H6IDS: Introduction to Data Science

Module Code:	H6IDS
Long Title	Introduction to Data Science APPROVED
Title	Introduction to Data Science
Module Level:	LEVEL 6
EQF Level:	5
EHEA Level:	Short Cycle
Credits:	10
Module Coordinator:	Arghir Moldovan
Module Author:	Arghir Moldovan
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	Master's degree in computing or cognate discipline.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
#	Learning Outcome Description
LO1	Explain pertinent applications in data science
LO2	Discuss key data science methodologies
LO3	Search for, identify and document relevant sources of data
LO4	Highlight and discuss the application of key technologies in data science
LO5	Identify potential issues with respect to privacy, ethics and data protection

Dependencies	
Module Recommendations	
No recommendations listed	
Co-requisite Modules	
No Co-requisite modules listed	
Entry requirements	See section 4.2 Entry procedures and criteria for the programme including procedures recognition of prior learning

H6IDS: Introduction to Data Science

Module Content & Assessment

Indicative Content

Defining Data Science
Overview of key data science applications:. Data Exploration. Pattern Recognition. Classification. Regression. Forecasting. Artificial Intelligence. Knowledge Representation

Data Science Methodologies
An introduction to KDD, CRISP-DM, their core use cases, similarities and differences. Examples of seminal works using KDD and/or CRISP-DM.

Types of Data
Categorical vs. Numerical Data . Observational vs. Longitudinal vs. Time Series . Representations of Data . Structured vs. Unstructured Data

Core Data Science Methods
Introduction to:. Numerical Summaries of Data . Statistics. Data Mining. Visualisation .

Data Protection Policies
Introduce learners to key aspects of data protection laws and frameworks, e.g.: GDPR, Right to be forgotten, etc.

Ethics in Data Science
General Introduction to ethics in data science:. Correlation vs. Causation. Informed Consent. Privacy. Data Anonymity. Availability of Data vs. Ethical Uses. (Un)Ethical Questions in Data Science. Hawthorn Effects and Observer Bias. Sampling Issues (E.g. González-Bailón et al. (2014))

Data Quality
Characteristics of high/low quality data. How data quality impacts data science, ethical data science, and decision making within data applications

IRB and Ethical Content
How to file an ethical review form for a data science study

Assessment Breakdown	%
Coursework	100.00%

Assessments

Full Time

Coursework

Assessment Type:	Continuous Assessment	**% of total:**	Non-Marked
Assessment Date:	n/a	**Outcome addressed:**	1,2,5
Non-Marked:	Yes		

Assessment Description:
Ongoing independent and group problem solving activities and feedback.

Assessment Type:	Continuous Assessment	**% of total:**	30
Assessment Date:	n/a	**Outcome addressed:**	1,3
Non-Marked:	No		

Assessment Description:
Search for Data: : Students will be asked to locate appropriate data sets for a fictitious study that met set criteria and requirements (e.g. size, quality, collection method(s), etc.). prepare a short report entailing how and why their discovered data sources are relevant and accessible for their given problem. Noting key details of the sources, application areas, and how they could contribute to a study design.

Assessment Type:	Project	**% of total:**	70
Assessment Date:	n/a	**Outcome addressed:**	2,3,5
Non-Marked:	No		

Assessment Description:
Students will document appropriate sources of their own personal data footprint (without providing the data), their structure (if online), and prepare an ethical review form for a fictitious data science study in accordance to either KDD, or CRISP-DM on their personal data footprint(s). Emphasis will be placed on how they will ensure adherence to relevant laws and legal frameworks of their study, and how participant risks are mitigated.

No End of Module Assessment

No Workplace Assessment

Reassessment Requirement

Coursework Only
This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.

Reassessment Description
Learners who fail this module will be required to sit a repeat module assessment where all learning outcomes will be examined.

H6IDS: Introduction to Data Science

Module Workload				
Module Target Workload Hours 0 Hours				
Workload: Full Time				
Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Classroom & Demonstrations (hours)	24	Per Semester	2.00
Tutorial	Other hours (Practical/Tutorial)	24	Per Semester	2.00
Independent Learning	Independent learning (hours)	202	Per Semester	16.83
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

Hasselbalch, G. & Tranberg, P.. (2016), Data Ethics: The New Competitive Advantage, PubliShare.

Nielsen, L. & Burlingame, N.. (2012), A Simple Introduction to Data Science, New Street Communications, LLC.

Nielsen, L. & Burlingame, N.. (2015), A Simple Introduction to Data Science: Book Two, New Street Communications, LLC.

Peng, R. & Matsui, E.. (2016), The Art of Data Science, LeanPub.

Supplementary Book Resources

Blum, A., Hopcroft, J. & Kannan, R.. (2017), Foundations of Data Science, cs, Retrieved from https://www.

MacIntyre, A.. (2003), Short History of Ethics: A History of Moral Philosophy from the Homeric Age to the 20th Century, Routledge.

O'Neil, C. & Schutt, R.. (2013), Doing Data Science Straight Talk from the Frontline, O'Reilly.

Satlz, J., & Stanton, J. (2017), An Introduction to Data Science, SAGE.

This module does not have any article/paper resources

Other Resources

Boyd, D. & Crawford, K. (2012).. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society.

[Journal], Crawford, K., Gray, M. L., & Miltner, K.. (2014), Big Data Critiquing Big Data: Politics, Ethics, Ep.

Ess, C. (2002), Ethical decision-making and Internet research: Recommendations from the aoir ethics working committee..

Hall, M., & Caton, S.. (2017), Am I who I say I am? Unobtrusive self-representation and personality recognition on Facebook..

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y.. (2014), Assessing the bias in samples of large online networks. Social Networks, 38, 16-27..

Kramer, A. D.. (2012), The spread of emotion via Facebook. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 767-770). ACM..

Lewis, K., Kaufman, J.Gonzalez, M., Wimmer, A., & Christakis, N.. (2008), Tastes, ties, and time: A new social network dataset using Facebook. com. Social networks, 30(4), 330-342..

Zimmer, M.,. (2010), "But the data is already public": on the ethics of research in Facebook. Ethics and information technology, 12(4), pp.313-325..

Zwitter, A.. (2014), Big Data ethics. Big Data & Society, Sage.

[Website], MIT Moral Machine,
http://moralmachine.mit.edu

Discussion Note: