

# BT4222 Project Deliverable Guideline

October 2023

## Table of Contents

|   |          |
|---|----------|
| <b>Deliverable List .....</b>                                     | <b>1</b> |
| <b>Requirement about your source code .....</b>                   | <b>2</b> |
| <b>Format of the Project Final Report .....</b>                   | <b>2</b> |
| <b>Content of the Project Final Report .....</b>                  | <b>2</b> |
| <b>Abstract</b> (within 300 words in a separate page) .....       | 2        |
| <b>Section 1 Proposal Review</b> (within one page) .....          | 2        |
| <b>Section 2 Data Description</b> (within two pages) .....        | 3        |
| <b>Section 3 Models and Performance</b> (within two pages) .....  | 3        |
| <b>Section 4 Contribution and Justification</b> .....             | 3        |
| <b>Section 5 Reference</b> (within one page) .....                | 4        |
| <b>Assessment Rubrics (35 marks) .....</b>                        | <b>5</b> |
| <b>BT4222 Project Presentation Schedule (2023/24-Term1) .....</b> | <b>5</b> |

## Deliverable List

| Deliverable                        | Format   | Due Time                |
|------------------------------------|--|-------------------------|
| Project Final Presentation         | .pptx uploaded to Canvas (we will upload your files into a laptop (with windows OS) in advance for your presentation purpose.)   | 11:59 PM Nov.16 (Thur.) |
| Project Final Report               | .pdf or .docx uploaded to Canvas   | 11:59 PM Nov.19 (Sun)   |
| Project source code                | .ipynb files uploaded to Canvas or accessible via a public link  | 11:59 PM Nov.15 (Wed)   |
| Datasets                           | .csv files including both raw datasets and intermediary datasets if applicable). Please provide the links via which the datasets can be accessed. Do not upload your data to Canvas. | 11:59 PM Nov.15 (Wed)   |
| Project group peer evaluation form | .doc or docx file uploaded to Canvas   | 11:59 PM Nov.19 (Sun)   |

Note:

- Both the source code links (or file names) and dataset links should be listed in a pdf file with *clear, precise, simple and short explanation about their purpose and content*. This pdf file should be also uploaded into Canvas by 11:59 PM Nov.15.
- Late submission of source code and datasets will result in the absence of the corresponding evaluation.

## Requirement about your source code

1. Source code file name should be as self-explainable as possible. For example, step1\_data\_preprocess.ipynb, step2\_classification\_models.ipynb, step3\_rnn\_models.ipynb.
2. Source code files should be able to directly run via Google colab, including loading relevant raw datasets or intermediary datasets. If it is indeed not feasible, please provide a readme file to specify the running environment.
3. Source code files should be able to reproduce your reported model performance (reasonable deviation is acceptable). To serve this purpose, please highlight in your source code the scripts that are used to generate the reported performance. Please attach or keep such performance report in the source code for verification purpose.

## Format of the Project Final Report

- Cover page: Project Topic, Group No., List of group members
- Line spacing: 1.5 lines
- Font: Arial, Calibri, or Times New Roman; Size 11
- Margins: Normal; Columns: One

## Content of the Project Final Report

Your project report should include the following sections.

### Abstract (within 300 words in a separate page)

- Summarize the business issue and the corresponding machine learning problem that have been addressed in this project. (please make sure the business issue and the machine learning problem are consistent with each other. Inaccuracy or mismatching between them will be penalized.)
- Acknowledge the existing datasets that have been utilized and the established models (or source code) that have been adapted into your project. (details about the sources should be provided in reference, not here ).
- Summarize in bullet points any achievement or noteworthy highlights directly attributed to your group's efforts

### Section 1 Proposal Review (within one page)

- A **table** comparing the potential contributions listed in your proposal with what you have achieved in the final project in terms of satisfaction levels, using a column to indicate the satisfactory level, i.e., Exceeds Expectations/ Satisfactory/ Under Expectations/ Not Achieved.
- Provide **justification** if some of the proposed contribution or outcome have not been achieved.

*This section serves as a self-reflection and/or also help us to better understand the challenges you faced.*

## Section 2 Data Description (within two pages)

- For each dataset, provide the original source, the number of features and the number of data points
- If your model input include features, please provide a table that consists of the following information for each feature (including new features created in your project)
  1. Feature name
  2. Feature brief description
  3. Value type and the unit for numeric value, e.g., days, minutes, count of words, etc.
  4. For numeric value, provide descriptive statistics including min, max, mean, median, standard deviation, the number of missing data points
  5. For categorical value, provide the list of distinct categories. For example, feature “gender” can take the value, female, or male, etc. and the number of missing data points

Note that, if your datasets include many features (more than 20), instead of listing all of them, you should group them by categories or aspects, for each category or aspect, pick up 1-2 representative features and provide their information as described in points 1-5.

*This section stresses the importance of understanding data. The required information is the preliminary requirement for a data analyst/scientist.*

## Section 3 Models and Performance (within two pages)

- Use diagrams to illustrate the alternative models and/or model architecture that have been used in your project. Precise and succinct explanation should be provided if necessary.
- Use table and/or figures to report the model performance on training set and testing sets (all these outcome can be reproduced by running your source code). Precise and succinct explanation should be provided if necessary.

