



# PHÂN TÍCH CẢM XÚC QUA BÀI ĐĂNG TRÊN MẠNG XÃ HỘI “CNN-Sentiment-Analysis”.

Nhóm 8: Đỗ Văn Tuyên

Bùi Việt Tiến

Nguyễn Thành Trung

# Giới Thiệu

Phân tích cảm xúc là một lĩnh vực quan trọng trong xử lý ngôn ngữ tự nhiên và học máy, tập trung vào việc xác định và phân loại cảm xúc biểu hiện trong văn bản. Khi áp dụng cho các bài đăng trên mạng xã hội như Facebook, Twitter, Instagram, và nhiều nền tảng khác, phân tích cảm xúc mang lại nhiều giá trị thực tiễn và ứng dụng phong phú.

Mục đích:

Hiểu rõ khách hàng và cộng đồng

Nghiên cứu và phát triển xã hội

Giám sát bảo vệ người dùng và đưa ra phương án xử lý kịp thời tội phạm mạng



# Phương Pháp Thực Hiện

Sử dụng mạng no-ron tích chập để phân biệt giữa các ngôn ngữ trên mạng xã hội.

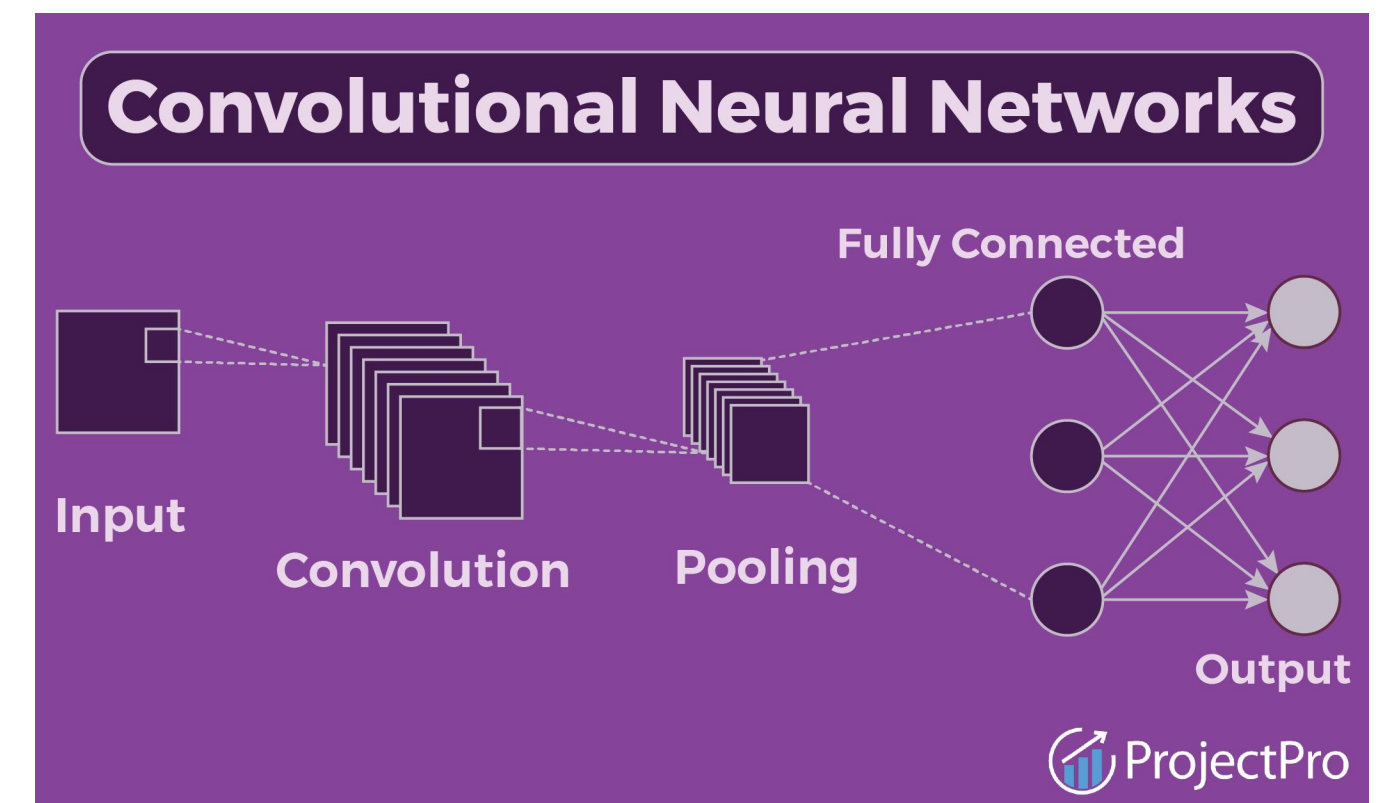
Conv2D được thiết kế để trích xuất các đặc trưng không gian từ dữ liệu đầu vào hai chiều. Nó giúp mô hình học các mẫu cục bộ và các đặc trưng không gian quan trọng trong dữ liệu.

