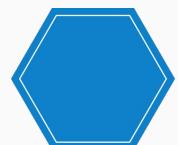
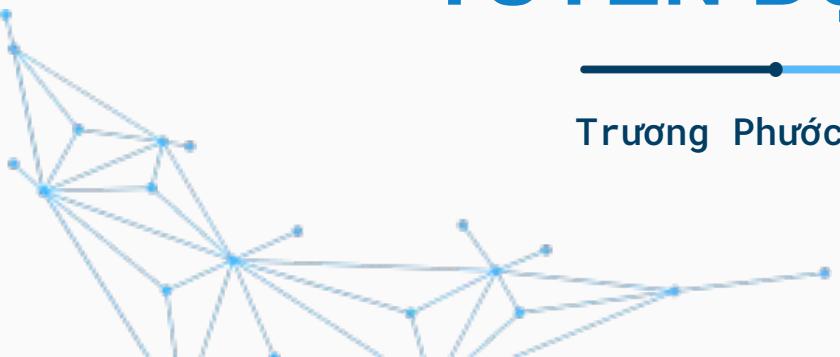




# PHÂN LOẠI VÀ XÁC ĐỊNH TỰ ĐỘNG YẾU TỐ BẤT THƯỜNG TRONG TIN TUYỂN DỤNG TIẾNG VIỆT

---

Trương Phước Bảo Khanh - 250101033  
CS2205.RM



# Tóm tắt

---



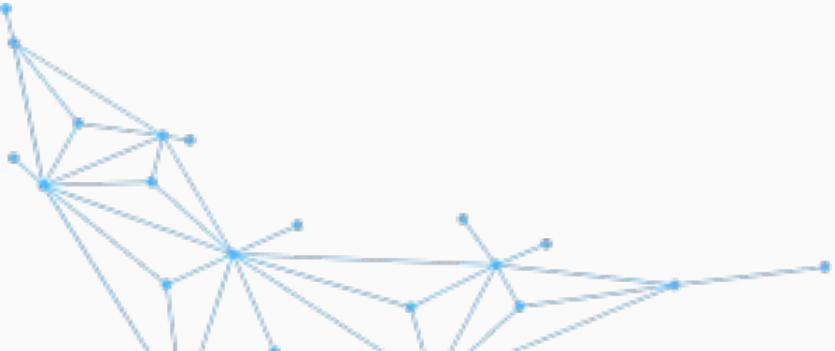
Lớp: CS2205.RM



Trương Phước Bảo Khanh - 250101033



[https://github.com/khanhtrphuocbao  
/Research-Methodology.git](https://github.com/khanhtrphuocbao/Research-Methodology.git)



# Giới thiệu

## LÝ DO THỰC HIỆN

- Sự bùng nổ của nhu cầu tuyển dụng trực tuyến -> Sự gia tăng của các tin tuyển dụng lừa đảo.
  - Những hậu quả:
    - Đánh cắp thông tin cá nhân
    - Lừa đảo, chiếm đoạt tín dụng, tài sản
    - Tổn hại đến uy tín các công ty tuyển dụng bị mạo danh
- > ViReCAX (Vietnamese RECruitment Abnormal Explanation): Bộ dữ liệu phát hiện tin tuyển dụng bất thường kèm phân tích mức độ xác thực và câu giải thích.

## TÍNH ỨNG DỤNG

- Giảm thiểu tình trạng lừa đảo tuyển dụng.
- Công cụ hỗ trợ và đánh giá tin tuyển dụng trực tuyến.

## CÁC THÁCH THỨC

- Ít các nghiên cứu liên quan.
- Ít các tiêu chuẩn đánh giá tin tuyển dụng.
- Thời gian gán nhãn.



# Mục tiêu

Xây dựng một bộ dữ liệu nhận biết các tin tuyển dụng bất thường ở mức độ khía cạnh giúp nhận biết và ngăn chặn các tin tuyển dụng có khả năng lừa đảo trên các cổng thông tin tìm việc làm ở Việt Nam.

→ Nhằm bảo vệ quyền lợi của người tìm việc và người tuyển dụng, cũng như cung cấp các hỗ trợ để đưa ra quyết định chính xác hơn.