*This section is a general requirement for any machine learning related project. Let us keep it simple and straightforward with only necessary information.*

## Section 4 Contribution and Justification

(Among the four aspects identified below or your unique aspects that are not listed here, please provide your justification for *each aspect* within 1-2 pages. This excludes the self-evaluation table.)

- Complete the following table and assess your own contribution taken into account both the extent of efforts and the effectiveness of the outcomes.

*Regarding the four aspects of the contribution, you are not required to cover all of them. It should be your own decision depending on your interest, and an optimization of your studying efforts and outcome. For instance, you may focus on only one aspect*

*that you are most interested. Our evaluation will assess your efforts and effectiveness in each aspect, considering the competitiveness among groups.*

| Contribution   |               | Low | Medium | High |
|--|---------------|-----|--------|------|
| Use valuable and high-quality new datasets, including integrating existing datasets, scrawling or retrieving data, etc.  | Effort        |     |        |      |
|  | Effectiveness |     |        |      |
| Creativity in feature engineering in a way that <ul style="list-style-type: none"> <li>• overcomes the limitation of current data or model;</li> <li>• or increases model performance;</li> <li>• or increases the interpretability of the models;</li> <li>• or reduces computing cost</li> </ul> | Effort        |     |        |      |
|  | Effectiveness |     |        |      |
| Design or adaptation of new ML methods/architecture or the integration of existing methods with a balance of resource and cost   | Effort        |     |        |      |
|  | Effectiveness |     |        |      |
| Creativity or insights in understanding or further explaining the prediction results and performance   | Effort        |     |        |      |
|  | Effectiveness |     |        |      |
| Any aspects that are distinct from the above...  | Effort        |     |        |      |
|  | Effectiveness |     |        |      |

- Provide justification for your self-evaluation with necessary details and objective evidences. Without any justification, the claimed contribution cannot be supported. For each aspect of your contribution, your justification should include but not limit itself to the following.
  - Necessary information for people to understand what you have actually done.
  - How does your approach differ from existing methods? You may cite established projects, papers, or the provided example source code and highlight the deviations from them. When using existing projects or papers as benchmarks, ensure they are widely recognized and not considered as noise in the field.
  - Any other objective and convincing evidence that can be reproduced from your models

#### Section 5 Reference (within one page)

Reference must be comprehensive with correct information and valid links if applicable.

## Assessment Rubrics (35 marks)

- Ontime submission of all the deliverables following required format (**2 marks**)
- Completion of the proposed tasks with reasonable performance and provision of required Abstract, Sections 1-3 and 5 in your report with clarity and simplicity (**15 marks**)
- Understanding your data and source code to the extent that any member can answer questions related to your data or source code during your presentation (**2 marks**)
- Presentation (**4 marks**) Your presentation should communicate the important information relevant to Section 1, 2, 3 and 4.
- Contribution (**12 marks**)

Presentation evaluation:

- Basic requirement (1 mark)
  - 10 minutes per group excluding Q&A;
  - Every group member must present;
  - Any group member can reply any questions related to the project;
  - Reading scripts throughout the presentation is not allowed.
- Presentation quality (3 marks)
  - The presentation is well organized with clarity, logic, and succinctness

**Note:** your individual project marks = your group project proposal marks + your group project final marks \* your average group peer evaluation (%)

## BT4222 Project Presentation Schedule (2023/24-Term1)

| Venue/Time Slots    | Final Presentation 17 Nov. 2023 (Fri.)       |
|---------------------|--|
|                     | <b>COM3-02-59 - Meeting Rm 20 (40 seats)</b> |
| 8:55 AM - 9:15 AM   | Sentimental Analysis Group 1                 |
| 9:15 AM - 9:30 AM   | Sentimental Analysis Group 3                 |
| 9:30 AM - 9:45 AM   | Sentimental Analysis Group 4                 |
| 9:45 AM - 10:00 PM  | Sentimental Analysis Group 6                 |
| 10:00 AM - 10:15 AM | Fraud Detection Group 2                      |
| 10:15 AM - 10:30 AM | Fraud Detection Group 3                      |
| 10:30 AM - 10:45 AM | Recommendation System Group 2                |
| 10:45 AM - 11:00 PM | Recommendation System Group 3                |
| 11:00 AM - 11:15 AM | Recommendation System Group 4                |
| 11:15 AM - 11:30 AM | Recommendation System Group 5                |
| 11:30 AM - 11:45 AM | Sentimental Analysis Group 7                 |
| 11:45 AM - 12:00 PM | Fraud Detection Group 4                      |
|                     | <b>COM1-03-19 MR1 (30 seats)</b>             |
|                     |  |
| 3:15 PM - 3:30 PM   | Fraud Detection Group 5                      |
| 3:30 PM - 3:45 PM   | Recommendation System Group 10               |
| 3:45 PM - 4:00 PM   | Recommendation System Group 6                |
| 4:00 PM - 4:15 PM   | Recommendation System Group 7                |
| 4:15 PM - 4:30 PM   | Recommendation System Group 8                |
| 4:30 PM - 4:45 PM   | Recommendation System Group 9                |
|                     |  |
| 5:00 PM - 5:15 PM   | Sentimental Analysis Group 8                 |
| 5:30 PM - 5:45 PM   | Text Summarization Group 1                   |
| 6:15 PM - 6:30 PM   | Sentimental Analysis Group 2                 |