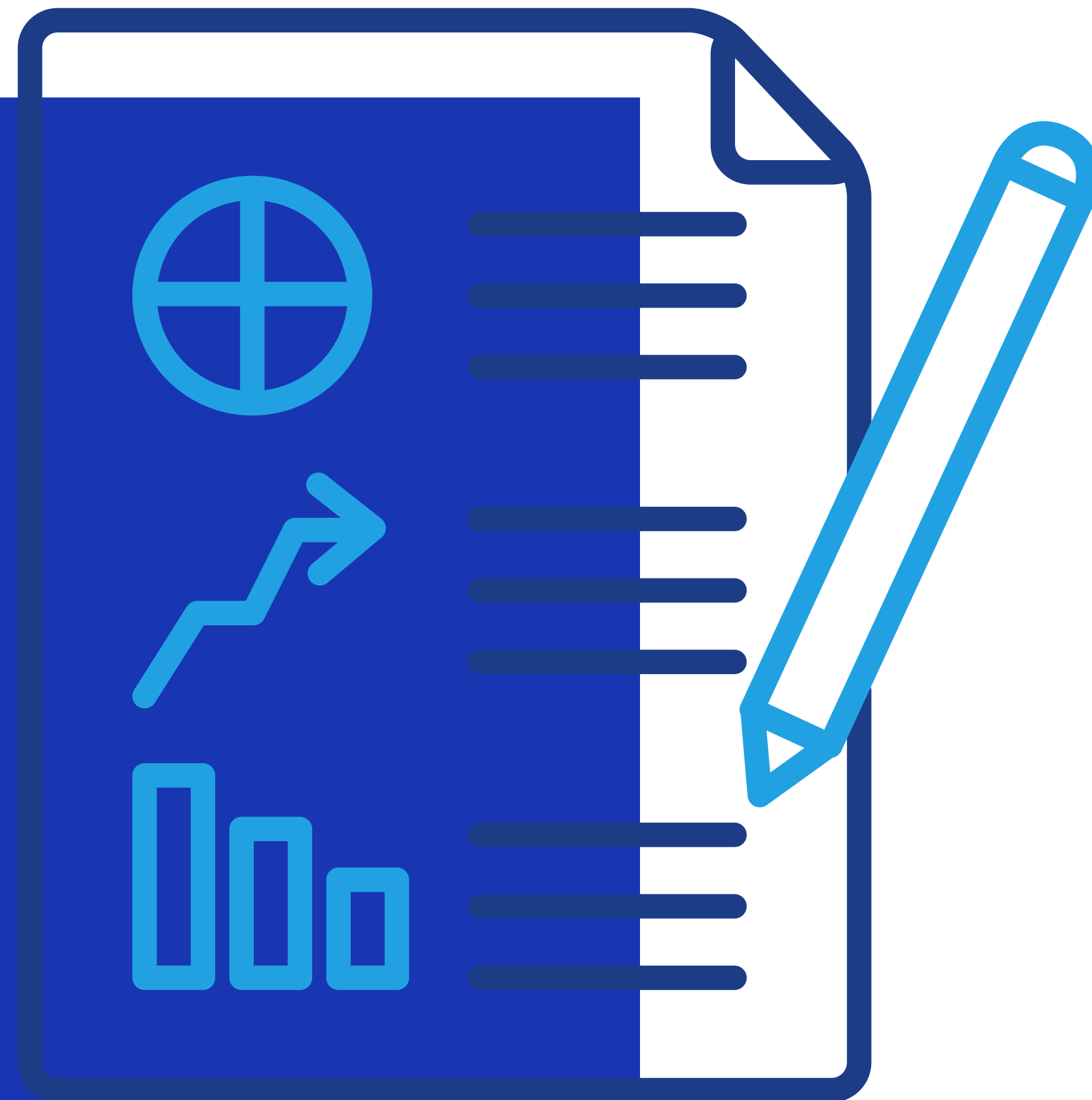




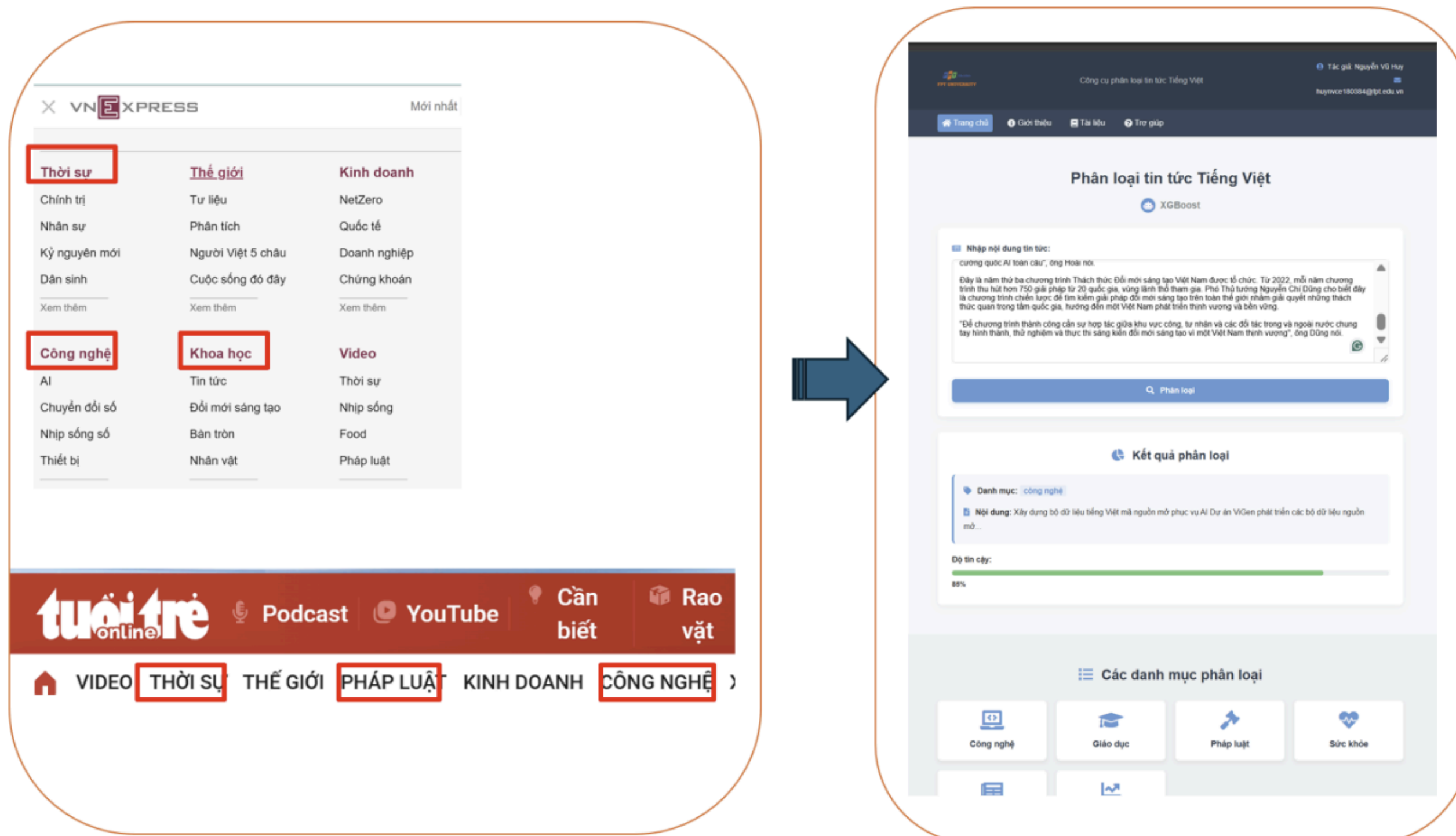
[https://github.com/nhunguet/SearchEngine/tree/main/Group%203/NguyenVuHuy\\_CE180384\\_PersonalProject](https://github.com/nhunguet/SearchEngine/tree/main/Group%203/NguyenVuHuy_CE180384_PersonalProject)

