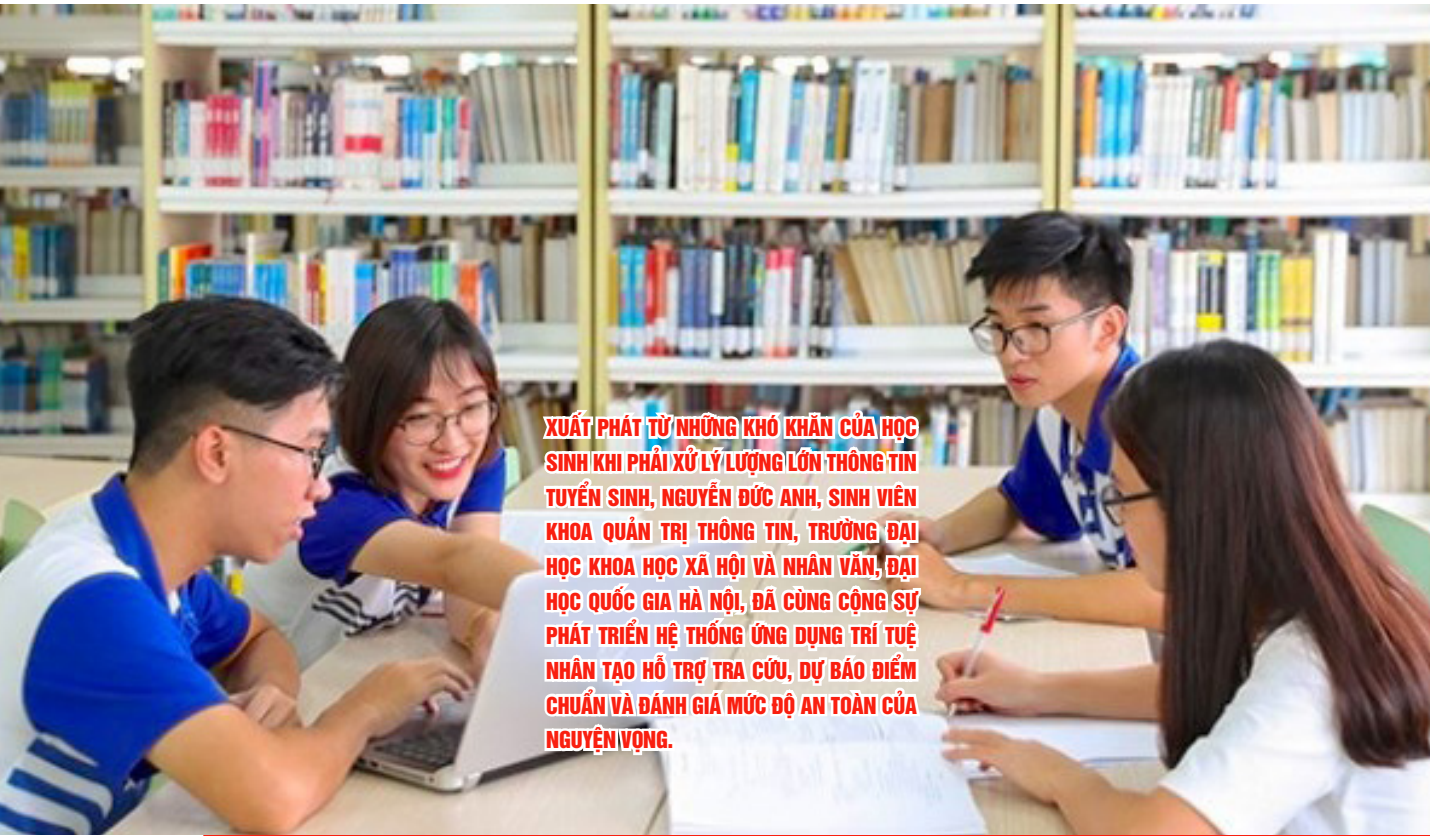




XUẤT PHÁT TỪ NHỮNG KHÓ KHĂN CỦA HỌC SINH KHI PHẢI XỬ LÝ LƯỢNG LỚN THÔNG TIN TUYỂN SINH, NGUYỄN ĐỨC ANH, SINH VIÊN KHOA QUẢN TRỊ THÔNG TIN, TRƯỜNG ĐẠI HỌC KHOA HỌC XÃ HỘI VÀ NHÂN VĂN, ĐẠI HỌC QUỐC GIA HÀ NỘI, ĐÃ CÙNG CỘNG SỰ PHÁT TRIỂN HỆ THỐNG ỨNG DỤNG TRÍ TUỆ NHÂN TẠO HỖ TRỢ TRA CỨU, DỰ BÁO ĐIỂM CHUẨN VÀ ĐÁNH GIÁ MỨC ĐỘ AN TOÀN CỦA NGUYỄN VỌNG.

