

# THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):  
<https://youtu.be/MDBoQ3-xnJQ>  
(ví dụ: <https://www.youtube.com/watch?v=AWq7uw-36Ng>)
- Link slides (dạng .pdf đặt trên Github của nhóm):  
<https://github.com/TaiTranTrongNgoc/CS2205.CH183/blob/main/Slide.pdf>  
(ví dụ: <https://github.com/mynameuit/CS2205.xxx/TenDeTai.pdf>)
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

- Họ và Tên: Trần Trọng Ngọc Tài
- MSSV: 240202011



- Lớp: CS2205.CH2023-01
- Tự đánh giá (điểm tổng kết môn): 7.5/10
- Số buổi vắng: 0
- Link Github:  
<https://github.com/TaiTranTrongNgoc/CS2205.CH183/>

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

PHÁT HIỆN TIN GIẢ BẰNG CÁCH SỬ DỤNG MÔ HÌNH BERT

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

FAKE NEWS DETECTION USING BERT MODEL

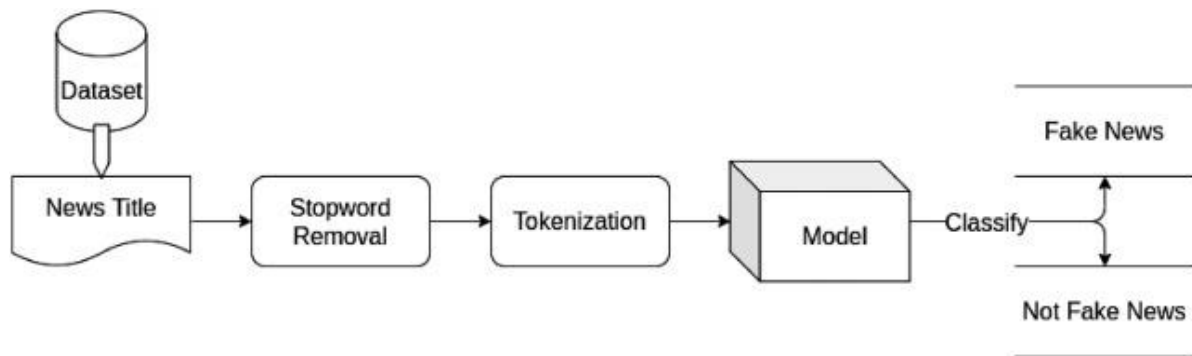
## TÓM TẮT *(Tối đa 400 từ)*

- Trong kỷ nguyên điện toán hiện đại, hệ sinh thái tin tức đã chuyển đổi từ phương tiện in ấn truyền thống cũ sang các phương tiện truyền thông xã hội. Các nền tảng truyền thông xã hội cho phép chúng ta tiêu thụ tin tức nhanh hơn nhiều, với ít hạn chế hơn về mặt biên tập dẫn đến việc lan truyền tin tức giả với tốc độ và quy mô đáng kinh ngạc. Trong nghiên cứu gần đây, nhiều phương pháp hữu ích để phát hiện tin tức giả sử dụng mạng nơ-ron tuần tự để mã hóa nội dung tin tức và thông tin cấp ngữ cảnh xã hội trong đó chuỗi văn bản được phân tích theo cách đơn hướng. Hai tập dữ liệu tin tức giả mới được giới thiệu, một tập thu thập thông qua cộng đồng và bao gồm sáu miền tin tức, và một tập khác thu thập thông qua web bao gồm những người nổi tiếng. Các mô hình hoạt động tốt nhất của chúng tôi đạt được độ chính xác tương đương với khả năng phát hiện nội dung giả của con người. Phương pháp đào tạo hai chiều là ưu tiên để mô hình hóa thông tin có liên quan của tin tức giả có khả năng cải thiện hiệu suất phân loại với khả năng nắm bắt các phụ thuộc ngữ nghĩa và khoảng cách xa trong các câu. Trong nghiên cứu này, chúng tôi đề xuất mô hình BERT-based (Bidirectional Encoder Representations from Transformers) và mô hình học sâu FakeBERT bằng cách kết hợp các khối song song khác nhau của Mạng nơ-ron Convolutional Neural Network (CNN) có kích thước hạt nhân và bộ lọc khác nhau với BERT.

## GIỚI THIỆU *(Tối đa 1 trang A4)*

- Nhiều người tin vào bất cứ điều gì họ đọc trên internet và những người nghiệp dư hoặc mới làm quen với công nghệ số có thể dễ dàng bị lừa. Chỉ khi đó, người

ta mới có thể xem xét các kỹ thuật và lĩnh vực khác nhau của máy học (ML), xử lý ngôn ngữ tự nhiên (NLP) và trí tuệ nhân tạo (AI) có thể giúp chúng ta chống lại tình trạng này. Tin giả trên mạng xã hội và nhiều phương tiện truyền thông khác đang lan truyền rộng rãi và là vấn đề đáng quan ngại nghiêm trọng do khả năng gây ra nhiều thiệt hại về mặt xã hội và quốc gia với những tác động tàn phá. Bài báo này phân tích nghiên cứu liên quan đến phát hiện tin giả và khám phá các mô hình học máy truyền thống để lựa chọn mô hình tốt nhất, nhằm tạo ra mô hình sản phẩm có thuật toán học máy có giám sát, có thể phân loại tin giả thành đúng hoặc sai, bằng cách sử dụng các công cụ như python scikit-learn, NLP để phân tích văn bản.



Biểu đồ quy trình làm việc của mô hình BERT

## MỤC TIÊU *(Viết trong vòng 3 mục tiêu)*

- Đề xuất mô hình (FakeBERT) kết hợp giữa BERT và mạng neural tích chập CNN để đối phó với sự lan truyền nhanh chóng của tin giả.
- Cải thiện độ chính xác trong việc phân loại tin thật/tin giả bằng cách phân tích ngữ cảnh hai chiều của văn bản.
- Giảm thiểu tác động tiêu cực của tin giả đến xã hội.

## NỘI DUNG VÀ PHƯƠNG PHÁP

### *Nội dung:*

Mô hình học sâu dựa trên BERT (FakeBERT) bằng cách kết hợp các khối song song khác nhau của CNN một lớp với Bidirectional Encoder Representations from Transformers (BERT). Chúng tôi sử dụng BERT như một bộ mã hóa câu, có thể lấy

chính xác biểu diễn ngữ cảnh của một câu. Công trình này trái ngược với các công trình nghiên cứu trước đây, trong đó các nhà nghiên cứu xem xét chuỗi văn bản theo cách đơn hướng (từ trái sang phải hoặc từ phải sang trái để đào tạo trước). Nhiều phương pháp hiện có và hữu ích đã được trình bày với mạng nơ-ron tuần tự để mã hóa thông tin có liên quan. Chúng tôi sử dụng các mô hình sau: Logistic Regression, Decision Tree, Gradient Booster, Random Forests, BERT.

### ***Phương Pháp***

- Tổng hợp các kỹ thuật cùng với tập hợp đặc điểm Ngôn ngữ học và Đếm từ (LIWC). Liên kết với các trang web khác như PolitiFact và Snopes được sử dụng để kiểm tra thông tin tiềm năng có thể giúp chống lại sự lan truyền tin tức sai lệch.
- Chọn ba tập dữ liệu cho các thí nghiệm, trong đó có tin tức từ các lĩnh vực như chính trị, giải trí, công nghệ và thể thao và chứa hỗn hợp cả các bài báo trung thực và giả mạo. Kết hợp các tập dữ liệu có sẵn trực tuyến và được trích xuất từ World Wide Web. Ngữ liệu thu thập được từ được xử lý trước khi được sử dụng làm đầu vào để đào tạo các mô hình.
- Các bài viết nhiều cột được chuyển đổi thành các bài viết một cột để thống nhất về định dạng và cấu trúc. Sau khi các thuộc tính có liên quan được chọn sau giai đoạn làm sạch và khám phá dữ liệu, bước tiếp theo là trích xuất các đặc điểm ngôn ngữ.
- Các đặc điểm ngôn ngữ liên quan đến một số đặc điểm văn bản được chuyển đổi thành dạng số để chúng có thể được sử dụng làm đầu vào cho các mô hình đào tạo.
- Để thực hiện trích xuất các tính năng từ ngữ liệu, chúng tôi đã sử dụng công cụ LIWC2015 để phân loại văn bản thành các biến rời rạc và liên tục khác nhau, trích xuất 93 tính năng khác nhau từ bất kỳ văn bản nào. Mỗi tập dữ liệu được chia thành tập huấn luyện và tập kiểm tra với tỷ lệ phân chia tương ứng là 70/30.
- Các bài viết được xáo trộn để đảm bảo phân bố công bằng các bài viết giả và thật trong các trường hợp huấn luyện và kiểm tra.

- Các thuật toán học được đào tạo với các siêu tham số khác nhau để đạt được độ chính xác tối đa cho một tập dữ liệu nhất định, với sự cân bằng tối ưu giữa phương sai và độ lệch.
- Mỗi mô hình được đào tạo nhiều lần với một tập hợp các tham số khác nhau bằng cách sử dụng tìm kiếm lưới để tối ưu hóa mô hình để có kết quả tốt nhất. Sử dụng tìm kiếm lưới để tìm các tham số tốt nhất tốn kém về mặt tính toán.
- Đánh giá hiệu suất trên nhiều tập dữ liệu bằng tổng hợp khác nhau như phân loại bagging, increasing và voting bao gồm ba mô hình học: hồi quy logistic, random forest và KNN, trong khi phân loại vote thứ hai bao gồm hồi quy logistic, linear SVM và cây phân loại và hồi quy (CART). trong khi hai thuật toán tập hợp boost được sử dụng, XGBoost và AdaBoost. Mô hình xác thực chéo k-fold ( $k = 10$ ) được sử dụng cho tất cả các mô hình học tập thể.
- Nghiên cứu thêm về hiện tượng "echo chambers" (những nhóm người có cùng quan điểm) trên mạng xã hội
- Xem xét mối quan hệ giữa người dùng trong cấu trúc đồ thị thay vì đơn lẻ

## KẾT QUẢ MONG ĐỢI

- Trích xuất 93 tính năng khác nhau từ bất kỳ văn bản nào để phát hiện tin giả.
- Mô hình được đề xuất có thể xác định chính xác tin giả và phù hợp để phát hiện tin tức trong các tình huống phức tạp, với độ chính xác cao nhất đạt 98,90%.
- Đề xuất hướng nghiên cứu tiếp theo về hiện tượng “echo chambers” trên mạng xã hội

## TÀI LIỆU THAM KHẢO (Định dạng DBLP)

- [1] Crestani F, Rosso P (2020) The role of personality and linguistic patterns in discriminating between fake news spreaders and fact-checkers.” In Natural language processing and information systems: 25th international conference on applications of natural language to information systems, NLDB 2020, Saarbrücken, Germany. Proceedings, vol 181. Springer Nature
- [2] Alkhodair S A, Ding S H.H, Fung B C M and Liu J 2020 Detecting breaking news

- rumors of emerging topics in social media” *Inf. Process. Manag.* 57 102018 2020 [3]
- Kaur Prabhjot et al 2019 Hybrid Text Classification Method for Fake News Detection *Inf.” International Journal of Engineering and Advanced Technology (IJEAT)* 2388-2392
- [4] Bondielli A, Marcelloni F (2019) A survey on fake news and rumor detection techniques.” *Inform Sci* 497:38–55
- [5] Chen W, Zhang Y, Yeo CK, Lau CT, Sung Lee B (2018) Unsupervised rumor detection based on users’ behaviors using neural networks.” *Pattern*
- [6] De S, Sohan FY, Mukherjee A (2018) Attending sentences to detect satirical fake news. ” In: *Proceedings of the 27th international conference on computational linguistics*, pp 3371–3380
- [7] Ahmed H, Traore I, Saad S (2017) Detection of online fake news using N-gram analysis and machine learning techniques. systems.” : *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*. Springer, Cham, pp 127–138
- [8] Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election.” *Econ Perspect* 31(2):211–36
- [9] Granik Mykhailo and Mesyura Volodymyr 2017 First Ukraine Conference on Electrical and Computer Engineering (UKRCON) (Ukraine: IEEE) Fake news detection using naive Bayes classifier” *Journal of Computational and Theoretical Nanoscience.*, 12. 6334-6342. 10.1166/jctn.2015.4675.
- [10] Greff K, Srivastava RK, Koutn’ík J, Steunebrink BR, Schmidhuber J (2016) LSTM” *IEEE Trans Neural Netw Learn Syst* 28(10):2222–2232, [7] De S, Sohan FY, Mukherjee A (2018) Attending sentences to detect satirical fake news. ” In: *Proceedings of the 27th international conference on computational linguistics*, pp 3371–3380