

I. Đặt vấn đề

- Trong thời đại mạng xã hội phát triển mạnh mẽ, phân tích cảm xúc người dùng trên các nền tảng như Facebook và Twitter là yếu tố quan trọng giúp doanh nghiệp hiểu hơn về mức độ hài lòng của khách hàng đối với sản phẩm/dịch vụ.
- Case 7 yêu cầu nhóm thực hiện phân tích dữ liệu cảm xúc dựa trên các tệp "facebook_sentiment.csv" và "twitter_sentiment.csv" với thang điểm cảm xúc từ -1 (rất tiêu cực) đến 1 (rất tích cực), chỉ sử dụng thư viện NumPy, không dùng Pandas.

II. Mục tiêu

1. Ứng dụng NumPy để xử lý và phân tích dữ liệu cảm xúc dạng mảng.
2. So sánh xu hướng cảm xúc giữa Facebook và Twitter.
3. Sử dụng slicing, boolean mask, fancy indexing để phân tách và trích xuất dữ liệu.
4. Phân tích biến động cảm xúc theo thời gian bằng np.diff().
5. Tạo bảng tổng hợp kết quả và xuất file CSV.
6. Phát hiện giai đoạn có biến động cảm xúc mạnh.

III. Quy trình thực hiện

1. Đọc dữ liệu và tách cột

- Dùng np.loadtxt() và np.genfromtxt() để đọc hai file CSV.
- Tách cột time và sentiment thành mảng riêng bằng slicing.

2. Chuyển đổi kiểu dữ liệu

- Chuyển cột thời gian sang datetime64 bằng astype('datetime64[m]').
- Kiểm tra dtype, shape, lấy 5 giá trị đầu/cuối.

3. Phân tích bằng NumPy

- Slicing: lấy 20% đầu – 20% giữa – 20% cuối của dữ liệu.
- Boolean mask:
 - +sentiment > 0.5 → tích cực mạnh
 - +sentiment < -0.5 → tiêu cực mạnh
 - +còn lại → trung tính
- Broadcasting: tính chênh lệch sentiment giữa Facebook và Twitter.
- Fancy indexing: chọn top 20 bình luận tích cực / tiêu cực.
- np.diff(): xác định biến động cảm xúc theo thời gian.
- Thống kê mô tả: mean, std, median, min, max.

4. Xuất kết quả

- Tổng hợp các giá trị thống kê vào mảng NumPy.
- Xuất file kết quả: ket_qua_case_7.csv bằng np.savetxt().

IV. Kết quả phân tích (tóm tắt)

- Facebook có mức cảm xúc trung bình là:-0.-195
- Twitter có mức cảm xúc trung bình là:0.0187
- Độ lệch chuẩn cho thấy mức dao động cảm xúc của Facebook/Twitter là:Facebook:0.5849,Twitter:0.5869

- Các giai đoạn có biến động mạnh nhất ($|\Delta \text{sentiment}| > 0.6$) xuất hiện tại thời điểm:
Facebook: 469 biến động, 457 biến động
- Tỷ lệ bình luận tích cực – tiêu cực trên mỗi nền tảng:
 - +Facebook: 38.70% tích cực, 42.10% tiêu cực
 - +Twitter: 43.56% tích cực, 38.44% tiêu cực
- Chênh lệch cảm xúc giữa 2 nền tảng: Mức cảm xúc trung bình của Twitter cao hơn Facebook 0.0382

V. Kết luận

- NumPy có thể xử lý tốt dữ liệu dạng mảng mà không cần Pandas, đặc biệt khi dùng slicing, boolean mask và broadcasting.
- Twitter/Facebook có xu hướng cảm xúc khác nhau theo từng thời điểm (nhóm điển cụ thể).
- Việc sử dụng np.diff() giúp phát hiện các giai đoạn biến động mạnh trong nhận xét của người dùng.
- Dữ liệu cảm xúc mang lại thông tin hữu ích cho doanh nghiệp trong việc cải thiện sản phẩm, dịch vụ.

VI. Nhật ký đóng góp nhóm 11

Thành viên	Nhiệm vụ
Bùi Quang Huy	1. Đọc dữ liệu từ hai file facebook_sentiment.csv và twitter_sentiment.csv. 2. Tách time và sentiment thành các mảng riêng. 3. Chuyển time sang datetime64.
Vương Quốc Thái Bình	4. Kiểm tra dtype, shape; in 5 giá trị đầu/cuối. 5. Slicing 20% đầu – 20% giữa – 20% cuối. 6. Boolean mask: sentiment > 0.5, sentiment < -0.5, sentiment trung tính.
Vũ Trần Duy Anh	7. Broadcasting chênh lệch sentiment giữa Facebook và Twitter. 8. Fancy indexing lấy top-20 bình luận tích cực và tiêu cực nhất mỗi nền tảng. 9. np.diff phân tích biến động $ \Delta \text{sent} > 0.6$.

Đào Yến Nhi

10. Thống kê tổng hợp: mean, std,
median, min, max.
11. So sánh sentiment 2 nền tảng.
12. Xuất file kết quả
ket_qua_case_7.csv

Hứa Hùng Duy

Viết báo cáo, kiểm tra lỗi, chỉnh sửa, up
bài lên github