

# MÔ HÌNH ĐỀ XUẤT DỰA TRÊN TIÊU CHÍ HỒ SƠ CỦA ỨNG VIÊN VÀ VỊ TRÍ VIỆC LÀM CỦA NGÀNH CÔNG NGHỆ THÔNG TIN

MODEL RECOMMEND BY TECHNICAL SKILLS, SOFT SKILLS, TOOLS OF JOBS FOR INFORMATION TECHNOLOGY

Tên: Trần Hữu Ân  
Đơn vị: Khoa công nghệ thông tin  
Trường: Trường Đại học Giao Thông Vận Tải TP HCM  
tranhuuan0202@gmail.com

Tên: Bùi Dương Thế  
Đơn vị: Khoa công nghệ thông tin  
Trường: Trường Đại học Giao Thông Vận Tải TP HCM  
duongthe@ut.edu.vn

**Abstract**—Đây là bài báo nghiên cứu mô tả các công việc về ngành công nghệ thông tin được thực hiện thu thập dữ liệu dựa trên trang web Indeed.com là một trang web tuyển dụng được thu thập vào ngày 25 tháng 01 năm 2023. Kiểm định giả thuyết Chi-square được sử dụng để đánh giá mức độ tương quan với nhau. Nhằm thống kê các loại hình vị trí công việc trong ngành công nghệ thông tin, các kỹ năng hiện đang được các công ty yêu cầu, các công nghệ sử dụng. Dựa vào đó ứng viên có thể tìm ra được các loại kỹ năng và công nghệ cho bản thân trước khi ứng tuyển.

Nghiên cứu đã ứng dụng mô hình BERTopic để xây dựng mô hình đề xuất các công việc theo các loại kỹ năng và công nghệ sử dụng của ứng viên trong ngành công nghệ thông tin. Mô hình còn có khả năng phục vụ cho các nhà tuyển dụng có thể kiểm tra CV ứng viên có mức độ phù hợp với các vị trí công việc cụ thể nào. Với đầu vào là một đoạn văn bản nói về kỹ năng và công nghệ có thể sử dụng của ứng viên.

**Keywords**— information technology, skill, tool, BERTopic, statistical.

## I. ĐẶT VẤN ĐỀ.

Công nghệ thông tin là một trong những ngành có chuyển biến tích cực nhất trong những năm qua khi đã triển khai ứng dụng công nghệ thông tin trong đời sống, xã hội và doanh nghiệp. Việt Nam là một trong những nước có số lượng người tham gia internet thuộc top cao nhất trên thế giới hiện nay. Việt Nam có hơn 61,37 triệu người dùng, tương đương tỉ lệ 64% dân số đang sở hữu smartphone nằm trong top 10 quốc gia có lượng người dùng smartphone lớn nhất thế giới [1]. Nguồn lực công nghệ thông tin cũng được chú trọng để đáp ứng xu thế phát triển của xã hội. Số lượng các công việc trong các ngành nghề vô cùng đa dạng. Theo quan sát các chỉ số về tỷ lệ doanh nghiệp và nhân lực ngành CNTT trong báo cáo VNICT Index 2016, số lượng nhu cầu nhân lực trong ngành vẫn tiếp tục tăng trưởng ở mức 8%. Lượng nhân sự thiếu hụt khoảng 78.000 nhân lực CNTT mỗi năm và đến năm 2020 sẽ thiếu hơn 500.00 nhân lực CNTT, chiếm hơn 78% tổng số nhân lực mà thị trường này cần [2].

Theo báo cáo thị trường công nghệ thông tin Việt Nam 2021 Developers Recruitment State, chất lượng sinh viên sau khi tốt nghiệp đáp ứng yêu cầu tuyển dụng của doanh nghiệp còn thấp. Cụ thể, trong số hơn 55.000 sinh viên công nghệ thông tin (CNTT) tốt nghiệp mỗi năm chỉ có khoảng 16.500 sinh viên (30%) đáp ứng được những kỹ năng và chuyên môn mà doanh nghiệp cần. Tuy

nhien nhu cầu cho ngành công nghệ thông tin tại Việt Nam vẫn đang tăng cao liên tục. Năm 2021 Việt Nam cần đến 450.000 nhân lực trong ngành công nghệ thông tin [3].

Với nhu cầu tuyển dụng trong lĩnh vực công nghệ thông tin (CNTT) luôn trên đà tăng trưởng mạnh. Việc thiếu hụt nhân sự trong ngành công nghệ thông tin tại Việt Nam là một vấn đề nghiêm trọng. Đối tượng sinh viên là một trong những đối tượng cần được quan tâm trong việc giải quyết vấn đề này. Nghiên cứu này mong muốn giải quyết được vấn đề thiếu hụt nhân sự ở các công ty và tình trạng thất nghiệp của ứng viên.

Nghiên cứu này sẽ đưa ra các thống kê về các vị trí, các kỹ năng cũng như các công cụ trong ngành công nghệ thông tin. Sau đó các ứng viên có thể dựa vào các công nghệ, kỹ năng thống kê của nghiên cứu để chọn ra các kỹ năng, công nghệ phù hợp với vị trí tuyển chọn cho bản thân. Nghiên cứu đã xây dựng một mô hình gợi ý vị trí công việc theo kỹ năng và công nghệ của ứng viên để đề xuất ra các vị trí công việc phù hợp.

## II. PHƯƠNG PHÁP NGHIÊN CỨU.

### 1. Đối tượng nghiên cứu.

Đối tượng nghiên cứu được lấy từ các bài viết tuyển dụng nhân sự của các công ty tại Việt Nam được lấy từ website Indeed. Từ ngày 25 tháng 01 năm 2023.

### 2. Thời gian và địa điểm nghiên cứu.

#### 1) Thời gian nghiên cứu.

Trong tháng 1 năm 2023.

#### 2) Địa điểm nghiên cứu.

Trường Đại học Giao Thông Vận Tải TP HCM.

### 3. Thiết kế nghiên cứu.

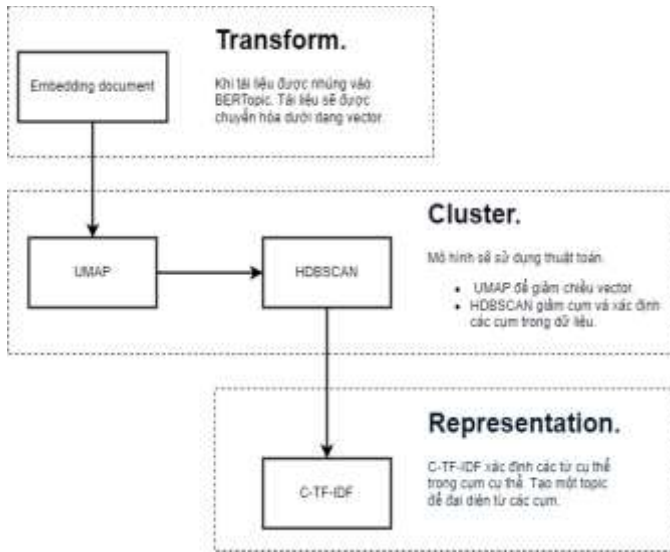
Đây là một nghiên cứu mô tả được thực hiện thu thập từ ngày 25 tháng 01 năm 2023 đến ngày 26 tháng 01 năm 2023. Bộ dữ liệu được thực hiện thu thập dựa trên các thông tin được công khai từ các bài tuyển dụng trên website. Dữ liệu được lấy từ 4379 bài tuyển dụng khác nhau về các vị trí được tuyển dụng trong ngành công nghệ thông tin.

