Institiúid Teicneolaíochta Chorcaí
Cork Institute of Technology

Book of Modules

CR_SDAAN_9 - Master of Science in Data Science and Analytics

13 modules listed.

13:03 on Tuesday, 25 July 2017

APPROVED

Institiúid Teicneolaíochta Chorcaí
Cork Institute of Technology

Master of Science in Data Science and Analytics

Awards
MSc

Programme Code:	CR_SDAAN_9	**Mode of Delivery:**	Full Time, Part Time, ACCS	**No. of Semesters:**	3
NFQ Level:	9				
Embedded Award:	No	**Programme Credits:**	90		

Valid From:	Semester 1 - 2017/18 (September 2017)
Next Review Date:	June 2020
Department:	MATHEMATICS
Field of Study:	Mathematics
Programme Sponsor:	AINE NI SHE
Educational Aim of Programme:	The aim of this programme is to prepare graduates for analytical Data Scientist roles in the interdisciplinary field of Data Science and Analytics.

Programme Outcomes

Upon successful completion of this programme the graduate will be able to demonstrate... :

PO1	Knowledge - Breadth	
	(a)	Demonstrate comprehensive and systematic understanding of the facts, principles, theories and methods from the fields of Mathematics, Statistics, Computer Science and Business Intelligence which are relevant to the Data Analyst.
	(b)	Identify and articulate the key considerations of a "Big Data" problem; draw complex information together; critically comment on the technical, social, economic, environmental and political implications of own work and the work of others in Data Science, including an appreciation of the philosophical and ethical issues involved.
PO2	Knowledge - Kind	
	(a)	Demonstrate comprehensive knowledge and detailed understanding of: the theories, paradigms, defining concepts and underlying principles of the rapidly evolving Data Science and Analytics field; demonstrate knowledge and experience of advanced and new methods and technologies for acquiring, interpreting and analysing big data, with a critical awareness of the appropriate contexts for their use through the study of original papers, reports, journals, and data sets; demonstrate comprehensive knowledge and understanding of: the identification, definition and resolution of novel, complex research problems; relevant legal and regulatory frameworks; aspects of the defining elements and the inter-relationships of Data Science & Analytics as a result of in-depth study and research; demonstrate comprehensive knowledge and appreciation of the current limits of theoretical and applied knowledge in interdisciplinary field of Data Science and Analytics.
PO3	Skill - Range	
	(a)	Demonstrate mastery of standard and specialised research tools in Statistics, Mathematics, Computer Science and Business Intelligence; use these to proactively model, troubleshoot and solve original technical problems in the "Big Data" space; source relevant information, critically interpret and apply appropriate referenced literature from a wide range of information sources; maintain detailed records of activities; present and defend scientific research findings in a variety of forms to data scientists, "data savvy" practitioners, and non-specialists; formulate a hypothesis and design a relevant programme of investigation; work independently within defined time and resource boundaries; write accurately and in a manner consistent with scientific publications in Data Science or related disciplines.
PO4	Skill - Selectivity	
	(a)	Design, develop and test novel hypotheses; design experiments; select from a range of scientific skills, in particular those which draw from Mathematics/Statistics and Computer Science, so as to apply the most appropriate in a range of situations; think independently and make informed effective decisions; make decisions in the Data Science work setting; develop new skills either independently or with minimal mentoring.
PO5	Competence - Context	
	(a)	Apply advanced research skills and Big Data technologies; act autonomously and think independently; constructively criticise, draw conclusions and offer recommendations in a wide range of contexts, including unpredictable situations; formulate and communicate judgements, with incomplete or limited information.
PO6	Competence - Role	
	(a)	Act effectively, demonstrate initiative, lead and take responsibility in a complex interdisciplinary team environment with data scientists and qualified practitioners of other disciplines; develop and implement novel technical solutions for "Big Data" problems; reflect on own practices.
PO7	Competence - Learning to Learn	
	(a)	Learn to act in variable and unfamiliar learning contexts; identify knowledge gaps through effective self-evaluation; source and undertake self-learning as necessary for continued academic and professional self-development as a Data Scientist.
PO8	Competence - Insight	
	(a)	Identify and articulate the key considerations of a "Big Data" problem; draw complex information together; critically comment on the technical, social, economic, environmental and political implications of own work and the work of others in Data Science, including an appreciation of the philosophical and ethical issues involved.

Semester Schedules

Stage 1 / Semester 1

Mandatory								
Mod Code	**Module Title**	**Co-ordinator**	**Level**	**Credits**	**FT Contact Hours**	**PT Contact Hours**	**Course Work**	**Formal Exam**
DATA8001	Data Science and Analytics (Approved)	AINE NI SHE	Advanced	5.0	4.00	3.00	40.0	60.0
STAT8006	Applied Stats & Probability (Approved)	AINE NI SHE	Advanced	5.0	4.00	3.00	50.0	50.0
MATH8009	Maths Methods and Modelling (Approved)	AINE NI SHE	Advanced	5.0	4.00	4.00	50.0	50.0
COMP8042	Analytical and Scientific Prog (Approved)	TIM HORGAN	Advanced	5.0	4.00	4.00	100.0	0.0
DATA8002	Data Management Systems (Approved)	TIM HORGAN	Advanced	5.0	4.00	4.00	50.0	50.0
DATA8003	Unstructured Data & Visualis'n (Approved)	AINE NI SHE	Advanced	5.0	4.00	3.00	100.0	0.0

Stage 1 / Semester 2

Mandatory								
Mod Code	**Module Title**	**Co-ordinator**	**Level**	**Credits**	**FT Contact Hours**	**PT Contact Hours**	**Course Work**	**Formal Exam**
STAT9004	Statistical Data Analysis (Approved)	AINE NI SHE	Expert	5.0	4.00	3.00	50.0	50.0
STAT9005	Time Series & Factor Analysis (Approved)	AINE NI SHE	Expert	5.0	4.00	3.00	100.0	0.0
DATA9001	Data Visualisation & Analytics (Approved)	AINE NI SHE	Expert	5.0	4.00	4.00	100.0	0.0
COMP9060	Applied Machine Learning (Approved)	TIM HORGAN	Expert	5.0	4.00	4.00	100.0	0.0
DATA9002	Distributed Data Management (Approved)	TIM HORGAN	Expert	5.0	4.00	4.00	100.0	0.0
MATH9001	Research Methods (Approved)	AINE NI SHE	Expert	5.0	3.00	3.00	100.0	0.0

Stage 1 / Semester 3

Mandatory								
Mod Code	**Module Title**	**Co-ordinator**	**Level**	**Credits**	**FT Contact Hours**	**PT Contact Hours**	**Course Work**	**Formal Exam**
DATA9003	Research Project -Data Science (Approved)	AINE NI SHE	Expert	30.0	1.00	1.00	100.0	0.0

PO Delivery Using DETAILED Mappings

Programme Outcomes		PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
M/E	Supporting Modules								
M	**DATA8001: Data Science and Analytics**	4	4	4	4	4	2	2	1
M	**STAT8006: Applied Stats & Probability**	6	6	6	6	4	6	3	3
M	**MATH8009: Maths Methods and Modelling**	4	4	4	4	4	2	2	1
M	**COMP8042: Analytical and Scientific Prog**	4	4	4	4	4	2	2	1
M	**DATA8002: Data Management Systems**	5	5	5	5	4	2	1	2
M	**DATA8003: Unstructured Data & Visualis'n**	4	4	4	4	4	2	2	1
M	**STAT9004: Statistical Data Analysis**	6	6	6	6		2	2	2
M	**STAT9005: Time Series & Factor Analysis**	4	4	4	4	4	1	1	4
M	**DATA9001: Data Visualisation & Analytics**	5	5	5	5	3	2	3	2
M	**COMP9060: Applied Machine Learning**	6	6	6	6	4	1	2	4
M	**DATA9002: Distributed Data Management**	5	5	5	5	5	5	5	5
M	**MATH9001: Research Methods**	5	5	5	5	5	5	5	5
M	**DATA9003: Research Project -Data Science**	6	6	6	6	6	6	6	6

PO1.: Knowledge - Breadth
(a) Demonstrate comprehensive and systematic understanding of the facts, principles, theories and methods from the fields of Mathematics, Statistics, Computer Science and Business Intelligence which are relevant to the Data Analyst.
(b) Identify and articulate the key considerations of a “Big Data” problem; draw complex information together; critically comment on the technical, social, economic, environmental and political implications of own work and the work of others in Data Science, including an appreciation of the philosophical and ethical issues involved.

Supporting Modules	
DATA8001 - Data Science and Analytics	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT8006 - Applied Stats & Probability	• LO 1: Graphically display and numerically summarise data using methods of descriptive statistics. • LO 2: Apply the rules of probability and use probability models for data analysis. • LO 3: Compute and interpret point and interval estimates of population parameters. Determine required sample sizes. • LO 4: Describe the structure and compute statistical tests of hypothesis. • LO 5: Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction. • LO 6: Analyse statistical output from statistical packages such as IBM SPSS and R.
MATH8009 - Maths Methods and Modelling	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
COMP8042 - Analytical and Scientific Prog	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
DATA8002 - Data Management Systems	• LO 1: Explain the concepts of Database Management Systems and Data Models, such as Relational and NoSQL • LO 2: Implement and query relational databases using SQL Data Definition and Manipulation commands • LO 3: Evaluate the suitability of data models for a given data management requirement • LO 4: Devise solutions to NoSQL database queries using interactive commands • LO 5: Compare and contrast approaches to the distribution of data
DATA8003 - Unstructured Data & Visualis'n	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.

STAT9004 - Statistical Data Analysis	• LO 1: Explore data sets and establish a data analysis protocol for data science problems. • LO 2: Explain and apply the statistical concepts relevant to experimental design and data analysis with an emphasis on large data sets. • LO 3: Build and validate statistical models with continuous response variables and multiple predictors (both categorical and continuous) using ANOVA, multiple regression and ANCOVA. • LO 4: Distinguish between parametric and non-parametric methods and decide when the most commonly used non-parametric methods should be applied. • LO 5: Build and validate statistical models with categorical response variables using logistic regression. • LO 6: Interpret the results of statistical analyses performed by a software package or presented in research papers.
STAT9005 - Time Series & Factor Analysis	• LO 1: Implement factor analysis techniques on a large dataset and interpret the resulting models. • LO 2: Apply the theoretical principles that govern a time series. • LO 3: Apply regression and time series model for prediction. • LO 5: Use statistical packages to generate and analyse models.
DATA9001 - Data Visualisation & Analytics	• LO 1: Describe the concepts, principles and methods of data visualisation. • LO 2: Select and apply a variety of data explorative and pre-processing techniques to a range of data visualisation problems. • LO 3: Design and implement appropriate data visualisation techniques to solve data analytical problems. • LO 4: Interpret, critique and communicate patterns and knowledge discovered as a result of applying data visualisation techniques to a variety of data sets and analytical problems. • LO 5: Research and appraise a variety of data analytics solutions to current challenges in the area.
COMP9060 - Applied Machine Learning	• LO 1: Apply appropriate machine learning methodologies to facilitate pre-processing, dimensionality reduction and model selection. • LO 2: Select and apply appropriate machine learning algorithms to datasets from a specific application domain. • LO 3: Evaluate the accuracy of predictive models using standard methods. • LO 4: Develop and implement machine learning algorithms for building predictive models. • LO 5: Apply neural networks and deep learning methods for solving real-world problems. • LO 6: Implement and apply optimization algorithms for solving complex problems with a high dimensional search space.
DATA9002 - Distributed Data Management	• LO 1: Appraise the challenges posed by big data and the new infrastructure, data models and processing techniques it demands. • LO 2: Compare and contrast the main NoSQL-based data models, discriminating the best fit for different use-cases. • LO 3: Combine document-oriented and graph-based data models for a fit for purpose multi-component system. • LO 4: Demonstrate the scalability, flexibility and reliability of a distributed data cluster supporting large data sets. • LO 5: Compare and contrast the MapReduce and Spark large-scale analytics libraries in terms of their expressiveness and efficiency.
MATH9001 - Research Methods	• LO 1: Propose a research question, develop the research methodology and project a plan for the research project. • LO 2: Undertake preliminary experimental/design/analytical/modelling work as appropriate. • LO 3: Evaluate critically a number of solutions to the identified problem. • LO 4: Communicate effectively the idea and contribution of the proposed research project. • LO 5: Select a potential peer-review conference/journal paper for the research work and identify how this work may contribute to furthering knowledge in the specific field.

DATA9003 - Research Project -Data Science	• LO 1: Undertake a data science research project in a specialised area.
	• LO 2: Conduct a literature review of the up-to-date methodologies and techniques appropriate to the specified area of research.
	• LO 3: Research and detail appropriate and effective objectives and final deliverables for a data science/analytics project. Conduct a feasibility study and plan for the project.
	• LO 4: Systematically review and adapt the employed data science methodologies during implementation in response to practical, real-world data considerations and constraints.
	• LO 5: Produce a final written thesis detailing the work undertaken, methodologies used, findings and recommendations of the research work.
	• LO 6: Present the project findings in person and via a poster using appropriate presentation and visual communication skills.

