**Đề bài (120 phút)**

*Không sử dụng tài liệu*

**Câu hỏi 1:** Liệt kê các thuật toán trong học có giám sát và học không giám sát?

**Câu hỏi 2:** Các cách thức nào được sử dụng để tính các điểm tương tự nhau trong Clustering? Hierarchical Clustering là gì (giải thích theo ý hiểu)?

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

**Câu hỏi 4:** Phân biệt linear và non-linear. Ví dụ về cả hai?

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

**Câu hỏi 6:** Liệt kê 5 vấn đề em nghĩ có thể sử dụng linear regression và 5 vấn đề sử dụng classification?

**Câu hỏi 7:** Trong một dự án về bệnh vùng kín, chúng ta phải nhận diện 16 bệnh lý liên quan đến vùng kín. Dữ liệu chúng ta có trên 1000 bức ảnh cho mỗi bệnh. Đầu vào của chúng ta sẽ là một ảnh chứa các bệnh cần nhận diện. Chúng ta phải trả về kết quả là những bệnh nào xuất hiện trong ảnh đó.

* 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?