Gồm có ba mục tiêu nhỏ hơn:

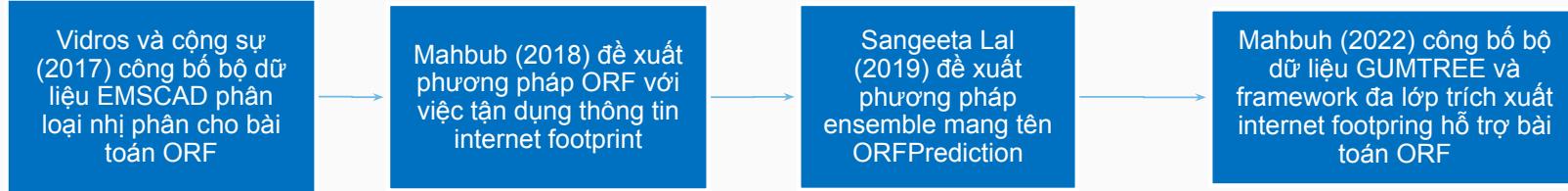
1. Xây dựng một quy trình gán nhãn cụ thể cho bài toán.
2. Xây dựng bộ dữ liệu về các tin tuyển dụng bất thường, có khả năng lừa đảo cho tiếng Việt.
3. Xây dựng các mô hình baseline và phân tích lỗi trên bộ dữ liệu.



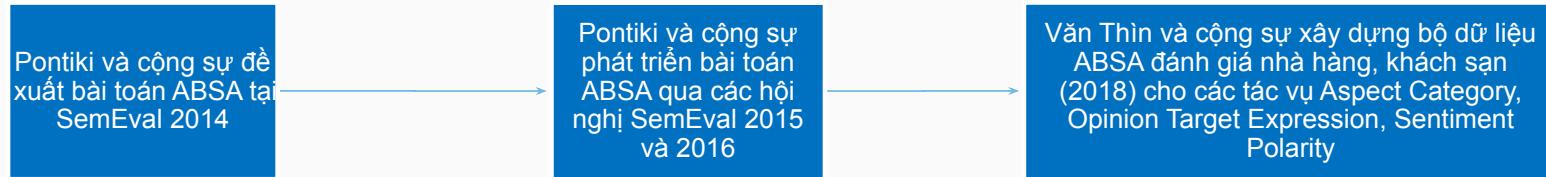
# Nội dung và Phương pháp

## Nội dung 1: Các công trình nghiên cứu liên quan

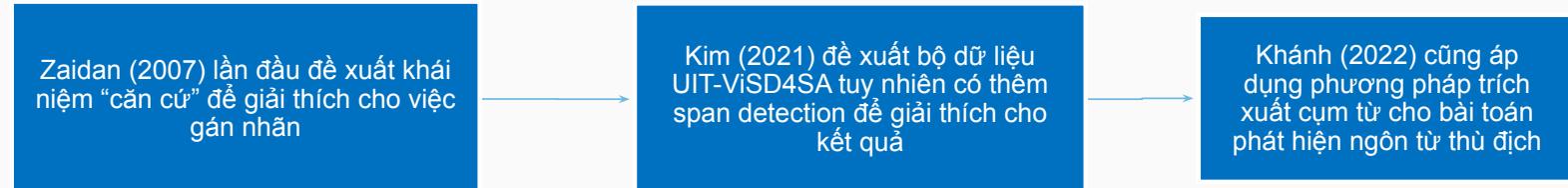
### VỀ BÀI TOÁN PHÁT HIỆN TIN TUYỀN DỤNG LỪA ĐẢO