### 4. Thu thập dữ liệu.

#### 1) Quy trình thu thập dữ liệu.

Thu thập dữ liệu sẽ tiếp cận website Indeed vào ngày 25 tháng 01 năm 2023 thông qua sử dụng thư viện Selenium, pandas, datetime của ngôn ngữ Python để thu thập dữ liệu thô. Dữ liệu

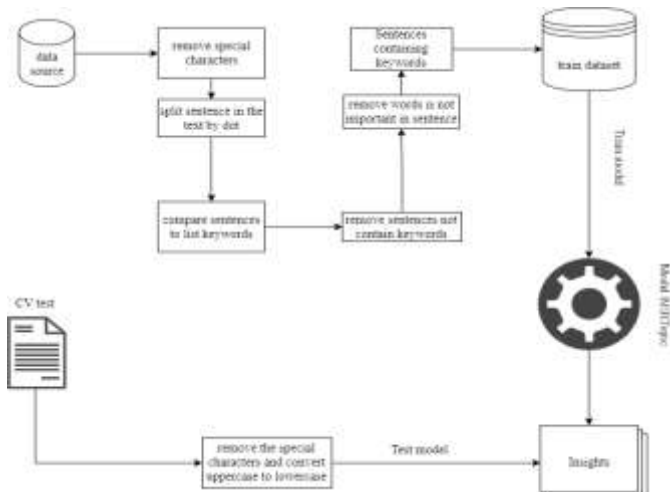




Sơ đồ 2: Biểu diễn quy trình xử lý của thuật toán BERTopic.

b) Quy trình xử lý công việc và xây dựng mô hình gợi ý vị trí công việc.

Nghiên cứu đã xây dựng mô hình gợi ý theo phương pháp sử dụng BERTopic model của một hồ sơ ứng viên như sau. Đầu vào là đoạn mô tả các kỹ năng và công nghệ của ứng viên. Các kỹ năng và công nghệ mà ứng viên có sẽ được so sánh với các kỹ năng tương ứng mà vị trí công việc có trong bộ dữ liệu.



Sơ đồ 3: Mô hình gợi ý dựa theo mô hình BERTopic kết hợp phương pháp thống kê.

Quy trình xử lý như sau: Nguồn dữ liệu thu thập được làm sạch thông qua các bước như loại bỏ các ký tự đặc biệt, phân tách đoạn văn bản thành các câu, chọn lọc các câu có chứa các từ khóa và các câu không chứa các từ khóa sẽ bị loại bỏ đi. Các câu có chứa từ khóa sẽ được tiếp tục loại bỏ các từ không quan trọng trong câu. Các câu có chứa từ khóa sẽ được gán một label về tên vị trí công việc. Mô hình BERTopic (mô hình phân cụm) sẽ được ứng dụng để train bộ dữ liệu. Kết quả đầu ra là những danh sách vị trí và mức độ phù hợp với ứng viên sẽ phụ thuộc vào mật độ xuất hiện của vị trí công việc trong danh sách.

### 3) Kết quả thực nghiệm.

Nghiên cứu đã thực hành thử nghiệm với 10 bộ CV ứng viên. Bộ CV sẽ được lọc các thông tin về kỹ năng và công nghệ sử dụng. Mô hình sẽ căn cứ vào các thông tin về kỹ năng và công nghệ của

ứng viên. Hai mô hình sẽ thống kê kết quả về mức độ phù hợp về vị trí của ứng viên trong ngành công nghệ thông tin.

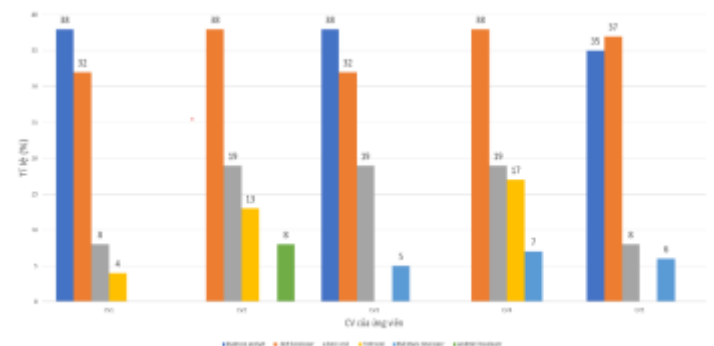
#### a) Phương pháp thống kê.

Mô hình gợi ý theo phương pháp thủ công sẽ thống kê công việc dựa theo danh sách kỹ năng và danh sách công nghệ của ứng viên. Sau đó truy vấn vào bộ dữ liệu để tìm các số lượng loại công việc phù hợp. Mô hình thống kê sẽ gợi ý 5 công việc tốt nhất dành cho ứng viên. Kết quả thực nhiệm được thể hiện chi tiết ở bảng 3.

	Output1	Output2	Output3	Output4
Cleaning, visualization, SQL, python, micorsoft, excel	Business analyst (38%)	.Net Developer (32%)	Back end (8%)	Front End Developer (4%)
Teamwork, design, reactjs, javascript, git	.Net developer (38%)	Back-end developer (19%)	Front end developer (13%)	Android Developer (8%)
SQL, visualization	Business analyst (38%)	Net Developer (32%)	Back-end developer (19%)	Full stack developer (5%)
Coding, test, design, support, JavaScript, typescript	.Net developer (38%)	Back-end developer (19%)	Font End Developer (17%)	Full stack developer (7%)
English, SQL, databases, visualization, teamwork, marketing, python	.Net developer (37%)	Business analyst (35%)	Back end Developer (8%)	Full stack developer (6%)

Bảng 3: Kết quả của mô hình gợi ý theo phương pháp thống kê.

Kết quả được trực quan hóa dưới dạng biểu đồ sau. Chi tiết được thể hiện ở biểu đồ 1.



Biểu đồ 1: Biểu đồ thể hiện kết quả khi sử dụng mô hình thống kê thủ công.

#### b) Phương pháp áp dụng mô hình BERTopic.

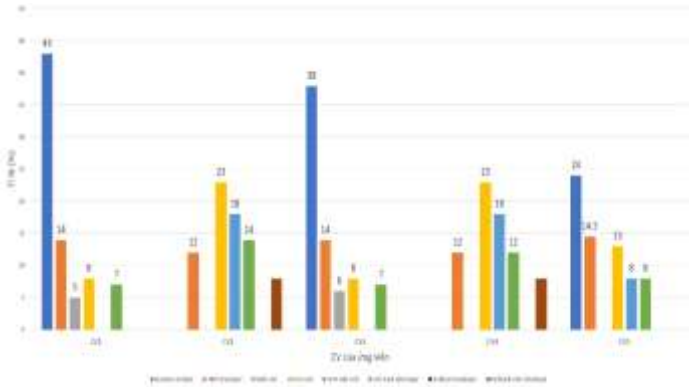
Mô hình gợi ý theo phương pháp BERTopic dựa theo một chuỗi mô tả về kỹ năng và công nghệ sử dụng của ứng viên trong CV xin việc. Mô hình sẽ dựa theo phân cụm chủ đề của mô hình để đưa ra gợi ý cho ứng viên về năm công việc phù hợp nhất cho ứng viên dựa theo kỹ năng và công nghệ sử dụng của họ. Kết quả thực nghiệm được thể hiện chi tiết ở bảng 4.

	Output1	Output2	Output3	Output4	Output5
Cleaning, visualization, SQL, python, Microsoft, excel	Business analyst (43%)	.Net Developer (14%)	Front end Developer (8%)	Full stack developer (7%)	Back-end developer (5%)

Teamwork, design, reactjs, JavaScript, git	Front end developer (23%)	Front end web developer (18%)	Full stack developer (14%)	.net developer (12%)	Full stack web developer (8%)
SQL, visualization	Business analyst (38%)	.Net Developer (14%)	Front end Developer (8%)	Full stack developer (7%)	Back-end developer (6%)
Coding, test, design, support, JavaScript, typescript	Front end developer (23%)	Front end web developer (18%)	Full stack developer (14%)	.net developer (12%)	Full stack web developer (8%)
English, SQL, databases, visualization, teamwork, marketing, python	Business analyst (24%)	.Net Developer (14.5%)	Front end Developer (13%)	Full stack developer (8%)	Front end web developer (8%)

Bảng 4: Kết quả thử nghiệm trên mô hình gợi ý BERTopic.

