**ĐẠI HỌC BÁCH KHOA HÀ NỘI**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**



**BÁO CÁO BÀI TẬP LỚN**

Môn học: Các hệ thống thông minh

Giảng viên hướng dẫn: GS.TS Trần Đình Khang

Hoàng Việt Bách - 20212236M

Cao Minh Sơn - 20212625M

Nguyễn Trường Thành - 20212610M

**Hà Nội – 2022**

# TRỢ GIÚP NGƯỜI LAO ĐỘNG TÌM KIẾM VIỆC LÀM

**Môn học**: Các Hệ Thống Thông Minh

Hoàng Việt Bách - 20212236M

Cao Minh Sơn - 20212625M

Nguyễn Trường Thành - 20212610M

**Giảng viên hướng dẫn**: GS.TS Trần Đình Khang

# Abstract

Trong cuộc sống hiện nay, một trong những vấn đề thiết yếu của con người là vấn đề công ăn việc làm. Và việc làm chính là nhu cầu cơ bản, thiết yếu của người lao động. Nhưng làm sao để người lao động có thể tìm kiếm được công việc như mong muốn, thích hợp với năng lực của bản thân thì đó lại là một vấn đề cần được giải quyết. Trong thời đại 4.0 khi mà mọi nhu cầu tìm kiếm của con người có thể được thực hiện thông qua internet. Việc giúp người lao động có thể dễ dàng tìm kiếm được công việc với cho mình tốt hơn, phù hợp nhất với bản thân chính là vấn đề mà nhiều người quan tâm hiện nay. Vì lý do đó, đề tài “Trợ giúp người lao động tìm kiếm việc làm” của nhóm chúng em sẽ phân tích các yếu tố khác nhau, thông qua những mô hình toán học để đưa ra những kết quả phù hợp nhất cho mỗi người lao động.

**Từ khóa:** Việc làm, Trợ giúp người lao động, cây quyết định, phân loại, Topsis, AHP

# Giới thiệu bài toán

**Mục tiêu bài toán**: giúp người lao động có thể tìm kiếm được công việc phù hợp với năng lực, mức lương, địa điểm...

Sau khi mô hình và ra được phương pháp giải, ra được bảng hỗ trợ quyết định về những công việc phù hợp nhất với mong muốn tìm việc của người lao động.

Với bài toán này, ta có thể dễ dàng nhận ra các yếu tố quyết định, và các yếu tố môi trường (ảnh hưởng tới bảng quyết định) như sau:

*Các yếu tố quyết định*:

* + Lĩnh vực công việc của ứng viên
  + Năm kinh nghiệm hiện tại của ứng viên
  + Bằng cấp của ứng viên
  + Mức lương mong muốn của ứng viên
  + Tỉnh/thành mong muốn làm việc của ứng viên
  + Quận/huyện mong muốn làm việc của ứng viên

*Các yếu tố môi trường*:

* + Lĩnh vực công ty cần tuyển dụng
  + Năm kinh nghiệm công ty yêu cầu
  + Mức lương công ty đề xuất
  + Bằng cấp công ty yêu cầu
  + Tỉnh/thành của công ty
  + Quận/huyện của công ty

Ở các bước sau, ta sẽ tiến hành phân tích các tiêu chí này để phục vụ việc giải bài toán.

# Background

Phần này giới thiệu các kiến thức nền tảng để giải quyết bài toán. Bao gồm việc giới thiệu bài toán phân loại và các cách giải quyết nó. Phương pháp phân loại dùng phương pháp TOPSIS cũng được miêu tả kỹ về mặt lý thuyết ở chương này. Trong quá trình đánh giá matrix trọng số cho các tiêu chí của TOPSIS, có áp dụng phương pháp AHP để đảm bảo độ chính xác và chấp nhận được của bộ trọng số. Cuối cùng, việc giải bài toán phân bổ sẽ được quy về giải bài toán CUNG - CẦU tối ưu.

## 1. Phân loại dùng phương pháp TOPSIS

Giới thiệu về TOPSIS: TOPSIS là một phương pháp phân tích quyết định đa tiêu chí, được phát triển bởi Ching-Lai Hwang và Yoon vào năm 1981, sau đó được Yoon phát triển thêm vào năm 1987, cuối cùng được Hwang, Lai và Liu hoàn thiện vào năm 1993. TOPSIS dựa trên quan điểm rằng phương án được chọn phải có khoảng cách hình học ngắn nhất tới kết quả lý tưởng, và có khoảng cách hình học dài nhất tới kết quả tồi nhất.

