



DUBLIN INSTITUTE OF TECHNOLOGY

DT228A MSc. in Computing

WINTER EXAMINATIONS 2016/2017

**DATA MINING [DATA9900]
[DT228A]**

MR. BRENDAN TIERNEY
DR. DEIRDRE LILLIS
MR. CONOR SAYLES

DATE & TIME TBA.

TWO HOURS

INSTRUCTIONS TO CANDIDATES e.g.

Answer **TWO** Questions.

All questions carry equal marks.

Illustrate your answers with appropriate examples and diagrams.

1. (a) Explain the differences between *classification* and *prediction*. In your answer provide a sample application of each. [10 Marks]
- (b) Discuss the process of using *classification*. In your discussion you will need to discuss how the data is divided between training, test and apply data sets. [10 Marks]
- (c) Explain the importance of having a representative sample of your data in the training data set to avoid the *overfitting* of the model. [10 marks]
- (d) Explain the following terms and their importance for evaluating classification models. For the following table calculate the value for the following terms:
- Precision
 - Recall / Sensitivity
 - Accuracy

Predicted	Class = N	Class = Y
Actual		
Class = N	221125	6305
Class = Y	28881	13094

[20 Marks]

2. (a) There are a number of definitions of what *data mining* is. Select one definition and discuss why you think it is a suitable definition of *data mining*. [15 Marks]
- (b) Discuss the role of Descriptive Analytics in your Data Science projects, and its importance as a precursor to Data Preparation stage of the CRISP-DM lifecycle. [10 Marks]
- (c) Explain what is an Ensemble Model. Give an example, based on a business scenario, of how and when you would build an Ensemble Model and what you would like to achieve by using the Ensemble Model. [10 marks]
- (d) Explain what is meant by Prescriptive Analytics. Explain the steps that an organisation must go through before they can have Prescriptive Analytics. What are the major drawbacks to having Prescriptive Analytics? [15 Marks]

3. (a) Explain the data mining technique *Clustering* and discuss its application in a suitable problem domain. [14 marks]
- (b) Explain what is meant by Segmentation and how it is different to Clustering.
Illustrate your answer using DIT Student registration data. [14 marks]
- (c) Explain how the K-means clustering method works. [12 marks]
- (d) Explain, at least 4, of the tasks that are involved in the post-processing of the outputs from clustering. [10 marks]