

# PHÂN TÍCH QUAN ĐIỂM ĐÁNH GIÁ CHẤT LƯỢNG DỊCH VỤ KHÁCH SẠN TRÊN NỀN TẢNG TRIPADVISOR

Môn học: **Xử Lí Ngôn Ngữ Tự Nhiên**

Giảng viên hướng dẫn: **TS. Đặng Ngọc Hoàng Thành**

Nhóm thực hiện: **Nhóm 4:**

**Đỗ Thị Mỹ Khánh**  
**Trần Nguyễn Thảo Nguyên**  
**Nguyễn Hữu Thanh**  
**Nguyễn Văn Phi Yến**

# NỘI DUNG

01 TỔNG QUAN BỘ DỮ LIỆU

02 TIỀN XỬ LÝ DỮ LIỆU

03 TRỰC QUAN HÓA

04 MÔ HÌNH HỌC MÁY

05 DEMO

06 KẾT LUẬN

# TỔNG QUAN BỘ DỮ LIỆU

	Số quan sát	Số giá trị unique	Kiểu dữ liệu
id	62887	62887	int64
language	62887	14	object
rating	62887	5	int64
additionalRatings	62887	2546	object
createdDate	62887	4659	object
helpfulVotes	62887	52	int64
username	62424	54130	object
userId	62432	56836	object
title	62885	51910	object
text	62887	62863	object
locationId	62887	1497	int64
parentGeoid	62887	185	int64
hotelName	62887	1470	object
stayDate	62887	202	object
tripType	62887	6	object

Bộ dữ liệu 'TripAdvisor Vietnam Hotel Reviews', thu thập các đánh giá của người dùng trên nền tảng du lịch trực tuyến TripAdvisor, cung cấp trải nghiệm, ý kiến và mức độ hài lòng của khách hàng khi lưu trú tại nhiều khách sạn khác nhau trên khắp Việt Nam.

**62 887 quan sát**

**14 biến quan sát**

# TIỀN XỬ LÝ DỮ LIỆU

## Chỉnh dạng dữ liệu

- Lọc các đánh giá bằng tiếng Việt
- Xóa các cột không cần thiết
- Gán nhãn '*sentiment*'
- Giảm kích thước dữ liệu

## Text mining

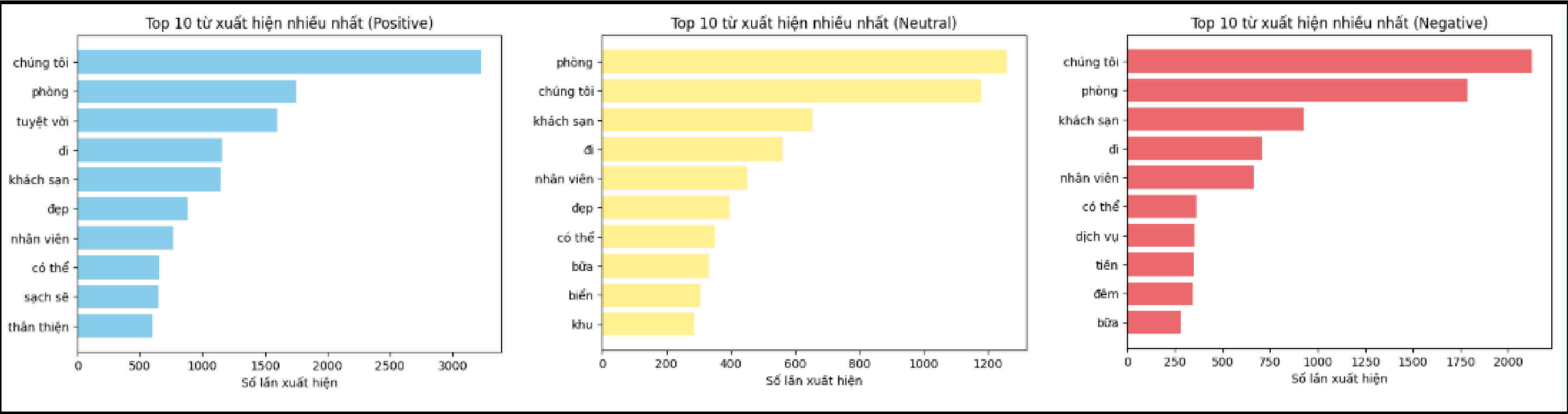
- Gộp văn bản.
- Xử lý văn bản thô.
- Xử lý từ không có trong tiếng Việt.
- Sửa lỗi chính tả, tách từ và loại bỏ stopwords.