Đây là một phương pháp thu nạp tuyến tính các khoản bù trừ để so sánh một tập hợp các lựa chọn thay thế bằng cách xác định trọng số cho từng tiêu chí, chuẩn hóa chúng và tính toán khoảng cách hình học giữa từng phương án với phương án lý tưởng. Một giả định của TOPSIS là các tiêu chí tăng hoặc giảm một cách đơn điệu. Chuẩn hóa thường được yêu cầu vì các tham số hoặc tiêu chí thường có kích thước không giống nhau trong các bài toán đa tiêu chí. Các phương pháp bù trừ như TOPSIS cho phép cân bằng giữa các tiêu chí, trong đó kết quả kém ở một tiêu chí có thể bị phủ định bởi kết quả tốt ở tiêu chí khác. Điều này cung cấp một mô hình mang tính thực tiễn hơn so với các phương pháp phổ thông không hỗ trợ bù trừ.

*Quy trình TOPSIS có thể được tiến hành trong 7 bước:*

**Step 1:** Tạo một Matrix gồm m phương án (alternatives), và n tiêu chí (criteria). Matrix này gọi là “Matrix đánh giá”

**Step 2:** Chuẩn hóa Matrix đánh giá

**Step 3:** Tính toán Matrix quyết định chuẩn hóa có trọng số. Cần lưu ý là mỗi tiêu chí phải có trọng số riêng để tất cả chúng sẽ có tổng bằng 1. **Trọng số** có thể được lấy ngẫu nhiên (không được khuyến nghị) hoặc dựa trên kiến thức chuyên môn (tiêu chuẩn ngành).

và:

**Step 4:** Xác định phương án thay thế tốt nhất và kém nhất cho từng tiêu chí:

*=*

*=*

**Step 5:** Tính toán khoảng cách Euclide giữa phương án mục tiêu và phương án thay thế tốt nhất / xấu nhất:

**Step 6:** Đối với mỗi phương án, hãy tính toán độ tương đồng với phương án xấu nhất. Kết quả là điểm TOPSIS

**Step 7:** Xếp hạng các lựa chọn thay thế theo điểm TOPSIS theo thứ tự tăng dần của

**Trọng số** ở Step 3 nên dựa xin ý kiến của chuyên gia. Hơn thế nữa, trọng số này có đủ đảm bảo chính xác hay không, cần phải dựa theo “Thuật toán AHP (Analytic Hierarchy Process)”

## 2. Phương pháp AHP

AHP là một phương pháp tính toán **trọng số** áp dụng cho các bài toán ra quyết định đa tiêu chuẩn.

**Step 1**: Xây dựng cây phân cấp AHP

Sau khi phân rã vấn đề thành các thành phần nhỏ, cây phân cấp AHP sẽ được xây dựng dựa trên các tiêu chí và các khả năng lựa chọn.

Diagram

Description automatically generated

**Step 2**: Xây dựng Matrix so sánh cặp giữa các tiêu chí (Pairwise comparison Matrix)

Với n tiêu chí, ta có Matrix vuông:

Các chuyên gia sẽ đánh giá các tiêu chí này (mức độ ưu tiên, độ quan trọng giữa chúng)

Các mức độ đánh giá được T.Saaty đề xuất từ 1~9:

A picture containing text, clock

Description automatically generated

**Step 3**: Chuẩn hóa Matrix so sánh cặp (Normalized pairwire comparision matrix)

**Step 4**: Tính trọng số của các tiêu chí (Criterion weights) = Average of row

**Step 5**: Kiểm tra tính nhất quán (Check the consistency)

a. Tính nhất quán:

b. Vector nhất quán:

c. Tính hệ số riêng lớn nhất của Matrix so sánh cặp

*=*

càng gần số tiêu chí (n), thì mức độ phù hợp càng cao)

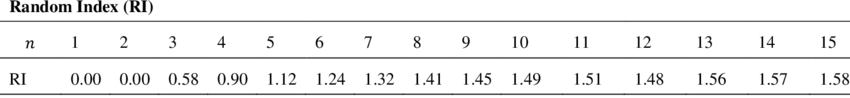
d. Tính chỉ số nhất quán (Consistency Index)

CI =

e. Tính tỷ lệ nhất quán (Consistency Ratio), theo CI và chỉ số ngẫu nhiên (Random consistency index - RI)

CR =

Chỉ số RI được T.Saaty đề xuất theo như bảng sau (Random index for n=15):



* Nếu tỷ lệ nhất quán (Consistency Ratio – CR) < 10%, tức là các phương án đánh giá của chuyên gia là nhất quán

và có thể chấp nhận được.

Tức là ta cố thể dùng trọng số của các tiêu chí () để tính toán các phương án tiếp theo.

