

**CLO-3 Assessment - Team Based Project**  
**CLO-3 Learning: Model selection, Decision Tree, kNN, Naïve Bayes**  
**Intro to AI**  
**Faculty of Informatics –Telkom University**

*Deadline for report submission (including program code link, presentation slide, presentation video recording) is on Sunday, 4 June 23, 23:59 on LMS*

*There is no class in Week-14 as it is reserved for doing this assignment.*

CLO-3 assessment is an evaluation of the student's group understanding to describe, implement, design and analyze learning algorithms, including Decision Tree/kNN/Naïve Bayes using the given dataset.

This is a group project, with 2-3 students per group.

**A. Scenario**

Your team is given a project to help a Cardiologist to explore arrhythmia data (<https://archive.ics.uci.edu/ml/datasets/Arrhythmia>). Arrhythmia is a condition where the heartbeat is unregulated, either faster or slower.

The first challenge is to perform **data pre-processing** using understanding on EDA-*Exploratory Data Analytics* which you can follow here: <https://youtu.be/E3n1h9sNAJ0>.

*To help you to understand this, there will be a session with the assistant on Wednesday, 31 May 19.00-20.00, or will be announced in the chat group.*

After observing the data, you must **choose two learning algorithms** given during the course (DT/kNN/Naïve Bayes). The chosen algorithms must be based on the characteristics of the dataset to achieve good accuracy.

The chosen algorithms must be explained in detail, along with the analysis of the results. Only 1 of the team members need to **submit via LMS**.

**In-person or online presentations** are scheduled for week-15 and week-16, allocated later by the lecturer.

**B. Tasks and responsibilities**

In this project, your team is responsible for describing, implementing, analyzing, and designing two of the learning techniques stated above. The following are the general guidelines of the tasks and responsibilities:

1. Investigate the **quality of the given dataset**. Explain your decision on the data pre-processing method. Explore the dataset by summarizing and using statistical method to

identify the data quality. There is no limit to the data summary, but please only report relevant information.

2. **Choose two learning methods.** You HAVE TO choose two methods you are going to implement, and then compare the results between the two. For each of the algorithms, use different hyperparameters, and then analyze the results. **DO NOT provide only screenshots of the code without any explanation.** Give complete explanations of each the function and the workflow of your code.
3. Write a report using the **Scientific Report** schema (given in section C). Thereafter, present your project in the class. You are supposed to record a presentation video to show the work of each team member (grading of the presentation is different per student). *This recording is meant to be a practice run, as your team will present this also in class.*

### C. Report structure

**The report has to be informative so it can provide complete explanations in a sufficient way. A perfect score for the report is 70.**

The following template is recommended, but you can adjust and add more information.

#### Section-0:

Cover page consists of each team member information (NIM, Name, Class, Group No, and contributions) and write down the following statement:

***"Our team completed this assignment by following the course's rules and academic ethics code. If we did any plagiarism or other violation, then we accept to have a final grade of E for this course."***

#### Chapter-I: Introduction (10 points)

This chapter consists of the background of the data pre-processing method, and the explanation of the results of the pre-processing. You also need a brief explanation of your chosen learning methods. You can use a structured table format to explain the background of the dataset that you choose and give explanation of the content of the dataset.

Describe the investigation that you did regarding the data quality and the points leading to the chosen data pre-processing method. Provide some plots/diagrams that support the decision-making process, and some explanations related to them.

#### Chapter-II: Background Theory (40 points)

Explain in great detail the TWO chosen algorithms (DT/kNN/Naïve Bayes). Write down the definition, logic, and implementation details. Give your expected results from the chosen algorithms. Provide analysis that describes the varying hyperparameter choice for each of the algorithms.

Explain how the chosen algorithms can produce results, by using a flowchart, in addition to the program code. Write down an explanation for each screenshot of your code. You are not allowed to only put screenshots, without explaining the meaning of the *Input-Process-Output*.

Include the link of google colab or github for your team, make sure that the link is working properly. Give comments for each of the code blocks. **Using libraries is allowed, however, there is a 20 bonus points if you code the algorithm from scratch.**

[Note: If there is any error or failure, you can still prove the mistake and get some scores for that].

### Chapter-III: Result evaluation and discussions (10 points)

Write down the results and the formula to evaluate the performance. Additionally, give an evaluation of the results which you may or may not have found. You can write it down using tables or graphs, depending on your results. Furthermore, analyze the results and explain.

Compare the results from the two chosen algorithms, which one gives the best accuracy. Provide explanations and analysis why that is the case.

### Chapter-IV: Conclusions and Suggestions (5 points)

Using paragraphs (not bullet points), write down the conclusions about the results and implementations of the scientific project. Give suggestions for the improvement of the course.

**In the last paragraph, include the link to access the slides, presentation vide, and report.**

### Chapter-V: References (5 points)

Include all the references that you use using the following format:

- [1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
- [2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
- [3] M. Wegmüller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.
- [4] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
- [5] (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [6] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: <http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/>
- [7] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
- [8] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.
- [9] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
- [10] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.

There is no limit to the number of references, a minimum of 3 references. Make sure you refer to all the sources and figures you used online.

#### **D. Presentation**

**Present your scientific report project using the proper presentation techniques that you know. Project that is presented well can expect a full 25 points.**

To support your presentation, you are expected to.

(1) prepare presentation slides which consist of all the important points from the report. Slides that give complete insights of the goal of the project will be given 10 points.

(2) presentation video +/- 15 minutes. The video starts with the introduction of each of the team members on-camera. Proceed the presentation using the slides, demo of the program, and show the results. The recording counts as 15 points.

**E. Additional: Summary and Grading for other groups (5 points)**

To help your understanding in implementation of this CLO-3 assessment, you **MUST** grade other teams' presentations with brief description of their presentations (see the form below).

Scores are the following:

**NOT READY/READY/OUTSTANDING** followed by your reasoning.

*Note: Print the form (or use Word doc) to bring to class during week-15 and week-16*

*Prepare the form by adding the group nr listed in your class. The form should be submitted via LMS after the final week (week-16).*

<b>Reviewer NIM:</b>		
<b>Reviewer Nama:</b>		
<b>Presentation Score Form</b>		
<b>Team</b>	<b>Summary</b>	<b>Score &amp; reasoning</b>
(0) with team nr/name <b>Bayangan AI</b> (just an example)	They implemented <b>kNN</b> really well and reached <b>90% accuracy</b> , which is based on the good choice of k. (just an example)	<b>READY</b> , this team presented well, and the talk can be followed really easily (just an example)
....		
....		
....		
....		