****

**Trinity Business School**

# Business Data Mining 2021/22

# MODULE CODE: BU7143

# MODULE NAME: Business Data Mining

**Guidelines for Assignments and Team Project**

**Individual Homework**

* Complete assignments independently, without discussing with other students
* Submit hardcopy at start of the class meeting; Uploaded files to course website before class.
* Late assignments accepted, subject to a penalty of 25% on assignment grade per late day

**Team Proposal Presentation (10 minutes)**

Each team presents their project proposal to the class. Create slides based on the 6 components in the proposal report.

**Team Proposal Report (2 pages)**

Each team should come up with a business idea that can be approached using data mining.

Format: 2-pages, 11-point font, 1-inch margins, 1.15 line space, PDF format.

The proposal should include the following six components:

|  |  |
| --- | --- |
| Suitable Title | Should reflect the business and data mining goals |
| Business Goal | * Brief description of the company/context * Who is the stakeholder or client? What is the action triggered by the prediction? * A description of the business goal. What are the business benefits of implementing this idea? What opportunity is it creating? What shortcoming does it address? * Ethical/social/human/environmental implications? * What would be considered a success? |
| Data Mining Goal | * A description of the analytics objective. * Is this a supervised or unsupervised task? Predictive or descriptive? Retrospective or forward-looking? * What is the main outcome variable(s) of interest? |
| Data | * Brief description of the available data. * Some guidance on the data subset that will be used and the re-processing or preparation that might be needed based on your past experience. * Sample of ten rows (records) with ten columns (variables) that will be used, including the outcome column. |
| Methods | * What are some data mining methods to consider? * Which performance measures are appropriate? How do they map to the business goal? |
| Implementation & Production | Operational requirements or constraints (who exactly will use the system and how? will the solution run in real-time? will it require collecting new data? one-time analysis or ongoing?) |

**Project Presentation (10-15 minutes present + 3-5 min Q&A)**

Each team presents their project to the class on session #17. The project is based on the proposal submission, where the idea is then developed into a data mining solution with final recommendations.

Create the following slides (10 slides max):

1. Cover: Informative title, team number and member names
2. Business problem (stakeholder, challenge/opportunity, humanity considerations)
3. Data mining problem (supervised/unsupervised, explanatory/predictive, how to be deployed)
4. Data description (what is a row? Output and input variables; partitioning)
5. Methods (methods, relevant outputs)
6. Evaluation (metrics of interest, benchmark, comparison)
7. Recommendations (what should the client be aware of? problems you encountered, suggestions for future data collection or analysis, etc.)

**Project Report (5 pages)**

Describe  the team project, from business problem through data mining goal and solution, to recommendations.

Format: 5 pages, 11-point font, 1-inch margins, 1.15-line space, PDF.

The report should be written clearly and professionally and include the following sections:

* **Cover Page** with informative title, team number and member names
* **One-page** **Executive Summary**: summarizes the entire report for a non-technical manager (the business problem, data, the analytics solution and recommendations)
* **Detailed Repor**t:

1. **Problem description** (business goal and data mining goal).
2. **Data description** (the data that you end up using: size and dimension, what is a record, list of output and input variables, sample of 5 rows).
3. **Brief data preparation** details (how your data were created from the raw data) and key charts. Details can be provided in an Appendix.
4. **Data mining solution**: Methods applied (with sufficient detail and screenshots; use Appendix if needed) and appropriate performance evaluation (proper choice of measures, benchmarking). Post your R code on github and share the link.

**Conclusions** (advantages and limitations; potential ethical issues) and operational recommendations.