

## Mid-Term

**Lecture: Bui Thanh Hung**

**Director of Data Analytics & Artificial Intelligence Laboratory - DAAI Lab**

**Director of Artificial Intelligence & Information System Programme**

**Institute of Engineering -Technology**

**Thu Dau Mot University**

**Email: [hungbt.cntt@tdmu.edu.vn](mailto:hungbt.cntt@tdmu.edu.vn)**

**Website: <https://sites.google.com/site/hungthanhbui1980>**

### **Time Schedule**

Submit to E-learning, **24h 14/03/2021**

Demo + Answer the question: **Thứ 2 ngày 15/03/2021**

## **PROBLEM: Sentiment Analysis**

Tải dữ liệu từ foody.vn (ít nhất 200 bình luận)

Chia thành 3 lớp: Tích cực (1), Tiêu cực (-1), Trung lập (0)

Thực hiện các yêu cầu:

- Tiền xử lý
- Tách câu
- Tách từ
- Chuyển chữ hoa thành chữ thường
- Loại bỏ stopword
- Loại bỏ số
- Chuyển thành đặc trưng TF-IDF
- Chia dữ liệu theo tỉ lệ 9:1
- Huấn luyện bằng 3 mô hình học máy: SVM, Bayes, KNN
- Đánh giá bằng 3 độ đo Accuracy, F1-score, Confusion Matrix
- Chọn mô hình có kết quả cao nhất
- Đánh giá dữ liệu bằng cross-validation (K-fold=5) với mô hình có kết quả cao nhất ở trên
- Lưu model với giải thuật cho kết quả cao nhất
- Xây dựng ứng dụng: load model đã lưu, vào 1 câu bình luận, ra câu đó là Tích cực (1), Tiêu cực (-1), Trung lập (0)