TRẢ LỜI CÂU HỎI

1. Q: *Sử dụng CNN và LSTM sau khi sử dụng BERT khác nhau chính là gì?*
   A:
   - CNN (Convolutional Neural Network): Phát hiện nhanh các mẫu từ ngữ cục bộ, cụm từ quan trọng trong văn bản ngắn, nhờ có các lớp convolution giúp thu thập các phụ thuộc cục bộ trong ngữ cảnh có kích thước cố định, các lớp pooling giúp giảm chiều dữ liệu và tập trung vào các đặc trưng nổi bật, nâng cao khả năng phát hiện các đặc trưng cục bộ quan trọng, từ đó giúp nhận diện nhanh chóng các mẫu từ ngữ có liên quan đến hate speech hoặc offensive language. Thế nên, CNN sẽ tốt cho các đoạn văn bản ngắn như tweet, bình luận ngắn.
   LSTM (Long Short-Term Memory): LSTM có khả năng ghi nhớ và hiểu ngữ cảnh dài hạn trong văn bản dài, lý do bởi LSTM được thiết kế để xử lý dữ liệu tuần tự và thu thập các phụ thuộc dài trong văn bản, có thể ghi nhớ các từ trước đó trong một chuỗi, nhờ có các cell memory giúp giữ lại thông tin qua các chuỗi dài, các cổng như forget gate, input gate và output gate giúp kiểm soát luồng thông tin, từ đó cho phép mô hình giữ lại hoặc loại bỏ thông tin khi cần, hiểu rõ ngữ cảnh và chi tiết cả văn bản. Thế nên, LSTM sẽ phù hợp hơn với các đoạn văn dài như bài viết, bình luận chi tiết.
   Tóm lại:
   CNN: Hiệu quả với văn bản ngắn, nhận diện cụm từ nhanh chóng.
   LSTM: Hiệu quả với văn bản dài, nắm bắt ngữ cảnh và mối quan hệ từ ngữ qua thời gian.
2. Q: *Sử dụng random swap làm ảnh hưởng tới cấu trúc của câu, từ đó có bị ảnh hưởng tới kết quả sau khi đưa qua BERT hay không?*
   A: Việc sử dụng random swap có thể làm ảnh hưởng đến cấu trúc của câu, từ đó có thể tác động đến kết quả sau khi xử lý bằng BERT. Bởi lẽ nếu cấu trúc câu bị thay đổi quá nhiều so với cách sắp xếp tự nhiên, trong khi BERT lại dựa vào ngữ cảnh để hiểu ý nghĩa của từ, mối quan hệ ngữ nghĩa giữa các từ có thể bị phá vỡ, khiến cho mô hình khó hiểu ý nghĩa đúng thực sự của câu.
   Ví dụ: Câu ban đầu: “He always says negative things and hurts others”
   Câu gốc: BERT dựa vào các mối quan hệ ngữ cảnh để hiểu nghĩa của từ. Trong câu gốc, BERT có thể hiểu rõ "He always says negative things" và "hurts others", nhận diện đây có thể là ngôn từ thù địch.

Random Swap: Khi các từ bị hoán đổi ngẫu nhiên, các dấu hiệu ngữ cảnh bị phá vỡ, khiến BERT khó hiểu nghĩa thực sự của câu. Câu "others hurts always and says He negative things" trở nên khó hiểu và mất ngữ cảnh gốc.

3. Q: *Lợi ích của việc sử dụng CNN và LSTM trong bài toán này khi kết hợp với BERT là gì?*

   A:

   Lợi ích khi sử dụng BERT + CNN:

   - CNN có khả năng phát hiện các mẫu cục bộ trong dữ liệu văn bản. Khi kết hợp với BERT, CNN có thể giúp phát hiện các cụm từ và từ ngữ liên quan đến ngôn từ thù địch một cách hiệu quả và nhanh chóng.

   Ví dụ: Các cụm từ như "negative words" và "hurts others" có thể được nhận diện rõ ràng hơn khi CNN được áp dụng.

   - CNN có thể xử lý văn bản ngắn một cách hiệu quả, giảm thời gian tính toán và tài nguyên cần thiết so với các mô hình phức tạp hơn, thứ đặc biệt hữu ích trong các hệ thống cần phân loại văn bản nhanh chóng, như kiểm duyệt bình luận trên mạng xã hội.

   - Cải thiện độ chính xác: Sự kết hợp này giúp tăng cường khả năng mô hình nhận diện các đặc điểm cục bộ trong văn bản, cải thiện độ chính xác tổng thể của hệ thống phát hiện ngôn từ thù địch.

   Lợi ích khi sử dụng LSTM với BERT:

   - Hiểu ngữ cảnh dài hạn: LSTM có khả năng nhớ dài hạn và duy trì ngữ cảnh trong các chuỗi văn bản. Khi kết hợp với BERT, LSTM giúp mô hình hiểu rõ hơn về ngữ cảnh của ngôn từ thù địch trong các đoạn văn dài.

   Ví dụ: Câu "He always says negative things and then hurts others" có thể được hiểu một cách toàn diện nhờ khả năng giữ ngữ cảnh của LSTM.

   - Xử lý văn bản dài: LSTM có khả năng xử lý hiệu quả các văn bản dài, duy trì ngữ cảnh và mối quan hệ giữa các từ qua nhiều câu, điều rất quan trọng trong việc phát hiện ngôn từ thù địch trong các bài viết dài, bình luận chi tiết hoặc các đoạn hội thoại.

   - Nâng cao khả năng dự đoán: Sự kết hợp này giúp mô hình có khả năng dự đoán chính xác hơn bằng cách học được các mẫu ngữ nghĩa phức tạp và các mối quan hệ dài hạn giữa các từ trong văn bản.

4. Q: *Có cần lọc các văn bản quá ngắn hoặc quá dài không?*

A: Có, cần phải lọc các văn bản quá ngắn, hoặc quá dài. Bởi nếu văn bản quá ngắn có thể không cung cấp đủ ngữ cảnh, thông tin và có thể gây nhiễu mô hình, khiến mô hình khó có thể hiểu rõ ý nghĩa và đưa ra dự đoán chính xác. Tương tự với văn bản quá dài có thể chứa quá nhiều thông tin, gây quá tải cho mô hình, ngữ cảnh dài hạn khiến cho mô hình có thể gặp khó khăn trong việc duy trì ngữ cảnh và hiểu rõ tất cả các mối quan hệ từ ngữ, từ đó làm giảm hiệu suất tính toán.

5. Q: *Sự kết hợp giữa BERT với CNN và LSTM có những ưu điểm gì so với chỉ sử dụng BERT đơn giản trong bài toán này?*
   A:
   Khả năng phát hiện mẫu cục bộ (CNN):
   Với CNN: Nhận diện các cụm từ và mẫu ngữ nghĩa ngắn hiệu quả, giúp nhanh chóng phát hiện từ ngữ liên quan đến ngôn từ thù địch.
   Chỉ với BERT đơn giản: BERT mạnh về ngữ cảnh toàn diện nhưng có thể bỏ lỡ các mẫu cục bộ ngắn.
   Khả năng nhớ dài hạn (LSTM):
   LSTM: Giữ và hiểu ngữ cảnh trong các đoạn văn dài, nâng cao khả năng dự đoán chính xác ngôn từ thù địch trong các văn bản chi tiết.
   Chỉ BERT: BERT có thể gặp khó khăn trong việc duy trì ngữ cảnh dài hạn so với LSTM.
6. Q: *Mô hình có thể phân biệt ngôn từ thù địch gián tiếp (implicit hate speech) và trực tiếp (explicit hate speech) tốt đến mức nào?*
   A:
   - Ngôn ngữ thù địch trực tiếp:
   Dễ nhận diện hơn: Ngôn ngữ thù địch trực tiếp thường bao gồm các từ ngữ xúc phạm, phân biệt chủng tộc, hoặc đe dọa "rõ ràng". Các từ khóa và cụm từ này dễ dàng được mô hình học và nhận diện.
   Ví dụ: "You are a hoe" hoặc "I hate you nigga" dễ dàng được nhận diện là ngôn ngữ thù địch trực tiếp.
   - Ngôn ngữ thù địch gián tiếp:
   Khó nhận diện hơn: Ngôn ngữ thù địch gián tiếp thường sử dụng ngôn ngữ ẩn dụ, ám chỉ hoặc các cách diễn đạt tinh vi, làm cho mô hình khó phát hiện hơn.

Cần ngữ cảnh sâu rộng: Mô hình cần hiểu rõ ngữ cảnh và mối quan hệ giữa các từ trong văn bản để xác định các biểu hiện thù địch gián tiếp.
Ví dụ: "They always get special treatment" có thể là ngôn ngữ thù địch gián tiếp nếu ngữ cảnh ám chỉ một nhóm cụ thể, chẳng hạn trong ngữ cảnh phân biệt chủng tộc, tôn giáo,...

7. Q: *Sự khác biệt giữa ngôn từ Hate và Offensive là gì? Tại sao lại cần chia thành 2 label riêng?*
   A:
   HATE: Là những lời lẽ, văn bản nhằm mục đích kích động thù hận, bạo lực hoặc phân biệt đối xử với một nhóm người dựa trên các đặc điểm như chủng tộc, tôn giáo, giới tính,... Ví dụ: "That black guy does not deserve to live".
   OFFENSIVE: Là những lời lẽ, văn bản có thể gây xúc phạm, làm tổn thương hoặc gây khó chịu cho người khác, nhưng không nhằm mục đích kích động thù hận hoặc bạo lực. Ví dụ: "You are stupid".
   Cần phải chia thành 2 label riêng bởi vì:
   - Mục đích và tác động khác nhau: Hate có tác động nghiêm trọng hơn, có thể dẫn đến bạo lực hoặc phân biệt đối xử, trong khi Offensive lại không đến mức như thế.
   - Phân loại và xử lý: Ngôn từ Hate cần phải được xử lý nghiêm ngặt hơn, có thể dẫn đến các biện pháp pháp lý hoặc hành động mạnh mẽ từ các nền tảng trực tiếp. Ngôn từ Offensive cũng cần được xử lý để duy trì môi trường giao tiếp lành mạnh, nhưng không gắt gỏng nghiêm ngặt như ngôn từ Hate
8. Q: *Nếu bình luận mang ý nghĩa tích cực nhưng trong một ngữ cảnh nào đó thì nó lại mang ý nghĩa thù ghét thì sao? Tại sao lại thay đổi ngẫu nhiên vị trí các từ trong câu trong khi điều đó có thể thay đổi ý nghĩa của câu đó?*
   A:
   Ví dụ:
   - Câu gốc: "You arte so talented and hardworking"
   - Trong ngữ cảnh tích cực, ví dụ như trong một buổi lễ trao giải, câu này mang ý nghĩa khen ngợi.
   - Trong ngữ cảnh thù ghét, ví dụ như trong một cuộc tranh cãi, với giọng điệu mỉa mai, câu này có thể mang ý nghĩa thù ghét, ám chỉ rằng người kia thông thực sự tài năng và chăm chỉ.

- Khi thay đổi ngẫu nhiên vị trí các từ: “Talented you are so and hardworking”, câu này trở nên khó hiểu và mất đi ý nghĩa khen ngợi ban đầu.

Ban đầu bộ dataset bị mất cân bằng nghiêm trọng, để cải thiện tình trạng trên, chúng em sử dụng kĩ thuật random swap để tăng số lượng và đa dạng hóa dữ liệu huấn luyện, giúp mô hình có thể xử lý tốt hơn trong các câu có cấu trúc không chuẩn hoặc bị xáo trộn, cải thiện khả năng tổng quát hóa trên dữ liệu mới, mặc dù có thể gây thay đổi hoặc mất ngữ cảnh của câu ban đầu. Thế nên, để hạn chế vấn đề đó, chúng em có kết hợp với kĩ thuật “synonym replacement”: thay để từ đồng nghĩa để bảo toàn ngữ cảnh.

9. Q: *Dựa vào đâu để nói CNN là mô hình tốt để phân loại văn bản ngắn?*
   A:

   Phát hiện mẫu cục bộ: CNN có khả năng nhận diện các cụm từ hoặc từ ngữ quan trọng trong văn bản ngắn. Điều này rất hữu ích vì các cụm từ này thường mang nhiều thông tin quan trọng.
   - Ví dụ: Trong câu "This product is amazing," CNN có thể dễ dàng nhận diện cụm từ "product is amazing" như một mẫu cục bộ quan trọng để phân loại câu này là tích cực.

   Hiệu suất tính toán cao:
   - Phép tính chập (Convolution): CNN sử dụng các phép tính chập để xử lý dữ liệu, giúp giảm thời gian tính toán và tài nguyên cần thiết. Điều này đặc biệt hữu ích khi xử lý văn bản ngắn.
   - Phép gộp (Pooling): CNN sử dụng phép gộp để giảm kích thước dữ liệu, giúp tăng tốc độ xử lý mà không làm mất đi thông tin quan trọng.

