**Đề xuất DATN (Đề tài 1)**

1. Tên đề tài:  
   - Phân loại và tóm tắt nội dung văn bản (NLP).
2. Thành viên:  
   - Một thành viên thực hiện là: Trần Long Duy – 19TCLC\_DT2
3. Mô tả:
   1. **Đầu vào:** nội dung của 1 văn bản (100-2000 từ) chủ yếu là các bài post, báo chí, tin tức trong ngày.
   2. **Đầu ra:** thể loại của văn bản đó và những nội dụng quan trọng.
4. Hướng triển khai:  
   - Chia chương trình thành 2 phase: (coi như các bước thu thập dữ liệu + feature engineering là như nhau)
   1. Phase 1:  
      - Áp dụng các thuật toán ML, DL(SVM, LSTM,.) để phân loại nhãn của các bài viết.  
      - So sánh và đánh giá giữa 2-3 thuật toán.  
      - Lựa chọn thuật toán ổn định nhất.
   2. Phase 2:  
      - Sử dụng các thuật toán để trích xuất keywords.  
      - So sánh và đánh giá giữa các thuật toán.  
      - Đưa ra kết luận.