PO2.: Knowledge - Kind
(a) Demonstrate comprehensive knowledge and detailed understanding of: the theories, paradigms, defining concepts and underlying principles of the rapidly evolving Data Science and Analytics field; demonstrate knowledge and experience of advanced and new methods and technologies for acquiring, interpreting and analysing big data, with a critical awareness of the appropriate contexts for their use through the study of original papers, reports, journals, and data sets; demonstrate comprehensive knowledge and understanding of: the identification, definition and resolution of novel, complex research problems; relevant legal and regulatory frameworks; aspects of the defining elements and the inter-relationships of Data Science & Analytics as a result of in-depth study and research; demonstrate comprehensive knowledge and appreciation of the current limits of theoretical and applied knowledge in interdisciplinary field of Data Science and Analytics.

Supporting Modules	
DATA8001 - Data Science and Analytics	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT8006 - Applied Stats & Probability	• LO 1: Graphically display and numerically summarise data using methods of descriptive statistics. • LO 2: Apply the rules of probability and use probability models for data analysis. • LO 3: Compute and interpret point and interval estimates of population parameters. Determine required sample sizes. • LO 4: Describe the structure and compute statistical tests of hypothesis. • LO 5: Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction. • LO 6: Analyse statistical output from statistical packages such as IBM SPSS and R.
MATH8009 - Maths Methods and Modelling	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
COMP8042 - Analytical and Scientific Prog	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
DATA8002 - Data Management Systems	• LO 1: Explain the concepts of Database Management Systems and Data Models, such as Relational and NoSQL • LO 2: Implement and query relational databases using SQL Data Definition and Manipulation commands • LO 3: Evaluate the suitability of data models for a given data management requirement • LO 4: Devise solutions to NoSQL database queries using interactive commands • LO 5: Compare and contrast approaches to the distribution of data
DATA8003 - Unstructured Data & Visualis'n	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.

STAT9004 - Statistical Data Analysis	• LO 1: Explore data sets and establish a data analysis protocol for data science problems. • LO 2: Explain and apply the statistical concepts relevant to experimental design and data analysis with an emphasis on large data sets. • LO 3: Build and validate statistical models with continuous response variables and multiple predictors (both categorical and continuous) using ANOVA, multiple regression and ANCOVA. • LO 4: Distinguish between parametric and non-parametric methods and decide when the most commonly used non-parametric methods should be applied. • LO 5: Build and validate statistical models with categorical response variables using logistic regression. • LO 6: Interpret the results of statistical analyses performed by a software package or presented in research papers.
STAT9005 - Time Series & Factor Analysis	• LO 1: Implement factor analysis techniques on a large dataset and interpret the resulting models. • LO 2: Apply the theoretical principles that govern a time series. • LO 3: Apply regression and time series model for prediction. • LO 5: Use statistical packages to generate and analyse models.
DATA9001 - Data Visualisation & Analytics	• LO 1: Describe the concepts, principles and methods of data visualisation. • LO 2: Select and apply a variety of data explorative and pre-processing techniques to a range of data visualisation problems. • LO 3: Design and implement appropriate data visualisation techniques to solve data analytical problems. • LO 4: Interpret, critique and communicate patterns and knowledge discovered as a result of applying data visualisation techniques to a variety of data sets and analytical problems. • LO 5: Research and appraise a variety of data analytics solutions to current challenges in the area.
COMP9060 - Applied Machine Learning	• LO 1: Apply appropriate machine learning methodologies to facilitate pre-processing, dimensionality reduction and model selection. • LO 2: Select and apply appropriate machine learning algorithms to datasets from a specific application domain. • LO 3: Evaluate the accuracy of predictive models using standard methods. • LO 4: Develop and implement machine learning algorithms for building predictive models. • LO 5: Apply neural networks and deep learning methods for solving real-world problems. • LO 6: Implement and apply optimization algorithms for solving complex problems with a high dimensional search space.
DATA9002 - Distributed Data Management	• LO 1: Appraise the challenges posed by big data and the new infrastructure, data models and processing techniques it demands. • LO 2: Compare and contrast the main NoSQL-based data models, discriminating the best fit for different use-cases. • LO 3: Combine document-oriented and graph-based data models for a fit for purpose multi-component system. • LO 4: Demonstrate the scalability, flexibility and reliability of a distributed data cluster supporting large data sets. • LO 5: Compare and contrast the MapReduce and Spark large-scale analytics libraries in terms of their expressiveness and efficiency.
MATH9001 - Research Methods	• LO 1: Propose a research question, develop the research methodology and project a plan for the research project. • LO 2: Undertake preliminary experimental/design/analytical/modelling work as appropriate. • LO 3: Evaluate critically a number of solutions to the identified problem. • LO 4: Communicate effectively the idea and contribution of the proposed research project. • LO 5: Select a potential peer-review conference/journal paper for the research work and identify how this work may contribute to furthering knowledge in the specific field.

DATA9003 - Research Project -Data Science	• LO 1: Undertake a data science research project in a specialised area. • LO 2: Conduct a literature review of the up-to-date methodologies and techniques appropriate to the specified area of research. • LO 3: Research and detail appropriate and effective objectives and final deliverables for a data science/analytics project. Conduct a feasibility study and plan for the project. • LO 4: Systematically review and adapt the employed data science methodologies during implementation in response to practical, real-world data considerations and constraints. • LO 5: Produce a final written thesis detailing the work undertaken, methodologies used, findings and recommendations of the research work. • LO 6: Present the project findings in person and via a poster using appropriate presentation and visual communication skills.

PO3.: Skill - Range
(a) Demonstrate mastery of standard and specialised research tools in Statistics, Mathematics, Computer Science and Business Intelligence; use these to proactively model, troubleshoot and solve original technical problems in the “Big Data” space; source relevant information, critically interpret and apply appropriate referenced literature from a wide range of information sources; maintain detailed records of activities; present and defend scientific research findings in a variety of forms to data scientists, “data savvy” practitioners, and non-specialists; formulate a hypothesis and design a relevant programme of investigation; work independently within defined time and resource boundaries; write accurately and in a manner consistent with scientific publications in Data Science or related disciplines.

Supporting Modules	
DATA8001 - Data Science and Analytics	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT8006 - Applied Stats & Probability	• LO 1: Graphically display and numerically summarise data using methods of descriptive statistics. • LO 2: Apply the rules of probability and use probability models for data analysis. • LO 3: Compute and interpret point and interval estimates of population parameters. Determine required sample sizes. • LO 4: Describe the structure and compute statistical tests of hypothesis. • LO 5: Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction. • LO 6: Analyse statistical output from statistical packages such as IBM SPSS and R.
MATH8009 - Maths Methods and Modelling	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
COMP8042 - Analytical and Scientific Prog	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
DATA8002 - Data Management Systems	• LO 1: Explain the concepts of Database Management Systems and Data Models, such as Relational and NoSQL • LO 2: Implement and query relational databases using SQL Data Definition and Manipulation commands • LO 3: Evaluate the suitability of data models for a given data management requirement • LO 4: Devise solutions to NoSQL database queries using interactive commands • LO 5: Compare and contrast approaches to the distribution of data
DATA8003 - Unstructured Data & Visualis'n	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.

STAT9004 - Statistical Data Analysis	• LO 1: Explore data sets and establish a data analysis protocol for data science problems. • LO 2: Explain and apply the statistical concepts relevant to experimental design and data analysis with an emphasis on large data sets. • LO 3: Build and validate statistical models with continuous response variables and multiple predictors (both categorical and continuous) using ANOVA, multiple regression and ANCOVA. • LO 4: Distinguish between parametric and non-parametric methods and decide when the most commonly used non-parametric methods should be applied. • LO 5: Build and validate statistical models with categorical response variables using logistic regression. • LO 6: Interpret the results of statistical analyses performed by a software package or presented in research papers.
STAT9005 - Time Series & Factor Analysis	• LO 1: Implement factor analysis techniques on a large dataset and interpret the resulting models. • LO 2: Apply the theoretical principles that govern a time series. • LO 3: Apply regression and time series model for prediction. • LO 5: Use statistical packages to generate and analyse models.
DATA9001 - Data Visualisation & Analytics	• LO 1: Describe the concepts, principles and methods of data visualisation. • LO 2: Select and apply a variety of data explorative and pre-processing techniques to a range of data visualisation problems. • LO 3: Design and implement appropriate data visualisation techniques to solve data analytical problems. • LO 4: Interpret, critique and communicate patterns and knowledge discovered as a result of applying data visualisation techniques to a variety of data sets and analytical problems. • LO 5: Research and appraise a variety of data analytics solutions to current challenges in the area.
COMP9060 - Applied Machine Learning	• LO 1: Apply appropriate machine learning methodologies to facilitate pre-processing, dimensionality reduction and model selection. • LO 2: Select and apply appropriate machine learning algorithms to datasets from a specific application domain. • LO 3: Evaluate the accuracy of predictive models using standard methods. • LO 4: Develop and implement machine learning algorithms for building predictive models. • LO 5: Apply neural networks and deep learning methods for solving real-world problems. • LO 6: Implement and apply optimization algorithms for solving complex problems with a high dimensional search space.
DATA9002 - Distributed Data Management	• LO 1: Appraise the challenges posed by big data and the new infrastructure, data models and processing techniques it demands. • LO 2: Compare and contrast the main NoSQL-based data models, discriminating the best fit for different use-cases. • LO 3: Combine document-oriented and graph-based data models for a fit for purpose multi-component system. • LO 4: Demonstrate the scalability, flexibility and reliability of a distributed data cluster supporting large data sets. • LO 5: Compare and contrast the MapReduce and Spark large-scale analytics libraries in terms of their expressiveness and efficiency.
MATH9001 - Research Methods	• LO 1: Propose a research question, develop the research methodology and project a plan for the research project. • LO 2: Undertake preliminary experimental/design/analytical/modelling work as appropriate. • LO 3: Evaluate critically a number of solutions to the identified problem. • LO 4: Communicate effectively the idea and contribution of the proposed research project. • LO 5: Select a potential peer-review conference/journal paper for the research work and identify how this work may contribute to furthering knowledge in the specific field.

DATA9003 - Research Project -Data Science	• LO 1: Undertake a data science research project in a specialised area. • LO 2: Conduct a literature review of the up-to-date methodologies and techniques appropriate to the specified area of research. • LO 3: Research and detail appropriate and effective objectives and final deliverables for a data science/analytics project. Conduct a feasibility study and plan for the project. • LO 4: Systematically review and adapt the employed data science methodologies during implementation in response to practical, real-world data considerations and constraints. • LO 5: Produce a final written thesis detailing the work undertaken, methodologies used, findings and recommendations of the research work. • LO 6: Present the project findings in person and via a poster using appropriate presentation and visual communication skills.

PO4.: Skill - Selectivity
(a) Design, develop and test novel hypotheses; design experiments; select from a range of scientific skills, in particular those which draw from Mathematics/Statistics and Computer Science, so as to apply the most appropriate in a range of situations; think independently and make informed effective decisions; make decisions in the Data Science work setting; develop new skills either independently or with minimal mentoring.

Supporting Modules	
DATA8001 - Data Science and Analytics	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT8006 - Applied Stats & Probability	• LO 1: Graphically display and numerically summarise data using methods of descriptive statistics. • LO 2: Apply the rules of probability and use probability models for data analysis. • LO 3: Compute and interpret point and interval estimates of population parameters. Determine required sample sizes. • LO 4: Describe the structure and compute statistical tests of hypothesis. • LO 5: Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction. • LO 6: Analyse statistical output from statistical packages such as IBM SPSS and R.
MATH8009 - Maths Methods and Modelling	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
COMP8042 - Analytical and Scientific Prog	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
DATA8002 - Data Management Systems	• LO 1: Explain the concepts of Database Management Systems and Data Models, such as Relational and NoSQL • LO 2: Implement and query relational databases using SQL Data Definition and Manipulation commands • LO 3: Evaluate the suitability of data models for a given data management requirement • LO 4: Devise solutions to NoSQL database queries using interactive commands • LO 5: Compare and contrast approaches to the distribution of data
DATA8003 - Unstructured Data & Visualis'n	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.