**Kernel/Filter:** Là một ma trận nhỏ di chuyển qua dữ liệu đầu vào để thực hiện phép tích chập. Kích thước thường là 3x3, 5x5 hoặc các kích thước khác.

**Stride:** Bước nhảy của kernel qua dữ liệu đầu vào, xác định khoảng cách di chuyển của kernel trong mỗi lần tính toán.

**Padding:** Thêm các giá trị vào biên của dữ liệu đầu vào để kiểm soát kích thước đầu ra.

**Activation Function:** Hàm kích hoạt được áp dụng sau khi thực hiện phép tích chập để thêm phi tuyến tính vào mô hình. Hàm kích hoạt phổ biến là ReLU .





# MÔ HÌNH CNN 2D

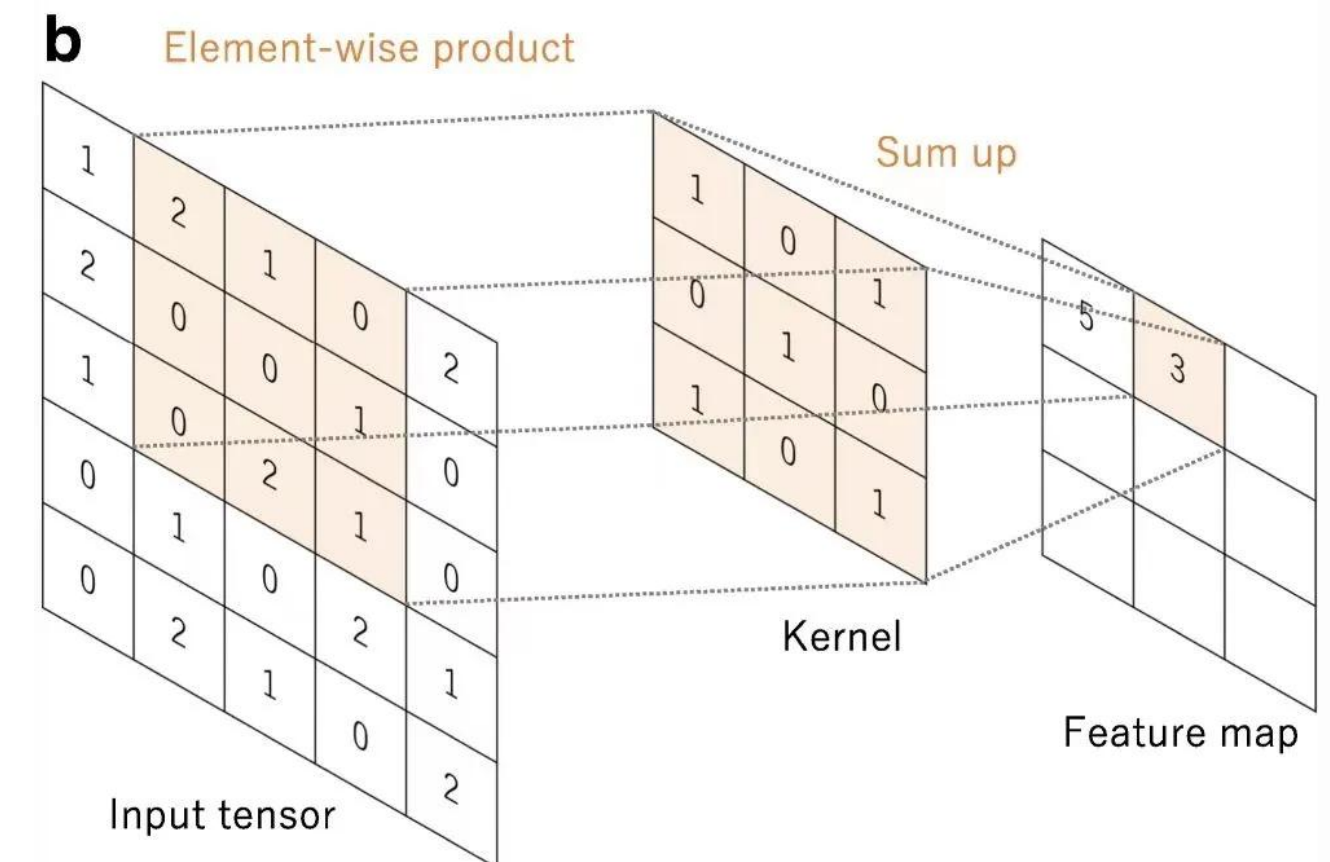
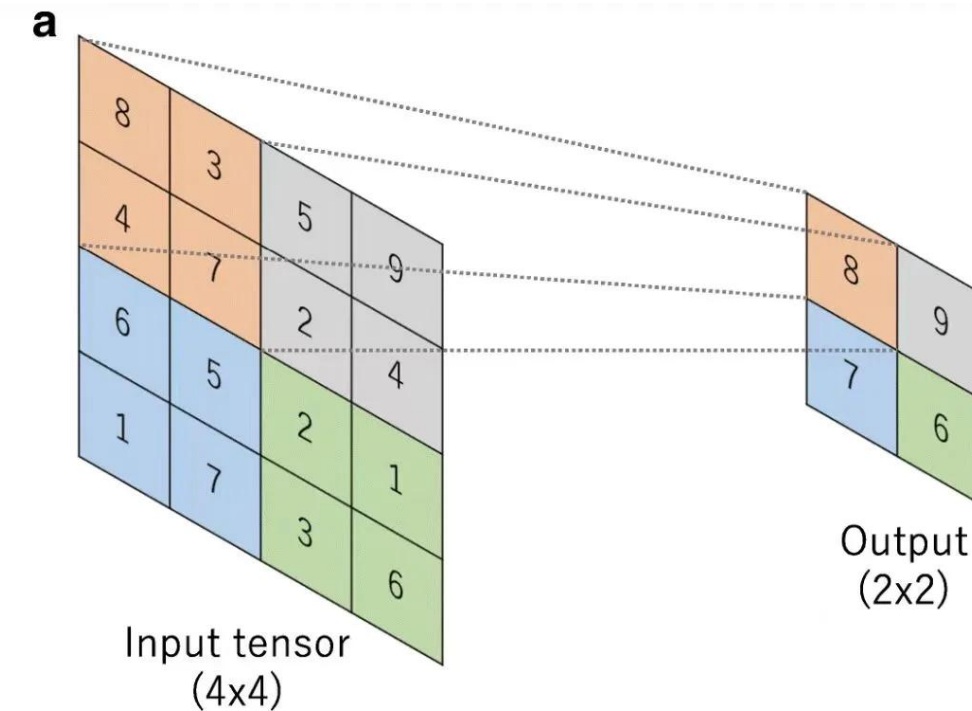
2D Convolutional Neural Networks (CNN) là một loại mạng nơ-ron nhân tạo được thiết kế đặc biệt để xử lý dữ liệu có cấu trúc dạng lưới, chẳng hạn như giọng nói hình ảnh, văn bản.

