

Module Code	STU44003
Module Name	Data Analytics
ECTS Weighting¹	10 ECTS
Semester taught	Semester 1
Module Coordinator/s	Professor Myra O' Regan
Module Learning Outcomes	<p>On successful completion of this module, students will be able to:</p> <p>To understand the theory and apply the following techniques to a set of data</p> <p>LO1. Classification and Regression trees</p> <p>LO2. Ensemble methods including Bagging, Boosting, Random Forests, Gradient Boosting, Extreme Gradient Boosting</p> <p>LO3. To evaluate all of the above models.</p>
Module Content	<p>The aim of the course is to introduce the students to a set of techniques including classification and regression trees, and ensemble methods. Methods to evaluate models will also be discussed. The following topics will be addressed</p> <p>Overview of Data Analytics</p> <p>Handling data</p> <p>Missing data</p> <p>Derived Variables</p> <p>Detailed discussion of Classification and Regression Trees</p> <p>General Overview of Ensemble methods</p> <p>Detailed discussion of</p> <p> Bagging</p> <p> Random forests</p> <p> Boosting including Gradient Boosting and Extreme Gradient Boosting</p> <p>Detailed discussion of Evaluating models</p> <p>Handling unbalanced datasets</p> <p>Stacking</p>

¹ [TEP Glossary](#)

4 lectures and 1 lab per week

Assessment Details²

Assessment Component	Brief Description	Learning Outcomes Addressed	% of total	Week set	Week due
Assignment	Application of techniques to a dataset	LO1, LO2, LO3	30%	Week 8 of term	Revision week
Exam	Examination (3 hours)	LO1, LO2, LO3	70%	Exam week	8

Reassessment Details

Same as above (Assignment (30%) and Final Exam (70%))

Contact Hours and Indicative Student Workload

Contact Hours (scheduled hours per student over full module), broken down by:	55 hours
lecture	44 hours
laboratory	11 hours
tutorial or seminar	0 hours
other	0 hours
Independent study (outside scheduled contact hours), broken down by:	126 hours
preparation for classes and review of material (including preparation for examination, if applicable)	66 hours
completion of assessments (including examination, if applicable)	60 hours
Total Hours	181 hours

Recommended Reading List

Hastie Trevor, Tibshirani, R., Friedman, J. The Elements of Statistical Learning, 2nd Edition, Springer Series, 2009

Kuhn, Max & Johnson, K. Applied Predictive Modeling, Springer, 2013

Seni, G. and Elder J. Ensemble methods in Data Mining, Morgan & Claypool, 2010

Detailed list will be handed out.

Module Pre-requisites

A course on Multivariate Analysis covering principal components multiple regression, clustering techniques and logistic regression (ST3011). A good working knowledge of R is also required.

² [TEP Guidelines on Workload and Assessment](#)

Module Co-requisites
Module Website
Last Update

On Blackboard

29-07-2019 Myra O' Regan