# COLLECT NEWS DATA AND BUILD NEWS CLASSIFICATION MODEL AND DEPLOY WEB APPLICATION

CE180384 - Nguyen Vu Huy



# Project Overview

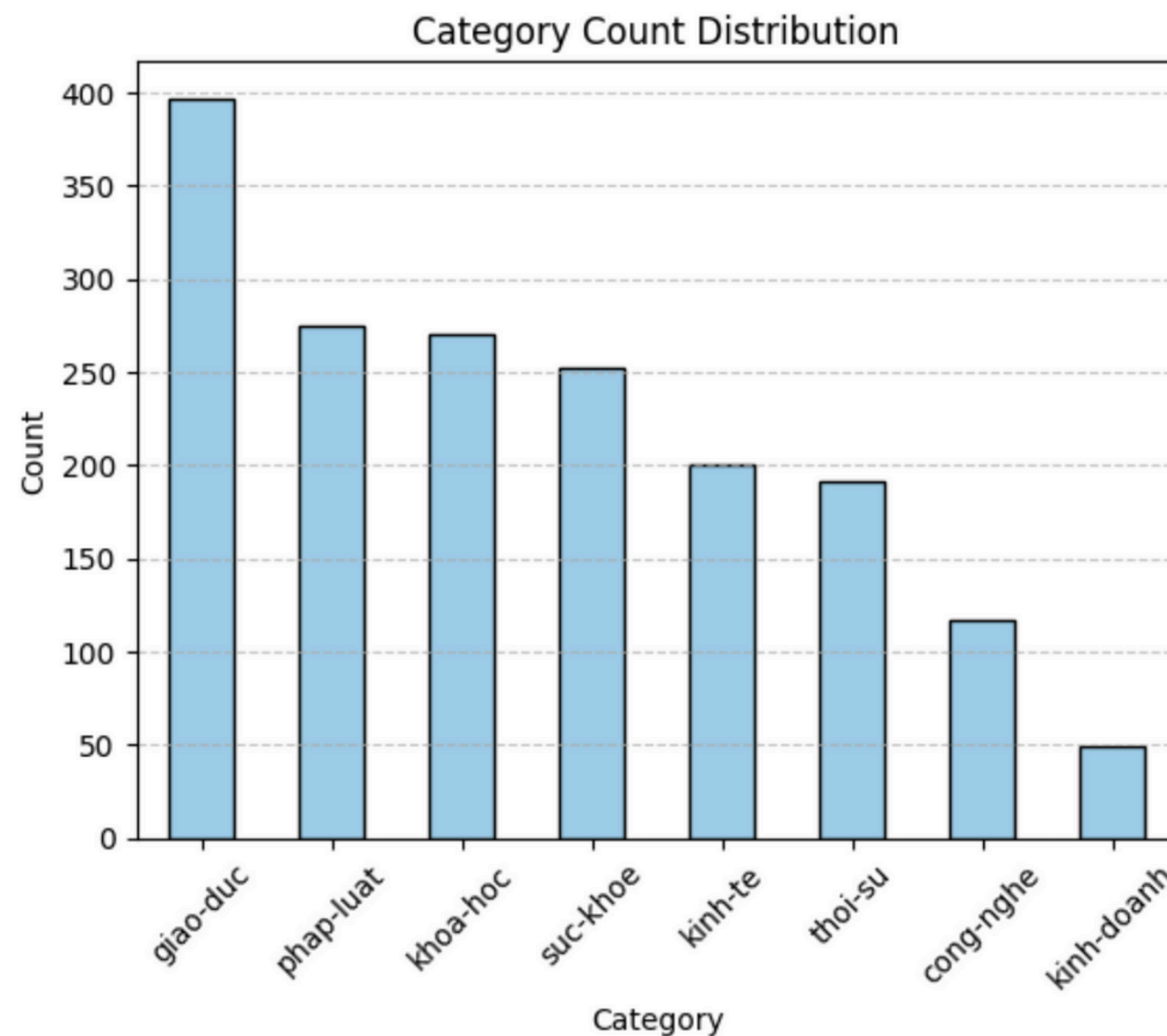


**Figura 1:** Project Overview

# Datasets

url	source	crawl_time	category	title	description	publish_time	content	author
https://tuoitre.vn/tong-bi-thu-to-lam-xay-dung...	tuoitre	2025-03-12 22:28:11	cong-nghe	Tổng Bí thư Tô Lâm: Xây dựng, ban hành danh mục...	NaN	NaN	Tổng Bí thư Tô Lâm chủ trì phiên họp - Ảnh: TT...	THÀNH CHUNG
https://tuoitre.vn/cau-noi-viettel-xay-nen-gia...	tuoitre	2025-03-12 22:28:40	cong-nghe	'Cầu nối' Viettel xây nên giấc mơ giáo dục 'kh...	NaN	NaN	Trong gần 2 thập kỷ, Internet trường học đã gi...	VIETTEL - T.HÀ
https://tuoitre.vn/chi-dung-y-nghi-nguoi-dan-o...	tuoitre	2025-03-12 22:29:09	cong-nghe	Chỉ dùng ý nghĩ, người đàn ông bị liệt có thể ...	NaN	NaN	Giáo sư thần kinh học Karunesh Ganguly (giữa) ...	TTXVN
https://tuoitre.vn/bat-ngo-vi-lam-ly-lich-tu-p...	tuoitre	2025-03-12 22:29:40	cong-nghe	Bất ngờ vì làm lý lịch tư pháp qua VNeID thuận...	NaN	NaN	Một số người dân vẫn đến trực tiếp làm thủ tục...	ÁI NHÂN