STAT9004 - Statistical Data Analysis	• LO 1: Explore data sets and establish a data analysis protocol for data science problems. • LO 2: Explain and apply the statistical concepts relevant to experimental design and data analysis with an emphasis on large data sets. • LO 3: Build and validate statistical models with continuous response variables and multiple predictors (both categorical and continuous) using ANOVA, multiple regression and ANCOVA. • LO 4: Distinguish between parametric and non-parametric methods and decide when the most commonly used non-parametric methods should be applied. • LO 5: Build and validate statistical models with categorical response variables using logistic regression. • LO 6: Interpret the results of statistical analyses performed by a software package or presented in research papers.
STAT9005 - Time Series & Factor Analysis	• LO 1: Implement factor analysis techniques on a large dataset and interpret the resulting models. • LO 2: Apply the theoretical principles that govern a time series. • LO 3: Apply regression and time series model for prediction. • LO 5: Use statistical packages to generate and analyse models.
DATA9001 - Data Visualisation & Analytics	• LO 1: Describe the concepts, principles and methods of data visualisation. • LO 2: Select and apply a variety of data explorative and pre-processing techniques to a range of data visualisation problems. • LO 3: Design and implement appropriate data visualisation techniques to solve data analytical problems. • LO 4: Interpret, critique and communicate patterns and knowledge discovered as a result of applying data visualisation techniques to a variety of data sets and analytical problems. • LO 5: Research and appraise a variety of data analytics solutions to current challenges in the area.
COMP9060 - Applied Machine Learning	• LO 1: Apply appropriate machine learning methodologies to facilitate pre-processing, dimensionality reduction and model selection. • LO 2: Select and apply appropriate machine learning algorithms to datasets from a specific application domain. • LO 3: Evaluate the accuracy of predictive models using standard methods. • LO 4: Develop and implement machine learning algorithms for building predictive models. • LO 5: Apply neural networks and deep learning methods for solving real-world problems. • LO 6: Implement and apply optimization algorithms for solving complex problems with a high dimensional search space.
DATA9002 - Distributed Data Management	• LO 1: Appraise the challenges posed by big data and the new infrastructure, data models and processing techniques it demands. • LO 2: Compare and contrast the main NoSQL-based data models, discriminating the best fit for different use-cases. • LO 3: Combine document-oriented and graph-based data models for a fit for purpose multi-component system. • LO 4: Demonstrate the scalability, flexibility and reliability of a distributed data cluster supporting large data sets. • LO 5: Compare and contrast the MapReduce and Spark large-scale analytics libraries in terms of their expressiveness and efficiency.
MATH9001 - Research Methods	• LO 1: Propose a research question, develop the research methodology and project a plan for the research project. • LO 2: Undertake preliminary experimental/design/analytical/modelling work as appropriate. • LO 3: Evaluate critically a number of solutions to the identified problem. • LO 4: Communicate effectively the idea and contribution of the proposed research project. • LO 5: Select a potential peer-review conference/journal paper for the research work and identify how this work may contribute to furthering knowledge in the specific field.

DATA9003 - Research Project -Data Science	• LO 1: Undertake a data science research project in a specialised area. • LO 2: Conduct a literature review of the up-to-date methodologies and techniques appropriate to the specified area of research. • LO 3: Research and detail appropriate and effective objectives and final deliverables for a data science/analytics project. Conduct a feasibility study and plan for the project. • LO 4: Systematically review and adapt the employed data science methodologies during implementation in response to practical, real-world data considerations and constraints. • LO 5: Produce a final written thesis detailing the work undertaken, methodologies used, findings and recommendations of the research work. • LO 6: Present the project findings in person and via a poster using appropriate presentation and visual communication skills.

PO5.: Competence - Context
(a) Apply advanced research skills and Big Data technologies; act autonomously and think independently; constructively criticise, draw conclusions and offer recommendations in a wide range of contexts, including unpredictable situations; formulate and communicate judgements, with incomplete or limited information.

Supporting Modules	
DATA8001 - Data Science and Analytics	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT8006 - Applied Stats & Probability	• LO 1: Graphically display and numerically summarise data using methods of descriptive statistics. • LO 4: Describe the structure and compute statistical tests of hypothesis. • LO 5: Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction. • LO 6: Analyse statistical output from statistical packages such as IBM SPSS and R.
MATH8009 - Maths Methods and Modelling	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
COMP8042 - Analytical and Scientific Prog	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
DATA8002 - Data Management Systems	• LO 1: Explain the concepts of Database Management Systems and Data Models, such as Relational and NoSQL • LO 3: Evaluate the suitability of data models for a given data management requirement • LO 4: Devise solutions to NoSQL database queries using interactive commands • LO 5: Compare and contrast approaches to the distribution of data
DATA8003 - Unstructured Data & Visualis'n	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 2: Apply programming principles to develop and implement a program design from a specification to solve data driven problems. • LO 3: Apply programming techniques to transform, manage, mine and visualise data sets. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT9005 - Time Series & Factor Analysis	• LO 1: Implement factor analysis techniques on a large dataset and interpret the resulting models. • LO 2: Apply the theoretical principles that govern a time series. • LO 3: Apply regression and time series model for prediction. • LO 5: Use statistical packages to generate and analyse models.

DATA9001 - Data Visualisation & Analytics	• LO 1: Describe the concepts, principles and methods of data visualisation. • LO 3: Design and implement appropriate data visualisation techniques to solve data analytical problems. • LO 4: Interpret, critique and communicate patterns and knowledge discovered as a result of applying data visualisation techniques to a variety of data sets and analytical problems.
COMP9060 - Applied Machine Learning	• LO 2: Select and apply appropriate machine learning algorithms to datasets from a specific application domain. • LO 4: Develop and implement machine learning algorithms for building predictive models. • LO 5: Apply neural networks and deep learning methods for solving real-world problems. • LO 6: Implement and apply optimization algorithms for solving complex problems with a high dimensional search space.
DATA9002 - Distributed Data Management	• LO 1: Appraise the challenges posed by big data and the new infrastructure, data models and processing techniques it demands. • LO 2: Compare and contrast the main NoSQL-based data models, discriminating the best fit for different use-cases. • LO 3: Combine document-oriented and graph-based data models for a fit for purpose multi-component system. • LO 4: Demonstrate the scalability, flexibility and reliability of a distributed data cluster supporting large data sets. • LO 5: Compare and contrast the MapReduce and Spark large-scale analytics libraries in terms of their expressiveness and efficiency.
MATH9001 - Research Methods	• LO 1: Propose a research question, develop the research methodology and project a plan for the research project. • LO 2: Undertake preliminary experimental/design/analytical/modelling work as appropriate. • LO 3: Evaluate critically a number of solutions to the identified problem. • LO 4: Communicate effectively the idea and contribution of the proposed research project. • LO 5: Select a potential peer-review conference/journal paper for the research work and identify how this work may contribute to furthering knowledge in the specific field.
DATA9003 - Research Project -Data Science	• LO 1: Undertake a data science research project in a specialised area. • LO 2: Conduct a literature review of the up-to-date methodologies and techniques appropriate to the specified area of research. • LO 3: Research and detail appropriate and effective objectives and final deliverables for a data science/analytics project. Conduct a feasibility study and plan for the project. • LO 4: Systematically review and adapt the employed data science methodologies during implementation in response to practical, real-world data considerations and constraints. • LO 5: Produce a final written thesis detailing the work undertaken, methodologies used, findings and recommendations of the research work. • LO 6: Present the project findings in person and via a poster using appropriate presentation and visual communication skills.

PO6.: Competence - Role
(a) Act effectively, demonstrate initiative, lead and take responsibility in a complex interdisciplinary team environment with data scientists and qualified practitioners of other disciplines; develop and implement novel technical solutions for “Big Data” problems; reflect on own practices.

Supporting Modules	
DATA8001 - Data Science and Analytics	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT8006 - Applied Stats & Probability	• LO 1: Graphically display and numerically summarise data using methods of descriptive statistics. • LO 2: Apply the rules of probability and use probability models for data analysis. • LO 3: Compute and interpret point and interval estimates of population parameters. Determine required sample sizes. • LO 4: Describe the structure and compute statistical tests of hypothesis. • LO 5: Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction. • LO 6: Analyse statistical output from statistical packages such as IBM SPSS and R.
MATH8009 - Maths Methods and Modelling	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
COMP8042 - Analytical and Scientific Prog	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
DATA8002 - Data Management Systems	• LO 1: Explain the concepts of Database Management Systems and Data Models, such as Relational and NoSQL • LO 5: Compare and contrast approaches to the distribution of data
DATA8003 - Unstructured Data & Visualis'n	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT9004 - Statistical Data Analysis	• LO 1: Explore data sets and establish a data analysis protocol for data science problems. • LO 2: Explain and apply the statistical concepts relevant to experimental design and data analysis with an emphasis on large data sets.
STAT9005 - Time Series & Factor Analysis	• LO 5: Use statistical packages to generate and analyse models.
DATA9001 - Data Visualisation & Analytics	• LO 1: Describe the concepts, principles and methods of data visualisation. • LO 4: Interpret, critique and communicate patterns and knowledge discovered as a result of applying data visualisation techniques to a variety of data sets and analytical problems.
COMP9060 - Applied Machine Learning	• LO 2: Select and apply appropriate machine learning algorithms to datasets from a specific application domain.
DATA9002 - Distributed Data Management	• LO 1: Appraise the challenges posed by big data and the new infrastructure, data models and processing techniques it demands. • LO 2: Compare and contrast the main NoSQL-based data models, discriminating the best fit for different use-cases. • LO 3: Combine document-oriented and graph-based data models for a fit for purpose multi-component system. • LO 4: Demonstrate the scalability, flexibility and reliability of a distributed data cluster supporting large data sets. • LO 5: Compare and contrast the MapReduce and Spark large-scale analytics libraries in terms of their expressiveness and efficiency.

MATH9001 - Research Methods	• LO 1: Propose a research question, develop the research methodology and project a plan for the research project. • LO 2: Undertake preliminary experimental/design/analytical/modelling work as appropriate. • LO 3: Evaluate critically a number of solutions to the identified problem. • LO 4: Communicate effectively the idea and contribution of the proposed research project. • LO 5: Select a potential peer-review conference/journal paper for the research work and identify how this work may contribute to furthering knowledge in the specific field.
DATA9003 - Research Project -Data Science	• LO 1: Undertake a data science research project in a specialised area. • LO 2: Conduct a literature review of the up-to-date methodologies and techniques appropriate to the specified area of research. • LO 3: Research and detail appropriate and effective objectives and final deliverables for a data science/analytics project. Conduct a feasibility study and plan for the project. • LO 4: Systematically review and adapt the employed data science methodologies during implementation in response to practical, real-world data considerations and constraints. • LO 5: Produce a final written thesis detailing the work undertaken, methodologies used, findings and recommendations of the research work. • LO 6: Present the project findings in person and via a poster using appropriate presentation and visual communication skills.

PO7.: Competence - Learning to Learn
(a) Learn to act in variable and unfamiliar learning contexts; identify knowledge gaps through effective self-evaluation; source and undertake self-learning as necessary for continued academic and professional self-development as a Data Scientist.

Supporting Modules	
DATA8001 - Data Science and Analytics	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT8006 - Applied Stats & Probability	• LO 1: Graphically display and numerically summarise data using methods of descriptive statistics. • LO 5: Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction. • LO 6: Analyse statistical output from statistical packages such as IBM SPSS and R.
MATH8009 - Maths Methods and Modelling	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
COMP8042 - Analytical and Scientific Prog	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
DATA8002 - Data Management Systems	• LO 5: Compare and contrast approaches to the distribution of data
DATA8003 - Unstructured Data & Visualis'n	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment. • LO 4: Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.
STAT9004 - Statistical Data Analysis	• LO 1: Explore data sets and establish a data analysis protocol for data science problems. • LO 6: Interpret the results of statistical analyses performed by a software package or presented in research papers.
STAT9005 - Time Series & Factor Analysis	• LO 3: Apply regression and time series model for prediction.
DATA9001 - Data Visualisation & Analytics	• LO 1: Describe the concepts, principles and methods of data visualisation. • LO 3: Design and implement appropriate data visualisation techniques to solve data analytical problems. • LO 4: Interpret, critique and communicate patterns and knowledge discovered as a result of applying data visualisation techniques to a variety of data sets and analytical problems.
COMP9060 - Applied Machine Learning	• LO 2: Select and apply appropriate machine learning algorithms to datasets from a specific application domain. • LO 4: Develop and implement machine learning algorithms for building predictive models.
DATA9002 - Distributed Data Management	• LO 1: Appraise the challenges posed by big data and the new infrastructure, data models and processing techniques it demands. • LO 2: Compare and contrast the main NoSQL-based data models, discriminating the best fit for different use-cases. • LO 3: Combine document-oriented and graph-based data models for a fit for purpose multi-component system. • LO 4: Demonstrate the scalability, flexibility and reliability of a distributed data cluster supporting large data sets. • LO 5: Compare and contrast the MapReduce and Spark large-scale analytics libraries in terms of their expressiveness and efficiency.