### VỀ BÀI TOÁN PHÂN TÍCH CẢM XÚC DỰA TRÊN DANH MỤC KHÍA CẠNH



### VỀ BÀI TOÁN PHÁT HIỆN VỚI KHẢ NĂNG GIẢI THÍCH



# Nội dung và Phương pháp

## Nội dung 2: Xác định bài toán

### Đầu vào

Tuyển nhân viên đi làm ngay không cần kinh nghiệm bằng cấp  
có theo ca

7,8 triệu - 13,6 triệu/tháng  Tín dụng, lương tối thiểu  Kinh phí ngày  Việt Nam

TP. Thủ Đức - Quận 9, TP.HCM

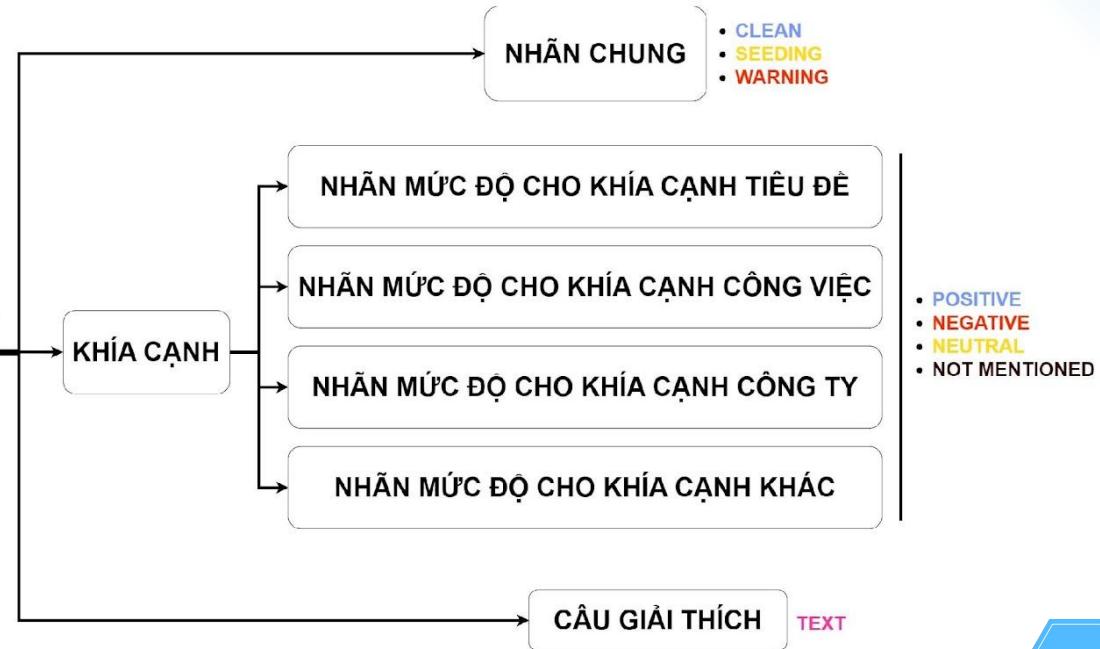
Ngày đăng: Hôm nay - Mã tin: 69376440

Việc làm hàng ngày : Sắp xếp hàng hóa, tách hàng  
Soạn hàng hóa, kiểm đếm hàng tồn  
Kiểm tra số lượng hàng tồn  
Đóng gói theo mẫu mã khách tron .  
\*THỜI GIAN LÀM VIỆC:  
-Sáng 7h30-11h30  
-Chiều 13h00-16h30  
-Làm từ thứ 2 đến thứ 7 (chủ nhật được nghỉ)  
Lương : 2.5triệu/t 13triệu 000  
YÊU CẦU:  
-Nhanh nhẹn, có trách nhiệm với công việc  
-Nam/Nữ từ 18 đến 55 tuổi có sức khỏe,vui vẻ, hòa đồng.  
\*QUYỀN LỢI:  
-Được hưởng Bảo hiểm đầy đủ quyền lợi (Theo Luật nhà nước.  
-Chỗ để phụ chuyển cẩn thận  
\*LIÊN HỆ: Trực tiếp bách tuyển  để trao đổi nhận việc làm  
ngày gửi Zalo cũng được,không nhận tin nhắn !  
-TUYỂN TRỰC TIẾP KHÔNG QUA TRUNG GIAN, KHÔNG THU PHÍ  
Liên hệ: 0337\*\*\*\*\*  Hiện số

Thông tin cơ bản

|                                     |                                 |
|-------------------------------------|---------------------------------|
| Nhà tuyển dụng: Cá nhân đăng tuyển  | Ngành nghề: Bán hàng            |
| Lý do tìm công việc: Toan thời gian | Hình thức trả lương: Theo tháng |
| Lương tối thiểu: 7.800.000 đ        | Lương tối đa: 13.600.000 đ      |
| Kinh nghiệm: < 1 năm                | Giới tính: Không yêu cầu        |
| Số lượng tuyển dụng: 5              | Độ tuổi: 18 - 56 tuổi           |
| Học vấn tối thiểu: Trung học        |                                 |

