



## **TBA3102 – Text Analytics**

### **AY 2024/25 Semester 2**

### **Group Project**

#### **Overview of Project**

This project is to be done on a group basis and is worth a total of 50%. The final deliverable for the project consists of a text analytics project plan and a separate project report.

Your deliverable will be submitted to Turnitin for a plagiarism check. Thus, please ensure that the final report that your group eventually submits is entirely original work.

At the completion of this group project, you should develop a better understanding of the following text analytics issues within the context of a typical real-world organisation:

- How is text analytics applied in a challenging real-world business domain/industry.
- How to initiate and undertake a new analytics project that involves unstructured text data.
- How to employ text preprocessing, feature engineering and modelling techniques to i) derive useful patterns and insights from text data; and ii) solve real-world practical problems that bring tangible benefits to businesses.

The project is divided into two main phases of planning and implementation.

#### **Project Task A – Project Planning (15 marks)**

Suppose your group is an analytics team in a typical real-world organisation. Your group may choose to work on any domain/industry. Propose an analytics project involving unstructured text data in collaboration with one or more other business or functional units in your organisation to solve a real-world business problem in the chosen domain/industry. Prepare a project plan for the Chief Analytics Officer to review, with the aim of getting his/her approval to initiate the project.

The project plan should contain the following elements:

- a. One-page executive summary of the project.
- b. Intended audiences and their primary activities. **(1 marks)**
- c. Business question **(3 marks)**:
  - i. State the business question.

- ii. Provide a description of the relevant business challenge/problem.
  - iii. Identify the affected business processes.
- d. Data requirement (**3 marks**):
  - i. State the data that are required.
  - ii. Provide relevant business and operational definitions of the data.
  - iii. Data collection plan for existing and/or new data.
- e. Data analysis plan (**6 marks**):
  - i. Explain in detail the type of analysis work that will be undertaken in order to answer the business question.
  - ii. Provide the details for the prediction involved.
- f. Type and format of deliverables (**1 marks**):
  - i. State what are the reporting, dashboarding and modelling output that would be created from the data analysis work.
  - ii. List the frequency and periods when the analysis should be delivered.
- g. Outline the initial presentation plan as well as the review and follow-up process. (**1 marks**)

### **Project Task B – Project Implementation (35 marks)**

Suppose your group's project plan has been approved by the Chief Analytics Officer. Implement the project and document the tasks and results for each of the following steps in the text analytics process:

- a. Construction of text corpus (**5 marks**) – Include the crawling of data from the Internet (if applicable) and ingestion of data into a suitable data storage engine such as a relational database or csv file.
- b. Text preprocessing (**10 marks**) – Understand and clean the text data.
- c. Feature engineering (**10 marks**) – Represent the unstructured text data in a structured format using various feature engineering models.
- d. Actual text analysis using machine learning (**10 marks**):
  - i. Train models using various supervised learning, unsupervised learning, reinforcement learning and deep learning algorithms.
  - ii. Evaluate the performance of the models.
  - iii. Deploy the final chosen model to predict new unseen observations.

### **Document Formatting Requirements**

Please format your document based on the following settings:

- Paper size: A4
- Margins: 1 inch all around.
- Orientation: Portrait.
- Line Spacing: Double spacing.
- Font Type for Main Text: Times New Roman.
- Font Size for Main Text: 12 pt.
- Page Limit: There is no page limit. Please include references when citing sources other than those provided. You may use any established reference style such as APA.
- File Format: DOCX. Do not convert the document into PDF.
- File Name: Name the files as “GPXX\_Plan.docx” and “GPXX\_Report.docx” where “XX” is your group number.

### **Deliverables Submission Instructions**

Compress your group’s written deliverable (i.e., “GPXX\_Plan.docx” and “GPXX\_Report.docx”) and all datasets as well as source files into a single zip archive file and name the file as “GPXX.zip” where “XX” is your group number. Upload your zip file to the Canvas Assignment: Assignments > Group Project - Deliverables by **Sunday, 20 April, 11:59 pm.**