

Chủ đề cho Assignment (3 tuần)

Chủ đề 1 (3 weeks): Streaming/Big data clustering (tập trung vào large/very large dataset clustering)

- Kỹ thuật: improved versions/variants k-means, BIRCH, CLARAN, etc. (key idea, mô tả kỹ thuật, đưa ra Pros/Cons, tính phổ dụng).
- Hiện thực (implementation), hoặc học và sử dụng một số well-know tools, frameworks.
- Một số ứng dụng trong việc phát hiện bất thường.
- Thực nghiệm (benchmark datasets, artificially generating datasets, e.g. HAWKS generator).
- Phân tích, đánh giá (Adjust Rand Index).

Chủ đề 2 (3 weeks): Image classification with augmentation

- Model: k -Nearest Neighbor, Neural Nets, etc (key idea, mô tả kỹ thuật, đưa ra Pros/Cons, tính phổ dụng).
- Tìm hiểu một số kỹ thuật về data augmentation.
- Một số ứng dụng trong việc phát hiện bất thường.
- Hiện thực.
- Thực nghiệm (labeled datasets, e.g. <https://lionbridge.ai/datasets/20-best-image-datasets-for-computer-vision>)
- Phân tích và đánh giá (accuracy, confusion matrix)

Nhóm: 3-4 người, tự chia, gửi lại danh sách cho tuam.tran1713810@hcmut.edu.vn (**Trần Hậu Tuấn**), **muộn nhất vào ngày 16/11/2020**

Báo cáo:

- Report (Giới thiệu, mô tả, thực nghiệm và kết quả, phân), Presentation (ngắn gọn: mô tả, kết quả thực nghiệm, phân tích).
- Oral presenting: 15min (Q&A) – *Có thể không thực hiện do việc rút ngắn thời lượng học.*