# TIỀN XỬ LÝ DỮ LIỆU

sentiment		review
0	positive	đêm ngủ nằm tôn đức thắng khu cao_cấp đối_diện...
4	positive	công_việc tiến_triển nghĩa_đen khách_sạn đẹp n...
8	positive	tuyệt_vời nghỉ_dưỡng quảng_ngãi dịp công_tác q...
9	positive	nghỉ_dưỡng tuyệt_vời khai_trương vắng_vẻ yên_t...
10	positive	địa_điểm lý_tưởng quảng_ngãi vô_tình ghé khách...
...	...	...
61927	neutral	ồn_ào gia_đình phòng thời_gian chúng_tôi tổ_ch...
62502	neutral	chỗ ok chỗ sạch_sẽ hòa không_khí tuy_nhiên chỗ...
62516	negative	kiểm_tra đi chúng_tôi du_lịch đồng_bằng sông c...
62611	negative	tiền vì_vậy giới_thiệu kiểm_tra đối_tác sống h...
62856	neutral	thức_ăn phòng đào_tạo kém dịch_vụ khách_hàng t...
3003 rows × 2 columns		

*Kết quả minh họa sau khi xử lý dữ liệu*

# TRỰC QUAN HÓA DỮ LIỆU



*Biểu đồ thể hiện top 10 từ xuất hiện nhiều nhất theo các nhãn*

# TRỰC QUAN HÓA DỮ LIỆU

## WordCloud cho toàn bộ dữ liệu



# MÔ HÌNH HỌC MÁY



**MULTINOMIAL LOGISTIC REGRESSION**



**MULTINOMIAL NAIVE BAYES**



**SUPPORT VECTOR MACHINE**



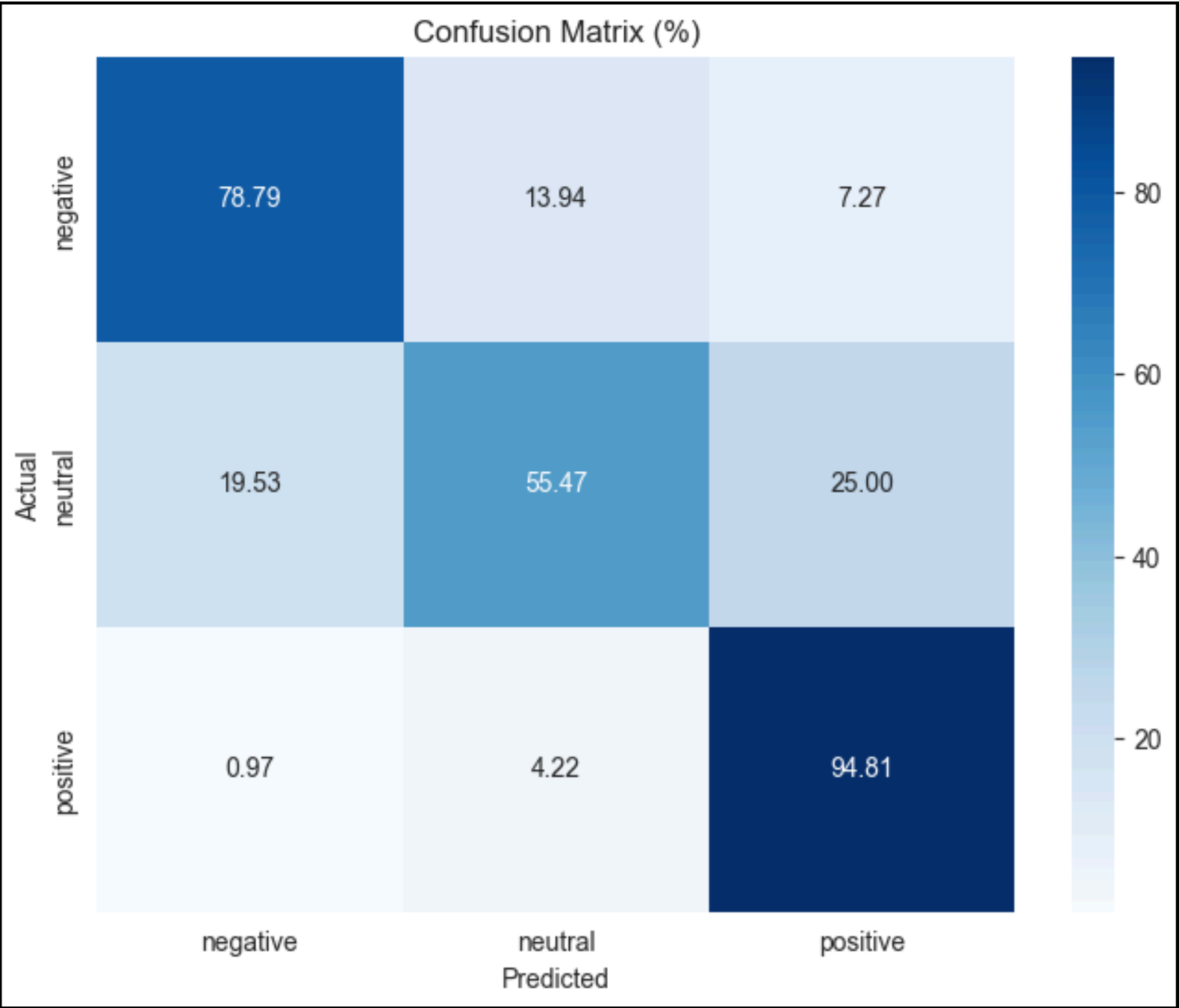
**FEEDFORWARD NEURAL NETWORK (FNN)**





# MÔ HÌNH HỌC MÁY

## MULTINOMIAL LOGISTIC REGRESSION



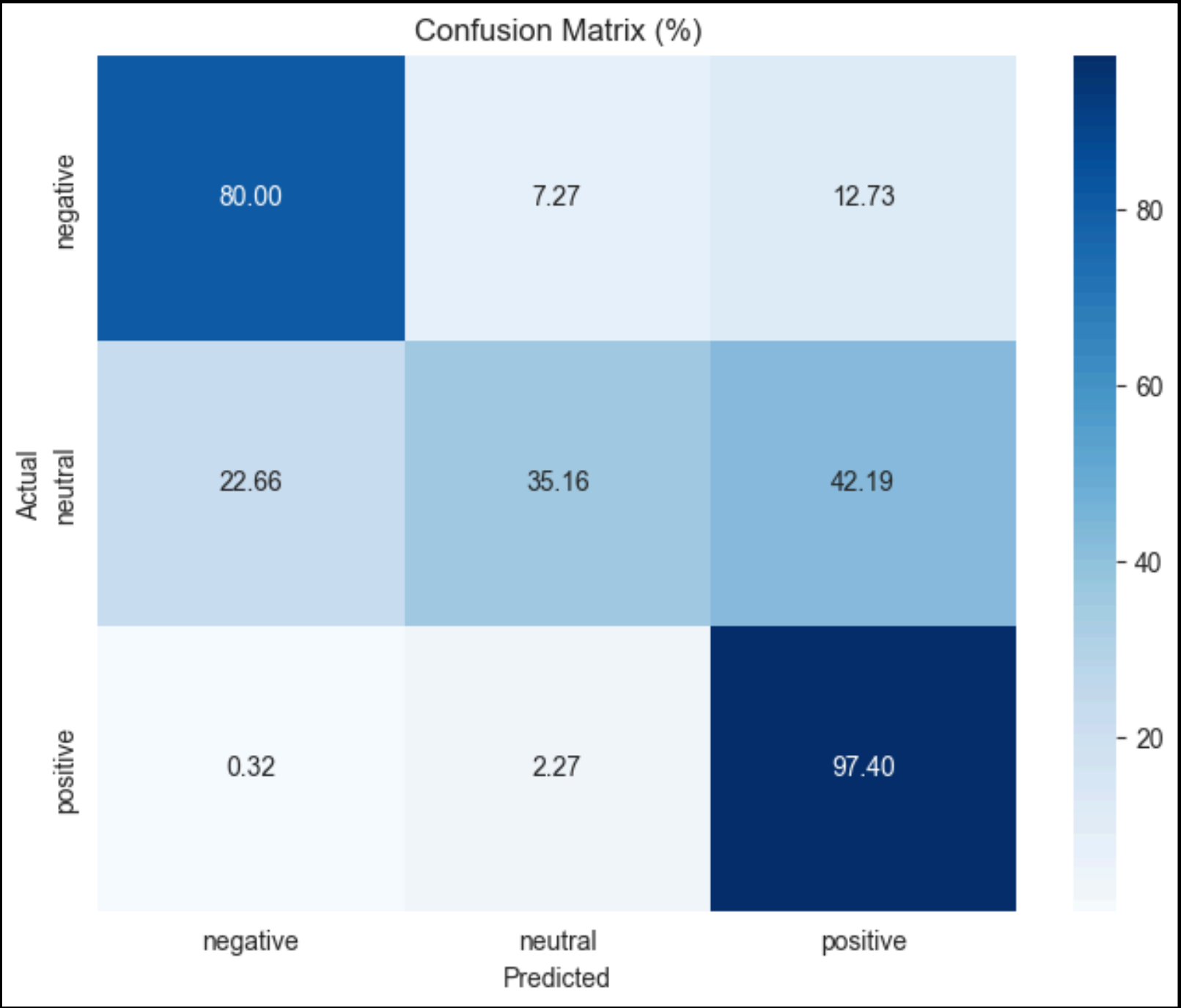
Multinomial Logistic Regression:  
Best Parameters: {'C': 10, 'max\_iter': 100, 'solver': 'lbfgs'}

