

1. Hoàn thành các code đính kèm trong thư mục

Có 2 assignment là phân loại email rác và phân loại nhóm tin tức (tương ứng với 2 file jupyter notebook). Học viên được yêu cầu làm hết những câu hỏi, trừ 2 câu nâng cao là tự chọn. Học viên nào muốn tìm hiểu sâu có thể làm 2 câu nâng cao.

2. Có thể cải tiến các assignment trong bài toán phân loại văn bản để thành project cuối kỳ:

Hướng 1: Áp dụng những thuật toán giới thiệu trong bài giảng, so sánh đánh giá sự hiệu quả của các thuật toán đó và đưa ra những thuật toán cải tiến mới tốt hơn cho bài toán phân loại văn bản và áp dụng cho các tập dữ liệu đã được xuất bản quốc tế.

Hướng 2: Crawl các bài báo trên các trang báo điện tử như vnexpress, dân trí (kể cả chuyên mục tương ứng) và áp dụng phân loại văn bản mới vào các mục chính của các trang báo đó, hoặc phân loại các loại email tiếng việt.