

H9DAI: Data Analytics for Artificial Intelligence

Module Code:	H9DAI
Long Title	Data Analytics for Artificial Intelligence APPROVED
Title	Data Analytics for Artificial Intelligence
Module Level:	LEVEL 9
EQF Level:	7
EHEA Level:	Second Cycle
Credits:	10
Module Coordinator:	Rejwanul Haque
Module Author:	Shauni Hegarty
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	PhD/Master's degree in a computing or cognate discipline. May have industry experience also.
Learning Outcomes	
<i>On successful completion of this module the learner will be able to:</i>	
#	Learning Outcome Description
LO1	Retrieve, extract, manipulate, synthesise, explore, and visualise data in preparation for data analysis and machine learning
LO2	Demonstrate expert knowledge of the theory, concepts and methods associated with the analysis of data using numerical and statistical techniques to assist on decision-making.
LO3	Use fundamental machine learning concepts and techniques to build and evaluate machine learning models on various problem domains.
LO4	Evaluate and employ graphical tools for building comprehensive analytics processes and dashboards.
LO5	Critically analyse, compare, summarise, and present results to support decision making and address requirements in real-world problems.
Dependencies	
Module Recommendations	
No recommendations listed	
Co-requisite Modules	
No Co-requisite modules listed	
Entry requirements	Applicants are required to hold a minimum of a Level 8 honours qualification (2.2 or higher) or equivalent on the National Qualifications Framework in either STEM (e.g., Information Management Systems, Information Technologies, Computer Science, Computer Engineer) or Business (e.g., Business Information Systems, Business Administration, Economics) discipline and a minimum of three years of relevant work experience in industry, ideally but not necessarily, in management. Previous numerical and computer proficiencies should be part of their work experience or formal training. Graduates from disciplines which do not have technical or mathematical problem-solving skills embedded in their programme will need to be able to demonstrate technical or mathematical problem-solving skills in addition to their level 8 programme qualifications (Certifications, Additional Qualifications, Certified Experience and Assessment Tests). All applicants for the programme must provide evidence that they have prior Mathematics and Computing module experience (e.g., via academic transcripts or recognised certification) as demonstrated in one mathematics/statistics module and one computing module or statement of purpose must specify numerical and computing work experience. NCI also operates a prior experiential learning policy where graduates with lower, or no formal qualifications, currently working in a relevant field, may be considered for the programme. Applicants must also be able to have their own laptop with the minimum required specification that will be communicated to each applicant through both the admissions and marketing departments.

H9DAI: Data Analytics for Artificial Intelligence

Module Content & Assessment			
Indicative Content			
Introduction to data analytics, nature of data Introduction to data analytics, nature of data, data analysis process/spectrum (descriptive, diagnostic, predictive, prescriptive). Measures of central tendency (mode, median, mean) Measures of dispersion (range, variance, standard deviation) Data mining methodologies (e.g., CRISP-DM, KDD)			
Data Collection and Data Manipulation Sources of data, data repositories, gather and Import data. Learn different file formats, relational and no relational databases, APIs, web scraping Selecting columns, rows, grouping, aggregation, filtering, joining datasets, remove duplicates, string manipulation, regular expressions, data cleaning			
Data Preprocessing and Transformation Strategies Discretisation and binning, feature normalisation, filtering outliers, handling missing values, handling class imbalance, handling categorical data, scaling, feature selection techniques			
Data Presentation (Visualisations) -Reporting Communicating and sharing Data Analysis Findings. Understand trends, outliers, and patterns in data through appropriate visualisations such as scatter plots, histograms, boxplots, pie charts, bar charts, overlaid bar charts, clustered bar charts, line charts, etc.			
Statistical Analysis -Hypothesis & Inference Statistical analysis, different kinds of hypothesis tests, Standard Errors Hypothesis Testing, Parametric Tests (e.g., T-Test, ANOVA, regression), Non-parametric Tests (e.g., chi-square tests) Correlation, Z-statistic, Distributions, Sample size, Confidence intervals, significance levels, p-values, effect size			
Dimension Reduction methods Need for dimension reduction, Principal Component Analysis, Singular Value Decomposition, Eigenvalues Criterion, Factor analysis, Backward Feature Elimination, Cross correlation			
Prediction (Regression) Simple Linear regression, p-value, F-statistics, residual standard error, Multiple Linear Regression, Logistic Regression, Forecasting			
Classification Binary Classification, Multi Class Classification, Multi-Label Classification, k-Nearest Neighbour, choosing k, Decision Trees, Random Forests, SVM, Logistic Regression			
Clustering What is clustering, distances (e.g., Euclidean, Manhattan, Minkowski). Normalising distances Hierarchical clustering methods, K-Means, K-means++, distortion cost function, choosing value of k, Density Based Clustering (DBScan)			
Modelling, Evaluation Splitting a dataset, training, testing and validation, cross validation. Resampling methods. Confusion matrix, Accuracy, Precision, Recall, F1 score, Roc curve. Sample size. Sampling methods (e.g., random, cluster)			
Time series Analysis Smoothing data, Analysing time series, curve fitting, seasonality. Moving averages, ARIMA (Seasonal, Non-seasonal)			
Content analysis Document classification, entity extraction, tokenizing, Filtering of Tokens, topic modelling, language modelling, Term Frequencies, Document Frequencies. Bayesian classification Handling unstructured Data, Stemming, syntax and semantics, word-embedding vectors;			
Assessment Breakdown			%
Coursework			100.00%
Assessments			
Full Time			
Coursework			
Assessment Type:	Formative Assessment	% of total:	Non-Marked
Assessment Date:	n/a	Outcome addressed:	1,2,3,4,5
Non-Marked:	Yes		
Assessment Description: Formative assessment will be provided on the in-class individual or group activities. Feedback will be provided in written or oral format, or on-line through Moodle. In addition, in class discussions will be undertaken as part of the practical approach to learning.			
Assessment Type:	Continuous Assessment	% of total:	30
Assessment Date:	n/a	Outcome addressed:	1,2,4
Non-Marked:	No		
Assessment Description: Assessment will be through an in-class, open book test, that will require learners to retrieve, extract, manipulate and present data. Learners will be also asked to make statistical inferences and draw conclusions about a population.			
Assessment Type:	Project	% of total:	70
Assessment Date:	n/a	Outcome addressed:	1,2,3,4,5
Non-Marked:	No		
Assessment Description: The terminal assessment will consist of a project that will evaluate all learning outcomes. Learners will have to identify and carry out a series of analytic tasks upon a large dataset (or a collection of datasets that are somehow related or complement each other), utilising appropriate tools and techniques for data extraction, processing, analysis and critical evaluation. The final submission will consist of an academic research paper style report as well as the implemented data analytics artefact. It is also expected students to present and communicate the results/insights of their study.			
No End of Module Assessment			
No Workplace Assessment			
Reassessment Requirement			
Repeat examination Reassessment of this module will consist of a repeat examination. It is possible that there will also be a requirement to be reassessed in a coursework element.			

