INFS4203/7203 Project Phase II (20 marks)

Semester 2, 2021

Due date:

16:00 on 29th October 2021(Brisbane Time) (Phase II, 20%)

All assignments should be submitted to UQ Blackboard only. If any assignment is failed to be submitted appropriately before due, a penalty will be applied according to ECP. It is your responsibility to ensure your submission is successful before due time. Email submission will not be accepted.

Overview

In Phase II, you will implement your proposal submitted in Phase I, with necessary adjustment according to the empirical performance and the feedback from the proposal. This is an individual assignment. The completion of the assignment should be based on your own design and feedback from the proposal.

Track 1: Data-oriented project

In Phase II, you will be provided with the test data named *Ecoli_test.csv*. The first row describes features' names. Except the first row, each row in the data file corresponds to one data point. There are 917 test data points in this file, and each column represents the same feature as the training data *Ecoli.csv*. Note that the test data only has 116 columns, without labels, i.e., the final column "Target (Column 117)" in *Ecoli.csv*.

In this phase, you will need to implement the ideas in your proposal and use them to classify the test data. You need to submit

- A *result report* on
  - Test result: the prediction on test data and
  - Evaluation result: the evaluated accuracy and F1 on the training data using cross-validation.
- A *readme* file with clear and thorough *description of your coding environment* (operation system, programming language and its version, additional packages installed etc.) and instructions on how to run the code such that your reported test and evaluation results can be reproduced.
- Your *implemented code* which has a main function to generate the test and evaluation results of classification.

Format

- The result report should be named as *sxxxxxxx.csv (sxxxxxxx is your student username)*. For example, if your student username is *s1234567*, then the result report should be named as *s1234567.csv*.
- The result report should be composed of 918 rows. For the first 917 rows, the *i*th row gives the prediction of the *i*th test instance, either 1 or 0. The last row (row 918) gives the accuracy (first column, rounded to the nearest **3rd decimal place**) and F1 (second column, rounded to the nearest **3rd decimal place**) evaluated by yourself through cross-validation on the training data. You could refer to *result_report_example.csv,* which provides an example (NOT groundtruth) of the result report.

Note that result report submitted in other forms or names will not be accepted or marked.

- Together with the result report, you need to submit a *readme* file and all your *code*.
- The *readme* file and your *code* should be compressed into **one** zip file named *sxxxxxxx.zip (sxxxxxxx is your student username)*.

Note that code and readme file submitted in other forms or names will not be accepted or marked.

We recommend you follow the Google Style Guides (https://google.github.io/styleguide/) for the programming style, which is not mandatory for this assignment, but using it may benefit your future career as a data scientist.

Submission

Only your submitted version will be marked. All required files need to be submitted before due. Otherwise, penalty will be applied according to ECP, i.e.,

10% of the maximum possible mark for the assessment item will be deducted per calendar day (or part thereof), up to a maximum of seven (7) days. After seven days, no marks will be awarded for the item. A day is considered to be a 24-hour block from the assessment item due time. Negative marks will not be awarded.

- Result report should be submitted through the "Report submission" Turnitin link provided on Blackboard -> Assessment -> Project Phase II -> Report submission before the deadline.
- Compressed file of readme and code should be submitted through the "Readme and code submission" Turnitin link provided on Blackboard -> Assessment -> Project Phase II -> Readme and code submission before the deadline.

Marking standard

Submissions satisfying both the following two will be accepted and marked

1. The test and evaluation results can be reproduced by the submitted readme file and code.
2. The test and evaluation results are generated by applying classification techniques to the data.

When the above two conditions are satisfied, the result report will be marked according to the F1 result on the test data in the following way (rounded to the nearest 1st decimal place)

- For F1 less than or equal to 0.2: Mark = F1 / 0.2 * 8
- For F1 greater than 0.2 but less than 0.7: Mark = (F1-0.2) / 0.5 * 5 + 8
- For F1 greater than or equal to 0.7 but less than 0.9: Mark = (F1-0.7) / 0.2 * 7 + 13
- For F1 greater than or equal to 0.9: Mark = 20
- Please see the example below

F1	Mark
0.1	4
0.2	8
0.3	9
0.4	10
0.5	11
0.6	12
0.7	13
0.8	16.5
0.9	20

Pre-submission feedback (optional)

If you want to get evaluation feedbacks before the due time, you could submit a *pre-submission report file* (**with the same format and file name as the result report file**) before Oct 19th 16:00 pm through Blackboard -> Assessment -> Project Phase II -> Pre-submission feedback (optional).

Oct 19th 16:00 pm is a **firm** deadline. Late submission will not be assessed and provided with feedback. File submitted in other forms or names will not be accepted.

Feedback will be provided based on the submitted *pre-submission report file*.

Track 2: Competition-oriented project

In this phase, you need to submit:

- A *result report* of the Public Leader Board results, including a screenshot and an URL of the Public Leader Board.
- A *readme* file with clear and thorough *description of your coding environment* (operation system, hardware requirement, programming language and its version, additional packages installed etc.) and instructions on how to run the code such that your final submission to Kaggle can be reproduced
- Your *implemented code* which has a main function to generate the final submission to Kaggle.

Your submission will be marked according to the marking standard specified in "Project Specification" released in Week 2.

Format

- The result report should be named as *sxxxxxxx.pdf or sxxxxxxx.doc/docx (sxxxxxxx is your student username)*. For example, if your student username is s1234567, then the result report should be named as *s1234567.pdf/doc/docx*.

Note that result report submitted in other forms or names will not be accepted or marked.

- Together with the report, you need to submit all your *code* and a *readme* file. The *readme* file and your *code* should be compressed into **one** zip file named *sxxxxxxx.zip (sxxxxxxx is your student username)*.

Note that code and readme file submitted in other forms or names will not be accepted or marked.

Submission

Only your submitted version will be marked. All required files need to be submitted before due. Otherwise, penalty will be applied according to ECP.

- Result report should be submitted through the "Report submission" Turnitin link provided at Blackboard -> Assessment -> Project Phase II -> Report submission before the deadline.
- Compressed readme file and code should be submitted through the "Readme and code submission" Turnitin link provided at Blackboard -> Assessment -> Project Phase II -> Readme and code submission before the deadline. Note that the zip file should be smaller than *100MB*. **If your file is larger than *100MB*, please contact Zijian Wang (zijian.wang@uq.edu.au) before due time by email in case there is any penalty applied to later submission.**