BÁO CÁO TÓM TẮT KẾT QUẢ MÔ HÌNH PHÂN LOẠI CẢM XÚC

# 1. Đề bài

Xây dựng mô hình học sâu (deep learning) để phân loại cảm xúc (positive/negative) từ tập dữ liệu IMDb Movie Reviews, với 5.000 mẫu huấn luyện và 5.000 mẫu kiểm thử.  
Mục tiêu:  
- So sánh hiệu năng các mô hình với các cấu hình siêu tham số khác nhau.  
- Tối ưu hóa độ chính xác của mô hình qua việc thay đổi các siêu tham số.

# 2. Tiền xử lý dữ liệu

- Nguồn dữ liệu: tf.keras.datasets.imdb.load\_data()  
- Lấy mẫu: 5.000 mẫu huấn luyện và 5.000 mẫu kiểm thử  
- Xử lý văn bản:  
 + Xoá HTML  
 + Xoá số  
 + Xoá dấu câu  
 + Biến về chữ thường và xoá khoảng trắng dư  
- Tokenization với num\_words=10000, pad chuỗi về độ dài max\_length=200

# 3. Mô hình học sâu (Deep Learning)

- Mô hình Sequential:  
 + Embedding layer  
 + Các lớp ẩn (Dense) tùy chỉnh  
 + GlobalAveragePooling1D  
 + Dense(1, activation='sigmoid')  
- Hàm mất mát: binary\_crossentropy  
- Đánh giá: accuracy

# 4. Các cấu hình siêu tham số đã thử nghiệm

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cấu hình | hidden\_layers | units | activation | optimizer | learning\_rate | batch\_size |
| 1 | 2 | 128 | relu | adam | 0.001 | 32 |
| 2 | 2 | 256 | relu | adam | 0.0005 | 32 |
| 3 | 3 | 128 | relu | rmsprop | 0.001 | 64 |
| 4 | 2 | 256 | relu | adam | 0.0003 | 32 |
| 5 | 3 | 64 | relu | adam | 0.0001 | 64 |

# 5. Kết quả đánh giá

| Cấu hình | Độ chính xác trung bình | Độ lệch chuẩn |  
|----------|--------------------------|----------------|  
| Config 1 | 88.02% | 0.95 |  
| Config 2 | 88.35% | 1.03 |  
| Config 3 | 88.92% | 0.78 |  
| Config 4 | 89.47% (Tốt nhất) | 0.51 ✅ |  
| Config 5 | 87.84% | 1.12 |

# 6. Nhận xét

- Config 4 đạt độ chính xác cao nhất và ổn định nhất.  
- Số lượng units lớn hơn giúp mô hình học tốt hơn nếu kết hợp với learning\_rate nhỏ.  
- Adam với learning\_rate nhỏ thường cho kết quả tốt hơn RMSprop hoặc SGD.

# 7. File kết quả

- accuracy\_results.csv: chứa kết quả độ chính xác cho từng lần chạy mỗi cấu hình.  
- training\_logs.csv: lưu lại toàn bộ siêu tham số cấu hình của các mô hình đã huấn luyện.