Cấu trúc cơ bản:

**Lớp Convolutional:** Đây là lớp chính của CNN, nơi các bộ lọc (filters) được áp dụng để trích xuất các đặc trưng từ hình ảnh, như cạnh, góc, hoặc các mẫu phức tạp hơn.

**Lớp Pooling:** Giảm kích thước của dữ liệu đầu ra từ lớp convolutional, giúp giảm số lượng tham số và tăng hiệu quả tính toán.

**Lớp Fully Connected:** Kết nối toàn bộ các nút từ lớp trước để đưa ra dự đoán cuối cùng.



# MÔ HÌNH CNN 2D

Cách hoạt động:

Một hình ảnh đầu vào được chia thành các vùng nhỏ.

Các bộ lọc thực hiện phép tích chập trên từng vùng để tạo ra một bản đồ đặc trưng.

Các bản đồ đặc trưng này sau đó được xử lý qua các lớp pooling và fully connected để đưa ra kết quả.

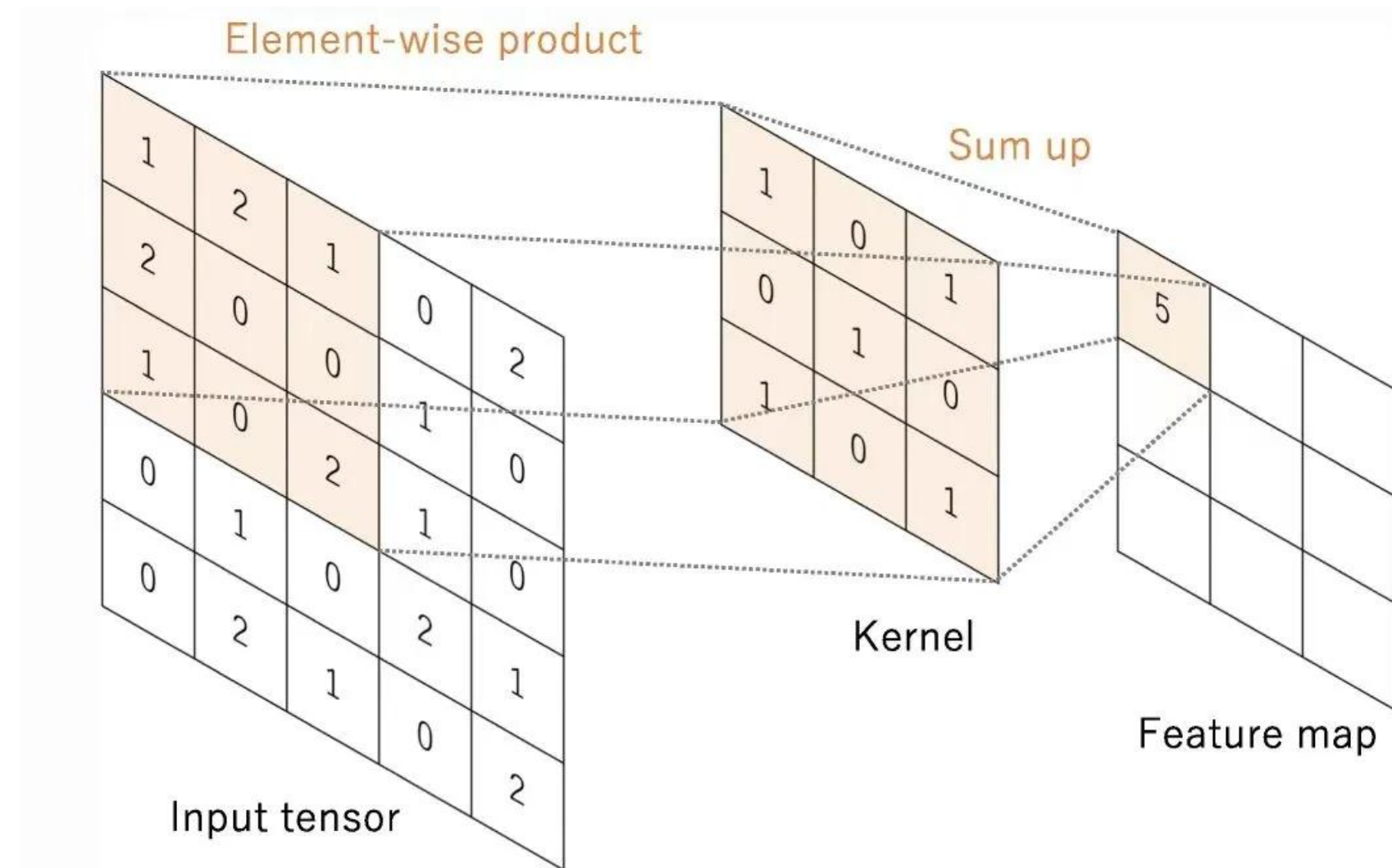
Ứng dụng:

Nhận diện hình ảnh.

Phân loại hình ảnh

Phát hiện đối tượng .

Xử lý giọng nói, phân tích văn bản..



$$\begin{aligned} & (1 \times 1) + (2 \times 0) + (1 \times 1) + (0 \times 0) + (0 \times 1) + (2 \times 0) + (1 \times 1) + (0 \times 0) + (2 \times 1) \\ &= 1 + 0 + 1 + 0 + 0 + 0 + 1 + 0 + 2 \\ &= 5 \end{aligned}$$

# TensorFlow - Framework xây dựng và huấn luyện

## *TensorFlow là gì?*

Là một mã nguồn mở khung của Google, giúp xây dựng và huấn luyện các cấu hình deep learning nhanh chóng và hiệu quả.

Lợi ích TensorFlow:

Hỗ trợ GPU/TPU giúp tăng tốc độ luyện tập.

