**ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**KHOA CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO MÔN HỌC**

**ỨNG DỤNG KHOA HỌC DỮ LIỆU**

**Đề tài :**

**HỆ THỐNG GIỚI THIỆU VIỆC LÀM THÔNG QUA CV NGƯỜI DÙNG**

**GIẢNG VIÊN HƯỚNG DẪN :**

Thầy Lê Ngọc Thành

Thầy Lê Nhựt Nam

**KHÓA :** K20

**NGÀNH : Khoa học dữ liệu**

**NHÓM :** Hopeless

**Thành phố Hồ Chí Minh, ngày 30 tháng 03 năm 2024**

**Mục Lục**

**Thành viên nhóm**

|  |  |
| --- | --- |
| **Họ và tên** | **Mã số** |
| Nguyễn Quang Gia Bảo | 20120136 |
| Huỳnh Tuấn Nam | 20120136 |
| Trần Hoàng Anh Phi | 20120158 |

**Bảng tóm tắt**

**Danh sách bảng**

**Danh sách hình ảnh minh họa**

1. **Giới thiệu đề tài**
   1. **Mục tiêu đề tài**

Mục tiêu của đề tài này hướng đến tạo ra một hệ thống đánh giá phần trăm (%) phù hợp giữa CV của một ứng viên so với tất cả các công ty đang tuyển dụng trên thị trường hiện tại.

* 1. **Lí do chọn đề tài**

Đối với các bạn sinh viên năm 4 chuẩn bị ra trường, hoặc những bạn sinh viên có ý định thực tập sớm để có kinh nghiệm, việc chọn một công việc phù hợp với khả năng bản thân hiện tại là vô cùng khó khăn, vì có rất nhiều thông tin ứng tuyển trên nhiều nên tảng khác nhau. Sự cần thiết lúc này đó là một hệ thống giúp các ứng viên đó lọc, chọn lựa ra các công việc thật sự phù hợp với mình, nhằm tránh chọn phải các công việc yêu cầu các kĩ năng nâng cao, mà các ứng viên có thể bỏ qua trong quá trình đọc thông tin ứng tuyển.

Bên cạnh các bạn sinh viên mới thì hệ thống cũng giúp đỡ các bạn có kinh nghiệm khác đang có như cầu tìm kiếm một môi trường làm việc mới có bản thân mình. Để các bạn có thể tìm được các công việc có mức lương phù hợp với khả năng, năng lực của mình.

* 1. **Ý tưởng thực hiện đồ án**

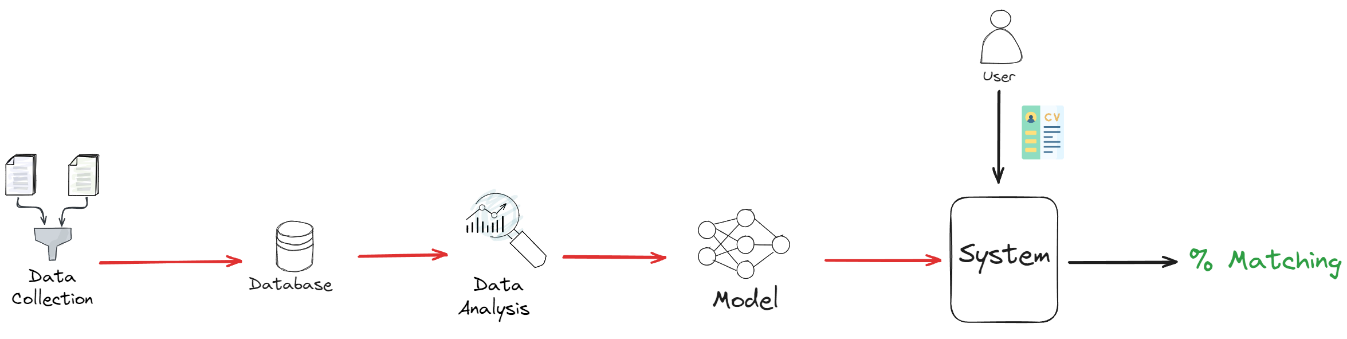
****

Figure 1: Luồng xử lí

Đối với đề tài này, vấn đề chính nằm ở bước xây dựng Model. Để có thể đáp ứng đầu ra là đánh giá mức phù hợp giữa CV ứng viên và yêu cầu tuyển dụng, mô hình phải xử lí tốt ngôn ngữ tự nhiên.