* Nếu tỷ lệ nhất quán CR >10%, tính nhất quá không đảm bảo, ta cần chỉnh sửa Matrix so sánh cặp và tính toán lại

**Step 6:** Tính mức độ ưu tiên của các phương án theo từng tiêu chí

Ở bước này sẽ tính toán cho từng tiêu chí, cách tính toán giống như các bước 2→5, nhưng số liệu đưa vào đánh giá là kết quả so sánh mức độ ưu tiên của các phương án xem xét theo từng tiêu chí (theo ý kiến các chuyên gia). Như thế, đánh giá phải thực hiện n ma trận cho n tiêu chí khác nhau. Kết quả là ta có n ma trận 1 cột (n-1) hàng (n-1 phương án). Cũng cần tiến hành kiểm tra tỷ số nhất quán để đảm bảo kết quả thu được có độ tin cậy phù hợp. Sau đó tính điểm cho các phương án và lựa chọn. Đây là bước cuối cùng trong quá trình đánh giá và đưa ra phương án. Từ kết quả của bước 6 này, tổng hợp được ma trận trọng số các phương án theo các tiêu chí. Nhân ma trận này với ma trận trọng số các tiêu chí là kết quả của bước 2→5, được kết quả là một ma trận (n-1) hàng (n-1 phương án) 1 cột (giá trị trọng số). Ma trận kết quả sẽ cho biết phương án tốt nhất nên chọn, là phương án có giá trị trọng số cao nhất.

*Chú ý*: khi giải bài toán trợ giúp tìm kiếm việc làm, ta chỉ tính toán bước 2→5 để thu được giá trị, bước 1 và bước 6 là không cần thiết vì đã áp dụng phương pháp TOPSIS.

# Trình bày phương pháp giải bài toán

Với bài toán trợ giúp người lao động tìm kiếm việc làm ta có thể tiến hành theo các bước sau đây:

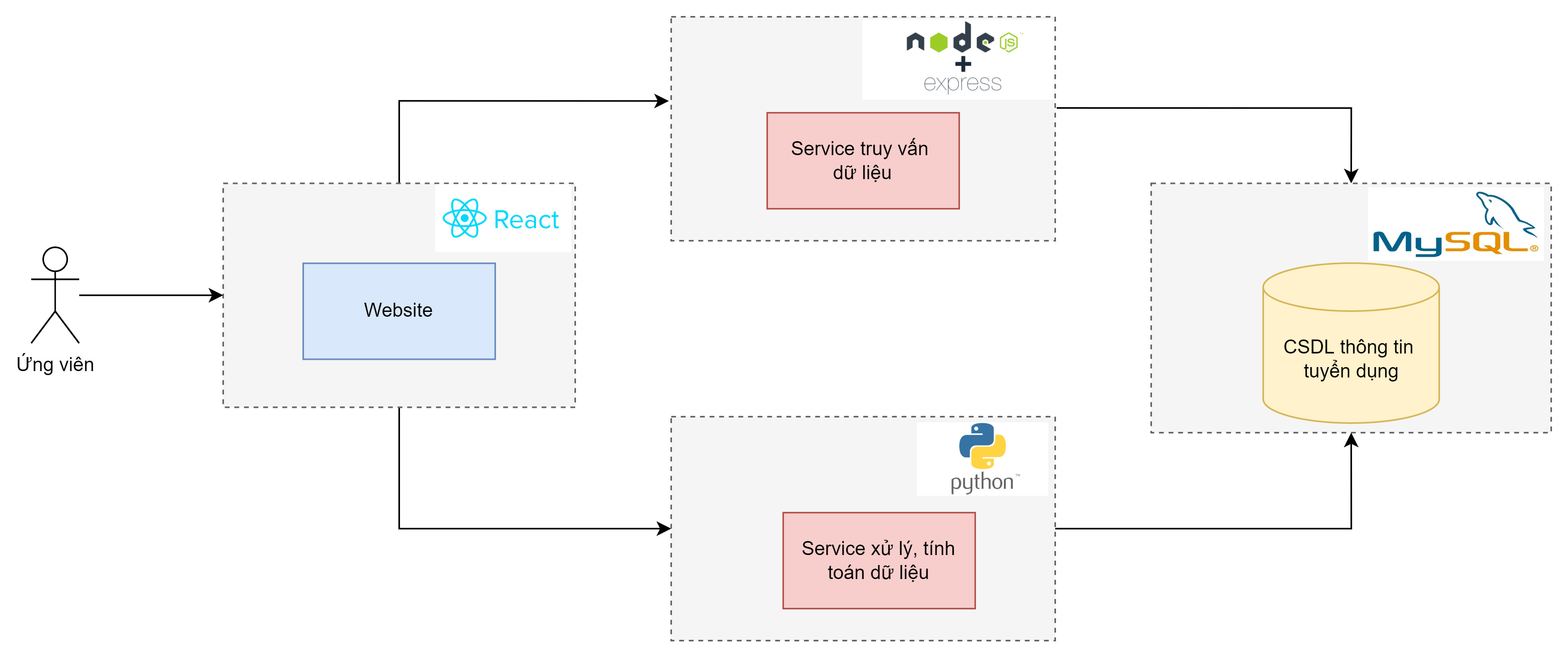
## Thu thập và xử lý dữ liệu

Thu nhận dữ liệu về việc làm từ các nguồn thống kê chất lượng:

* Địa chỉ công ty
* Các loại bằng cấp yêu cầu(chưa tốt nghiệp, đã tốt nghiệp, bằng đại học, bằng thạc sĩ)
* Các loại kỹ năng
* Hình thức làm việc(partime, fulltime)
* Khoảng lương

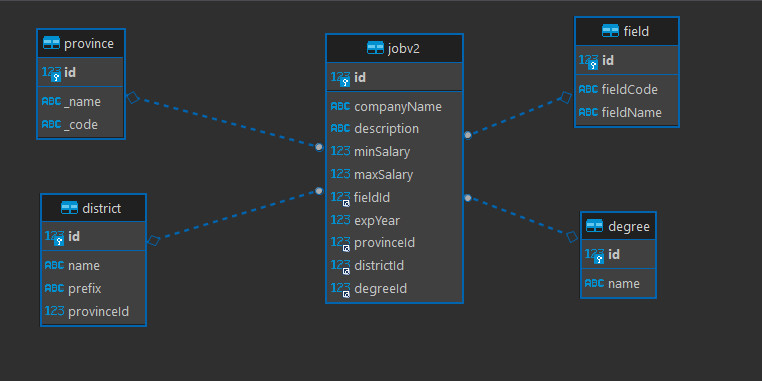
Xây dựng cấu trúc dữ liệu:

* Ứng viên đang muốn tìm kiếm công việc gì?
* Địa điểm làm việc tại Tỉnh/thành, Quận/Huyện nào?
* Mức lương mong muốn là bao nhiêu
* Đang quan tâm đến hình thức làm việc fulltime/partime?



*Hình 3: Sơ đồ thu thập và xử lý dữ liệu*

Tiến hành xử lý dữ liệu để lọc các thông tin cần thiết, tiến hành lưu trữ tới Database local.

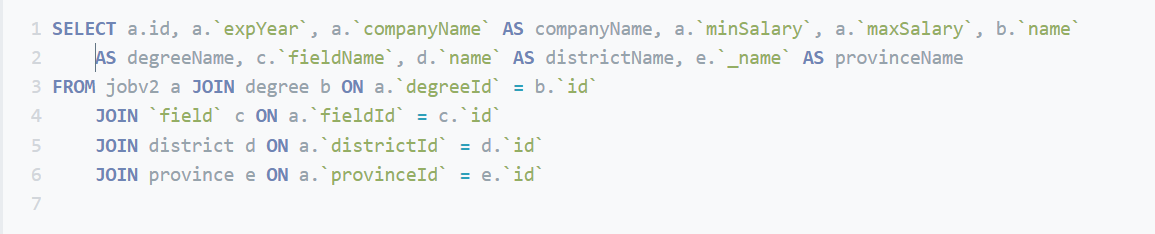


*Hình 4: Sơ đồ quan hệ giữa các thực thể (ERD)*

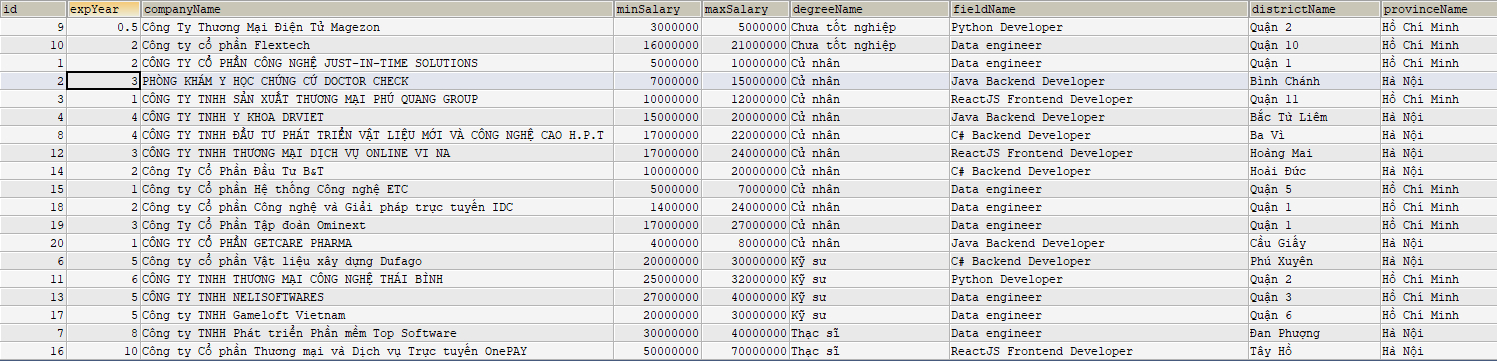
## Tính toán tìm ra phương án việc làm tối ưu nhất

