|  |  |
| --- | --- |
| **Course: AIL303m MACHINE LEARNING** | **Contribution: 30% of the course** |
| This project should take an average student who is up to date with tutorial work approximately 5 weeks | |
| **Learning Outcomes:** CLO10 | |

**Plagiarism** is presenting somebody else’s work as your own. It includes copying information directly from the Web or books without referencing the material; submitting joint coursework as an individual effort; copying another student’s coursework; stealing or buying coursework from someone else and submitting it as your own work.  Suspected plagiarism will be investigated and if found to have occurred will be dealt with failure of the course.

**All material copied or amended from any source (e.g. internet, books) must be referenced correctly according to the reference style you are using.**

**Assignment Submission Requirements**

* Source code zipped in .zip file
* A technical report to present the details of your work
* Slides for *x* minutes presentation (*x* = 5 x No. of members)
* If one of the above-mentioned requirements is missing, the student will not be allowed to do the assignment’s demonstration.

**Detailed Specification**

In this assignment (mini project), you will learn how to use ML in practice by doing an ML project that includes steps: data collection, data wrangling, exploratory data analysis, model development, model evaluation, and reporting:

* Your team may choose a use case for yourself and study it with at least 3 ML models of your choice, but you need my approval before proceeding. That project's requirements will align with the requirements for the topic below. Or,
* Your team have to do the following project to predict hospital Length of Stays following [Huynh Huy’s B. Eng thesis](https://feedu-my.sharepoint.com/:b:/g/personal/khuongna2_fe_edu_vn/ETegFYhbwwNLk_xFVCTIM0UBEsU21b2rQMkwT_lDnK5C1A?e=nFF6tQ) provided by the instructor (with [dataset available here).](https://paperswithcode.com/dataset/mimic-iii) You must read the thesis, at least to reproduce his results, write your technical report and reimplement his proposal to obtain a maximum of 8.0 points. If you can prepare your own dataset, improve his proposal, improve his results and generalize his problem setting, you will have a chance to obtain a perfect score of 10 points.

Your technical report should include:

1. **A team introduction:** A brief introduction about the project group (3-> 5 members)
2. **An introduction:** motivation, objectives, scope, challenges and solutions, report structure.
3. **Backgrounds and related works:** Present domain knowledge in relevant domain (see relevant sections in the [provided thesis](https://feedu-my.sharepoint.com/:b:/g/personal/khuongna2_fe_edu_vn/ETegFYhbwwNLk_xFVCTIM0UBEsU21b2rQMkwT_lDnK5C1A?e=0q9C1h)) and related work in approach to the problem (see relevant sections in the [provided thesis](https://feedu-my.sharepoint.com/:b:/g/personal/khuongna2_fe_edu_vn/ETegFYhbwwNLk_xFVCTIM0UBEsU21b2rQMkwT_lDnK5C1A?e=0q9C1h))
4. **Proposed model:** A detailed description of how you analyse requirements and design the model and the system. You should explain your model and data (see relevant chapters in the [provided thesis](https://feedu-my.sharepoint.com/:b:/g/personal/khuongna2_fe_edu_vn/ETegFYhbwwNLk_xFVCTIM0UBEsU21b2rQMkwT_lDnK5C1A?e=0q9C1h)).
5. **Implementation and evaluation**: Present your data preparation, data processing, experiments, metrics, and evaluations in detail.
6. **Conclusion and Discussion**: the pros and cons of the ML approach. What have you learned through the process of doing this assignment? If you had more time, how would you improve it in the future?
7. **Contribution:** Evaluate the contribution of each member during the project

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic** | **Team**  **Effort** | **Member 1** | **Member 2** | **Member …** |
| **Backgrounds of** relevant domain. | 100% | Ex: 40% | Ex: 30% | Ex: 30% |
| **Backgounds and related works of RL techniques you used** | 100% |  |  |  |
| **Your proposed model** | 100% |  |  |  |
| **Implementation and evaluation** | 100% |  |  |  |
| **Documentation (technical report, slides)** | 100% |  |  |  |

Your implementation:

* All source code must be zipped and uploaded to EduNext.
* Code comments are required

Your demonstration (*x* minutes presentation, *x* = 5 x No. of members)):

* You must briefly demonstrate your result (slide should be prepared). Prepare to answer the lecturer’s questions

|  |  |  |
| --- | --- | --- |
| **Rubrics** | | |
| **Task** | **Score** | **Condition** |
| Backgrounds of relevant domain. | 10% | Understand some basic backgrounds in relevant domain. |
| Backgounds and related works of RL techniques you used | 20% | Detail backgrounds of M techniques you used, and related works of ML approach to the problem, |
| Model design | 20% | Problem statement, model design, data preparation |
| Experiment and evaluation | 30% | Data processing, experiements, metrics, evaluation, analyse he result |
| Technical report quality and in-class presentation | 20% | Content , Structure, Clarity, Delivery, Professionalism |

* **The mini capstone can be seen as a shortened version of** the [provided thesis](https://feedu-my.sharepoint.com/:b:/g/personal/khuongna2_fe_edu_vn/ETegFYhbwwNLk_xFVCTIM0UBEsU21b2rQMkwT_lDnK5C1A?e=0q9C1h)**.**