* 1. **Lợi ích sau khi thực hiện đề tài**

Các ứng viên sẽ dễ dàng trong việc tìm ra công việc phù hợp với bản thân mình thông qua sự tương thích giữa CV của họ và yêu cầu công việc. Hệ thống sẽ phân tích CV và cho ra các bài đăng hợp lí, giúp các ứng viên tiết kiệm thời gian tìm kiếm, truy xuất từng bài. Mặc khác các công ty ứng tuyển sẽ có được nguồn nhân sự cần thiết cho vị trí đang thiếu.

1. **Nguồn dữ liệu**

Dữ liệu của bài toán này có thể nói là rất nhiều hiện tại. Để tìm được nguồn dữ liệu uy tín cũng không phải là khó. Các nguồn dữ liệu mà nhóm tin dùng là:

* **Facebook**: đây là nền tảng mạng xã hội rất phổ biến hiện tại, và có rất nhiều người dùng. Có rất nhiều các nhóm do một số người dùng thành lập nhầm đăng các thông tin tuyển dụng dưới hình thức là các bài “post”.
* **Các trang web uy tín:** TopCV, Vietnamworks, hiện tại đây là hai trang “web” chính mà nhóm đang sử dụng cho việc thu thập dữ liệu. Đây là hai trang “web” có độ uy tín rất cao, và hầu như các nhà tuyển dụng đều tin dùng. Vì thế sẽ an toàn đối với các bạn sinh viên mới.

1. **Những giới hạn, phạm vi thực hiện**

Trước tiên, vấn đề chính của đề tài là tính toán và đưa ra chỉ số phù hợp (% matching) giữa CV ứng viên và yêu cầu ứng tuyển của các công ty, đó là:

* Tìm kiếm sự tương đồng trong CV của ứng viên và thông tin ứng tuyển đề cập (ngôn ngữ lập trình được sử dụng, năm kinh nghiệm, vị trí ứng tuyển, trình độ tiếng anh,... )

Sau khi hoàn thành được mục tiêu chính của bài toán, nhóm sẽ tiếp tục nghiên cứu để đưa ra các mục tiêu để tăng độ chính xác hơn. Ví dụ như là:

* Sinh viên vừa ra trường chưa có việc làm, hoặc các bạn sinh viên còn đang đi học nhưng có ý định thực tập sớm, thì hệ thống sẽ ưu tiên lựa chọn các công ty tuyển “không yêu cầu năng lực, có đào tạo”, hoặc các từ khóa như “Intern, Thực tập sinh”.
* Những từ khóa đặc biệt sẽ tương thích với những năm kinh nghiệm. Ví dụ như để trở thành một Java Expired, cần phải có ít nhất vài năm kinh nghiệm làm việc với Java, giả sử là 3 năm. Hệ thống sẽ dựa vào một số tự khóa đặc biệt đó để chọn ra các công việc trên 3 năm kinh nghiệm.
* Một bạn điền vào CV về ngôn ngữ phổ biến của bản thân là Cpp. Vậy liệu bạn ấy có thể thực tập ở một công ty đào tạo bạn sử dụng C#, Java hay không ( vì Cpp, C#, java có điểm khá tương đồng với nhau về cách xây dựng).

…

1. **Đầu ra dự kiến**

Đối với đầu ra của đề tài là một sản phẩm “web”, hoặc “app” tương tác giữa người dùng và hệ thống. Người dùng sẽ “upload” CV cá nhân lên, hệ thống sẽ đưa ra cho người dùng một lượng công việc (có giới hạn số lượng) phù hợp với CV mà ứng viên đưa ra và tỉ lệ phù hợp với công việc đó.

1. **Thời gian thực hiện**

Thời gian dự kiến để thực hiện đề tài được chia ra: ( có 13 tuần để thực hiện đề tài )

* **2 tuần đầu tiên của đề tài**: nhóm sẽ xây dựng hệ thống thu thập dữ liệu từ các trang web trên, và việc thu thập này sẽ tiếp tục diễn ra cho đến khi báo cáo sản phẩm
* **2 - 3 tuần tiếp theo**: sẽ tiến hành xử lí dữ liệu thu thập ( tạo trường dữ liệu, làm sạch dữ liệu, …) và phân tích dữ liệu.
* **3 tuần tiếp theo**: sẽ ưu tiên cho việc xây dựng Model xử lí vấn đề chính của đề tài.
* **2 tuần tiếp theo**: sẽ xây dựng giao diện, hệ thống tương tác với người dùng.
* **Các tuần còn lại**: Cập nhật thêm các mục tiêu đề ra bên trên và xử lí các lỗi trước khi báo cáo.