Applied Statistical Analysis: Problem 3 (Team-Based Project)

Conduct the following subjects by using the data for Bank Marketing Data Set, which is uploaded to K-LMS (Canvas). There are two data sets: “train+test.csv” and “submit.csv”. See the document for the detail. This problem is performed by the team-based project.

1. Construct several models to predict the output variable ((has the client subscribed a term deposit? (binary: 'yes','no')). In order to do it, you can divide the data “train+test.csv” into training sample and test sample data.
2. As for the models, you are requested to use **logistic regression** and **classification tree** models. Furthermore, you can use other advanced models, such as Support Vector Machine, Random Forest, Deep Learning, LightGBM etc. (these models are not explained in this class.)
3. Show the ROC (Receiver Operating Characteristics) Curve and AUC (Area under the curve) for the several candidate models by using the test data set. Select the best model in terms of the AUC. For the selected model, examine the important input variable which could have an effect on the output variable.
4. As for the selected model, examine the classification limit (or threshold to determine the output variable ('yes','no')). Show the confusion matrix (recall, precision) and F-measure for the determined threshold of the selected model, from test data set.

- Deadline for the submitting file for each team: January 12 (Thursday), 23:59, 2023

  Please submit the “submit” file via K-LMS (Canvas).

  Note that the representative of each team is required to submit the file, which mean that all members do not have to submit it.

  The “submit” file is used to submit the **predicted output variable** (has the client subscribed a term deposit? (binary: 'yes','no')) and the **probability of ‘yes’** (or score). There are 19 input variables. The output variable is not provided. The number of instances is 1,000.

- Deadline for the presentation file of Power Point for each team:
  January 12 (Thursday) 23:59, 2023.

  Please submit the report file of Power Point via K-LMS (Canvas).

  Note that the representative of each team is required to submit the file, which mean that all members do not have to submit it.

  Write in English.
  Write the following items in the 1st page.
  - Title of your presentation
  - Team member name

Note that as for the problem 3, two aspects are considered:
- The AUC and F-measure scores on the basis of your submitted data
- The report explaining about the constructed models, analysis results, etc.