Subject Code	COMP6434
Subject Title	Big Data Analytics and Artificial Intelligence
Credit Value	3
Level	6
Normal Duration	1 semester
Pre-requisite/ Co-requisite/ Exclusion	Nil
Role and Purposes	The objectives of this subject are for students to: 1. introduce students the concept and challenge of big data (3 V's) 2. apply skills and tools to manage and analyze big data in the fintech context. 3. understand the fundamental concepts of artificial intelligence; 4. understand the contemporary techniques in machine learning; 5. appreciate the effectiveness of hybridization of different artificial intelligence techniques. This subject will contribute to the achievement of the DFintech program outcomes by • allowing students to acquire the ability to conduct original applied research in tech-related business areas. (Outcome 3)
Subject Learning Outcomes	Upon completion of the subject, students will be able to: a. acquire deep understanding of the sophisticated concepts and features of big data and AI technologies and applications; b. master advanced big data models, AI models, and their technical features, as well as build deep insights about what kinds of applications they can support; c. analyze the impact of advanced big data techniques for real-world business decisions and strategy applied in international companies. d. acquire a complete and in-depth landscape of the history, development and various applications of artificial intelligence in various real-world business sectors; e. master advanced AI techniques, including knowledge representation and reasoning process techniques, and be able to apply them in business applications; f. develop strong skills in machine learning, such as linear regression, decision tree induction, and artificial neural networks, and be able to devise new real-world solutions by applying the skills.

Subject Synopsis/ Indicative Syllabus	Topic 1. Introduction to and collection of Big Data • The 3 V's, their challenges and application domains. • Eventual Consistency and NoSQL systems MongoDB, Google BigTable • Data Visualization: Data types and dimensions; Visual encoding and perception Topic 2. Large-Scale Data Analytics Systems • Auto-Parallel Data Programming; MapReduce, Hive, and Parallel Databases • Sentiment Analysis Topic 3. Machine Learning Systems for Big Data Topic 4. Artificial Intelligence (AI), its roots and scope • Overview of AI application areas; • Expert systems, natural language understanding and semantics, planning and robotics, and machine learning. Topic 5. Artificial Intelligence as representation and search • The Propositional Calculus and Predicate Calculus; using inference rules to produce predicate calculus expressions; • Strategies and structures for state space search; heuristic search; recursion-based search; admissibility, monotonicity and informed-ness of search algorithms. Topic 6. Knowledge representation and reasoning • Rule-based production systems; case-based reasoning systems and model based reasoning systems; • Reasoning under uncertain situations: stochastic methods and fuzzy expert systems. Topic 7. Machine Learning • Linear regression; • Decision tree induction algorithms; • Artificial neural networks
Teaching/Learning Methodology	The course will be offered in a mode that combines seminars, case study, team presentations, and group discussions.

Assessment Methods in Alignment with Intended Learning Outcomes

(Note 4)

Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)					
		a.	b.	c.	d	e	f
Continuous Assessment*	**100%**						
1. Class participation	20 %	√	√	√	√	√	√
2. Group project & presentation	20 %	√	√	√	√	√	√
3. Lab exercise/Assignment	20%	√	√	√	√	√	√
4. Individual Assessment (e.g. Test)	40%	√	√	√	√	√	√
Total	100 %						

**Weighting of assessment methods/tasks in continuous assessment may be different, subject to each subject lecturer.*

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes: the various methods are designed to ensure that all students taking this subject –

1. Class participation aims to stimulate students to be exposed to various new potential applications of the smart city and urban informatics technologies in different business areas.
2. Lab exercise is designed to encourage students to understand the relevant knowledge and to practice, in order to enrich their hands-on experience with various software tools.
3. The group project is designed to enhance students' ability to acquire the understanding and using different knowledge, principles, techniques, tools to solve a real problem through team.
4. Individual assessment is used to assess individual students' ability to have an overall understanding of the inter-relationship of the various technologies and their individual characteristics.

Student Study Effort Expected	Class contact:	
	▪ Lectures	25 Hrs.
	▪ Laboratories	5 Hrs.
	Other student study effort:	
	▪ Preparation for the class	30 Hrs.
	▪ Preparation for Projects/Assignments	50 Hrs.
	▪ Preparation for Individual assessment	10 Hrs.
	Total student study effort	120 Hrs.
Reading List and References	Big Data Analytics 1. How Vertica Was the Star of the Obama Campaign, and Other Revelations 2. Cohen et al.“MAD Skills: New Analysis Practices for Big Data”, 2009 3. Ullman, Rajaraman, Mining of Massive Datasets, Chapter 2 4. Stonebraker et al., “MapReduce and Parallel DBMS’s: Friends or Foes?”, Communications of the ACM, January 2010. 5. Dean and Ghemawat, “MapReduce: A Flexible Data Processing Tool”, Communications of the ACM, January 2010. 6. Rick Cattell, “Scalable SQL and NoSQL Data Stores”, SIGMOD Record, December 2010 (39:4) 7. Elmagarmid, et. al. "Duplicate Record Detection: A Survey" 8. Koudas, et. al. "Record Linkage: Similarity Measures and Algorithms" 9. Chapter 3 of A Handbook of Statistical Analyses Using R 10. Gregory Park on overfitting to the leaderboard in a Kaggle Competition 11. Xindong Wu et al., Top 10 Algorithms in Data Mining, Knowledge and Information Systems, 14(2008), 1: 1-37. (read C4.5) 12. Ullman, Rajaraman, Mining of Massive Datasets , Chapter 1 13. Pedro Domingos, A Few Useful Things to Know about Machine Learning, CACM 55(10), 2012 14. Pat Hanaran, Tools for Data Enthusiasts 15. Jeffrey Heer, Michael Bostock, Vadim Ogievetsky, A Tour through the Visualization Zoo, Communications of the ACM, Volume 53 Issue 6, June 2010 16. Howard Wen, "Big Ethics for Big Data", O'Reilly Media Artificial Intelligence 1. George F. Luger, Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 6th Edition, Addison Wesley, 2009.	

	Reference Books: 1. Sankar K. Pal and Simon C. K. Shiu, Foundations of Soft Case-Based Reasoning, John Wiley, 2004. 2. Michael Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems. 2nd edition, Addison Wesley, 2005.