MATH9001 - Research Methods	• LO 1: Propose a research question, develop the research methodology and project a plan for the research project. • LO 2: Undertake preliminary experimental/design/analytical/modelling work as appropriate. • LO 3: Evaluate critically a number of solutions to the identified problem. • LO 4: Communicate effectively the idea and contribution of the proposed research project. • LO 5: Select a potential peer-review conference/journal paper for the research work and identify how this work may contribute to furthering knowledge in the specific field.
DATA9003 - Research Project -Data Science	• LO 1: Undertake a data science research project in a specialised area. • LO 2: Conduct a literature review of the up-to-date methodologies and techniques appropriate to the specified area of research. • LO 3: Research and detail appropriate and effective objectives and final deliverables for a data science/analytics project. Conduct a feasibility study and plan for the project. • LO 4: Systematically review and adapt the employed data science methodologies during implementation in response to practical, real-world data considerations and constraints. • LO 5: Produce a final written thesis detailing the work undertaken, methodologies used, findings and recommendations of the research work. • LO 6: Present the project findings in person and via a poster using appropriate presentation and visual communication skills.

PO8.: Competence - Insight
(a) Identify and articulate the key considerations of a “Big Data” problem; draw complex information together; critically comment on the technical, social, economic, environmental and political implications of own work and the work of others in Data Science, including an appreciation of the philosophical and ethical issues involved.

Supporting Modules	
DATA8001 - Data Science and Analytics	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment.
STAT8006 - Applied Stats & Probability	• LO 1: Graphically display and numerically summarise data using methods of descriptive statistics. • LO 5: Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction. • LO 6: Analyse statistical output from statistical packages such as IBM SPSS and R.
MATH8009 - Maths Methods and Modelling	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment.
COMP8042 - Analytical and Scientific Prog	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment.
DATA8002 - Data Management Systems	• LO 1: Explain the concepts of Database Management Systems and Data Models, such as Relational and NoSQL • LO 5: Compare and contrast approaches to the distribution of data
DATA8003 - Unstructured Data & Visualis'n	• LO 1: Appraise the role and applicability of programming languages within a data analysis environment.
STAT9004 - Statistical Data Analysis	• LO 1: Explore data sets and establish a data analysis protocol for data science problems. • LO 6: Interpret the results of statistical analyses performed by a software package or presented in research papers.
STAT9005 - Time Series & Factor Analysis	• LO 1: Implement factor analysis techniques on a large dataset and interpret the resulting models. • LO 2: Apply the theoretical principles that govern a time series. • LO 3: Apply regression and time series model for prediction. • LO 5: Use statistical packages to generate and analyse models.
DATA9001 - Data Visualisation & Analytics	• LO 1: Describe the concepts, principles and methods of data visualisation. • LO 4: Interpret, critique and communicate patterns and knowledge discovered as a result of applying data visualisation techniques to a variety of data sets and analytical problems.
COMP9060 - Applied Machine Learning	• LO 1: Apply appropriate machine learning methodologies to facilitate pre-processing, dimensionality reduction and model selection. • LO 2: Select and apply appropriate machine learning algorithms to datasets from a specific application domain. • LO 4: Develop and implement machine learning algorithms for building predictive models. • LO 6: Implement and apply optimization algorithms for solving complex problems with a high dimensional search space.
DATA9002 - Distributed Data Management	• LO 1: Appraise the challenges posed by big data and the new infrastructure, data models and processing techniques it demands. • LO 2: Compare and contrast the main NoSQL-based data models, discriminating the best fit for different use-cases. • LO 3: Combine document-oriented and graph-based data models for a fit for purpose multi-component system. • LO 4: Demonstrate the scalability, flexibility and reliability of a distributed data cluster supporting large data sets. • LO 5: Compare and contrast the MapReduce and Spark large-scale analytics libraries in terms of their expressiveness and efficiency.

MATH9001 - Research Methods	• LO 1: Propose a research question, develop the research methodology and project a plan for the research project. • LO 2: Undertake preliminary experimental/design/analytical/modelling work as appropriate. • LO 3: Evaluate critically a number of solutions to the identified problem. • LO 4: Communicate effectively the idea and contribution of the proposed research project. • LO 5: Select a potential peer-review conference/journal paper for the research work and identify how this work may contribute to furthering knowledge in the specific field.
DATA9003 - Research Project -Data Science	• LO 1: Undertake a data science research project in a specialised area. • LO 2: Conduct a literature review of the up-to-date methodologies and techniques appropriate to the specified area of research. • LO 3: Research and detail appropriate and effective objectives and final deliverables for a data science/analytics project. Conduct a feasibility study and plan for the project. • LO 4: Systematically review and adapt the employed data science methodologies during implementation in response to practical, real-world data considerations and constraints. • LO 5: Produce a final written thesis detailing the work undertaken, methodologies used, findings and recommendations of the research work. • LO 6: Present the project findings in person and via a poster using appropriate presentation and visual communication skills.

Title:	Data Science and Analytics APPROVED
Long Title:	Data Science and Analytics

Module Code:	DATA8001

Credits:	5

NFQ Level:	Advanced

Field of Study:	Data Format

Valid From:	Semester 1 - 2017/18 (September 2017)

Module Delivered In	**1 programme(s)**

Module Coordinator:	AINE NI SHE

Module Author:	Aengus Daly

Module Description:	This module will provide the learner with an overview of the important themes in the growing field of data science and analytics. The learner will study both the established methods and technologies used and also investigate new and emerging ones. Emphasis will be placed on the context and use of data analytics within organisations, within, for example, decision support systems, business performance management and knowledge management systems. Data analytics/mining software will be used, e.g. SAS, RapidMiner and R in both the lectures and labs.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Describe the field of data science and analytics, its concepts, technologies and its historical roots. Give a detailed overview of the main approaches to developing a data analytics/mining project lifecyle.
LO2	Perform exploratory data analysis using mining software packages.
LO3	Find patterns and solutions within a data set using data mining techniques.
LO4	Describe a number of data mining and business intelligence concepts and techniques.
LO5	Describe the area of big data, unstructured data and visualisation techniques and their relationship to data mining and analytics.

Pre-requisite learning

Module Recommendations
This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).

13373	DATA8001	Data Science and Analytics

Incompatible Modules
These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.

No incompatible modules listed

Co-requisite Modules

No Co-requisite modules listed

Requirements
This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.

No requirements listed

Co-Requisites

No Co Requisites listed

Module Content & Assessment

Indicative Content
Overview Data Science and Analytics landscape, terminology, technologies and historical development.
Data Analytics Project Life Cycle CRISP-DM etc., variety of actors, challenges. Investigate case studies in the field, looking at a variety of approaches, technologies with successes, failures, new developments and unusual applications of analytics.
Information and Business Systems for Data Analytics Where does the data come from? Information systems theory, and various systems architectures - relational databases, data warehouses, OLAP, NoSQL, main characteristics of cloud computing. How people and culture of an organisation impact on information systems and analytics.
Data Quality Cleaning/scrubbing data, data modelling, ETL (Extract, Transform, Load) systems and methods.
Data Mining Techniques and Software Technologies Introduction to various data mining techniques and methods. Use some data mining technologies e.g. SAS, RapidMiner and/or R. How to load data, and carry out initial data analysis and visualisation.
Technical Report Writing How to write a technical report - structure and narrative of documents, referencing, bibliography and awareness of expected audience.
Ethics, privacy and security Investigate ethics, privacy, security, data protection legistation and related topics in data governance.
Analytics in a Business Setting How data analytics is used within a business setting to monitor performance. Key performance indicators (KPIs), dashboards, balanced score cards, performance prism. How data analytics is incorporated into an organisation's strategy and vision.
Big Data Definitions, differences in structured/unstructured data. Parallel processing systems, NoSQL. Search and big data. Investigative visualisation technologies for data mining and anomaly detection.

Assessment Breakdown	%
Course Work	40.00%
End of Module Formal Examination	60.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Project	Solve a data analytics problem using a data mining software package and produce a report.	2,3,4	40.0	Week 9

End of Module Formal Examination

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Formal Exam	End-of-Semester Final Examination	1,4,5	60.0	End-of-Semester

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.*

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Theory and Case Studies	2.0	Every Week	2.00
Lab	Computer-based lab	1.0	Every Week	1.00
Tutorial	Theory/Practical	1.0	Every Week	1.00
Independent & Directed Learning (Non-contact)	Independent Study	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Theaory and case Studies	2.0	Every Week	2.00
Lab	Computer-based lab	2.0	Every Second Week	1.00
Independent Learning	Independent Study	4.0	Every Week	4.00
			Total Hours	8.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	3.00

Module Resources

Recommended Book Resources

- **Efraim Turban , Ramesh Sharda, Dursun Delen 2011, *Decision Support and Business Intelligence Systems*, 9th Ed., Pearson Prentice Hall New Jersey [ISBN: 013610729X]**
- **Andy Field, Jeremy Miles, *Discovering Statistics Using SAS*, 1st Ed. [ISBN: 1849200920]**

Supplementary Book Resources

- **Ramez Elmasri, Shamkant B. Navathe 2007, *Fundamentals of database systems*, 5th Ed., Pearson Addison Wesley Boston [ISBN: 0321369572]**

Recommended Article/Paper Resources

- **Watson, Hugh 2011, *Business Analytics Insight: Hype or Here to Stay?*, Business Intelligence Journal, vol. 16, No. 1, 1-8**

This module does not have any other resources

Module Delivered In

Programme Code	Programme	Semester	Delivery
CR_SDAAN_8	Higher Diploma in Science in Data Science & Analytics	1	Mandatory

Title:	Applied Stats & Probability APPROVED
Long Title:	Applied Statistics and Probability

Module Code:	STAT8006

Credits:	5

NFQ Level:	Advanced

Field of Study:	Statistics

Valid From:	Semester 1 - 2014/15 (September 2014)

Module Delivered In	1 programme(s)

Module Coordinator:	AINE NI SHE

Module Author:	Sean Lacey

Module Description:	This module will apply statistics and probability distributions to modern day problems. It will develop graphical visualisation methods, probability theory and distributions. The module will develop knowledge, skill and competence of sampling theory, hypothesis testing and linear regression.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Graphically display and numerically summarise data using methods of descriptive statistics.
LO2	Apply the rules of probability and use probability models for data analysis.
LO3	Compute and interpret point and interval estimates of population parameters. Determine required sample sizes.
LO4	Describe the structure and compute statistical tests of hypothesis.
LO5	Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction.
LO6	Analyse statistical output from statistical packages such as IBM SPSS and R.

Pre-requisite learning
Module Recommendations *This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).*
No recommendations listed
Incompatible Modules *These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.*
No incompatible modules listed
Co-requisite Modules
No Co-requisite modules listed
Requirements *This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.*
No requirements listed
Co-Requisites
No Co Requisites listed

Module Content & Assessment

Indicative Content
Data collection and presentation Collection and presentation of data. Basic descriptive statistics (both graphical and numerical).
Probability Relative frequency and axiomatic definitions. Laws of probability, conditional probability, independent and mutually exclusive events.
Probability distributions Random variables. Discrete and continuous distributions. Properties of probability density and cumulative density functions. Binomial, Poisson, normal and exponential distributions. Use of statistical tables.
Sampling Sampling distributions of proportions and means. Calculate the required sample size to obtain confidence intervals of required length for a single parameter. Confidence intervals and hypothesis tests for: one-sample mean and proportion; difference between two-sample means and proportions.
Linear regression Interpret scatterplots, correlation coefficients and the results from simple linear regression. Use the results of linear regression for prediction.
Statistical packages Analyse statistical output from statistical packages such as IBM SPSS and R.

Assessment Breakdown	%
Course Work	50.00%
End of Module Formal Examination	50.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Practical/Skills Evaluation	Descriptive statistics	1,2,6	25.0	Week 6
Practical/Skills Evaluation	Confidence intervals and hypothesis testing	3,4,5,6	25.0	Week 11

End of Module Formal Examination

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Formal Exam	End-of-Semester Final Examination	1,2,3,4,5,6	50.0	End-of-Semester

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.*

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Formal lecture	2.0	Every Week	2.00
Lab	Statistical package lab	2.0	Every Week	2.00
Independent & Directed Learning (Non-contact)	Worksheets	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Formal lecture	1.5	Every Week	1.50
Lab	Statistical package lab	1.5	Every Week	1.50
Independent & Directed Learning (Non-contact)	Worksheets	4.0	Every Week	4.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	3.00

Module Resources

Recommended Book Resources

- **Mario F. Triola,, *Elementary Statistics* [ISBN: 0321782682]**
- **Tadhg L. O'Shea 2013, *Essential Statistics fo Researchers* [ISBN: 0957505902]**
- **Perry R. Hinton 2004, *Statistics explained*, Routledge [ISBN: 9780415332859]**
- **Colin Gray, Paul R Kinnear, *IBM SPSS Statistics 19 Made Simple* [ISBN: 9781848720695]**
- **David S. Moore, George P. McCabe, Bruce Craig, *Introduction to the Practice of Statistics & CD-Rom* [ISBN: 978-1429286640]**

Supplementary Book Resources

- **Douglas C. Montgomery, George C. Runger, *Applied Statistics and Probability for Engineers* [ISBN: 978-0470053041]**
- **David S. Moore, *The basic practice of statistics* [ISBN: 978-1429224260]**
- **B. Burt Gerstman, *Basic Biostatistics Pkg w/ 2009 Formulas and Tables* [ISBN: 978-0763781347]**
- **Jim Fowler, Phil Jarvis, and Mel Chevannes, *Practical statistics for nursing and health care* [ISBN: 978-0471497165]**

This module does not have any article/paper resources

This module does not have any other resources

Module Delivered In

Programme Code	Programme	Semester	Delivery
CR_SDAAN_8	Higher Diploma in Science in Data Science & Analytics	1	Mandatory

Title:	Maths Methods and Modelling APPROVED
Long Title:	Maths Methods and Modelling
Module Code:	MATH8009
Credits:	5
NFQ Level:	Advanced
Field of Study:	Mathematics
Valid From:	Semester 1 - 2017/18 (September 2017)
Module Delivered In	1 programme(s)
Module Coordinator:	AINE NI SHE
Module Author:	Sean Lacey
Module Description:	This module will explore various mathematical techniques and will focus on mathematical models of real world processes, their formulation and methods of solution - both numerical and analytical. Central to the module will be practical problems that arise in industry and commerce.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Formulate well posed linear, exponential and statistical models, along with differential equations.
LO2	Carry out mathematical analysis on formulated models.
LO3	Select and develop numerical methods/algorithms to solve statistical models.
LO4	Write computer programs which yield sensible answers to linear, exponential and statistical models.
LO5	Develop programs to implement numerical algorithms to solve formulated models.