Kết quả được trực quan hóa dưới dạng biểu đồ sau. Chi tiết được thể hiện ở biểu đồ 2.



Biểu đồ 2: Biểu đồ thể hiện kết quả của mô hình BERTopic.

#### IV. THẢO LUẬN.

Với sự phát triển của kinh tế số, nhu cầu nhân lực trong lĩnh vực này ngày càng tăng. Tuy nhiên, việc tuyển dụng nhân lực chất lượng cao trong lĩnh vực công nghệ thông tin lại không phải là một điều dễ dàng. Theo một cuộc nghiên cứu của lực lượng lao động khoa học và kỹ thuật của Hoa Kỳ vào tháng 6 năm 2013, đã cho rằng vào năm 2011 tỷ lệ thất nghiệp ngành công nghệ thông tin lên đến 4.1% khoảng 3.293.020 người thất nghiệp đứng thứ hai trong khối ngành S&E (Science & Engineer) sau Mathematical [5].

Một trong những thách thức đối với tuyển dụng công nghệ thông tin ở Việt Nam hiện nay là thiếu hụt nhân lực chất lượng. Nhiều doanh nghiệp gặp khó khăn trong việc tìm kiếm và tuyển dụng các ứng viên có kinh nghiệm và kỹ năng đáp ứng được yêu cầu của công việc.

Các kỹ năng, công cụ và ngôn ngữ lập trình đóng vai trò rất quan trọng trong việc xây dựng và phát triển các ứng dụng và hệ thống. Đầu tiên, các kỹ năng về ngôn ngữ lập trình, quản lý dữ liệu, bảo mật thông tin, mạng. Ngoài ra, Kỹ năng giải quyết vấn đề, tư duy logic...cũng rất quan trọng trong việc xử lý các tình huống phát sinh và tối ưu hiệu suất của ứng dụng. Các công cụ phát triển phần mềm các trình quản lý mã nguồn và công cụ lưu

trữ dữ liệu là cần thiết để tạo ra một hệ thống chất lượng, đảm bảo tính ổn định. Kiểm định giả thuyết t-Test đã đánh giá, theo thống kê nghiên cứu của VietNam IT market report ngôn ngữ lập trình đang phổ biến là Javascript lên đến 70% vào năm 2021 so sánh tỷ lệ trong nghiên cứu này vào năm 2023 tỷ lệ đã giảm xuống [3].

Bài báo đã nghiên cứu và xây dựng mô hình gợi ý về việc làm công việc cho ứng viên và nhà tuyển dụng theo cả hai phương pháp và đưa ra kết quả so sánh như sau:

Đối với mô hình gợi ý theo phương pháp thống kê thủ công.

- Các kỹ năng và công nghệ sử dụng ứng viên phải được chuyển thành các danh sách riêng để so sánh từng giá trị của các trường kỹ năng và công nghệ của dữ liệu công việc. Điều rất khó khăn và đòi hỏi ứng viên phải cung cấp các giá trị kỹ năng phải đúng theo keyword trong trường dữ liệu.
- Các vị trí công việc đều có số lượng khác nhau trong bộ dữ liệu, điều có thể dẫn đến việc kết quả truy vấn dữ liệu sẽ không được chính xác.

Đối với mô hình gợi ý theo phương pháp BERTopic.

