

## THỰC HÀNH 5: Gaussian Mixture Model

### Hướng dẫn nộp bài:

Nội dung nộp sẽ bao gồm:

- 01 file báo cáo.
- 01 file zip chứa toàn bộ source code.

Đặt tên là **MSSV.zip** và nộp lên course.

**Bài 1:** (2 điểm) Sử dụng bộ ngữ liệu UIT-VSFC với bài toán phân loại cảm xúc dựa trên phản hồi của sinh viên, thực hiện phân đoạn từ (Word Segmentation) bằng công cụ VnCoreNLP.

**Bài 2:** (2 điểm) Sử dụng PhoBERT để trích xuất vector ngữ nghĩa của các câu phản hồi (sau khi đã được phân đoạn từ).

**Bài 3:** (6 điểm) Sử dụng mô hình Gaussian Mixture xây dựng 3 clusters cho 3 nhãn Positive, Neutral, và Negative với phương pháp tối ưu Expectation Maximization (EM). Đánh giá kết quả của mô hình trên tập test sử dụng độ đo Precision, Recall, và F1.

*Lưu ý: Chỉ sử dụng thư viện Numpy, không sử dụng các thư viện khác hỗ trợ Học máy và Học sâu.*