Pre-requisite learning
Module Recommendations *This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).*
No recommendations listed
Incompatible Modules *These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.*
No incompatible modules listed
Co-requisite Modules
No Co-requisite modules listed
Requirements *This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.*
No requirements listed
Co-Requisites
No Co Requisites listed

Module Content & Assessment

Indicative Content
Introduction to modelling Highlight the pattern in the modelling process. Examine linear and exponential functions – with models.
Statistical modelling Deriving and modelling situations using the normal, binomial and Poisson distributions functions.
Markov chains Demonstrate how effective Markov Chains are at modelling practical situations.
Differential equations Modelling with differential equations. Analysing specific models and trying to make “sense” of the reasoning behind the model while at the same time solving the equation.
Software Excel, VBA.

Assessment Breakdown	%
Course Work	50.00%
End of Module Formal Examination	50.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Practical/Skills Evaluation	Mathematical Analysis	1,2,4,5	25.0	Week 6
Practical/Skills Evaluation	Numerical Methods	1,3,4,5	25.0	Week 12

End of Module Formal Examination

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Formal Exam	End-of-Semester Final Examination	1,2,3	50.0	End-of-Semester

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.*

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Formal lecture	2.0	Every Week	2.00
Lab	Excel and VBA	2.0	Every Week	2.00
Independent & Directed Learning (Non-contact)	Review of lecture notes and solving problems from worksheets	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Formal lecture	1.5	Every Week	1.50
Lab	Excel and VBA	1.5	Every Week	1.50
Lecturer-Supervised Learning (Contact)	Solving problems from worksheets	1.0	Every Week	1.00
Independent & Directed Learning (Non-contact)	Review of lecture notes and solving problems from worksheets	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

- George F. Simmons 2016, ***Differential Equations with Applications and Historical Notes***, 3 Ed., Chapman and Hall/CRC [ISBN: 1498702597]
- Milton Abramowitz, Irene Stegun 2014, ***Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*** [ISBN: 978-161427617]
- Nicolas Privault 2013, ***Understanding Markov Chains: Examples and Applications***, Springer [ISBN: 9814451509]
- Daniel P. Maki, Maynard Thompson 2006, ***Mathematical modeling and computer simulation***, Thomson Brooks/Cole Belmont, CA [ISBN: 0534384781]

Supplementary Book Resources

- Gary N. Felder 2015, ***Mathematical methods for physics and engineering***, 1 Ed., John Wiley & Sons Cambridge [ISBN: 1118449606]
- Thomas Witelski 2015, ***Methods of Mathematical Modelling: Continuous Systems and Differential Equations***, 1 Ed. [ISBN: 978-331923041]
- Bennett, J and Briggs, W. 2014, ***Using and understanding mathematics: A quantitative reasoning approach***, 6 Ed., Pearson [ISBN: 1292062304]
- D. W. Jordan and P. Smith 2008, ***Mathematical techniques***, 4 Ed., OUP Oxford [ISBN: 0199282013]
- E. Joseph Billo 2007, ***Excel for scientists and engineers***, Wiley-Interscience Hoboken, N.J. [ISBN: 978-0471387343]
- Frank R. Giordano, Maurice Weir, 1997, ***First Course in Mathematical Modeling***, 2 Ed., Brooks/Cole [ISBN: 0534222482]
- J. Berry, K. Houston 1995, ***Mathematical modelling***, Edward Arnold London [ISBN: 978-0340614044]

This module does not have any article/paper resources

Other Resources

- Website: ***Wolfram Alpha***
  https://www.wolframalpha.com/
- Website: ***MathCentre***
  http://www.mathcentre.ac.uk

Module Delivered In

Programme Code	Programme	Semester	Delivery
CR_SDAAN_8	Higher Diploma in Science in Data Science & Analytics	1	Mandatory

Title:	Analytical and Scientific Prog APPROVED
Long Title:	Analytical and Scientific Programming

Module Code:	COMP8042
Credits:	5
NFQ Level:	Advanced
Field of Study:	Computer Science
Valid From:	Semester 2 - 2012/13 (February 2013)
Module Delivered In	1 programme(s)
Module Coordinator:	TIM HORGAN
Module Author:	AISLING O DRISCOLL
Module Description:	In this module the learner will use a programming language to manipulate, manage and process data using next generation technologies. More specifically, statistical and scientific libraries will be applied to analyse, mine and visualise complex data sets.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Appraise the role and applicability of programming languages within a data analysis environment.
LO2	Apply programming principles to develop and implement a program design from a specification to solve data driven problems.
LO3	Apply programming techniques to transform, manage, mine and visualise data sets.
LO4	Develop and document a program that uses open source scientific and statistical analysis libraries on a chosen data set.

Pre-requisite learning
Module Recommendations *This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).*
No recommendations listed
Incompatible Modules *These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.*
No incompatible modules listed
Co-requisite Modules
No Co-requisite modules listed
Requirements *This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.*
No requirements listed
Co-Requisites
No Co Requisites listed

Module Content & Assessment

Indicative Content
Programming Concepts: Categories of programming language, their typical application, programming in an analytical and scientific context, Language, Syntax, error checking and debugging, variables and basic data types, lists, dictionaries and sets. Processing data structures: Conditionals and Loops Efficient code structure: functions, modules, packages and files. Engineering code: objects, classes, and Object Oriented Programming (OOP).
Data Manipulation and Visualisation Overview of the standard programming library, basic mathematical libraries, Visualisation of Data (object oriented plotting in 2D and 3D, annotation, styles, legend, subplots), use of enhanced interactive shells for improved debugging, profiling code and interactive plotting.
Numerical Analysis and Scientific Computing: Use of open source numerical and scientific libraries for: Manipulation of arrays and matrix structures, indexing, slicing, broadcasting, sorting, searching and counting, basic linear algebra, basic Fourier transformation and random number generation routines. Statistical functions, integration, numerical optimization and interpolation tools, spatial computation, advanced mathematical routines.
Advanced Data Analysis Data gathering, analysis, processing, mining, plotting and storage using open source libraries, Efficient handling of large datasets.

Assessment Breakdown	%
Course Work	100.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Practical/Skills Evaluation	Laboratory practical/skills evaluations	1,2,3	50.0	Every Second Week
Project	Design, build, document, test and deploy a programme to analyse a given data set	2,3,4	50.0	Sem End

No End of Module Formal Examination

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

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Theory	1.0	Every Week	1.00
Lab	Laboratory Practical	3.0	Every Week	3.00
Independent Learning	Independent Student Learning	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Theory	1.0	Every Week	1.00
Lab	Laboratory practical	3.0	Every Week	3.00
Independent Learning	Independent Student Learning	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

- **Eli Bressert, *SciPy and NumPy: An Overview for Developers*, O' Reilly Media [ISBN: 978-1449305468]**
- **Wes McKinney, *Python for Data Analysis*, O' Reilly Media [ISBN: 978-1449319793]**
- **Hans Petter Langtangen,, *A Primer on Scientific Programming with Python* [ISBN: 9783642183652]**

Supplementary Book Resources

- **Philipp K. Janert, *A hands-on guide for programmers and data scientists Larger Cover Data Analysis with Open Source Tools*, O' Reilly [ISBN: ISBN 10: 0-596-80235-8]**

This module does not have any article/paper resources

Other Resources

- **Website: *Java Data Analytics Library***
  http://jwork.org/jhepwork/
- **Website: *Python Documentation***
  http://www.python.org/doc/
- **Website: *Python Data Analysis Library***
  http://pandas.pydata.org/
- **Website: *Python Scientific Library***
  http://www.scipy.org/
- **Website: Python Numerical Library*n/a***
  http://www.numpy.org/
- **Website: *Python Visualisation Library***
  http://matplotlib.org/
- **Website: *Python Interactive Computing***
  http://ipython.org/

Module Delivered In

Programme Code	Programme	Semester	Delivery
CR_SDAAN_8	Higher Diploma in Science in Data Science & Analytics	1	Mandatory

Title:	Data Management Systems APPROVED
Long Title:	Data Management Systems
Module Code:	DATA8002
Credits:	5
NFQ Level:	Advanced
Field of Study:	Data Format
Valid From:	Semester 1 - 2016/17 (September 2016)
Module Delivered In	no programmes
Module Coordinator:	TIM HORGAN
Module Author:	Larkin Cunningham
Module Description:	This module introduces students to the use of database management systems for applications. It includes an evaluation of the relational model and NoSQL data models, and how to query and manipulate data stored using these models. Students will learn how these data models are used in the distribution of data and the emerging "Big Data" paradigm.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Explain the concepts of Database Management Systems and Data Models, such as Relational and NoSQL
LO2	Implement and query relational databases using SQL Data Definition and Manipulation commands
LO3	Evaluate the suitability of data models for a given data management requirement
LO4	Devise solutions to NoSQL database queries using interactive commands
LO5	Compare and contrast approaches to the distribution of data

Pre-requisite learning
Module Recommendations *This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).*
No recommendations listed
Incompatible Modules *These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.*
No incompatible modules listed
Co-requisite Modules
No Co-requisite modules listed
Requirements *This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.*
No requirements listed
Co-Requisites
No Co Requisites listed

Module Content & Assessment

Indicative Content
Traditional Database Systems Concepts DBMS concepts: Data Integration and sharing, comparison with traditional data processing systems; DBMS architectures; Data Independence; The Relational Data Model.
Structured Query Language Manipulating data in SQL; Processing Single & Multiple Tables - SELECT commands. Functions & Group By; Database Definition in SQL - CREATE, DROP, ALTER, CHECK commands.
NoSQL Systems Motivation for NoSQL Data Models and Systems; Types of NoSQL systems / data models: MapReduce framework, Key-value stores, Document stores, Graph database systems. Creating and querying NoSQL Systems.
Distributed Databases Sharding; Master-Slave and Peer-to-Peer Replication; Distributed Filesystems; The Big Data Paradigm.

Assessment Breakdown	%
Course Work	50.00%
End of Module Formal Examination	50.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Practical/Skills Evaluation	SQL data definition and manipulation	1,2	25.0	Week 7
Practical/Skills Evaluation	Creating and manipulating data in a NoSQL system	1,4	25.0	Week 11

End of Module Formal Examination

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Formal Exam	End-of-Semester Final Examination	1,2,3,4,5	50.0	End-of-Semester

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.*

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Theory	2.0	Every Week	2.00
Lab	Lab	2.0	Every Week	2.00
Independent & Directed Learning (Non-contact)	No Description	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Theory	2.0	Every Week	2.00
Lab	Lab	2.0	Every Week	2.00
Independent & Directed Learning (Non-contact)	No Description	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

- **Thomas M. Connolly, Carolyn E. Begg 2009, *Database systems: A Practical Approach to Design, Implementation and Management*, 5th Ed. [ISBN: 978-0321523068]**
- **Pramod J. Sadalage, Martin Fowler 2012, *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence* [ISBN: 978-0321826626]**

Supplementary Book Resources

- **Eric Redmond, Jim Wilson 2012, *Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement* [ISBN: 978-1934356920]**

Supplementary Article/Paper Resources

- **Codd, E. F. 1970, *A Relational Model of Data for Large Shared Data Banks*, Communications of the ACM, 13:6, 377-387**

Other Resources

- **Website: *myNoSQL - NoSQL Databases and Polyglot Persistence: A Curated Guide***
  http://nosql.mypopescu.com/

Title:	Unstructured Data & Visualis'n APPROVED
Long Title:	Unstructured Data & Visualis'n

Module Code:	DATA8003
Credits:	5
NFQ Level:	Advanced
Field of Study:	Data Format
Valid From:	Semester 1 - 2017/18 (September 2017)
Module Delivered In	1 programme(s)
Module Coordinator:	AINE NI SHE
Module Author:	Aengus Daly
Module Description:	In data visualisation, data will be investigated using various visualisation and modelling techniques. More advanced visualisation concepts and tools for analysing multi dimensional data, large data sets and geospatial data will also be examined. In unstructured data analysis, the learner will examine how to organise and analyse both text based data forms and other unstructured data (e.g. web logs, web content, twitter). The learner will investigate the characteristics of unstructured data and how challenges in the area can be overcome using a variety of descriptive and analytical techniques.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Investigate data using various visualisation and modeling theories and techniques.
LO2	Use a variety of visualisation techniques and tools to solve data mining and analytics problems.
LO3	Make observations and build a body of evidence to support a case or project.
LO4	Organise and analyse non-numerical or unstructured data.
LO5	Examine relationships in data; and combine analysis with linking, searching and modelling.

