Programming assignment 03 – Learning
CII-2M3 Introduction to Artificial Intelligence
Even Semester 2021/2022

1. Description

Given file **traintest.xlsx** which contains 2 sheets: *train* and *test*, consists of a dataset for binary classification problem. Each record/row comprised of row number (id), input features (x1, x2, x3), and class output (y). Input features are integers within a certain range for each different feature. Class output is binary (either 0 or 1).

id	x1	x2	x3	y
1	60	64	0	1
2	54	60	11	0
3	65	62	22	0
4	34	60	0	1
5	38	69	21	0

The *train* sheet comprised of 296 data, complete with class target output (y). Use this sheet to model or training depending on the learning method you use. The *test* sheet comprised of 10 data, with hidden class output. Please use this sheet to test your model.
In the end, the output of your model will be compared with the target or actual classes.

You may choose one of the following learning methods:

- Decision Tree (ID3)
- KNN
- Naïve Bayes

Note: You must implement your chosen method completely for the given dataset. Please write down the complete steps in the report.

Implement the following in your code (can use function/procedure):

- Read training/test data
- Model training
- Save the trained model
- Model testing
- Model evaluation
- Save output to a file

Note: All processes related to training, testing, and evaluation MUST be made without any existing learning library.

2. Program Output

Your program mainly consists of 3 steps: training, testing, and evaluation. The training step results in a model, depending on your chosen method. The testing step results in classification, either class 0 or 1. Test may receive more than 1 item (can be a list/array). The evaluation step results in accuracy metrics, depending on your chosen metric and reasoning (e.g., accuracy, recall, precision, etc).

Rules and Grading

1. **Implementation**
   - This is a **group** assignment.
   - **1 group = 2 students**; If there is one student left, then he/she can join to any group in the class. Therefore, there can be **only one group** which has **three students**. The rest of the group should only consist of 2 students.
   - Every group member is expected to contribute equally. This contribution will be considered during grading.

2. **Submission**
   - Assignment should be submitted by **Saturday, 18 June 2022, 23.59 WIB** at the latest via LMS CeLOE in your respective classes.
   - **Submission requirements** (in 1 **.ZIP file**):
     - Program *Source Code*
     - Report
   - **Each student submits their own (group) work** (all 3 requirements above); For each group, the submitted files should be the same.
   - File naming convention (.zip): File format **CLASS_GROUP _NIM** (NIM – your own student ID):
     - Class code in 1 word, without any "-", "_", or other characters
     - Group code is written in 2-digit numbers.
     - NIM, your own student ID
     - Example: **IF44INT_03_1311281234.zip**

3. **Program Source Code**
   - Program should be written in **Python** (if you are not familiar with Python, other programming language is allowed). However, using Python is strongly recommended.
   - **Using external libraries which directly perform GA processes are not allowed**, as mentioned in the previous page. Any use of external GA library will reduce the grade of this assignment.
   - Please include an instruction on how to run the program in a **Readme.txt**; Please put it in the same folder as the main program.

4. **Report**
   - Content: (1) Problem description, and (2) **each point** which must be **analyzed** and **designed**, which were described in the previous page.
   - Free form of writing; submitted in **PDF** format.
   - Report must be in accordance with the submitted code, otherwise it will affect grading.
   - Include **screenshot** from the program run based on the chosen parameter. Also include program output: the best chromosome with the x and y decoded value.

- Include team member contribution. Also specify if any of the team member does not contribute.

5. **Presentation**
   - Prepare presentation slides to be presented in class. **Grade per group member can be different depending on the presentation!**
   - Slides' content: (1) understanding of the problem, (2) what you have done to solve the problem, and (3) result/output with your opinion/discussion point.
   - Max duration 10 minutes
   - Presentation to be scheduled later depending on your class.

6. **Grade criteria**
   - **Plagiarism/cheating** will be given **Grade = E**
     - Same program/report **up to 80%** (not including team member).
     - Submit other student's work (not including team member).
   - Grading component:
     - **Program = 40%**:
       - Originality (without Library) = 15%
       - Program works properly = 20% (no bug in the program)
       - Program output = 5% (according to accuracy with test data)
     - **Report = 40%**:
       - Problem understanding = 10%
       - Method design and analysis = 20%
       - In accordance with the code/program = 10%
     - **Presentation video = 20%**:
       - Presenting clearly = 5%
       - Understanding the topic = 10%
       - Ability and correctness in answering question = 25%

Note: If there is any question, please refer to the designated class lecturer.

Bandung, 26 May 2022

Lecturer Team – Introduction to Artificial Intelligence
S1 Informatics (Even Semester 2021/2022)