### Đầu ra

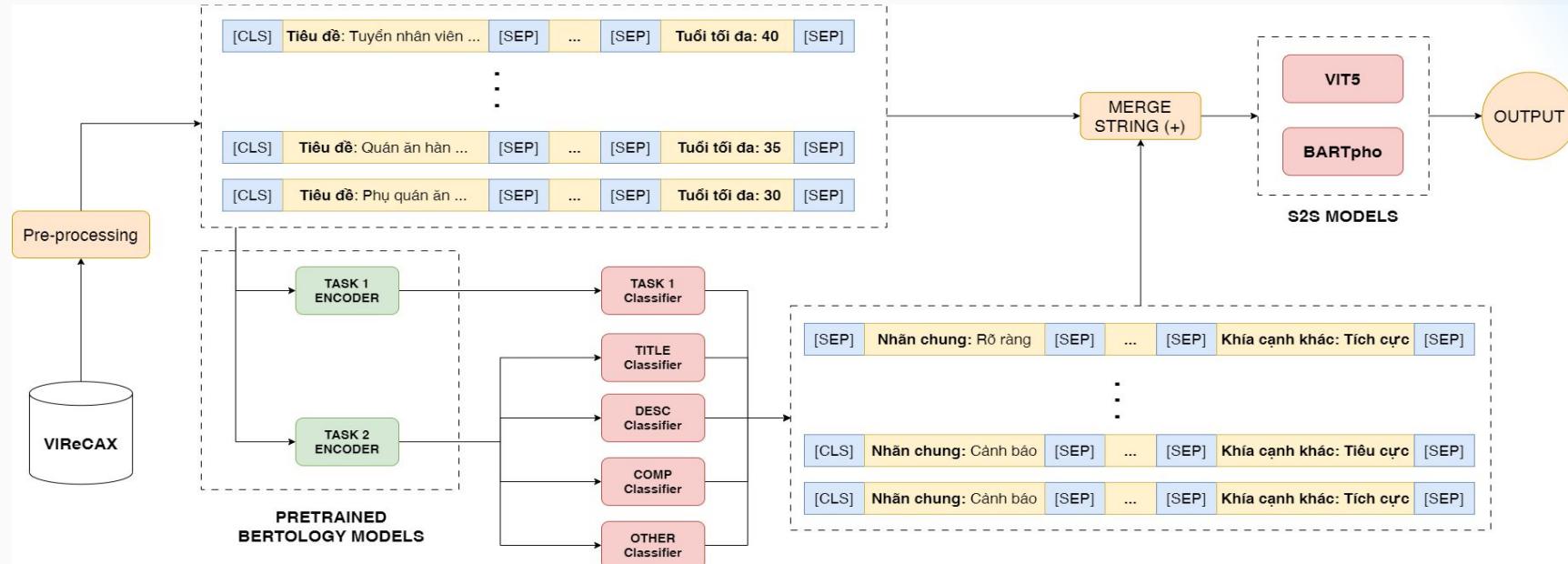


Hình 1. Định nghĩa bài toán phát hiện bất thường trong tin tuyển dụng.



# Nội dung và Phương pháp

## Nội dung 3: Phương pháp tiếp cận



- Mô hình kết hợp giữa các phương pháp BERT-based và LSTM hoặc CNN cho tác vụ 1 và tác vụ 2.
- Mô hình seq2seq cho tác vụ 3.

# Kết quả dự kiến

- Xây dựng guideline gán nhãn cho các mẫu tin tuyển dụng lừa đảo, cần cảnh giác có thể tham khảo.
- Xây dựng bộ dữ liệu đánh giá khía cạnh tin tuyển dụng với tối thiểu 10.000 mẫu tin.
- Xây dựng các mô hình baseline (gồm hướng tiếp cận đã nêu và thử nghiệm thêm các phương pháp XAI mới như LIME, SHAP, ANCHOR) trên tập dữ liệu đề xuất.



# Tài liệu tham khảo

---

1. Stefanos Vidros, Constantinos Koliас, Georgios Kambourakis: Online recruitment services: another playground for fraudsters. Comput. Fraud Secur. 2016(3): 8-13 (2016)
2. Stefanos Vidros, Constantinos Koliас, Georgios Kambourakis, Leman Akoglu: Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset. Future Internet 9(1): 6 (2017)
3. Dang Van Thin et al.: Multi-task learning for aspect and polarity recognition on Vietnamese datasets. PACLIC 2019
4. Tin Van Huynh, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, Anh Gia-Tuan Nguyen: Job prediction: From deep neural network models to applications. RIVF 2020: 1-6
5. Quoc-Nam Nguyen, Thang Chau Phan, Duc-Vu Nguyen, Kiet Van Nguyen: ViSoBERT: A Pre-Trained Language Model for Vietnamese Social Media Text Processing. CoRR abs/2310.11166 (2023)
6. Nguyen Luong Tran, Duong Minh Le, Dat Quoc Nguyen: BartPho: pre-trained sequence-to-sequence models for Vietnamese. CoRR abs/2109.09701 (2021)