Pre-requisite learning

Module Recommendations
This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).

13372	DATA8003	Unstructured Data & Visualis'n

Incompatible Modules
These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.

No incompatible modules listed

Co-requisite Modules

No Co-requisite modules listed

Requirements
This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.

No requirements listed

Co-Requisites

No Co Requisites listed

Module Content & Assessment

Indicative Content
Data visualisation concepts Understand the various categories used in the field e.g. Information/data/scientific visualisation, infographics, visual analytics.
Data Visualisation traditional statistical approaches Histograms, boxplots, scatter plots; Analysing correlations and patterns between variables. Univariate, bivariate and multivariate ways of presenting data.
Advanced visualisation techniques Investigate computer based tools for visualisation and their features - dashboards, drop-down menus, interactivity. Use software to display the data e.g. R Shiny, Qlikview, Tableau, Rapidminer.
Unstructured data Understand the characteristics of unstructured data and how this impacts on analysis.
Organise data Using a variety of software (e.g. R, Rapidminer, NVivo) to gather (e.g. web-scrape) and organise unstructured data and explore its characteristics. Investigate various data cleaning techniques so that the data can be analysed further.
Explore data Use statistical and data mining techniques to create models so that meaningful analyse can be performed on unstructured data, e.g.TF-IDF, bag of words, concepts; as well as K-means clustering, K-Nearest Neighbour, Bayesian inference.

Assessment Breakdown	%
Course Work	100.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Practical/Skills Evaluation	Lab examination in visualisation.	1,2,3	25.0	Week 4
Project	Analyse a data set using appropriate visualisation techniques.	1,2,3	25.0	Week 8
Project	Analyse an unstructured data set and write a report on the findings.	3,4,5	50.0	Week 13

No End of Module Formal Examination

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

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Formal lecture	2.0	Every Week	2.00
Lab	Laboratory sessions	2.0	Every Week	2.00
Independent & Directed Learning (Non-contact)	Review of lecture notes and preparing for labs	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Formal Lecture	2.0	Every Week	2.00
Lab	Laboratory sessions	2.0	Every Second Week	1.00
Independent & Directed Learning (Non-contact)	Review of lecture notes and preparing for labs	4.0	Every Week	4.00
			Total Hours	8.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	3.00

Module Resources

Recommended Book Resources

- **Julie Steele, Noah Iliinsky, 2011, *Designing Data Visualizations* [ISBN: 1449312284]**
- **Nathan Yau, 2011, *Visualize This* [ISBN: 0470944889]**
- **Ben Fry 2007, *Visualizing data*, O'Reilly Media Sebastopol, CA [ISBN: 0596514557]**

This module does not have any article/paper resources

Other Resources

- **Website: RStudio*R Shiny Tutorial***
  https://shiny.rstudio.com/tutorial/
- **Website: QlikView*Qlikview Tutorial***
  http://www.tutorialspoint.com/qlikview/
- **Website: *http://www.qsrinternational.com/***
- **Website: Brian Suda 2012, *The top 20 data visualisation tools***
  http://www.netmagazine.com/features/top- 20-data-visualisation-tools

Module Delivered In

Programme Code	Programme	Semester	Delivery
CR_SDAAN_8	Higher Diploma in Science in Data Science & Analytics	1	Mandatory

Title:	Statistical Data Analysis APPROVED
Long Title:	Statistical Data Analysis
Module Code:	STAT9004
Credits:	5
NFQ Level:	Expert
Field of Study:	Statistics
Valid From:	Semester 1 - 2017/18 (September 2017)
Module Delivered In	no programmes
Module Coordinator:	AINE NI SHE
Module Author:	Catherine Palmer
Module Description:	In this module, the learner will study statistical techniques, with particular emphasis on large data sets. Statistical analytical software such as R will be used in the labs.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Explore data sets and establish a data analysis protocol for data science problems.
LO2	Explain and apply the statistical concepts relevant to experimental design and data analysis with an emphasis on large data sets.
LO3	Build and validate statistical models with continuous response variables and multiple predictors (both categorical and continuous) using ANOVA, multiple regression and ANCOVA.
LO4	Distinguish between parametric and non-parametric methods and decide when the most commonly used non-parametric methods should be applied.
LO5	Build and validate statistical models with categorical response variables using logistic regression.
LO6	Interpret the results of statistical analyses performed by a software package or presented in research papers.

Pre-requisite learning		
Module Recommendations *This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).*		
10513	STAT8006	Applied Stats & Probability
Incompatible Modules *These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.*		
No incompatible modules listed		
Co-requisite Modules		
No Co-requisite modules listed		
Requirements *This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.*		
No requirements listed		
Co-Requisites		
No Co Requisites listed		

Module Content & Assessment

Indicative Content
Data Analysis Protocol Exploratory data analysis: graphical and numerical methods to explore categorical and continuous data sets, outlier detection, missing values, testing of assumptions and transformation of variables. Model fitting and model interpretation. Model diagnostics.
Design of Experiments Observational (vs) experimental data. The fundamentals of experimental design. Analysis of variance. Factorial design. Statistical power and multiple comparisons. Non-parametric alternatives.
Multiple Regression Assumptions, collinearity, interpreting coefficients, model fitting, model diagnostics, confidence intervals of coefficients, Analysis of covariance (ANCOVA).
Generalised Linear Models Definition of a generalized linear model: link functions. Overview of different types of generalised linear models and their uses with a focus on logistic regression for binary data.
Software analysis SPSS, R, Excel

Assessment Breakdown	%
Course Work	50.00%
End of Module Formal Examination	50.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Practical/Skills Evaluation	Solve and analyse problems in the laboratory setting.	1,2,3,6	25.0	Week 7
Project	Analyse data sets and report on the results.	1,3,5,6	25.0	Week 12

End of Module Formal Examination

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Formal Exam	End-of-Semester Written Examination	1,2,3,4,5,6	50.0	End-of-Semester

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.*

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Formal lectures describing the theory underpinning the statistical techniques covered by the learning outcomes.	2.0	Every Week	2.00
Lab	A series of laboratory exercises where the student will use a statistical software package to analyse data sets using the statistical techniques covered by the learning outcomes.	2.0	Every Week	2.00
Independent Learning	Independent learning	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Lecture	1.5	Every Week	1.50
Lab	Lab	1.5	Every Week	1.50
Lecturer Supervised Learning (Non-contact)	Lecturer Supervised Learning	4.0	Every Week	4.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	3.00

Module Resources

Recommended Book Resources

- **Michael J. Crawley 2012, *The R Book*, Wiley-Blackwell [ISBN: 978-0470973929]**
- **Peter Dalgaard 2008, *Introductory Statistics with R*, Springer New York [ISBN: 9780387790534]**

Supplementary Book Resources

- **Annette J. Dobson 2002, *An Introduction to Generalized Linear Models, Second Edition*, Chapman and Hall [ISBN: 978-1584881650]**
- **Andrew P. Beckerman, Dylan Childs and Owen Petchey 2017, *Getting Started With R*, Second Ed., Oxford University Press [ISBN: 978-019878784]**

Supplementary Article/Paper Resources

- **Ginsberg J, et al. 2009, *Detecting influenza epidemics using search engine query data*, Nature, 457**
- **Lazer D, Kennedy R, King G, et al. 2014, *The Parable of Google Flu: Traps in Big Data Analysis*, Science, 343**

This module does not have any other resources

Title:	Time Series & Factor Analysis APPROVED
Long Title:	Time Series & Factor Analysis

Module Code:	STAT9005

Credits:	5

NFQ Level:	Expert

Field of Study:	Statistics

Valid From:	Semester 1 - 2017/18 (September 2017)

Module Delivered In	no programmes

Module Coordinator:	AINE NI SHE

Module Author:	Sean Lacey

Module Description:	This module will provide the learner with the necessary tools to develop and critically evaluate structural equation modelling and time series models. In this module, data will be summarised using factor analysis, while the forecasting function of models is presented and evaluated, enabling the learner to create short and medium term forecasting models.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Implement factor analysis techniques on a large dataset and interpret the resulting models.
LO2	Apply the theoretical principles that govern a time series.
LO3	Apply regression and time series model for prediction.
LO4	Critically analyse and report on the paradigm under which forecasts are being made, along with their reliability. Perform residuals analysis and tests of fit.
LO5	Use statistical packages to generate and analyse models.

Pre-requisite learning
Module Recommendations *This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).*
No recommendations listed
Incompatible Modules *These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.*
No incompatible modules listed
Co-requisite Modules
No Co-requisite modules listed
Requirements *This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.*
No requirements listed
Co-Requisites
No Co Requisites listed

Module Content & Assessment

Indicative Content
Factor analysis Assumptions, Data screening, Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), Structural Equation Modelling (SEM).
Time series analysis Decomposition (trend, periodicity, seasonality, white noise), Smoothing Techniques, Autoregressive (AR), Moving Average (MA) and mixed (ARIMA) models.
Forecasting Forecast Error, Confidence Intervals, MAE, MAPE, RMSE, Ljung-Box statistic.
Software packages R, Minitab, Excel, SPSS

Assessment Breakdown	%
Course Work	100.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Practical/Skills Evaluation	Apply factor analysis to a dataset. Derive a model from the result.	1,5	30.0	Week 5
Short Answer Questions	Theory test	1,2,3,4	20.0	Week 9
Project	Analyse datasets and report on the results	1,2,3,4,5	50.0	Sem End

No End of Module Formal Examination

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.*

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Lecture	2.0	Every Week	2.00
Lab	Computer practical	2.0	Every Week	2.00
Independent & Directed Learning (Non-contact)	Work based on texts and class material	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Lecture	1.5	Every Week	1.50
Lab	Computer practical	1.5	Every Week	1.50
Independent & Directed Learning (Non-contact)	Work based on texts and class material	4.0	Every Week	4.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	3.00

Module Resources

Recommended Book Resources

- **Douglas C. Montgomery, Cheryl L. Jennings, Murat Kulahci 2015, *Introduction to Time Series Analysis and Forecasting*, John Wiley & Sons [ISBN: 1118745116]**
- **Niels J. Blunch 2013, *Introduction to Structural Equation Modeling Using IBM SPSS Statistics and Amos*, 2 Ed., Sage Publications Ltd [ISBN: 978-144624900]**

Supplementary Book Resources

- **Bruce L. Bowerman, Richard T. O'Connell, Anne B. Koehler 2005, *Forecasting, time series, and regression: An Applied Approach*, Thomson Brooks/Cole Belmont, CA [ISBN: 978-0534409777]**
- **Randall E. Schumacker 2016, *A Beginner's Guide to Structural Equation Modeling*, 4 Ed., Routledge [ISBN: 1138811939]**
- **Timothy A. Brown 2015, *Confirmatory Factor Analysis for Applied Research*, 2 Ed., Guilford Press [ISBN: 1462515363]**
- **Rex B. Kline 2015, *Principles and Practice of Structural Equation Modeling*, 4 Ed., Guilford Press [ISBN: 1462523344]**

This module does not have any article/paper resources

Other Resources

- **Online textbook: Rob J Hyndman and George Athanasopoulos *Forecasting: principles and practice* http://otexts.com/fpp/**
- **Online textbook: StatSoft *How To Identify Patterns in Time Series Data: Time Series Analysis* http://www.statsoft.com/textbook/time-se ries-analysis/**
- **Website: Gaskin, J. *http://statwiki.kolobkreations.com/wiki/ Main_Page***

Title:	Data Visualisation & Analytics APPROVED
Long Title:	Data Visualisation & Analytics

Module Code:	DATA9001

Credits:	5

NFQ Level:	Expert

Field of Study:	Data Format

Valid From:	Semester 1 - 2017/18 (September 2017)

Module Delivered In	no programmes

Module Coordinator:	AINE NI SHE

Module Author:	Aengus Daly

Module Description:	Data visualisation is of growing interest in the field of data science and analytics. In this module, the learner will investigate a variety of visualisation concepts, theories and techniques. More advanced visualisation concepts and tools for analysing multi dimensional data, large data sets and geospatial data will also be examined and appraised. The learner will also research and critique some of the major current challenges within data science and analytics.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Describe the concepts, principles and methods of data visualisation.
LO2	Select and apply a variety of data explorative and pre-processing techniques to a range of data visualisation problems.
LO3	Design and implement appropriate data visualisation techniques to solve data analytical problems.
LO4	Interpret, critique and communicate patterns and knowledge discovered as a result of applying data visualisation techniques to a variety of data sets and analytical problems.
LO5	Research and appraise a variety of data analytics solutions to current challenges in the area.

