H8DMVP: Data Mining and Visualisation Principles

Module Code:	H8DMVP
Long Title	Data Mining and Visualisation Principles APPROVED
Title	Data Mining and Visualisation Principles
Module Level:	LEVEL 8
EQF Level:	6
EHEA Level:	First Cycle
Credits:	10
Module Coordinator:	
Module Author:	Alex Courtney
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	MSc and/or PhD degree in computer science 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	Apply fundamental techniques in both descriptive and inferential statistics for real world problems
LO2	Propose and apply fundamental data mining methodologies such as KDD to IoT data sets
LO3	Evaluate the application of data mining methods to IoT data
LO4	Assemble representative visualisations of IoT data to derive and identify contextual understanding
LO5	Generalise and interpret IoT data through the application and evaluation of data mining and visualisation techniques

Dependencies

Module Recommendations

No recommendations listed

Co-requisite Modules

No Co-requisite modules listed

Entry requirements	Learners should have attained the knowledge, skills and competence gained from stage 3 of the BSc (Hons) in Computer Science.

H8DMVP: Data Mining and Visualisation Principles

Module Content & Assessment

Indicative Content
Descriptive Statistics Arrangement, pre-processing and representation of data. Measures of central tendency (mode, median, mean). Measures of dispersion (range, variance, standard deviation). Statistical graphics & visuals (e.g., box-plot, histograms). Ethics in statistics
Inferential Statistics Hypothesis Testing. Test for Normality. Sample Tests
Introduction to Data Mining Data mining methodologies: KDD, CRISP-DM. Data security and ethical implications of data mining. Supervised vs Unsupervised Learning. Regression vs Classification Problems. Introduction to data mining tools such as Python SciKit-Learn, R/RStudio, Weka, RapidMiner
Data Handling and Transformation Attribute selection and discretization. Sampling methods. Data cleaning. Understanding, Detecting and Handling (massive) class imbalance
Regression What is regression?. Simple Linear Regression. Multiple Linear Regression. Evaluating Regression Models
Classification What is classification?. Evaluating classification models (confusion matrix). Logistic Regression. K-Nearest Neighbours. Naïve Bayes
Visualisation Principles What is Data Visualisation?. Fundamentals of Visualisation (e.g. Weber's Law, Steven's Power Law, Gestalt Principles, Tufte's Principles of Information Design). Characteristics of Data, Data Types and Information. Communication through visualisation
Visualisation Design Principles of data visualization. Graphical integrity. Clarity of data representation. Elements of visual design (layout, colour, fonts, labelling etc.)
Data Visualisations (I) Vector fields and flow data. Time-varying data
Data Visualisations (II) High-dimensional data: dimension reduction, parallel coordinates. Non-spatial data: multi-variate, tree/graph structured, text
Evaluation of Visualisation Methods Small and large data sets. Suitable visualisation design. Data and application characteristics
Unsupervised and Association Rule Learning Clustering Methods: k-means, k-medoids, hierarchical. Clustering for outlier detection. Plotting and understanding clusters. Frequent Pattern Mining

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.

Assessment Type:	Project	**% of total:**	80
Assessment Date:	n/a	**Outcome addressed:**	2,3,4,5
Non-Marked:	No		

Assessment Description:
Learners should choose and acquire data sets related to the IoT domain, develop, and document a process for preparing and analysing the data through to implementing a number of data visualizations. They should then analyse the results and provide a comparative evaluation of the different data mining and visualisation methods leveraged in the project. Learners will also present the results of their project in a non-technical context, focusing on the code distillation of their applied methodology, core results and takeaways from the project.

End of Module Assessment

Assessment Type:	Terminal Exam	**% of total:**	20
Assessment Date:	End-of-Semester	**Outcome addressed:**	1
Non-Marked:	No		

Assessment Description:
Learners are presented with a series of IoT data sets and/or hypothetical data sets, to which they will apply descriptive statistics as well as three statistical tests. They will then prepare a brief report on their findings.

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.

Reassessment Description
Coursework Only This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination. The repeat strategy will assess all the learning outcomes. Learning EnvironmentLearning will take place in a classroom/lab environment with access IT resources. Learners will have access to library resources, both physical and electronic and to faculty outside of the classroom where required. Module materials will be placed on Moodle, the College's virtual learning environment

H8DMVP: Data Mining and Visualisation Principles

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	Every Week	24.00
Tutorial	Other hours (Practical/Tutorial)	24	Every Week	24.00
Independent Learning	Independent learning (hours)	202	Every Week	202.00
			Total Weekly Contact Hours	48.00

Module Resources

Recommended Book Resources

Andy Kirk. (2019), Data Visualisation: A Handbook for Data Driven Design, [ISBN: 978-1526468925].

Gareth James,Daniela Witten,Trevor Hastie,Robert Tibshirani. (2014), An Introduction to Statistical Learning, Springer, p.426, [ISBN: 9781461471370].

Andy Field. (2018), Discovering Statistics Using IBM SPSS Statistics, SAGE Publications Limited, p.1104, [ISBN: 9781526419521].

This module does not have any article/paper resources

This module does not have any other resources

Discussion Note:	