	precision	recall	f1-score	support
negative	0.82	0.79	0.80	165
neutral	0.65	0.55	0.60	128
positive	0.87	0.94	0.90	308
accuracy			0.82	601
macro avg	0.78	0.76	0.77	601
weighted avg	0.81	0.82	0.81	601



# MÔ HÌNH HỌC MÁY

## MULTINOMIAL NAIVE BAYES



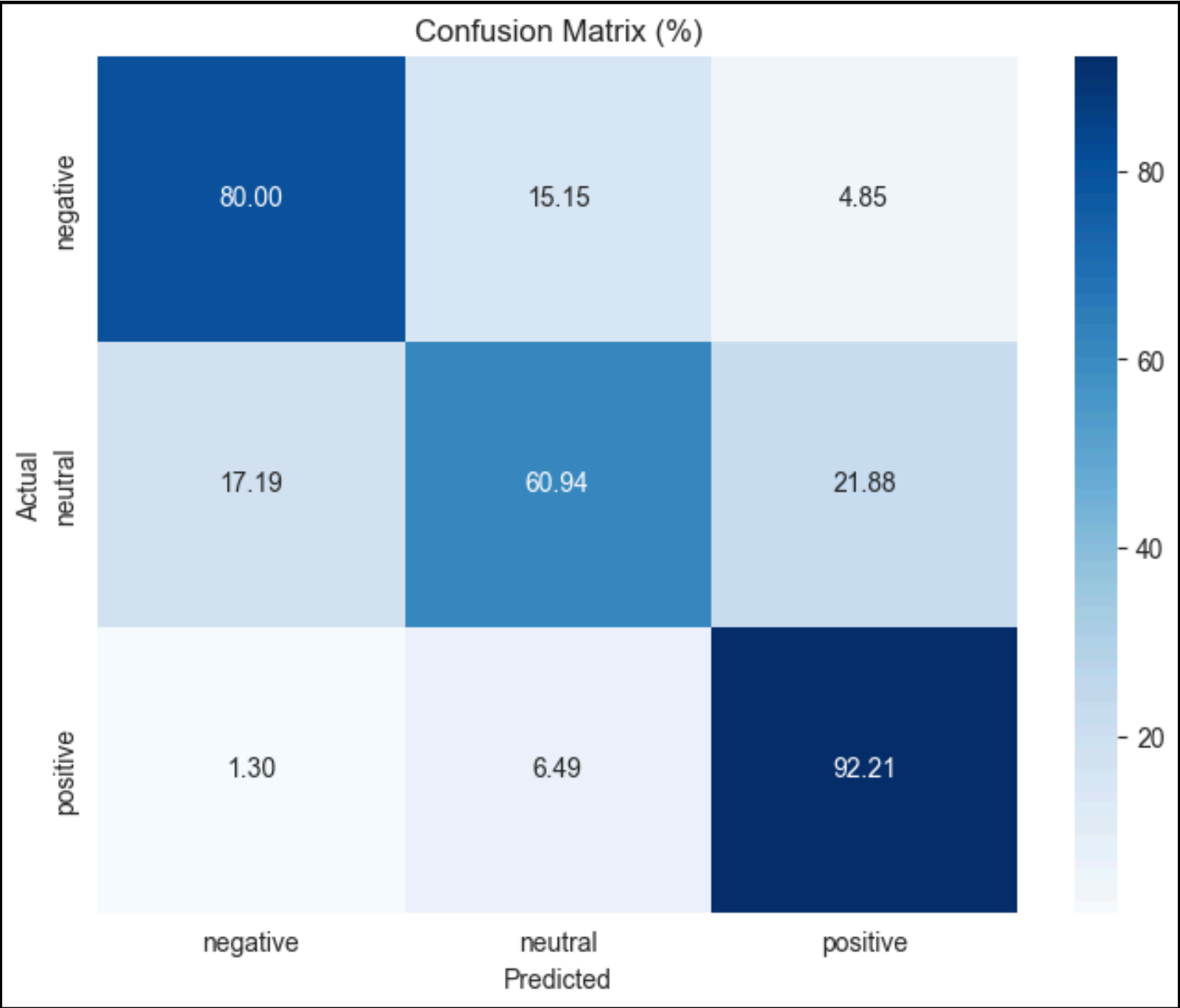
Multinomial Naive Bayes:  
Best Parameters: {'alpha': 0.1, 'fit\_prior': False}

