UNIVERSITAS INDONESIA
Faculty of Mathematics and Natural Sciences
Department of Mathematics
Building D, Kampus UI Depok 16424, Telp: 021 - 7863439,
Email: sekretariat.math@sci.ui.ac.id, website: https://www.math.ui.ac.id/

MODULE HANDBOOK

Module designation	*Introduction to Data Science*
Semester(s) in which the module is taught	*1*
Person responsible for the module	*Sarini Abdullah*
Language	*Indonesian*
Relation to curriculum	*Compulsory* *The module is shared with all of undergraduate study programs in FMIPA UI.*
Teaching methods	*Flipped Class and Problem based learning using E-learning*
Workload (incl. contact hours, self-study hours)	*(Estimated) Total workload: 2 × 170 minutes per-week, for 16 weeks* *Lecture time: Friday, 13.00-14.40 pm (100 minutes)* *Laboratory session: 60 minutes per 2- week* *Private study (weekly average):* • *120 minutes asynchronously before the lecture time* • *90 minutes for reflection and evaluation after the lecture time.*
Credit points	*2 SKS (3.18 ECTS)*
Required and recommended prerequisites for joining the module	*No pre-requisites*

Module objectives/intended learning outcomes	*After completing the course, students are expected to have the following capabilities:* *In terms of knowledge:* • *Able to identify and explain variety of data types and their relation to the option of methods of data analysis.* • *Aware of- and can explain the ethics in data gathering.* • *Explain the concept of probability (data distribution) for discrete and continuous data.* • *Explain the concept of sampling distribution* • *Able to explain the concept of big-data and data science, and their relationship (or the role of statistics) in those fields* *In terms of skills:* • *Able to extract insight from transactional data and provide the interpretation in relation to the problem represented by the data.* • *Able to perform data pre-processing and simple visualization.* • *Able to perform data processing using R/Python (or other open-source packages) for simple data: correlation, regression, clustering, hypothesis testing (1 sample, $K \geq 2$ samples, proportion, variance test)* *In terms of competence:* • *Able to identify and explain problems in the real-life application that can be solved using basic statistical methods, data science, and/or big-data technologies* • *Able to use the open-source packages (R and/or Python) for data analysis.*
Content	1. *Introduction to big data and data science: history and properties of big data and data science, data profession and competencies, variety, types and formats of data, types of learning algorithms, technology in/for data science and big data; challenge, trend and opportunity in data science and big data.* 2. *Probability and data distributions (discrete and continuous)* 3. *Hypothesis testing: 1 sample, $K \geq 2$ samples, proportion, variance* 4. *Correlation, simple linear regression, K-means clustering.*
Examination forms	*Post-tests : multiple-choice questions, Mid- and final-exams : multiple choice and essay, laboratory activities : data processing.*

Study and examination requirements	*The marking is based on the following components:* 1. *Post-tests: 15%* 2. *Group assignment & lab sessions: 30%* 3. *Mid-term exam: 25%* 4. *Final exam: 30%* *To successfully pass the module it requires minimum 55% of the total mark.* *Mark — Grade* *85—100 — A* *80—<85 — A-* *75—<80 — B+* *70—<75 — B* *65—<70 — B-* *60—<65 — C+* *55—<60 — C* *40—<55 — D* <40 — E
Reading list	1. *Baesens, B. (2014). Analytics in a big data world: The essential guide to data science and its applications. John Wiley & Sons. (general introduction to data science and big data)* 2. *Cielen, D., Meysman, A., & Ali, M. (2016). Introducing data science: big data, machine learning, and more, using Python tools. Manning Publications Co. (good technical reference to big data related problems)* 3. *Berman, J. J. (2018). Principles and Practice of Big Data: Preparing, Sharing, and Analyzing Complex Information. Academic Press.* 4. *Ratner, B. (2017). Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data. Chapman and Hall/CRC.* 5. *Furht, B., & Villanustre, F. (2016). Big data technologies and applications. Berlin, Germany: Springer.* 6. *Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (1993). Probability and statistics for engineers and scientists (Vol. 5). New York: Macmillan.* 7. *Freeman, L., & Peace, A. G. (Eds.). (2005). Information ethics: privacy and intellectual property. IGI Global.*