https://tuoitre.vn/xac-thuc-dien-tu-qua-vneid-...	tuoitre	2025-03-12 22:30:11	cong-nghe	Xác thực điện tử qua VNeID tăng cường an ninh,...	NaN	NaN	Trung tâm RAR (Bộ Công an) và Sacombank ký kết...	DANH TRỌNG
...	...	...	...	...	...	...	...	...
https://vnexpress.net/chi-em-sinh-doi-gianh-ho...	vnexpress	2025-03-12 15:03:55	giao-duc	Chị em sinh đôi giành học bổng hơn 17 tỷ đồng ...	Quỳnh Anh và Quỳnh Hương, 19 tuổi, giành học b...	Thứ ba, 4/3/2025, 11:09 (GMT+7)	Tin trúng tuyển và học bổng đến vào sáng 22/2,...	NaN
https://vnexpress.net/chuyen-gia-giai-dap-co-n...	vnexpress	2025-03-12 15:04:34	giao-duc	Chuyên gia giải đáp 'có nên cho con du học THPT...	Ông Hoàng Nam Tiến cùng các chuyên gia giáo dục...	Thứ ba, 11/3/2025, 10:00 (GMT+7)	Sự kiện diễn ra vào 15h - 18h, thứ Bảy, ngày 2...	NaN
https://vnexpress.net/vung-dat-cua-dai-may-tra...	vnexpress	2025-03-12 15:05:14	giao-duc	'Vùng đất của dải mây trắng' là tên gọi nước nào?	Quốc gia này ở bán cầu Nam, được người Maori g...	Thứ sáu, 28/2/2025, 19:00 (GMT+7)	Bình Minh (Tổng hợp)\nNồng độ bụi mịn PM 2.5 v...	NaN