**Bước 1:** Truy vấn dữ liệu

Truy vấn dữ liệu việc làm từ Database:



Kết quả sau khi truy vấn:



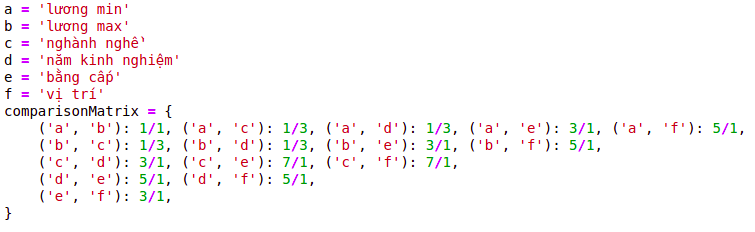
*Hình 5: Kết quả truy vấn dữ liệu từ DB*

**Bước 2:** Thiết lập vector trọng số giữa các tiêu chí

**Bước 3:** Dùng phương pháp AHP để đánh giá độ tin cậy của Vector trọng số trên

**Bước 4:** Đánh giá tỷ lệ nhất quán

Ta đánh trọng số độ quan trọng, tỷ lệ giữa các biến:



Sau khi tính toán dùng APH ta thu được vector trọng số



Trọng số tỷ lệ nhất quán được tính toán ra giá trị: 5.1% < 10%

--> Chấp nhận được tỷ lệ đánh giá giữa các đáp án.

**Bước 5:** Lập bảng quyết định theo phương pháp TOPSIS

Code thuật toán TOPSIS (ngôn ngữ lập trình Python):

Graphical user interface, text, email

Description automatically generated

*Hình 6: Ảnh chụp triển khai giải thuật Topsis bằng ngôn ngữ lập trình Python*

**Bước 6:** Tìm ra bảng xếp hạng, và quyết định công ty nào phù hợp nhất với ứng viên

Gọi module tính toán thuật toán TOPSIS ở trên:

Topsis(evaluation\_matrix, weights, criterias)

Trong đó:

evaluation\_matrix: Matrix định giá các tiêu chí (đã được chuẩn hóa)

weights: Vector trọng số giữa các tiêu chí

criterias: Vector phân loại thuộc tính

TRUE (hoặc 1): Thuộc tính có lợi - càng gần nó càng tốt

FALSE (hoặc 0): Thuộc tính bất lợi - càng xa nó càng tốt

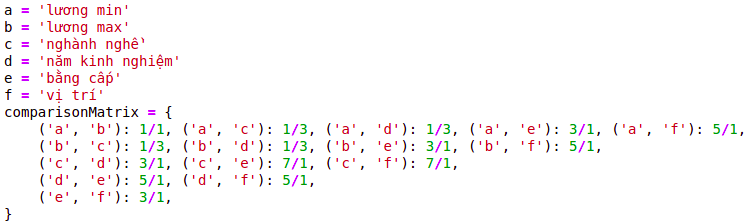
Kết quả ta sẽ nhận được Vector rank từ 1~10 trong đó:

* Càng gần 0 càng tốt, nghĩa là công việc đó phù hợp nhất với mong muốn của ứng viên tìm kiếm
* Càng gần 1 càng xấu, nghĩa là kết quả đó không phù hợp nhất với mong muốn của ứng viên tìm kiếm

Ta có thể mô hình hóa các bước giải bài toán như sơ đồ hình vẽ bên dưới:

# Cài đặt chương trình

## Đánh giá độ tin cậy của Vector trọng số

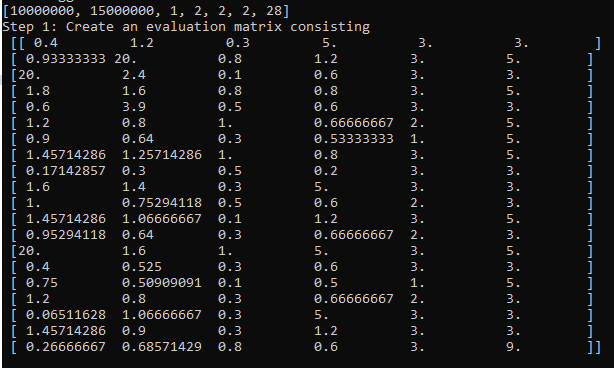
*Hình 12: Matrix độ tương quan giữa các tiêu chí*

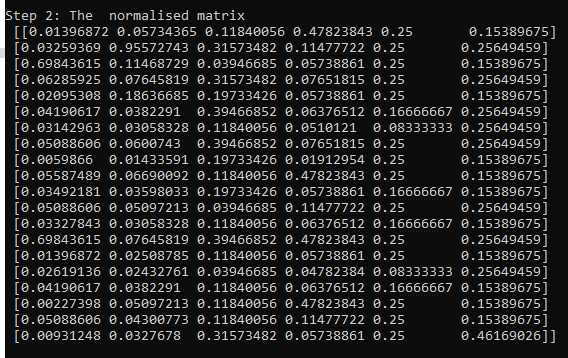
Ta có thể thấy, độ tin cậy của Vector trọng số là: 5.1% < 10%, nên hoàn toàn có thể sử dụng được

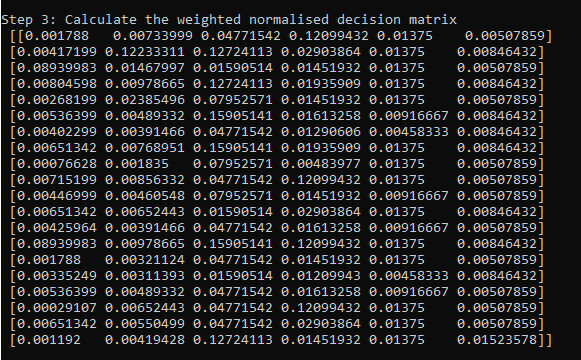
## Lập bảng quyết định

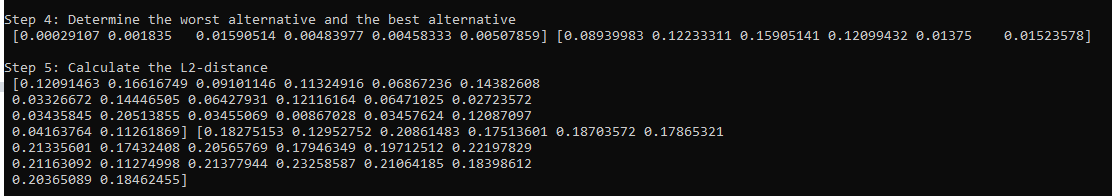
Gọi phương thức TOPSIS để tính toán khoảng cách L2 cho từng tiêu chí.

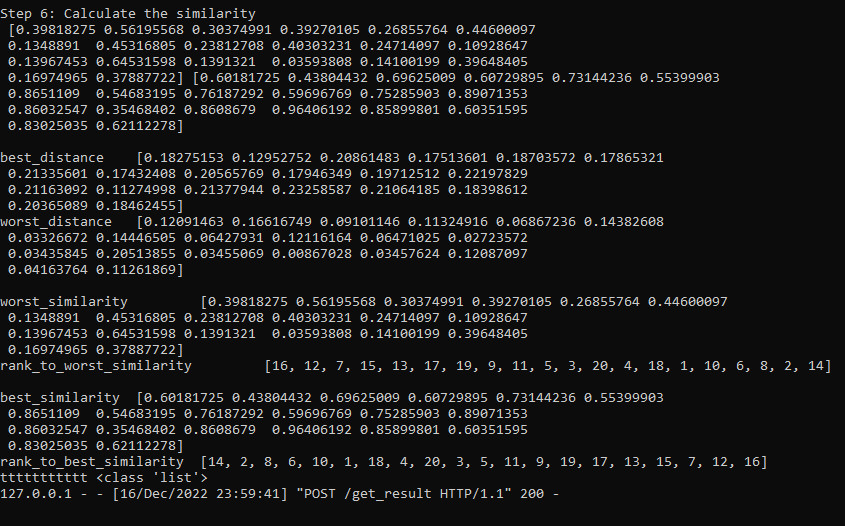
Kết quả thực nghiệm khi chạy tập dữ liệu với từng step của TOPSIS:











*Hình 14: Ảnh chụp kết quả các bước khi tính toán bằng phương pháp TOPSIS*

## Giao diện ứng dụng của bài toán

## Form điền thông tin tìm kiếm việc làm

## Thông tin tìm kiếm sẽ bao gồm:

## Ngành nghề

## Năm kinh nghiệm

## Bằng cấp

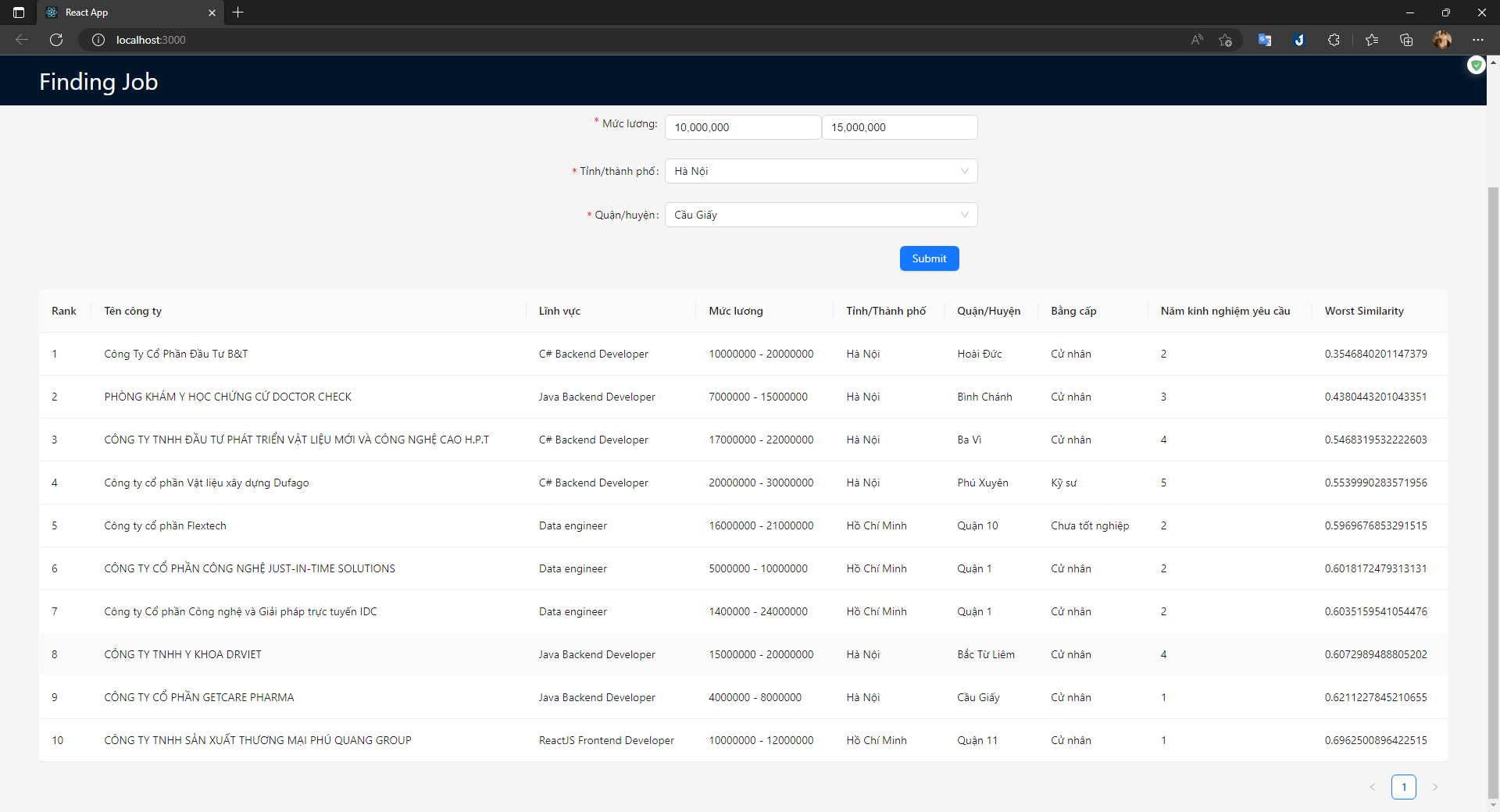
## Mức lương (Min,Max)

## Tỉnh/thành

## Quận/huyện

## 

Kết quả trả về sau khi submit:



Sau khi truy vấn các thông tin mà ứng viên gửi về, hệ thống sẽ tự động tính toán ra trọng số dựa vào phương pháp AHP, tính toán thuật toán TOPSIS và cuối cùng trả về kết quả bao gồm Rank, Tên công ty, Lĩnh vực, Mức lương, Tỉnh/Thành, Quận/huyện, Bằng cấp, Năm kinh nghiệm yêu cầu.