Pre-requisite learning
Module Recommendations *This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).*
No recommendations listed
Incompatible Modules *These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.*
No incompatible modules listed
Co-requisite Modules
No Co-requisite modules listed
Requirements *This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.*
No requirements listed
Co-Requisites

No Co Requisites listed

Module Content & Assessment

Indicative Content
Theory and Concepts of Data Visualisation History of data visualisation. Understand the various categories used in the field e.g. Information/data/scientific visualisation, infographics, visual analytics. Investigate theorists and best practice in these fields, e.g. cognitive amplification, perceptual enhancement and ways to encourage inferential processes.
Data visualisation pre-processing techniques Learn data cleaning techniques relevant to data visualisation - data aggregation, data sampling, impute missing data, find inconsistencies. Learn transformation techniques - data normalisation, construct new variables, Investigate how to use regular expressions and data manipulation techniques to pre-process data sets.
Data Visualisation traditional statistical approaches Histograms, boxplots, scatter plots; Analysing correlations and patterns between variables. Univariate, bivariate and multivariate ways of presenting data.
Advanced visualisation techniques Investigate computer based tools for visualisation and their features - interactivity, geospatial methods, hierarchical and networks solutions.
Visual Analytics Understand and critique the various visualisation methods used to solve data mining and data analytics problems, e.g. anomaly detection, pattern discovery.
Data Analytics Techniques Investigate the main pitfalls in data visualisation and data analytics in a real-world setting. Compare and contrast data analytics techniques investigating their theoretical principles, assumptions, strengths and weaknesses.
Data Visualisation Software Produce data visualisations using computing software e.g. R, RapidMiner, Excel, Qlikview, Tableau.

Assessment Breakdown	%
Course Work	100.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Project	Design and implement an appropriate data visualisation solution for a specified data mining problem.	1,2,3,4	50.0	Week 6
Project	Evaluate and implement a visualisation technique to solve a problem; research, critique and communicate the data analytics topic.	1,2,3,4,5	50.0	Week 12

No End of Module Formal Examination

Reassessment Requirement

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

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Theory, algorithms and models.	2.0	Every Week	2.00
Lab	Apply and research data pre-processing, data mining and model building techniques.	2.0	Every Week	2.00
Independent & Directed Learning (Non-contact)	Research, investigate and apply data mining concepts and techniques.	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Theory, algorithms and models.	2.0	Every Week	2.00
Lab	Apply and research data pre-processing, data mining and model building techniques.	2.0	Every Week	2.00
Independent & Directed Learning (Non-contact)	Research, investigate and apply data mining concepts and techniques.	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

- **Ben Fry 2007, *Visualizing Data*, O'Reilly Media Sebastopol, CA [ISBN: 0596514557]**
- **Nathan Yau 2011, *Vizualise This*, Wiley [ISBN: 0470944889]**
- **Matloff, Norman 2011, *The Art of R Programming*, No Starch Press San Francisco [ISBN: 9781593273842]**
- **Kabacoff, Robert I. 2015, *R in Action*, 2nd Ed., Manning New York [ISBN: 9781617291388]**

Supplementary Book Resources

- **Jiawei Han, Micheline Kamber, Jian Pei 2011, *Data Mining: Concepts and Techniques*, Morgan Kaufmann [ISBN: 9780123814807]**
- **Kotu, Vijay and Deshpande, Bala 2015, *Predictive Analytics and Data Mining*, Morgan Kaufmann, Elsevier MA, USA [ISBN: 9780128014608]**

Supplementary Article/Paper Resources

- **2011 *From abstract to actual: art and designer?like enquiries into data visualisation*, Kybernetes, Vol. 40 Issue: 7/8 [ISSN: 0368-492X]**
  https://doi-org.cit.idm.oclc.org/10.1108/03684921111160278

Other Resources

- **Website: Nathan Yau*FlowingData***
  http://flowingdata.com/
- **Website: *The Guardian***
  https://www.theguardian.com/technology/d ata-visualisation
- **Journal: *Journal of Big Data*, Springer**
  https://journalofbigdata.springeropen.co m
- **Journal: IEEE Computer Society*EEE Transactions on Knowledge and Data Engineering***
  http://ieeexplore.ieee.org.cit.idm.oclc. org/servlet/opac?punumber=69
- **Journal: *International Journal of Data Mining, Modelling and Management***
  http://www.inderscience.com/jhome.php?jc ode=ijdmmm
- **Journal: IEEE*IEEE Transactions on Big Data***
  http://ieeexplore.ieee.org.cit.idm.oclc. org/servlet/opac?punumber=6687317

Title:	Applied Machine Learning APPROVED
Long Title:	Applied Machine Learning
Module Code:	COMP9060
Credits:	5
NFQ Level:	Expert
Field of Study:	Computer Science
Valid From:	Semester 1 - 2017/18 (September 2017)
Module Delivered In	no programmes
Next Review Date:	June 2022
Module Coordinator:	TIM HORGAN
Module Author:	Ted Scully
Module Description:	This module will provide learners with a comprehensive knowledge of supervised and unsupervised machine learning techniques. It will also equip students with the skills to comprehensively evaluate models and apply appropriate pre-processing methods. The module will also focus on the application of neural networks and deep learning techniques to real-world problems such as image analysis.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Apply appropriate machine learning methodologies to facilitate pre-processing, dimensionality reduction and model selection.
LO2	Select and apply appropriate machine learning algorithms to datasets from a specific application domain.
LO3	Evaluate the accuracy of predictive models using standard methods.
LO4	Develop and implement machine learning algorithms for building predictive models.
LO5	Apply neural networks and deep learning methods for solving real-world problems.
LO6	Implement and apply optimization algorithms for solving complex problems with a high dimensional search space.

Pre-requisite learning

Module Recommendations
This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).

9098	COMP8042	Analytical and Scientific Prog

Incompatible Modules
These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.

No incompatible modules listed

Co-requisite Modules

No Co-requisite modules listed

Requirements
This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.

No requirements listed
Co-Requisites
No Co Requisites listed

Module Content & Assessment

Indicative Content
Pre-processing and Evaluation Application of a standard machine learning pre-processing methodology using techniques such as dimensionality reduction, model selection, feature selection and hyper-parameter optimization. Overview of evaluation methods such as precision, recall, confusion matrices, learning curves, ROC curves.
Classification Algorithms Building predictive models using Scikit-Learn for solving classification problems. In-depth understanding of algorithms such as linear classification, logistical regression, decision trees, naive bayes, bayesian networks and instance based learning.
Regression Algorithms Introduction to the area of regression. Produce predictive regression models using Scikit-Learn. In-depth understanding of algorithms such as uni-variate and multi-variate linear regression and ridge regression. Avoid overfitting by using regularization.
Unsupervised Learning Algorithms Overview of unsupervised learning techniques. Overview of K-Means, density-based and hierarchical clustering techniques. Optimization and distortion cost function. Random initialization and methods of selecting number of clusters such as silhouette plots.
Neural Networks and Deep Learning Overview of the concept of a neuron and neural networks. Introduction to concept of a neuron and a perceptron. Training a neural network using back-propagation. Introduction to deep neural networks. Building and deploying deep learning neural networks using TensorFlow. Building computational graphs, running tensor flow sessions, visualizing graphs using tensorboard, using optimizers, building, training and evaluating models using TensorFlow, implementing deep learning for problems such as image analysis, sentiment analysis, audio analysis.
Optimization Introduction to the area of optimization. Categories of optimization such as meta-heuristic and constraint-based optimization. Informed/Uninformed search strategies. Meta-heuristic optimization algorithms. Introduce the concept of heuristic algorithms such as hill climbing, simulated annealing, evolutionary, particle swarm optimization (PSO), etc.

Assessment Breakdown	%
Course Work	100.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Project	Apply a range of machine learning algorithms to a dataset. Appropriate pre-processing methodologies should be applied and a comprehensive reporting evaluating the performance of the algorithms should be submitted.	1,2,3,4	40.0	Week 6
Project	Apply deep learning neural networks to a complex real-world problem such as image analysis. The findings should be documented.	5	40.0	Week 10
Project	Implement an optimization algorithm for solving a complex problem with a high dimensional search space.	6	20.0	Week 13

No End of Module Formal Examination

Reassessment Requirement

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

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Delivers the concepts and theories underpinning the learning outcomes.	2.0	Every Week	2.00
Lab	Application of learning to case studies and project work.	2.0	Every Week	2.00
Directed Learning	Student reads recommended papers and practices implementation.	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Delivers the concepts and theories underpinning the learning outcomes.	2.0	Every Week	2.00
Lab	Application of learning to case studies and project work.	2.0	Every Week	2.00
Independent Learning	Student reads recommended papers and practices implementation.	2.0	Every Week	2.00
			Total Hours	6.00
			Total Weekly Learner Workload	6.00
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

- **Sebastian Raschka 2015, *Python Machine Learning*, 1st Ed., Packt Publishing [ISBN: 9781783555130]**
- **John Hearty 2016, *Advanced Machine Learning with Python*, 1st Ed., Packt Publishing [ISBN: 9781784398637]**

Supplementary Book Resources

- **Peter Flach 2012, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, 1st Ed., Cambridge University Press [ISBN: 9781107422223]**
- **Ethem Alpaydin 2016, *Machine Learning: The New AI*, 1st Ed., MIT Press [ISBN: 9780262529518]**

Recommended Article/Paper Resources

- **Pedro Domingos 2012, *A Few Useful Things to Know about Machine Learning*, Communications of the ACM, 55**

Other Resources

- **Website: *Data Science***
  https://www.kaggle.com
- **Website: *UCI Machine Learning Repository***
  https://archive.ics.uci.edu/ml/
- **Website: *Scikit-Learn***
  http://scikit-learn.org/stable/
- **Website: *TensorFlow***
  https://www.tensorflow.org/

Title:	Distributed Data Management APPROVED
Long Title:	Distributed Data Management
Module Code:	DATA9002
Credits:	5
NFQ Level:	Expert
Field of Study:	Data Format
Valid From:	Semester 1 - 2017/18 (September 2017)
Module Delivered In	no programmes
Module Coordinator:	TIM HORGAN
Module Author:	Ignacio Castineiras
Module Description:	Big data analytics turns big datasets into high-quality information, providing deeper insights enabling better decisions. However, big data requires novel data storage and data process techniques. In this module, the learner will be introduced to different NoSQL-based data models, their possible combination and the best use-cases for each of them. The learner will also compare and contrast different large scale analytics libraries, comparing them in terms of their expressiveness and efficiency.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Appraise the challenges posed by big data and the new infrastructure, data models and processing techniques it demands.
LO2	Compare and contrast the main NoSQL-based data models, discriminating the best fit for different use-cases.
LO3	Combine document-oriented and graph-based data models for a fit for purpose multi-component system.
LO4	Demonstrate the scalability, flexibility and reliability of a distributed data cluster supporting large data sets.
LO5	Compare and contrast the MapReduce and Spark large-scale analytics libraries in terms of their expressiveness and efficiency.

Pre-requisite learning
Module Recommendations *This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).*
No recommendations listed
Incompatible Modules *These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.*
No incompatible modules listed
Co-requisite Modules
No Co-requisite modules listed
Requirements *This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.*
No requirements listed

Co-Requisites
No Co Requisites listed

Module Content & Assessment

Indicative Content
The Big Data Revolution. Data storage and data process: Historical evolution. New infrastructure, data models and processing techniques required to deal with big data. Main challenges: Capture, store, search, analyse and visualise the data.
NoSQL Databases. Alternative to relational databases to address big data challenges. Impedance mismatch, scale-out vs. scale-up. Wide range of data models: Pure key/value, columnn-based, document-oriented and graph-based. Polyglot persistance. CAP theorem, partition tolerance, BASE vs. ACID transactions.
Document-oriented DBs. Efficient, scalable and resilient data storage: Replication and sharding. Clusters, configuration nodes, shards, chunk of data, shard key range, balancing backgroud operators. Expressive and efficient data queries: JSON-based document representation. Aggregation framework: Commands and pipelines.
Graph-based DBs. Efficient, scalable and resilient data storage: Property graph data model. Nodes, relationships, properties and labels. Expressive and efficient data queries: Cypher declarative SQL-like language. Graph formalism and optimal path-traversal algorithms. Polyglot persistance: On combining document-oriented and graph-based data models for a fit for purpose multi-component system.
Large-Scale Data Framework. Storage: Distributed File System. Data nodes vs. name nodes. Large files splitting and distribution algorithms. Analysis: Map-Recude. Divide and conquer algorithm schema. Map-sort-reduce process. Parallel processing. Key/value-based communication. Standard I/O file streaming. Spark: Resilient Distributed Dataset. Transformations and actions, basic API. Lazy evaluation. Context, cluster manager and worker nodes.