SINH VIÊN USSH PHÁT TRIỂN HỆ THỐNG AI DỰ BÁO ĐIỂM CHUẨN

 **TRẦN TRANG**

TỪ TRẢI NGHIỆM CÁ NHÂN ĐẾN MỘT ĐỀ TÀI NGHIÊN CỨU

Ý tưởng xây dựng hệ thống bắt đầu từ chính trải nghiệm của Nguyễn Đức Anh khi tìm hiểu thông tin tuyển sinh đại học. Số liệu điểm chuẩn được công bố ở nhiều nguồn, dưới những định dạng khác nhau, khiến việc tổng hợp, so sánh và nhận diện xu hướng trở nên khó khăn đối với không ít học sinh.

Ban đầu, Đức Anh chỉ dự định xây dựng một công cụ nhỏ nhằm tự động hóa quá trình thu thập và sàng lọc dữ liệu điểm chuẩn qua các năm. Dưới sự hướng dẫn của thầy Trần Đăng Ninh, giảng viên Học viện Kỹ thuật và Công nghệ An ninh, ý tưởng này dần được phát triển thành một đề tài nghiên cứu có hệ thống, kết hợp giữa công nghệ xử lý dữ liệu với nhu cầu thực tế của học sinh và phụ huynh.

Trong quá trình nghiên cứu, Đức Anh nhận thấy sự biến động của điểm chuẩn không chỉ phụ thuộc vào kết quả thi mà còn chịu tác động từ mức độ quan tâm của xã hội, xu hướng lựa chọn ngành học và nhu cầu của thị trường lao động.

Từ nhận định đó, nhóm nghiên cứu mở rộng phạm vi dữ liệu, thu thập và phân tích hàng nghìn bài viết về giáo dục nhằm đánh giá mức độ quan tâm đối với từng nhóm ngành. Các mô hình học máy được đưa vào hệ thống để chuyển công cụ tra cứu dữ liệu tĩnh thành nền tảng có khả năng phân tích xu hướng và đưa ra khoảng điểm dự báo.

Từ đề tài nghiên cứu khoa học ban đầu, Đức Anh và cộng sự đã xây dựng nền tảng trực tuyến diemthi.techtreesolution.com. Hệ thống cung cấp chức năng

Trang chủ Sản phẩm Tra điểm thi Việt Nam

Tra cứu Điểm chuẩn

Hệ sinh thái TechTree Solution

Tìm theo Trường Tìm theo Ngành

Nhập tên hoặc mã trường (VD: Ngoại thương, QHX)...

Tỉnh/Thành phố

Trường Đại Học Khoa Học Xã Hội và Nhân Văn Hà Nội - VNU-HN

Mã: QHX Hà Nội

Trường Đại Học Công Nghệ - Đại Học Quốc Gia Hà Nội - VNU-HN

Mã: QHI Hà Nội

Trường Đại Học Giáo Dục - ĐHQG Hà Nội - VNU-HN

Mã: QHS Hà Nội

Trường Đại Học Khoa Học Tự Nhiên Hà Nội - VNU-HN

Mã: QHT Hà Nội

Trường Đại Học Kinh Tế - ĐHQG Hà Nội - VNU-HN

Mã: QHE Hà Nội

Trường Đại học Luật - ĐHQG Hà Nội - VNU-HN

Mã: QHL Hà Nội

tra cứu dữ liệu, dự báo điểm chuẩn, đánh giá rủi ro và đề xuất khoảng điểm an toàn đối với nhiều ngành, trường đại học trên cả nước.

CHẤT LƯỢNG DỮ LIỆU QUYẾT ĐỊNH ĐỘ TIN CẬY CỦA DỰ BÁO

Theo Đức Anh, yếu tố quan trọng nhất của hệ thống không nằm ở việc sử dụng các thuật toán phức tạp, mà ở khả năng xây dựng nguồn dữ liệu ổn định, có cấu trúc và được cập nhật thường xuyên.

