

BÀI TẬP NHÓM

1. Danh sách đề tài:

Đề 01: Phân cụm

<https://www.machinelearningplus.com/pyspark/pyspark-mllib-k-means-clustering/>

Đề 02: Tìm luật kết hợp

<https://www.kaggle.com/code/megan3/market-basket-analysis-using-pyspark>

Đề 03: Tìm độ tương đồng văn bản

<https://datascience.fm/exploring-news-article-similarity-with-pyspark/>

Đề 04: Hệ khuyến nghị

<https://medium.com/analytics-vidhya/crafting-recommendation-engine-in-pyspark-a7ca242ad40a>

2. Yêu cầu:

Nhóm hoàn thành tài liệu sau, **đặc biệt chú trọng vào việc tạo video chất lượng**.

- Tập tin word (10%)
- Slide (20%)
- Cấu trúc trình bày video: đặt vấn đề, dữ liệu, giải thích phân lập trình PySpark (Âm thanh thật sự chất lượng + hình cá nhân trình bày ở góc phải dưới). [Video mẫu](#). (70%)

3. Cách nộp bài.

Link nộp bài sẽ gửi trong group.

4. Thời hạn: 21/04/2025 (cuối ngày thứ 2)