Tích hợp Keras API , dễ sử dụng.

Cung cấp TensorBoard để giám sát quá trình huấn luyện viên.

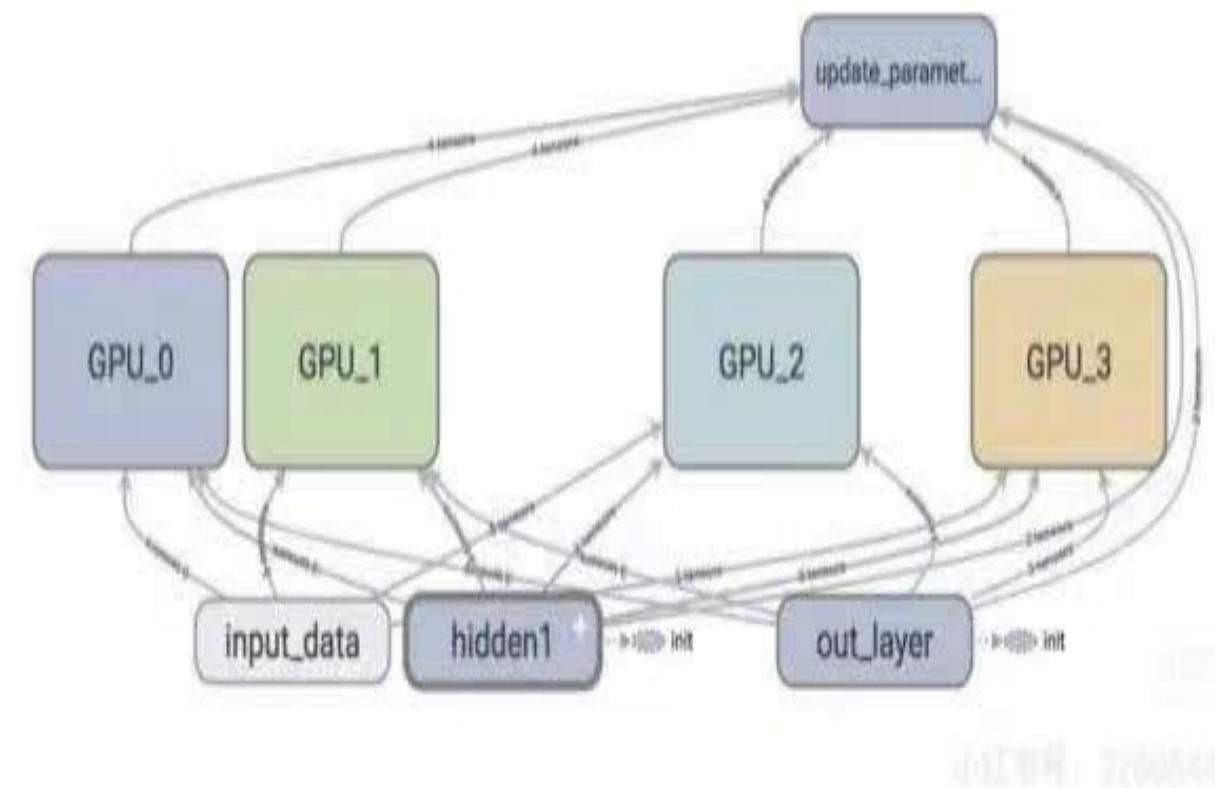
Ứng dụng trong phân tích :

Xây dựng mô hình mạng nơ-ron.

Xử lý dữ liệu tiền tệ : Thay đổi kích thước, chuẩn hóa video.

Huấn luyện & đánh giá : Sử dụng trình tối ưu hóa Adam, mất entropy chéo.

Tăng tốc độ GPU : Sử dụng `tf.device('/GPU:0')` hoặc Google Colab TPU.





## Dataset sử dụng

Bộ dữ liệu được thu thập từ các bài đăng trên mạng xã hội :

- + Bao gồm văn bản vui vẻ, không xác định, giận dữ.
- + Được sử dụng phổ biến để huấn luyện các đánh giá phát hiện khác nhau.

Bộ thu thập dữ liệu tự động:

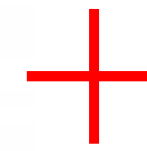
- + Tổng hợp từ nhiều nguồn với các tình huống đa dạng.

- + Giúp mô hình nhận dạng nhiều loại bạo lực thực tế hơn.





COLAB & GOOGLE DRIVER



VSCODE & PYTHON





*Thank You*