   Tính linh hoạt và khả năng tổng quát hóa: CNN có thể học được các đặc điểm quan trọng từ dữ liệu huấn luyện và áp dụng chúng vào dữ liệu mới, nơi các đặc điểm cục bộ có thể thay đổi nhưng vẫn mang ý nghĩa tương tự, từ đó giúp mô hình tổng quát hóa tốt hơn.
   Ví dụ: CNN có thể nhận diện các cụm từ tích cực như "great service" hoặc "excellent quality" trong các bình luận khác nhau và phân loại chúng là tích cực.

   Khả năng xử lý song song: CNN có khả năng xử lý song song các phần khác nhau của văn bản, giúp tăng tốc độ xử lý và cải thiện hiệu suất tổng thể.

Thế nên, CNN là một mô hình mạnh mẽ và hiệu quả trong việc phân loại văn bản ngắn nhờ vào khả năng phát hiện mẫu cục bộ, hiệu suất tính toán cao, tính linh hoạt và khả năng tổng quát hóa, cũng như khả năng xử lý song song. Những đặc điểm trên giúp CNN trở thành lựa chọn tốt cho các tác vụ phân loại văn bản ngắn.

Ngoài ra có thể tham khảo thêm ở mục [10] tài liệu tham khảo.

10. Q: *BERT chưa đủ mạnh sao, nó có những điểm yếu gì mà phải cần có CNN. CNN có những tính chất gì bổ trợ thêm cho mô hình BERT. Tương tự, LSTM có điểm mạnh gì mà BERT không có?*

A:

- Điểm yếu của BERT:
  - Khả năng phát hiện mẫu cục bộ hạn chế: BERT tập trung vào ngữ cảnh toàn diện, nhưng có thể bỏ lỡ các mẫu cục bộ quan trọng trong văn bản ngắn.
  - Khả năng duy trì ngữ cảnh dài hạn: BERT có thể gặp khó khăn trong việc duy trì ngữ cảnh dài hạn, đặc biệt trong các văn bản dài.
- Tính bổ trợ của CNN:
  - Khả năng phát hiện mẫu cục bộ: CNN có khả năng nhận diện các cụm từ và mẫu ngữ nghĩa cục bộ nhờ vào các lớp convolution, giúp bổ sung cho BERT trong việc phát hiện các đặc điểm quan trong trong văn bản ngắn.
  - Hiệu suất tính toán cao: CNN có khả năng xử lý dữ liệu nhanh chóng và hiệu quả nhờ vào các phép tích chập (convolution) và gộp (pooling), giúp giảm thời gian tính toán và tài nguyên cần thiết.
- Tính bổ trợ của LSTM:
  - Xử lý thông tin tuần tự dài hạn: LSTM giữ được thông tin từ các bước thời gian trước, giúp học được các phụ thuộc dài hạn.
  - Lưu trữ trạng thái ẩn qua các bước thời gian: Giúp duy trì thông tin qua các từ trong chuỗi, đặc biệt trong văn bản dài.

11. Q: *Dữ liệu của các em bị mất cân bằng, vậy có nên dùng các độ đo phù hợp hơn cho loại dữ liệu này không? Em có thể trình bày rõ hơn protocol của việc tăng cường dữ liệu để độ chính xác mà các em công bố là khách quan.*

A:

Em nghĩ là có, dù các độ đo như precision, recall, f1-score và Confusion matrix đã đủ để giúp đánh giá hiệu suất của mô hình một cách toàn diện, đặc biệt khi dữ liệu bị mất cân bằng, nhưng nếu muốn có góc nhìn toàn diện hơn về hiệu suất của mô hình thì ta có thể sử dụng độ đo **AUC-ROC**, giúp đo lường khả năng phân biệt giữa các lớp của mô hình mà không phụ thuộc vào ngưỡng phân loại cụ thể.

Trình bày cụ thể hơn về protocol của việc tăng cường dữ liệu:

Thay thế từ đồng nghĩa (Synonyms Replacement)

- **Cơ chế:** Thay thế một số từ trong câu gốc bằng các từ đồng nghĩa của chúng.
- **Tác dụng:** Giúp tạo ra nhiều phiên bản khác nhau của cùng một câu, tăng sự đa dạng của dữ liệu huấn luyện mà không làm thay đổi ý nghĩa cơ bản của câu.

