CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Autumn Examinations 2017/18

Module Title: Data Mining and Knowledge Discovery

Module Code:	DATA8004
School:	Science and Informatics
Programme Title:	Higher Diploma in Science in Data Science and Analytics Master of Engineering in Mechanical Engineering
Programme Code:	SDAAN_P8_Y5 EMENG_9_Y5
External Examiner(s):	Prof. Michael Wallace & Dr. Niall Fitzgerald
Internal Examiner(s):	Dr. David Hawe
Instructions:	Answer Question One and Any Other Two questions.
Duration:	2 hours
Sitting:	Autumn 2018

Requirements for this examination:

Note to Candidates: Please check the Programme Title and the Module Title to ensure that you are attempting the correct exaination. If in doubt please contact an Invigilator.

Q.1 Compulsory Question - Total 40 Marks – Answer any 4 parts.

a) Give a brief explanation of each of the following: 10 Marks

1. Cluster Analysis
2. The difference between knowledge and data
3. Heuristics
4. Evaluating a model

b) Outline CRISP-DM and explain how it can be useful in approaching a data mining problem. 10 Marks

c) Outline SEMMA and explain how it can be useful in approaching a data mining problem. 10 Marks

d) Explain with the aid of an example where ethical considerations may be an issue in data mining. 10 Marks

e) Explain the following: 10 Marks

1. C5.0 Algorithm
2. Cross Validation
3. Bias

[Total 40 Marks]

Answer any 2 of the remaining 3 questions (all questions carry equal marks)

Q.2 Total 30 Marks

a) Explain Adaptive Boosting. 10 Marks

b) Outline the assumptions associated with linear regression. 10 Marks

c) Explain reasons why one may choose to use an artificial neural network instead of a regression model. 10 Marks

[Total 30 Marks]

Q.3 Total 30 Marks

a) Outline how the K Means Cluster analysis algorithm works. 15 Marks

b) Explain how the Support Vector Machine (SVM) algorithm works. 15 Marks

[Total 30 Marks]

Q.4 Total 30 Marks

a) Explain the Support Vector machines technique. List the limitations of this method. 15 Marks

b) Give 3 advantages and 3 disadvantages of using Naïve Bayes. 15 Marks

[Total 30 Marks]