**DATA MINING ASSIGNMENT 2**

Select open data source(s); Determine what questions are to be

answered; Apply the CRISP-DM methodology to analyse the data.

Reports to be completed and submitted in stages with the full report End Week 12

• End Week 6 (28/03 23:55) Project Proposal

• End Week 8 (11/04 23:55) CRISP-DM Business and Data Understanding

• End Week 10 (25/04 23:55) CRISP-DM Data Preparation, Modelling, Evaluation

• End Week 12 (9/05 23:55) CRISP-DM Deployment, Full Report All Stages

A Presentation should be prepared for the lectures during Week 12. It should include a ten-minute formal presentation followed by five minutes of questioning time from lecturers and the class.

Have some slides prepared to cover any relevant back-up / detail for answering questions. Teams will be allocated time slots.

Time limits will be enforced to allow all teams adequate time to present. The project will be evaluated as:

• Presentation /Artefact (.10) GROUP (2-3)

• Report (.20) INDIVIDUAL

Two files should be loaded to Moodle on or before Sunday 9th May 2021 (23:55).

1. A SINGLE pdf file named CA02\_Surname\_First-Name\_Student-ID\_lessons\_learned\_report.

2. A zipped file including your team artefact contents.

Key measurable objectives of the INFORMS Certified Analytics Professional exam are included. (Note:- for information only).

**INFORMS CAP Certified Analytics Professional competencies**

**Business Problem Framing**

**(Ability to understand a business problem and determine whether the problem is suitable for an analytics solution)**

Objective 1: For the open data source selected, define the problem to be addressed

Objective 2: Identify the stakeholders

Objective 3: Determine whether the problem is suitable for an analytics solution

Objective 4: Refine the statement of the problem with any constraints

Objective 5: Define an initial set of business benefits

**Analytics Problem Framing**

**(Ability to reformulate a business problem into an analytics problem with a potential analytics solution)**

Objective 1: Reformulate problem statement as an analytics problem

Objective 2: Develop ABT attributes and outputs

Objective 3: State the set of assumptions related to the problem

Objective 4: Define the key metrics of success

**Data**

**(Ability to work effectively with data to help identify quality issues and identify potential relationships that will lead to refinement of the business and analytics problem)**

Objective 1: Identify data needs and sources

Objective 2: Acquire data

Objective 3: Explore data visually

Objective 4: Harmonize, rescale and clean data

Objective 5: Document and report findings (e.g. quality report, data insights)

Objective 6: Refine the business and analytics problem statements

**Methodology (Approach) Selection**

**(Ability to identify and select potential approaches/methods/algorithms for solving the business problem)**

Objective 1: Identify potential problem-solving approaches (methods)

Objective 2: Select software tools

Objective 3: Test approaches (methods)

Objective 4: Select approaches (methods)

**Model Building**

**(Ability to identify and build effective model structures to help solve the business problem)**

Objective 1: Identify model structures

Objective 2: Run and evaluate the models

Objective 3: Calibrate models and data

Objective 4: Document and communicate findings (incl. assumptions, limitations, constraints)

**Deployment**

**(Ability to deploy the selected model to help solve the business problem)**

Objective 1: Perform business validation of the model

Objective 2: Produce the report with findings and recommendations for deployment