- Các kỹ năng và công nghệ của ứng viên có thể để ở dạng đoạn văn bản.
- Các keyword về kỹ năng và công nghệ của ứng viên không yêu cầu phải đúng với keyword được lưu trong bộ dữ liệu. Mô hình sẽ đưa ra dự đoán dựa vào mô hình phân cụm. Điều này sẽ cho kết quả với độ chính xác cao hơn mô hình gợi ý theo phương pháp thống kê.

Hạn chế bài báo này vẫn chưa giải quyết được vấn đề trong quá trình thu thập dữ liệu thô từ các bài tuyển dụng của các công ty. Có một số bài tuyển dụng thuộc các công ty vừa và nhỏ yêu cầu làm nhiều công việc trong bài tuyển dụng của mình. Chưa tập trung vào công việc chuyên môn cụ thể. Hiện tại nhóm nghiên cứu vẫn chưa giải quyết được vấn đề đó. Vì thế quá trình xây dựng mô hình dự đoán có thể gặp sai sót.

#### V. KẾT LUẬN.

Hiện nay, ngành công nghệ thông tin đang phát triển với tốc độ vượt bậc và đã góp phần tạo nên sự thay đổi to lớn trong cuộc sống của chúng ta. Tuy nhiên, những thách thức của ngành công nghệ đang phải gặp là thiếu hụt nhân sự chuyên môn. Với tốc độ phát triển nhanh chóng về ngành công nghệ thông tin này. Các công ty công nghệ đòi hỏi các nhân viên trong lĩnh vực này phải liên tục cập nhật và thay đổi và trau dồi các kỹ năng.

Nghiên cứu xây dựng mô hình theo phương pháp thống kê để đưa ra các gợi ý công việc phù hợp cho ứng viên theo công nghệ sử dụng và kỹ năng của ứng viên, đồng thời mô hình còn có khả năng phụ vụ cho nhà tuyển dụng vào mục đích kiểm tra CV của ứng viên phù hợp với vị trí công việc nào. Các kỹ năng và công nghệ của ứng viên sẽ được so sánh với mặt bằng chung của các vị trí ngành công nghệ trên thị trường hiện nay. Để góp phần giảm thiểu về tình trạng thiếu hụt nhân sự hiện nay.

#### VI. TÀI LIỆU TRÍCH DẪN.

- [1] T. L. T. ANH, "Tiềm năng phát triển nền kinh tế số ở Việt Nam và một số hàm ý chính sách," *TẠP CHÍ CÔNG*

THƯƠNG - CƠ QUAN THÔNG TIN LÝ LUẬN CỦA BỘ CÔNG THƯƠNG, 2022.

- [2] B. T. T. V. T. THÔNG, "Báo cáo chỉ số sẵn sàng cho phát triển và ứng dụng công nghệ thông tin và truyền thông Việt Nam 2016 -VN Index 2016," THÔNG TIN VÀ TRUYỀN THÔNG, 2017.
- [3] TopDev, "VIET NAM IT REPORT MARKET Developers Recruitment State," Viet Nam, 2021.
- [4] M. C. Y. C. O. B. Mohamed Chiny, "LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model," (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, 2021.
- [5] J. F. S. Jr., "The U.S. Science and Engineering Workforce: Recent, Current, and Projected Employment, Wages, and Unemployment," Congressional Research Service, 2013.
- [6] V. InTECH, "báo cáo thị trường nhân lực ngành CNTT thập niên 2010 & năm 2020," Viet Nam, 02/06/2020.
- [7] J. S. G. G. Karen Kaminski, "Workforce readiness: A study of university students' fluency with information technology," September 2009.
- [8] D. G. S. Dr. Janet Bailey, "Preparing the Information Technology Workforce," *ACM Journals*, p. <https://dl.acm.org/doi/abs/10.1145/571475.571476>, 01 August 2002.
- [9] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 11 03 2022. [Online]. Available: <chrome-extension://gphandlahdpffmccakmbngmbjnjiihp/https://arxiv.org/pdf/2203.05794.pdf>.