H9DAB: Data Analytics for Business

Module Code:	H9DAB
Long Title	Data Analytics for Business APPROVED
Title	Data Analytics for Business
Module Level:	LEVEL 9
EQF Level:	7
EHEA Level:	Second Cycle
Credits:	5
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

On successful completion of this module the learner will be able to:

#	Learning Outcome Description
LO1	Demonstrate critical understanding of data analytics and machine learning concepts and methods.
LO2	Employ data analytics tools to manipulate, synthesise, explore, and visualise data.
LO3	Select and apply machine learning concepts and methods to assist on business decision-making.
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 business 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.

H9DAB: Data Analytics for Business

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)
Getting data from different Data sources Sources of data, data repositories, gather and Import data. Learn different file formats, relational and no relational databases, APIs, web scraping
Data Manipulation – working with data Selecting columns, rows, grouping, aggregation, filtering, transformation joining datasets, remove duplicates, string manipulation, regular expressions, data cleaning, discretisation and binning, feature normalisation, filtering outliers, handling missing values
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, overlayed bar charts, clustered bar charts, line charts, etc.
Statistical Analysis - Hypothesis & Inference Statistical analysis, high-level overview on different kinds of hypothesis tests, Standard Errors Hypothesis Testing, Difference between parametric (e.g., T-Test, ANOVA, regression) and non-parametric Tests (e.g., chi-square tests) Overview on 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) Prediction overview, Simple Linear regression, residual standard error, Multiple Linear Regression
Classification Binary Classification, Multi-Class Classification, Multi-Label Classification, Overview on classification algorithms: k-Nearest Neighbour and Decision Trees
Clustering What is clustering, distances (e.g., Euclidean, Manhattan, Minkowski). Normalising distances k-Means clustering, Choosing value of K, KMeans++ clustering
Modelling, Evaluation Splitting a dataset, training, testing and validation, cross validation. Confusionmatrix, 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, High-level overview on ARIMA (Seasonal, Non-seasonal)
Content analysis Document classification, entity extraction, tokenizing, TF-IDF scores, Topic Modelling, Bayesian classification Handling Unstructured Data Tokenizing, Stemming, Filtering of Tokens

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 online 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

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

H9DAB: Data Analytics for Business

Module Workload
Module Target Workload Hours 0 Hours

Module Resources

Recommended Book Resources

James T. McClave,Terry Sincich. (2016), Statistics, Pearson, p.896, [ISBN: 978-0134080215].

Peter Bruce,Andrew Bruce,Peter Gedeck. (2020), Practical Statistics for Data Scientists, O'Reilly Media, p.350, [ISBN: 978-1492072942].

Han, J., Pei, J., & Kamber, M.. (2012), Data Mining: Concepts and Techniques, [ISBN: 978-0123814791].

Ethem Alpaydin. (2020), Introduction to Machine Learning, fourth edition, MIT Press, p.712, [ISBN: 978-0262043793].

Supplementary Book Resources

Shai Shalev-Shwartz,Shai Ben-David. (2014), Understanding Machine Learning, Cambridge University Press, p.409, [ISBN: 978-1107057135].

Thomas A. Runkler. (2012), Data Analytics, Springer Science & Business Media, p.137, [ISBN: 978-3834825889].

Antony Davies. Understanding Statistics, [ISBN: 978-1944424350].

John H. Kranzler. (2017), Statistics for the Terrified, Rowman & Littlefield Publishers, p.224, [ISBN: 978-1538100288].

John D. Kelleher,Brian Mac Namee,Aoife D'Arcy. (2020), Fundamentals of Machine Learning for Predictive Data Analytics, second edition, MIT Press, p.856, [ISBN: 978-0262044691].

Nathan Marz,James Warren. (2015), Big Data, Manning Publications Company, p.328, [ISBN: 978-1617290343].

Markus Hofmann,Ralf Klinkenberg. (2013), RapidMiner, CRC Press, p.525, [ISBN: 978-1482205497].

This module does not have any article/paper resources

Other Resources

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/~mlearn/MLReposit ory.html

RapidMiner,
https://rapidminer.com/

Azure Machine Learning,
https://azure.microsoft.com/en-in/servic es/machine-learning/

Kaggle Competitions,
http://Kaggle Competitions

MySQL Tutorial,
https://www.mysqltutorial.org

mongoDB Tutorial,
https://www.mongodb.com/nosql-explained

JSON,
http://developer.mozilla.org/en-US/docs/ Learn/JavaScript/Objects/JSON

Discussion Note:	