Quy trình vận hành được tổ chức theo ba công đoạn chính, gồm thu thập dữ liệu; xử lý và lưu trữ; phân tích và dự báo. Các chương trình tự động định kỳ rà quét thông tin từ những nguồn chính thống. Dữ liệu sau khi thu thập được làm sạch, chuẩn hóa và đưa vào cơ sở dữ liệu có cấu trúc trước khi chuyển sang bước phân tích.

Tại công đoạn cuối, các mô hình xử lý ngôn ngữ tự nhiên được sử dụng để nhận diện mức độ quan tâm của truyền thông và xã hội đối với từng ngành học. Kết quả này được kết hợp với dữ liệu điểm thi, phổ điểm và điểm chuẩn những năm trước để đưa ra khoảng điểm dự báo phù hợp.

Thách thức lớn nhất của nhóm là dữ

liệu tuyển sinh giữa các cơ sở đào tạo chưa đồng nhất. Mỗi trường có thể công bố thông tin theo một cấu trúc, cách đặt tên ngành hoặc định dạng văn bản khác nhau. Đặc điểm đa dạng của tiếng Việt, cùng sự xuất hiện của các từ viết tắt và cách diễn đạt không thống nhất, cũng làm tăng độ phức tạp trong quá trình xử lý dữ liệu.

Thay vì đưa toàn bộ dữ liệu thô trực tiếp vào mô hình, Đức Anh dành phần lớn thời gian cho bước tiền xử lý. Nhóm xây dựng hệ thống từ khóa, phân loại thông tin theo từng cụm chủ đề và thử nghiệm mô hình trên những tập dữ liệu nhỏ, có độ chính xác cao trước khi mở rộng quy mô.

Cách tiếp cận này giúp hạn chế sai số và nâng cao độ ổn định của kết quả dự báo. Đây cũng là kinh nghiệm Đức Anh rút ra trong quá trình nghiên cứu: hiệu quả của một hệ thống trí tuệ nhân tạo phụ thuộc rất lớn vào chất lượng dữ liệu đầu vào.

CẬP NHẬT DỮ LIỆU ĐỂ HỖ TRỢ THÍ SINH LỰA CHỌN NGUYỆN VỌNG

Điểm đáng chú ý của dự án là khả năng cập nhật dữ liệu định kỳ thay vì chỉ cung cấp số liệu điểm chuẩn của các năm trước. Sau mỗi 12 giờ, thuật

toán tự động rà soát các thông tin mới, xử lý dữ liệu và cập nhật vào hệ thống.

Nhờ đó, khoảng điểm dự báo của từng ngành có thể được điều chỉnh theo những thay đổi về mức độ quan tâm, xu hướng đăng ký và các thông tin liên quan đến tuyển sinh. Thí sinh có thêm cơ sở để nhận diện mức độ cạnh tranh của ngành học, đánh giá khả năng trúng tuyển và cân đối danh sách nguyện vọng.

Tại Hội nghị Nghiên cứu khoa học sinh viên, giá trị của dự án được nhìn nhận từ khả năng ứng dụng công nghệ để giải quyết một vấn đề cụ thể của học sinh và phụ huynh. Thay vì chỉ dừng ở mô hình thử nghiệm, sản phẩm đã được phát triển thành một nền tảng có thể tiếp cận và sử dụng trực tiếp.

Trong thời gian tới, Đức Anh dự kiến tiếp tục hoàn thiện công cụ tra cứu, nâng cao khả năng lọc dữ liệu và cải thiện trải nghiệm người dùng. Nam sinh cũng định hướng mở rộng tập dữ liệu huấn luyện từ các nền tảng mạng xã hội, qua đó giúp mô hình phản ánh nhanh hơn những thay đổi trong sự quan tâm của thí sinh.

Một mục tiêu khác của nhóm là phát triển phiên bản dành cho thiết bị di động, tạo điều kiện để học sinh ở nhiều khu vực có thể tiếp cận hệ thống thuận tiện hơn.

Việc kết hợp kiến thức quản trị thông tin với các công nghệ học máy, xử lý ngôn ngữ tự nhiên và dữ liệu lớn cho thấy khả năng tiếp cận liên ngành của sinh viên Trường Đại học Khoa học Xã hội và Nhân văn. Từ nhu cầu thực tế của người học, Nguyễn Đức Anh và cộng sự đang từng bước hoàn thiện một công cụ hỗ trợ thí sinh tiếp cận thông tin tuyển sinh chủ động, có hệ thống và phù hợp hơn với quá trình lựa chọn ngành học.