**Chương 4. Thực nghiệm**

Bao gồm các nội dung như:

Dữ liệu thực nghiệm: trình bày về dữ liệu cho thực nghiệm (bao nhiêu cặp câu, chia ra Train bao nhiêu, dev bao nhiêu, test bao nhiêu, ..). Nếu bài báo đóng góp phần khởi tạo dữ liệu thì phải tách riêng mục này thành nội dung lớn như một chương chẳng hạn

Công cụ đánh giá: Trình bày các độ đo mà mình sử dụng để đánh giá kết quả thực nghiệm

Kết quả thực nghiệm: trình bày kết quả thực nghiệm, thường là bảng, hình

Thảo luận: Giải thích tại sao kết quả thực nghiệm lại như vậy, thường sử dụng các ví dụ trong bộ test để minh chứng cho kết quả thực nghiệm

**Chương 5. Ứng dụng minh hoạ**

Chương này có thể sáp nhập vào chương 4 nếu nó quá ngắn

Trình bày ứng dụng minh hoạ (có thể là web, windows, …) để thể hiện ứng dụng demo nhờ dựa vào các phương pháp trên.

**Chương 6. Kết luận và hướng phát triển**

Trình bày từ 2 đến 5 trang giấy, tóm tắt kết quả đạt được và dự định nghiên cứu tiếp theo mình sẽ làm gì

6.1. Kết luận

6.2. Hướng phát triển

**Tài liệu tham khảo**

Liệt kê các tài liệu tham khảo ở đây

Bảng 1 trình bày kết quả đánh giá bằng các độ đo trên của các mô hình khi được huấn luyện trên cùng một tập dữ liệu huấn luyện và kiểm tra.

● **Roberta**: Đạt hiệu suất phân loại khá kém

Tuy đây là một mô hình khá phổ biến trong xử lý ngôn ngữ tự nhiên, tuy nhiên trong trường hợp phân loại tin giả tiếng Việt đã không đạt được kết quả như mong đợi. Với Độ chính xác (Accuracy) đạt 0.741 và AUC là 0.835, thấp hơn tất cả các mô hình còn lại.

Mặc dù Precision đạt mức ổn là 0.845, cho thấy mô hình có khả năng phân loại đúng tin thật với tỉ lệ lớn. Tuy nhiên, Recall chỉ ở mức 0.604, cho thấy mô hình lại bỏ lỡ nhiều trường hợp tin giả. Điều này có nghĩa là mặc dù RoBERTa có khả năng xác định một số lượng lớn tin thật, nhưng lại không thể phát hiện được nhiều trường hợp tin giả, làm giảm độ tin cậy của mô hình trong thực tế.

● **BERT**: Đạt hiệu suất tốt hơn so với RoBERTa

BERT có độ chính xác đạt 0.787 và AUC là 0.858, cho thấy mô hình này có khả năng phân biệt giữa các loại tin tức tốt hơn RoBERTa. Precision của BERT đạt 0.850 gần giống với RoBERTa nhưng Recall của mô hình này đạt giá trị cao hơn là 0.708, , cho thấy BERT có khả năng phát hiện nhiều hơn các trường hợp tin giả so với RoBERTa.

Cùng với F1 Score đạt 0.773 cho thấy tổng quan BERT là mô hình có khả năng tốt hơn trong việc phân loại tin giả Tiếng Việt nhưng vẫn chưa phải là mô hình tốt nhất trong các mô hình đã thử nghiệm.

● **PhoBERT**: Đạt hiệu suất cao nhất với Độ chính xác 0.872 và AUC là 0.948, cho thấy khả năng phân biệt tin thật và tin giả rất xuất sắc. F1 Score là 0.879 cho thấy mô hình này cân bằng tốt giữa Precision và Recall, mặc dù hơi thấp hơn so với PhoBERT + TF-IDF.

● **PhoBERT + TF-IDF**: Đạt Độ chính xác cao nhất là 0.898 trong số các mô hình, với Precision (0.863), Recall (0.913), và AUC (0.923) đều rất tốt. Mô hình này cân bằng tốt giữa độ chính xác và khả năng phát hiện, khiến nó trở thành mô hình hiệu quả nhất cho nhiệm vụ này.