Hoán đổi ngẫu nhiên (Random Swap)

- **Cơ chế:** Hoán đổi vị trí của các từ trong câu một cách ngẫu nhiên.
- **Tác dụng:** Tạo cấu trúc câu đa dạng, giúp mô hình linh hoạt hơn trong việc xử lý các câu có cấu trúc không quen thuộc.

Lý do chọn 2 cơ chế trên:

1. Giữ nguyên ý nghĩa:

   Thay thế từ đồng nghĩa: Việc thay thế từ đồng nghĩa không làm thay đổi ý nghĩa cơ bản của câu, vì từ đồng nghĩa có ngữ nghĩa tương tự. Điều này đảm bảo rằng ý nghĩa của câu vẫn giữ nguyên và do đó, nhãn của câu cũng không cần thay đổi.

   Hoán đổi ngẫu nhiên: Việc hoán đổi vị trí của các từ trong câu vẫn giữ nguyên các từ gốc và không thay đổi ngữ nghĩa cơ bản của câu. Điều này giúp mô hình học cách nhận diện ngữ nghĩa từ các cấu trúc câu khác nhau mà không cần thay đổi nhãn.

   Ví dụ:Giả sử câu gốc là "The cat is sitting on the mat" và có nhãn là "neutral":

   Thay thế từ đồng nghĩa:

   - Câu gốc: "The cat is sitting on the mat"
   - Câu thay thế từ đồng nghĩa: "The cat is resting on the mat"
   - Ý nghĩa vẫn giữ nguyên, nên nhãn "neutral" vẫn hợp lý.

   Hoán đổi ngẫu nhiên:

   - Câu gốc: "The cat is sitting on the mat"

- Câu hoán đổi ngẫu nhiên: "On the mat the cat is sitting"
- Ý nghĩa vẫn giữ nguyên, nên nhãn "neutral" vẫn hợp lý.

2. Tăng sự đa dạng: Giúp tạo ra nhiều câu mới từ các câu gốc, tăng sự đa dạng của dữ liệu huấn luyện, giúp mô hình học tốt hơn từ các ngữ cảnh và cấu trúc câu khác nhau.
3. Duy trì tính nhất quán: Giữ nguyên nhãn cho các câu đã được tăng cường, đảm bảo tính nhất quán trong quá trình huấn luyện mô hình.

Quy Trình Cụ Thể:

1. Chuẩn bị dữ liệu:
   - Đọc dữ liệu văn bản và nhãn tương ứng.
   - Tách riêng các câu thuộc nhãn hate và nhãn neither (những nhãn chiếm tỷ lệ nhỏ, cụ thể nhãn Hate chiếm ~5,7%, nhãn Neither chiếm ~17%).
2. Tăng cường dữ liệu cho nhãn 0 và nhãn 2:
   - Áp dụng các phương pháp synonyms replacement và random swap để tạo ra nhiều câu mới từ các câu gốc thuộc nhãn 0 và nhãn 2.
   - Ví dụ, nếu có một câu gốc thuộc nhãn 0: "The cat is sitting on the mat", bạn có thể thay thế từ "sitting" bằng "resting" hoặc hoán đổi vị trí các từ để tạo ra các câu mới.
3. Kết hợp dữ liệu: Gộp các câu mới tăng cường vào dataset ban đầu để tạo thành bộ dữ liệu huấn luyện đầy đủ.

Kết quả trước khi thực hiện tăng cường và tiền xử lý dữ liệu:

Model	Acc	Hate Speech			Offensive Language			Neither		
		P	R	F1	P	R	F1	P	R	F1
BERT + CNN	0.88	0.74	0.94	0.83	0.94	0.80	0.87	0.92	0.94	0.93
BERT + LSTM	0.89	0.77	0.94	0.85	0.94	0.82	0.88	0.92	0.94	0.93
BERT	0.86	0.72	0.90	0.80	0.89	0.80	0.84	0.89	0.90	0.89

Kết quả sau khi thực hiện tăng cường và tiền xử lý dữ liệu:

Model	Acc	Hate Speech			Offensive Language			Neither		
		P	R	F1	P	R	F1	P	R	F1
BERT + CNN	0.90	0.76	0.96	0.85	0.96	0.83	0.89	0.94	0.96	0.95
BERT + LSTM	0.92	0.79	0.97	0.87	0.97	0.85	0.91	0.95	0.97	0.96
BERT	0.88	0.74	0.93	0.82	0.92	0.82	0.87	0.91	0.93	0.92