	precision	recall	f1-score	support
negative	0.81	0.80	0.81	165
neutral	0.70	0.35	0.47	128
positive	0.80	0.97	0.88	308
accuracy			0.79	601
macro avg	0.77	0.71	0.72	601
weighted avg	0.78	0.79	0.77	601



# MÔ HÌNH HỌC MÁY

## SUPPORT VECTOR MACHINE



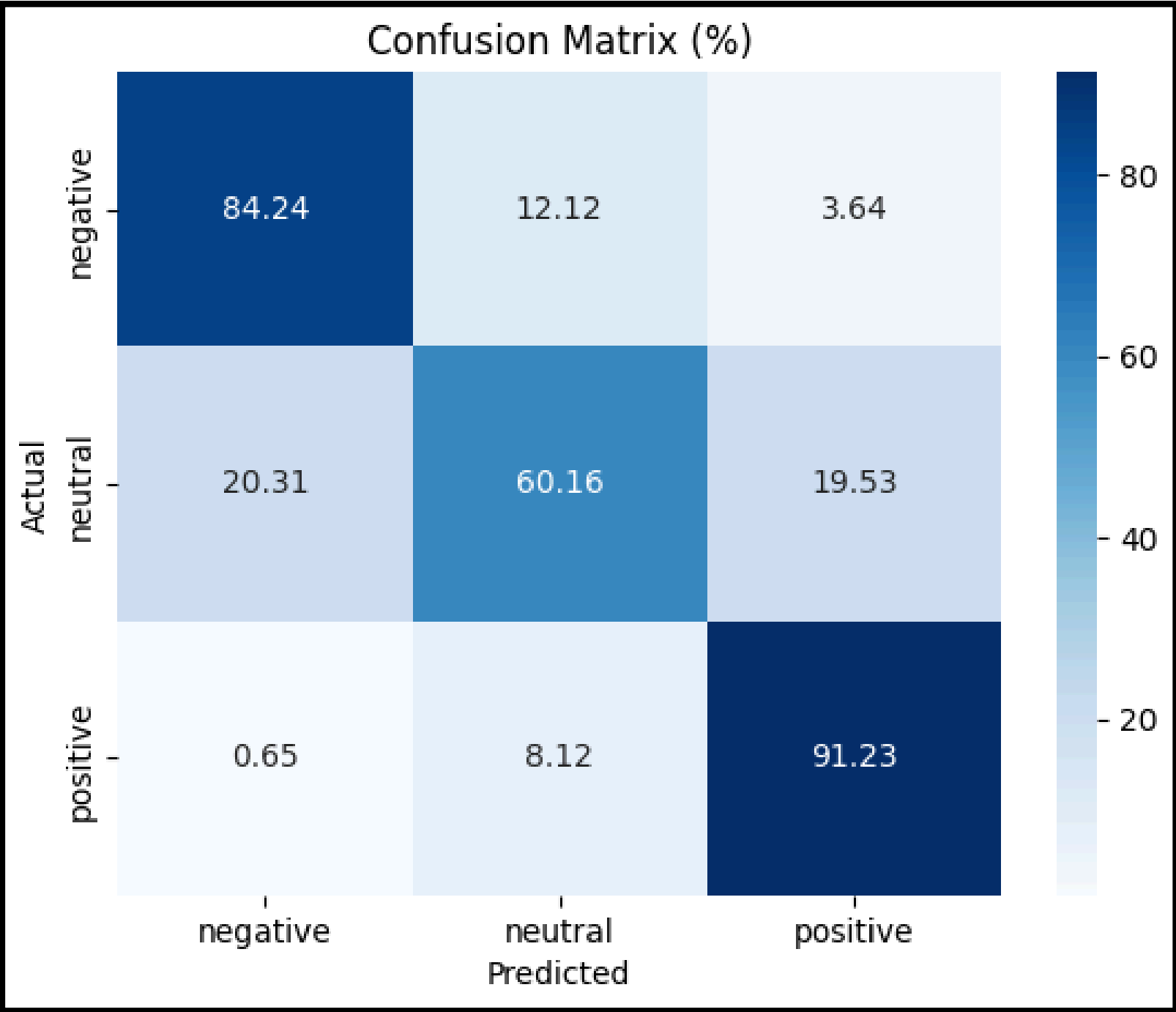
Support Vector Machine:  
Best Parameters: {'C': 1, 'kernel': 'linear'}

	precision	recall	f1-score	support
negative	0.84	0.80	0.82	165
neutral	0.63	0.61	0.62	128
positive	0.89	0.92	0.90	308
accuracy			0.82	601
macro avg	0.79	0.78	0.78	601
weighted avg	0.82	0.82	0.82	601



# MÔ HÌNH HỌC MÁY

## FEEDFORWARD NEURAL NETWORK



Feedforward Neural Network:

	precision	recall	f1-score	support
negative	0.83	0.84	0.84	165
neutral	0.63	0.60	0.62	128
positive	0.90	0.91	0.91	308
accuracy			0.83	601
macro avg	0.79	0.79	0.79	601