# Đánh giá kết quả

Trong giới hạn của project kết thúc môn học này, nhóm em đã đề xuất một bài toán trợ giúp người lao động tìm kiếm việc làm. Với kết quả đạt được, có thể rút ra kết luận, gồm những ưu điểm và nhược điểm như sau:

## Kết quả đạt được (ưu điểm)

* Đã áp dụng các kiến thức và phương pháp đã học để thực hiện phân tích, tính toán các thông tin thuộc tính
* Xây dựng được một giao diện website trực quan, dễ sử dụng
* Bài toán có tính ứng dụng cao, là nền tảng để mở rộng và triển khai trong thực tế

## Khó khăn (nhược điểm)

* Dữ liệu còn hạn chế, chưa thu thập đầy đủ
* Chưa có chức năng để nhà tuyển dụng tự nhập các thông số giữa các tiêu chí
* Chưa có chức năng để nhà tuyển dụng tự nhập các thông số giữa các tiêu chí

# Kết luận

Qua tìm hiểu để thực hiện và hoàn thành dự án này, nhóm em đã thực sự hiểu được việc áp dụng các kiến thức Toán học, cũng như CNTT vào giải quyết các bài toán/hệ thống thực tiễn một cách thông minh. Về hướng phát triển trong tương lai, nếu có thể tiến hành cài đặt hệ thống trong thực tiễn, đánh giá độ hiệu quả và tinh chỉnh để phù hợp với thực tiễn yêu cầu, sẽ phần nào góp phần giải quyết bài toán hỗ trợ người lao động tìm kiếm việc làm một cách hiệu quả nhất .  
 Chúng em xin gửi lời cảm ơn sâu sắc đến GS.TS Trần Đình Khang, người đã góp ý và giúp đỡ rất nhiều để em có cơ hội thực hiện đồ án tuyệt vời này.

**References**

[1] Programming with Mosh, “Python Tutorial - Python Full Course for Beginners”,

https://www.youtube.com/watch?v=\_uQrJ0TkZlc

[2] freeCodeCamp.org, “Data Analysis with Python - Full Course for Beginners (Numpy, Pandas, Matplotlib, Seaborn)”,   
https://www.youtube.com/watch?v=r-uOLxNrNk8

Mục lục

[TRỢ GIÚP NGƯỜI LAO ĐỘNG TÌM KIẾM VIỆC LÀM 1](#__RefHeading___Toc858_2270627991)

[Abstract 1](#__RefHeading___Toc860_2270627991)

[I. Giới thiệu bài toán 1](#__RefHeading___Toc862_2270627991)

[II. Background 2](#__RefHeading___Toc864_2270627991)

[1. Phân loại dùng phương pháp TOPSIS 2](#__RefHeading___Toc866_2270627991)

[2. Phương pháp AHP 4](#__RefHeading___Toc868_2270627991)

[III. Trình bày phương pháp giải bài toán 7](#__RefHeading___Toc870_2270627991)

[1. Thu thập và xử lý dữ liệu 7](#__RefHeading___Toc872_2270627991)

[2. Tính toán tìm ra phương án việc làm tối ưu nhất 8](#__RefHeading___Toc874_2270627991)

[IV. Cài đặt chương trình 11](#__RefHeading___Toc876_2270627991)

[1. Đánh giá độ tin cậy của Vector trọng số 11](#__RefHeading___Toc878_2270627991)

[2. Lập bảng quyết định 11](#__RefHeading___Toc880_2270627991)

[3. Giao diện ứng dụng của bài toán 13](#__RefHeading___Toc882_2270627991)

[3.1 Form điền thông tin tìm kiếm việc làm 13](#__RefHeading___Toc884_2270627991)

[Thông tin tìm kiếm sẽ bao gồm: 13](#__RefHeading___Toc886_2270627991)

[Ngành nghề 13](#__RefHeading___Toc888_2270627991)

[Năm kinh nghiệm 13](#__RefHeading___Toc890_2270627991)

[Bằng cấp 13](#__RefHeading___Toc892_2270627991)

[Mức lương (Min,Max) 13](#__RefHeading___Toc894_2270627991)

[Tỉnh/thành 13](#__RefHeading___Toc896_2270627991)

[Quận/huyện 13](#__RefHeading___Toc898_2270627991)

[V. Đánh giá kết quả 15](#__RefHeading___Toc900_2270627991)

[1. Kết quả đạt được (ưu điểm) 15](#__RefHeading___Toc902_2270627991)

[2. Khó khăn (nhược điểm) 15](#__RefHeading___Toc904_2270627991)

[VI. Kết luận 15](#__RefHeading___Toc906_2270627991)