**Đề bài (120 phút)**

*Không sử dụng tài liệu*

**Câu hỏi 1:** Giải thích ý hiểu của bọn về học có giám sát? Liệu kê 3 thuật toán học có giám sát và 3 thuật toán về học không giám sát

**Câu hỏi 2:** Tỷ lệ chia dữ liệu trong AI như thế nào? Mục đích của việc chia tập training và testing để làm gì?

**Câu hỏi 3:** Giải thích ý hiểu về SVM(Support Vector Machine)?

**Câu hỏi 4:** Phân biệt K-Means và K-Nearest. Ví dụ về cả hai?

**Câu hỏi 5:** Em hiểu thế nào về classification?

**Câu hỏi 6:** Liệt kê 5 bài toán thực tế em nghĩ có thể sử dụng linear regression, 5 bài toán thực tế sử dụng classification và 3 bài toán sử dụng clustering?

**Câu hỏi 7:** Trong một dự án về nông nghiệp, chúng ta phải đưa ra mức khuyến nghị tưới nước hợp lý trong ngày hôm đó cho cây trồng. Đầu vào chúng ta sẽ có thông tin về thời tiết hôm đấy, thời tiết dự báo của 2 ngày tới, độ ẩm của đất hiện tại, thời gian sinh trưởng của cây . Dữ liệu chúng ta có trên 1500 ngày dữ liệu về cây trồng đó. Đầu ra là số lượng nước cần nước trong ngày hôm đấy.

* Trong bài toán trên ta sẽ sử dụng mô hình machine learning nào?
* Cách phân chia tỉ lệ giữa tập train và tập test?
* Hãy nêu rõ ý tưởng và cách thực hiện bài toán trên của em?

**Test (120 mins)**

*Do not use material*

**Quiz 1:** Express your understanding about supervised learning? List 3 supervised learning and 3 unsupervised learning

**Quiz 2:** What is the data split ratio in AI? What is the purpose of splitting data into training sets and test sets?

**Quiz 3:** Explain your understanding about SVM(Support Vector Machine)?

**Quiz 4:** Distinguish K-Means and K-Nearest. Give examples of the two above?

**Quiz 5:** What is your understanding about classification?

**Quiz 6:** List 5 real problem could apply linear regression, 5 real problem could apply classification và 3 problem use clustering?

**Quiz 7:** In the project for Agriculture, we need to give advice on watering level on that day for all the plants. Our input is weather forecast on that day, weather forecast on the next two day, soil moisture at current time, plant growth time. Our data has one thousand five hundred days of that plant. Our desired output is how much water we need watering on that day.

* On that problem we should use which machine learning model?
* Ratio between training set and testing set?
* Write your idea about how to solve the problem above