weighted avg	0.82	0.83	0.83	601

# MÔ HÌNH HỌC MÁY

## ĐÁNH GIÁ

FEEDFORWARD NEURAL NETWORK  
(FNN) LÀ MÔ HÌNH TỐT NHẤT

**Accuracy: 83%**

**F1 - score cao nhất**  
cho cả Positive và Negative

---

# DEMO

---

# KẾT LUẬN

## 1 TÓM TẮT KẾT QUẢ

Áp dụng được 3 mô hình học máy và một mô hình học sâu vào phân tích quan điểm đánh giá.

Mô hình FNN là mô hình phù hợp nhất đối với dữ liệu nghiên cứu và trong bối cảnh của đề tài.

## 2 HẠN CHẾ

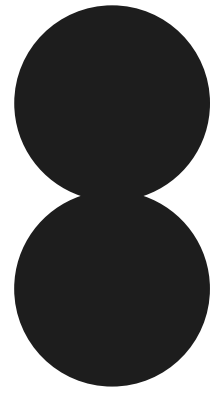
Số lượng dữ liệu hạn chế, đặc biệt là lớp Neural.

Mất cân bằng dữ liệu giữa các lớp.

## 3 HƯỚNG PHÁT TRIỂN NGHIÊN CỨU

- Tăng cường số lượng dữ liệu trung tính.
- Áp dụng phương pháp cân bằng.
- Điều chỉnh tham số.
- Thử với mô hình học sâu khác.
- Phân nhóm người dùng.





**THANKS FOR**  
**WATCHING**