Assessment Breakdown	%
Course Work	100.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Practical/Skills Evaluation	Given a large data set to be stored and queried, produce a report comparing and contrasting a document-oriented vs. graph-based solution for it. Implement a polyglot persistance-based solution combining two components using the document-oriented and graph-based approaches, respectively.	1,2,3	50.0	Week 7
Practical/Skills Evaluation	Given a large data set to be stored and analysed, produce a report comparing and contrasting a Map-Reduce vs. Spark-based solution for it. Implement the two solutions, comparing them in terms of their expressiveness and efficiency.	1,4,5	50.0	Week 12

No End of Module Formal Examination

Reassessment Requirement

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

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Lecture based on Indicative Content	1.0	Every Week	1.00
Lab	Lab based on Indicative Content	3.0	Every Week	3.00
Independent Learning	Independent student learning	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Lecture based on Indicative Content	1.0	Every Week	1.00
Lab	Lab based on Indicative Content	3.0	Every Week	3.00
Independent Learning	Independent student learning	3.0	Every Week	3.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

- **Pramod J. Sadalage and Martin Fowler 2013, *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*, Addison-Wesley [ISBN: 9780321826626]**
- **Ofer Mendelevitch, Casey Stella and Douglas Eadline 2017, *Practical Data Science with Hadoop and Spark: Designing and Building Effective Analytics at Scale*, Pearson Education [ISBN: 9780134024141]**

Supplementary Book Resources

- **John Sharp et. al 2013, *Data Access for Highly-Scalable Solutions: Using SQL, NoSQL, and Polyglot Persistence*, Microsoft patterns & practices [ISBN: 9781621140306]**
- **Kristina Chodorow 2013, *MongoDB: The Definitive Guide*, O'Reilly Media [ISBN: 9781449344689]**
- **Srinath Perera and Thilina Gunarathne 2013, *Hadoop MapReduce Cookbook*, Packt Publishing [ISBN: 9781849517294]**

Supplementary Article/Paper Resources

- **Sugam Sharma et. al. 2014, *A Brief Review on Leading Big Data Models*, Data Science Journal, 13**
- **A. B. M. Moniruzzaman and Syed Akhter Hossain 2013, *NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison*, CoRR/abs/1307.0191.**
- **Landset, S., Khoshgoftaar, T.M., Richter, A.N. et al. 2015, *A survey of open source tools for machine learning with big data in the Hadoop ecosystem*, Journal of Big Data, 2:24**
- **Kyong-Ha Lee et. al. 2012, *Parallel data processing with MapReduce: a survey*, ACM SIGMOD, 40:4**

Other Resources

- **Website: *MongoDB documentation***
  https://docs.mongodb.com/
- **Website: *Neo4j documentation***
  https://neo4j.com/docs/
- **Website: *Hadoop Cloudera Map-Reduce documentation***
  https://www.cloudera.com/documentation/e nterprise/5-5-x/categories/hub_mapreduce .html
- **Website: *Hadoop Cloudera Spark documentation***
  https://www.cloudera.com/documentation/e nterprise/5-5-x/categories/hub_spark.htm l

Title:	Research Methods APPROVED
Long Title:	Research Methods
Module Code:	MATH9001
Credits:	5
NFQ Level:	Expert
Field of Study:	Mathematics
Valid From:	Semester 1 - 2017/18 (September 2017)
Module Delivered In	no programmes
Module Coordinator:	AINE NI SHE
Module Author:	Noreen Quinn
Module Description:	This module requires the learner to plan and solve a real-world problem using big data from industry, academia or government, e.g. collecting and processing real data, designing and implementing big data methods and tools, applying and evaluating data techniques to solve a real problem. The learner will undertake a piece of research conducted largely independently with academic supervision and, where appropriate, in collaboration with industry.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Propose a research question, develop the research methodology and project a plan for the research project.
LO2	Undertake preliminary experimental/design/analytical/modelling work as appropriate.
LO3	Evaluate critically a number of solutions to the identified problem.
LO4	Communicate effectively the idea and contribution of the proposed research project.
LO5	Select a potential peer-review conference/journal paper for the research work and identify how this work may contribute to furthering knowledge in the specific field.

Pre-requisite learning

Module Recommendations
This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).

9098	COMP8042	Analytical and Scientific Prog
10513	STAT8006	Applied Stats & Probability

Incompatible Modules
These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.

No incompatible modules listed

Co-requisite Modules

No Co-requisite modules listed

Requirements
This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.

No requirements listed

Co-Requisites

No Co Requisites listed

Module Content & Assessment

Indicative Content
Research Skills and Planning Methods Review research skills including technical writing, research methodology and project planning. Identify a real-world problem using big data and create an appropriate research question. Principles of research and design. Role of hypotheses. Construction and planning of experiments. Plagiarism. Ethics. Intellectual Property. Referencing. Publication Types. Peer Review. Library Catalogue. E-book database and online databases.
Research Principles and Methodologies Research in a professional context, in support of planning, decision-making and policy analysis; the impact of research. Research principles (eg validity, reliability, generalisability); qualitative and quantitative approaches and rationale (eg empirical, basic, applied, practical and action research).
Reporting of Data Analysis and Management Data management, cleaning, analysis, interpretation, presentation and preparation.
Thesis Development and Publication Structure of a Masters Thesis and a peer-reviewed conference/journal. Preparation for conference. Identify an appropriate journal/conference and identify how it may contribute to the wider knowledge in the field.

Assessment Breakdown	%
Course Work	100.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Presentation	Communicate the work to be undertaken which includes the research question and a preliminary literature review in oral format.	1,2,4	30.0	Week 7
Written Report	Submit a report of the research to be undertaken which will include a project brief, work schedule, the analysis and outcomes of preliminary work, overall findings and a project plan for the realisation stage.	1,2,3,4,5	70.0	Week 13

No End of Module Formal Examination

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

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Lecture delivering theory underpinning learning outcomes	2.0	Every Week	2.00
Lab	Practical to develop individual proposal	1.0	Every Week	1.00
Independent & Directed Learning (Non-contact)	Independent Study/Research	4.0	Every Week	4.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	3.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Lecture delivering theory underpinning learning outcomes	2.0	Every Week	2.00
Lab	Practical to develop individual proposal	1.0	Every Week	1.00
Independent & Directed Learning (Non-contact)	Independent Study/Research	4.0	Every Week	4.00
			Total Hours	7.00
			Total Weekly Learner Workload	7.00
			Total Weekly Contact Hours	3.00

Module Resources

Recommended Book Resources

- **Martyn Denscombe 2014, *The Good Research Guide*, 5th Ed., Open University Press, McGraw-Hill Education [ISBN: 9780335264704]**
- **Prabhat Pandey, Meenu Mishra Pandey 2015, *Research Methodology: Tools and Techniques,*, 1st Ed., Bridge Centre [ISBN: 9786069350270]**

Supplementary Book Resources

- **Steven J. Taylor, Robert Bogdan, Marjorie DeVault 2016, *Introduction to Qualitative Research Methods: A Guidebook and Resource*, 4th Ed., Wiley [ISBN: 9781118767214]**
- **James D.Lester 2014, *Writing Research Papers: A Complete Guide*, 15th Ed., Pearson [ISBN: 9780321952950]**

This module does not have any article/paper resources

Other Resources

- **Website: Henrichsen, L. et al 2007, *Taming the Reseach Beast***
  http://linguistics.byu.edu/faculty/henri chsenl/ResearchMethods/
- **Website: *APA reference style: Tightening up your citations***
  http://linguistics.byu.edu/faculty/henri chsenl/APA/APA11.html

Institiúid Teicneolaíochta Chorcaí
Cork Institute of Technology

DATA9003: Research Project -Data Science
(Semester:3 Mandatory)

Title:	Research Project -Data Science APPROVED
Long Title:	Research Project -Data Science

Module Code:	DATA9003
Credits:	30
NFQ Level:	Expert
Field of Study:	Data Format
Valid From:	Semester 1 - 2017/18 (September 2017)
Module Delivered In	no programmes
Module Coordinator:	AINE NI SHE
Module Author:	Aengus Daly
Module Description:	In this module, the learner will undertake a research project in a specialised area of data science. The module will develop the learner's knowledge, skills, and competences required to research, develop and scope a data science project, and to successfully complete it in accordance with an approved plan. The learner is expected to be self motivated whilst working under direction of a project supervisor, and to communicate the process and outcomes of their work on a regular basis. The learner will disseminate their research work and findings via a written thesis and an oral and poster presentation.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
LO1	Undertake a data science research project in a specialised area.
LO2	Conduct a literature review of the up-to-date methodologies and techniques appropriate to the specified area of research.
LO3	Research and detail appropriate and effective objectives and final deliverables for a data science/analytics project. Conduct a feasibility study and plan for the project.
LO4	Systematically review and adapt the employed data science methodologies during implementation in response to practical, real-world data considerations and constraints.
LO5	Produce a final written thesis detailing the work undertaken, methodologies used, findings and recommendations of the research work.
LO6	Present the project findings in person and via a poster using appropriate presentation and visual communication skills.

Pre-requisite learning

Module Recommendations
This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).

9098	COMP8042	Analytical and Scientific Prog
10513	STAT8006	Applied Stats & Probability
13329	DATA9001	Data Visualisation & Analytics
13331	MATH9001	Research Methods
13336	COMP9060	Applied Machine Learning

Incompatible Modules
These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.

No incompatible modules listed

Co-requisite Modules
No Co-requisite modules listed
Requirements *This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.*
No requirements listed
Co-Requisites
No Co Requisites listed

Module Content & Assessment

Indicative Content
Literature Survey Gather, critically analyse research and other appropriate literature in the relevant area of data science and analytics.
Research Question Formulate the core research question, identifying data set(s) relevant to the chosen data science/analytics application area.
Methodology Development Formulate and assess viable methodologies and technologies to address the identified research question with a view to identifying the most appropriate methodologies and technologies.
Project Implementation Design and implement a data science solution using supervised self-directed learning and utilising the researched data science methodologies.
Written Report Write a thesis that details the project work, the research question, the methodologies used, the findings and recommendations arising from the research.
Oral Presentation and poster Present the project research findings in person; this will include an in-depth question and answer session. Design a poster detailing the project and its main findings using appropriate data visualisation techniques.

Assessment Breakdown	%
Course Work	100.00%

Course Work

Assessment Type	*Assessment Description*	*Outcome addressed*	*% of total*	*Assessment Date*
Written Report	Submit an interim report summarising a workplan, detailing the project objectives and work to date. A preliminary literature review should also be completed.	1,2,3	10.0	Week 4
Written Report	Submit a written thesis detailing the research question, methodologies used, project findings and recommendations.	1,2,5	70.0	Sem End
Presentation	Oral presentation of the project and research findings. This will include an in-depth question and answer session.	1,6	15.0	Sem End
Presentation	Create a poster describing the project and its main research findings and results.	6	5.0	Sem End

No End of Module Formal Examination

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Independent & Directed Learning (Non-contact)	Independent research and project work	41.0	Every Week	41.00
Lecturer-Supervised Learning (Contact)	Interaction with project supervisor	1.0	Every Week	1.00
			Total Hours	42.00
			Total Weekly Learner Workload	42.00
			Total Weekly Contact Hours	1.00

Workload: Part Time				
WorkLoad Type	*WorkLoad Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecturer-Supervised Learning (Contact)	Interaction with Project Supervisor	1.0	Every Week	1.00
Independent & Directed Learning (Non-contact)	Independent research and project work	41.0	Every Week	41.00
			Total Hours	42.00
			Total Weekly Learner Workload	42.00
			Total Weekly Contact Hours	1.00

Module Resources

Recommended Book Resources

- **Joyner, R. L.,Rouse, W. A., Glatthorn, A. A. 2013, *Writing the winning thesis or dissertation : a step-by-step guide*, 3rd Edition Ed., Corwin Press Thousand Oaks, Calif [ISBN: 9781452258782]**
- **Murray, Rowena 2006, *How to write a thesis*, 2nd Edition Ed., Open University Press Maidenhead [ISBN: 9780335219681]**

This module does not have any article/paper resources

Other Resources

- **Website: Writing a Thesis: *Thesis Guide***
  https://www.seas.harvard.edu/programs/co mputer-science/thesis-guide
- **Website: CIT Library and resource portal: *Library and resource portal***
  https://library.cit.ie/