# EDA

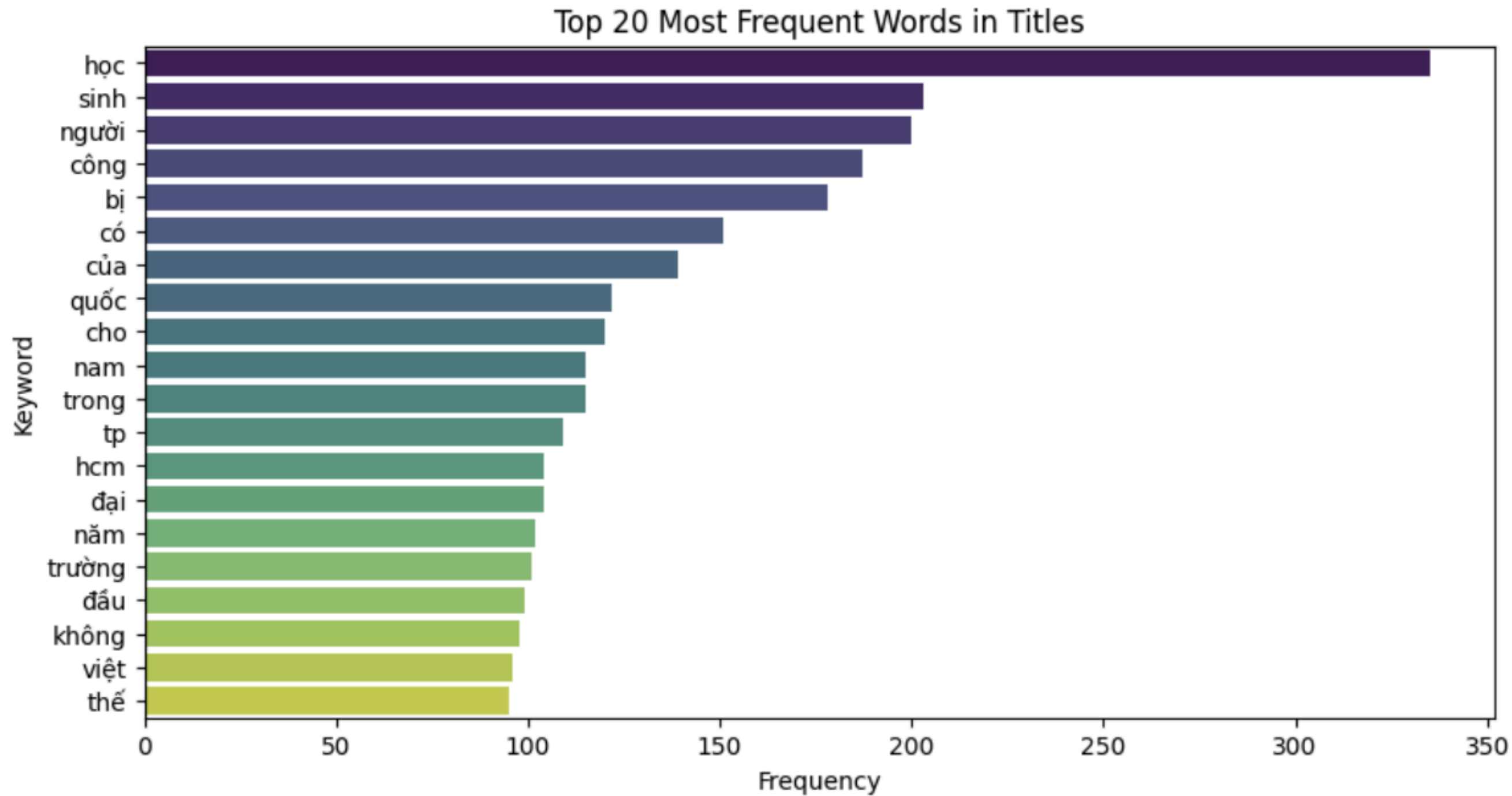


**Figura 3:** Category distribution

**Figura 4:** Word Cloud Representations for Titles, Content, and Descriptions



# EDA



**Figura 5:** Top 20 Most Frequent Words in Titles

# Pipeline Project

huynvce180384@fpt.edu.vn

AI1801 - SEG301

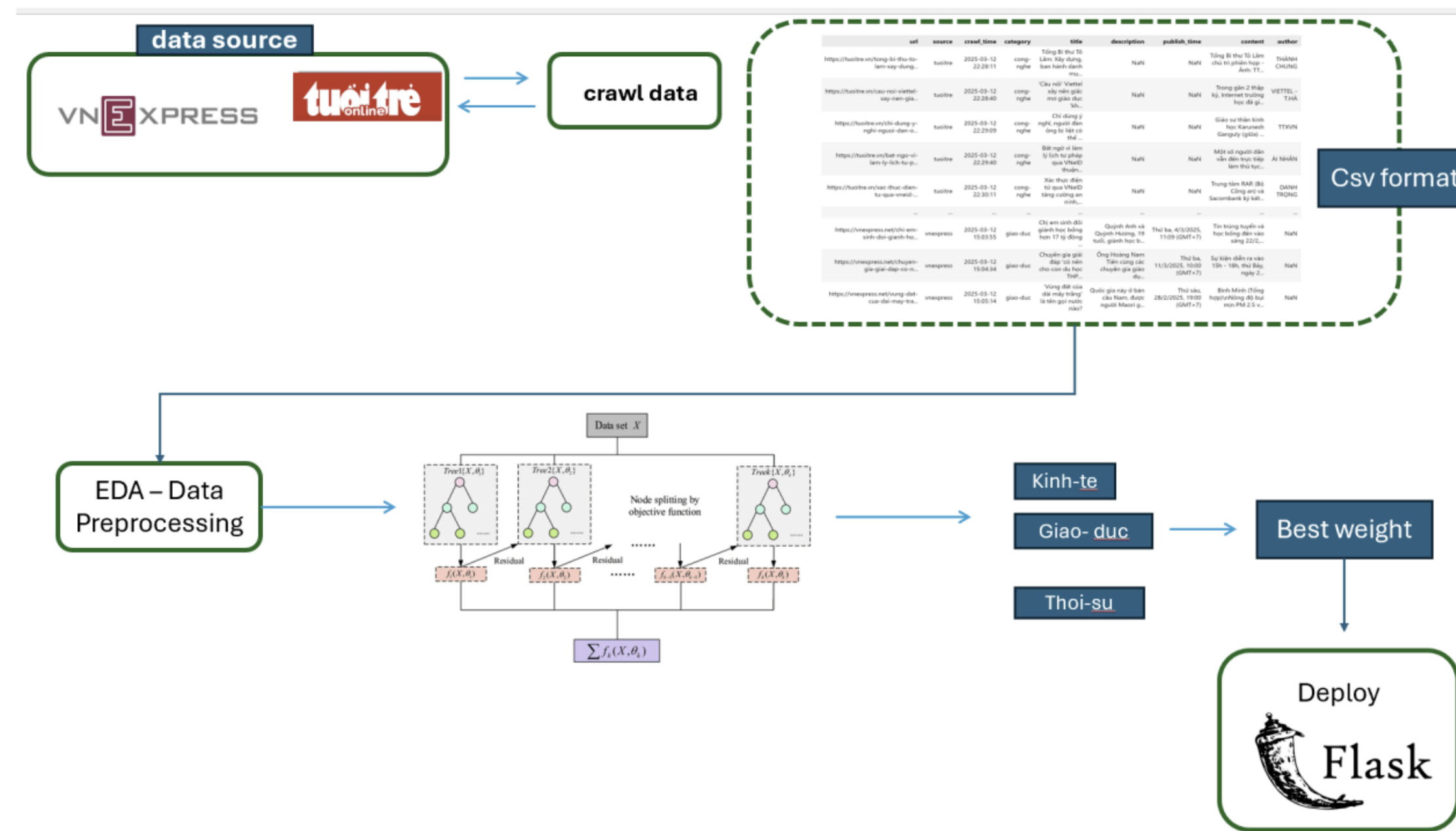



Figura 6: Complete processing pipeline

# Demo



Công cụ phân loại tin tức Tiếng Việt

Tác giả: Nguyễn Vũ Huy  
huynvce180384@fpt.edu.vn

Trang chủ

Giới thiệu

Tài liệu

Trợ giúp

## Phân loại tin tức Tiếng Việt

XGBoost

Nhập nội dung tin tức:

cường quốc AI toàn cầu", ông Hoài nói.

Đây là năm thứ ba chương trình Thách thức Đổi mới sáng tạo Việt Nam được tổ chức. Từ 2022, mỗi năm chương trình thu hút hơn 750 giải pháp từ 20 quốc gia, vùng lãnh thổ tham gia. Phó Thủ tướng Nguyễn Chí Dũng cho biết đây là chương trình chiến lược để tìm kiếm giải pháp đổi mới sáng tạo trên toàn thế giới nhằm giải quyết những thách thức quan trọng tầm quốc gia, hướng đến một Việt Nam phát triển thịnh vượng và bền vững.

"Để chương trình thành công cần sự hợp tác giữa khu vực công, tư nhân và các đối tác trong và ngoài nước chung tay hình thành, thử nghiệm và thực thi sáng kiến đổi mới sáng tạo vì một Việt Nam thịnh vượng", ông Dũng nói.

Phân loại

Kết quả phân loại

Danh mục: công nghệ

Nội dung: Xây dựng bộ dữ liệu tiếng Việt mã nguồn mở phục vụ AI Dự án ViGen phát triển các bộ dữ liệu nguồn mở...

Độ tin cậy:



THANK  
YOU