Module Code: STU44003
Module Name: Data Analytics
ECTS Weighting: 10 ECTS
Semester taught: Semester 1
Module Coordinator/s: Professor Myra O’ Regan

Module Learning Outcomes:
On successful completion of this module, students will be able to:
To understand the theory and apply the following techniques to a set of data

- LO1. Classification and Regression trees
- LO2. Ensemble methods including Bagging, Boosting, Random Forests, Gradient Boosting, Extreme Gradient Boosting
- LO3. To evaluate all of the above models.

Module Content:
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:

- Overview of Data Analytics
- Handling data
- Missing data
- Derived Variables
- Detailed discussion of Classification and Regression Trees
- General Overview of Ensemble methods
- Detailed discussion of
  - Bagging
  - Random forests
  - Boosting including Gradient Boosting and Extreme Gradient Boosting
- Detailed discussion of Evaluating models
- Handling unbalanced datasets
- Stacking

1 TEP Glossary
4 lectures and 1 lab per week

<table>
<thead>
<tr>
<th>Assessment Component</th>
<th>Brief Description</th>
<th>Learning Outcomes Addressed</th>
<th>% of total</th>
<th>Week set</th>
<th>Week due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment</td>
<td>Application of techniques to a dataset</td>
<td>LO1, LO2, LO3</td>
<td>30%</td>
<td>Week 8</td>
<td>Revision week</td>
</tr>
<tr>
<td>Exam</td>
<td>Examination (3 hours)</td>
<td>LO1, LO2, LO3</td>
<td>70%</td>
<td>Exam week</td>
<td>8</td>
</tr>
</tbody>
</table>

**Reassessment Details**

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

**Contact Hours and Indicative Student Workload**

<table>
<thead>
<tr>
<th>Contact Hours (scheduled hours per student over full module), broken down by:</th>
<th>55 hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>lecture</td>
<td>44 hours</td>
</tr>
<tr>
<td>laboratory</td>
<td>11 hours</td>
</tr>
<tr>
<td>tutorial or seminar</td>
<td>0 hours</td>
</tr>
<tr>
<td>other</td>
<td>0 hours</td>
</tr>
</tbody>
</table>

Independent study (outside scheduled contact hours), broken down by:

<table>
<thead>
<tr>
<th>Preparations for classes and review of material (including preparation for examination, if applicable)</th>
<th>66 hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>completion of assessments (including examination, if applicable)</td>
<td>60 hours</td>
</tr>
</tbody>
</table>

Total Hours 181 hours

**Recommended Reading List**

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

Detailed list will be handed out.

**Module Prerequisites**

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.
<table>
<thead>
<tr>
<th><strong>Module Co-requisites</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Module Website</strong></td>
<td>On Blackboard</td>
</tr>
<tr>
<td><strong>Last Update</strong></td>
<td>29-07-2019  Myra O’ Regan</td>
</tr>
</tbody>
</table>