

FUNDAMENTALS OF DATA SCIENCE

GENERAL INFORMATION

Course Title	In English: Fundamentals of Data Science		
	In Vietnamese: Khoa học dữ liệu đại cương		
Course Code	<i>DS3.001</i>	Credit points (ECTS)	3
Requirement	Required		
Prerequisites			
Time Commitment	Lecture	24 hrs	
	Exercises	0 hrs	
	Practical/Labwork	12 hrs	
	Tutorial	0 hrs	
	Total	36 hrs	

DESCRIPTION

Course Objectives	CO1	Provide a general overview of data science and point out its important features, purpose, taxonomy, and methods
	CO2	Provide essential concepts behind various data science techniques
Course learning outcomes (CLO)	CLO 1	Implement step-by-step data science methods to specific problems
	CLO 2	Develop data-analytic way of thinking through theory and practice
Course Description	Data science is an interdisciplinary field that uses scientific methods from statistics, data mining and machine learning etc. to extract knowledge and insights from data and apply them in a broad range of applications. This course provides an introduction to the essential concepts of data science and presents effective methods for data analysis. Most importantly, students will learn how to think data-analytically, and fully appreciate how data science methods can solve various real-world problems.	

Textbooks	(1) Vijay Kotu, & Bala Deshpande. (2019). Data Science Concepts and Practices. Morgan Kaufmann Publishers . (2) Chris Chapman, & Elea McDonnell Feit (2015). R for Marketing Research and Analytics. Springer.
------------------	---

ASSESSMENT/EVALUATION

	% kết quả/Percentage	Loại hình/Type
Attendance/Attitude	10%	
Assignment	40%	
Mid-term exam	0%	
Project / Presentation	0%	
Final exam	50%	Project Presentation

MAIN CONTENTS

No.	Contents	Hours			Resources
		L e c t .	E x r.	P r c.	
1	Introduction				
	- Data science way of thinking	1			
	- Data science components and data science process	1			
	- Some problems and data science solutions	1			
2	Fundamentals of data analysis				
	- Describing data	1		1	
	- Relationships between variables	2		1	
	- Comparing groups: tables, visualizations and statistical tests	3		2	
3	Identifying drivers of outcomes				
	- Linear regression model	1		1	
	- Linear model for binary outcomes: logistic regression	2		1	
	Confirmatory factor analysis and structural equation modeling	2		1	
4	Reducing data complexity	1		0	

	- Statistical techniques for feature selection Principal component analysis	2	1	
5	Segmentation analysis - Clustering-based segmentation Supervised segmentation with tree-structured model	2 2	2 2	
6	Final project presentation	3		