practical problems / scenarios will be discussed in the tutorial sessions to illustrate how the theory developed can be applied in practice. Students are encouraged to use R or Python to perform an exploratory analysis by using real-world data.

Assessment Methods in Alignment with Intended Learning Outcomes

Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)			
		a	b	c	d
1. Assignments	20%	✓	✓	✓	✓
2. Project	30%	✓	✓	✓	
3. Examination	50%	✓	✓	✓	✓
Total	100%				

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:

This subject focuses on the mathematical foundation of data science. Many of these topics are based on theory in statistics and machine learning. **Exam-based assessment** is an appropriate assessment method, including a 50% examination.

Since this subject also emphasizes on understanding the implementation of various numerical methods in data science, a mini-project that takes a weight of 30% is appropriate for assessing the intended learning outcomes (c) and (d), in which students will be encouraged to analyze large datasets using numerical methods and communicate their findings. A 20% worth of assignments are also included as a component of continuous assessment in order to keep students in progress.

Continuous Assessment comprises assignments, mini-project and test. A written examination is held at the end of the semester.

Student Study Effort Required

Class contact:	
▪ Lecture	26 Hrs.
▪ Tutorial	13 Hrs.
Other student study effort:	
▪ Assignments/Projects	58 Hrs.
▪ Self-study	40 Hrs.
Total student study effort	137 Hrs.

Reading List and References

Textbooks:

George Casella, Roger L. Berger	Statistical Inference	Duxbury / Thomson Learning, 2002
Trevor Hastier, Robert Tibshirani, Jerome Friedman	The Elements of Statistical Learning: Data Mining, Inference, and Prediction	Springer 2018, 2nd edition

	Sheldon Ross	A First Course in Probability	Pearson, 8th edition
	References:		
	Christopher Bishop	Pattern Recognition and Machine Learning	Springer 2006, 1st edition
	Jure Leskovec, Anand Rajaraman, Jeff Ullman	Mining of Massive Datasets	Available online, 2nd edition
	Amy N. Langville and Carl D. Meyer	Google's PageRank and Beyond: The Science of Search Engine Rankings	Princeton University Press, 2012
	Yangchang Zhao	R and Data Mining: Examples and Case Studies	Academic Press, 2012