H9DAI: Data Analytics for Artificial Intelligence

Module Workload				
Module Target Workload Hours 0 Hours				
Workload: Full Time				
Workload Type	Workload Description	Hours	Frequency	Average Weekly Learner Workload
Lecture	Lectures	24	Per Semester	2.00
Independent Learning	Independent Learning	202	Per Semester	16.83
Tutorial	Tutorials/Practicals	24	Per Semester	2.00
Total Weekly Contact Hours				4.00

Module Resources	
<i>Recommended Book Resources</i>	
McClave, J. T. & Sincich, T. (2017). Statistics(13th ed.). Pearson. [ISBN: 978-0134080215].. Bruce, P., Bruce, A., & Gedeck, P.(2020).Practical Statistics for Data Scientists(2nd ed.).O'Reilly Media.[ISBN: 978-1492072942].. Han, J., Pei, J., & Kamber, M.(2012).Data Mining: Concepts and Techniques(3rd ed.).Morgan Kaufmann. [ISBN: 978-0123814791].. Alpaydin, E. (2020). Introduction to Machine Learning. The MIT Press.[ISBN: 978-0262043793]..	
<i>Supplementary Book Resources</i>	
Shalev-Shwartz, S. & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms.Cambridge University Press.[ISBN: 978-1107057135].. Runkler, T.A. (2012).Data Analytics: Models and Algorithms for Intelligent Data Analysis.Springer.[ISBN: 978-3834825889].. Davies, A. (2017). Understanding Statistics: An Introduction.Cato Institute. [ISBN: 978-1944424350].. Kranzler, J.H. (2017). Statistics for the Terrified(6th ed.).Rowman & Littlefield Publishers. [ISBN: 978-1538100288].. Kelleher, J. D., MacNamee, B.,& D'Arcy, A. (2020). Fundamentals of Machine Learning for Predictive Data Analytics(2nd ed.). The MIT Press.[ISBN: 978-0262044691].. Marz, N. & Warren, J.(2015).Big Data: Principles and Best Practices of Scalable Realtime Data Systems. Manning Publications [ISBN: 978-1617290343]. Hofmann, M. & Klinkenberg, R. (2013).RapidMiner: Data Mining Use Cases and Business Analytics Applications.CRC Press.[ISBN: 978-1482205497]..	
<i>This module does not have any article/paper resources</i>	
<i>Other Resources</i>	
DataCamp, Learn R, Python & Data Science Online (https://www.datacamp.com/). Machine Learning Stanford(https://www.coursera.org/course/ml). UCI Repository(http://www.ics.uci.edu/~mllearn/MLRepository.html). DataCamp (www.datacamp.com). RapidMiner(https://rapidminer.com/). Azure Machine Learning(https://azure.microsoft.com/en-in/services/machine-learning/). KaggleCompetitions (https://www.kaggle.com/competitions). MySQL Tutorial(https://www.mysqltutorial.org). mongoDB Tutorial (https://www.mongodb.com/nosql-explained). JSON (https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Objects/JSON).	
Discussion Note:	