ĐẠI HỌC HUẾ



**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**

🙠🙟🕮🙝🙢

Icon

Description automatically generated

**BÁO CÁO ĐỒ ÁN**

**HỌC KÌ I, năm học 2022-2023**

**Học phần:**

**CẤU TRÚC DỮ LIỆU & GIẢI THUẬT**

**Số phách**

**(**Do hội đồng chấm ghi thi)

***Thừa Thiên Huế, tháng 04 năm 2023.***

ĐẠI HỌC HUẾ



**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**

🙠🙟🕮🙝🙢

Icon

Description automatically generated

**BÁO CÁO ĐỒ ÁN**

**HỌC KÌ I, năm học 2022-2023**

**Học phần:**

**CẤU TRÚC DỮ LIỆU & GIẢI THUẬT**

**Giảng viên hướng dẫn:** Nguyễn Thanh Nam.

**Lớp:** Khoa học dữ liệu và Trí tuệ nhân tạo khóa 3.

**Sinh viên thực hiện:** Phạm Phước Bảo Tín\_22E1020021.

(ký và ghi rõ họ tên)

**Số phách**

**(**Do hội đồng chấm ghi thi)

***Thừa Thiên Huế, tháng 04 năm 2023***

ĐẠI HỌC HUẾ

**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**



**PHIẾU ĐÁNH GIÁ ĐỒ ÁN/TIỂU LUẬN/BÀI TẬP LỚN**

**Học kỳ II, năm học 2022 - 2023**

|  |  |
| --- | --- |
| **Cán bộ chấm thi 1** | **Cán bộ chấm thi 2** |
| **Nhận xét:**  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  **Điểm đánh giá của CBCT1:**  Bằng số: .........................................  Bằng chữ: ...................................... | **Nhận xét:**  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  **Điểm đánh giá của CBCT2:**  Bằng số: .........................................  Bằng chữ: ....................................... |

Điểm kết luận: .........................................................................................................

Bằng số: ...................................................................................................................

Bằng chữ: ................................................................................................................

*Thừa Thiên Huế, ngày 07 tháng 06 năm 2023*

|  |  |
| --- | --- |
| **Cán bộ chấm thi 1**  *(Ký và ghi rõ họ và tên)* | **Cán bộ chấm thi 2**  *(Ký và ghi rõ họ và tên)* |

# LỜI CẢM ƠN

Được trở thành sinh viên Khoa Kỹ Thuật và Công Nghệ - Đại học Huế em rất hạnh phúc và biết ơn. Hạnh phúc vì mình đã đạt được mục tiêu mong muốn và biết ơn sự cống hiến, chỉ bảo tận tình sâu sắc của quý thầy cô trong khoa đồng thời đã tạo điều kiện học tập lí tưởng cho chúng em. Để hoàn thành đồ án một cách chỉnh chu nhất có thể em xin gửi lời cảm ơn đến thầy giáo bộ môn – Thầy Nguyễn Thanh Nam đã hướng dẫn tận tình, chi tiết cho chúng em trong quá trình hoàn thành đồ án lẫn quá trình học tập học của sinh viên chúng em. Hi vọng rằng thời gian sắp tới em sẽ luôn cố gắng, nổ lực hơn nữa trong học tập chuyên nghành của mình.

Trong quá trình hoàn thành đồ án mặc dù em đã chuẩn bị kĩ nhưng không thể tránh khỏi những sai sót, em mong nhận được sự góp ý từ quý thầy, cô. Lời cuối cùng em xin kính chúc quý thầy, cô thật nhiều sức khỏe để tiếp tục dẫn dắt chúng em và những thế hệ tiếp theo thành người.

# DANH MỤC HÌNH ẢNH

[Hình 1: Mô tả thuật toán đệ quy Tháp Hà Nội 1](#_Toc135861432)

[Hình 2: Kết quả bài toán tháp Hà Nội với 3 đĩa 2](#_Toc135861433)

[Hình 3: Kết quả bài toán mã đi tuần với bàn cờ 8x8 4](#_Toc135861434)

[Hình 4: Kết quả bài toán 8 quân hậu. 6](#_Toc135861435)

[Hình 5: Kết quả thực thi với danh sách liên kết đơn 10](#_Toc135861436)

[Hình 6: Kết quả thực thi với danh sách liên kết đôi 13](#_Toc135861437)

[Hình 7: Kết quả thao tác với ngăn xếp 14](#_Toc135861438)

[Hình 8: Kết quả thao tác với hàng đợi 16](#_Toc135861439)

[Hình 9: Mô tả cây 20](#_Toc135861440)

[Hình 10: Kết quả duyệt cây theo thứ tự trước và sau 21](#_Toc135861441)

[Hình 11: Kết quả bài toán đồ thị vô hướng 22](#_Toc135861442)

[Hình 12: Kết quả bài toán đồ thị có hướng 24](#_Toc135861443)

[Hình 13: Kết quả sắp xếp chọn 25](#_Toc135861444)

[Hình 14: Kết quả sắp xếp chèn 27](#_Toc135861445)

[Hình 15: Kết quả sắp xếp nổi bọt 28](#_Toc135861446)

[Hình 16: Kết quả Quick sort 29](#_Toc135861447)

[Hình 17: Kết quả sắp xếp vun đống 30](#_Toc135861448)

[Hình 18: Kết quả thuật toán sắp xếp trộn 32](#_Toc135861449)

# NHIỆM VỤ MỖI THÀNH VIÊN

Sau khi hội nhóm để phân công nhiệm vụ hoàn thành đồ án, Nhóm 5 xin được trình bày về công việc mỗi thành viên như sau:

1. Trần Tùng Dương: Giới thiệu dữ liệu, mô tả dữ liệu, viết báo cáo đồ án;
2. Trịnh Quốc Dân: Đọc dữ liệu và xử lí dữ liệu, vẽ biểu đồ, bài toán ước lượng;
3. Văn Khiêm Chương: Gộp nhóm, bài toán tương quan;
4. Phạm Phước Bảo Tín: Thống kê cơ bản, bài toán kiểm định.

# MỤC LỤC

[LỜI CẢM ƠN i](#_Toc137195505)

[DANH MỤC HÌNH ẢNH ii](#_Toc137195506)

[NHIỆM VỤ MỖI THÀNH VIÊN iii](#_Toc137195507)

[MỤC LỤC iv](#_Toc137195508)

[PHẦN 1: KHÁI QUÁT DỮ LIỆU CỦA NHÓM 1](#_Toc137195509)

[1.1 Giới thiệu dữ liệu 1](#_Toc137195510)

[1.2 Mô tả dữ liệu 1](#_Toc137195511)

[1.2.1 Thực thể 1](#_Toc137195512)

[1.2.2 Thuộc tính 1](#_Toc137195513)

[PHẦN 2: THAO TÁC VỚI DỮ LIỆU BẰNG PYTHON 3](#_Toc137195514)

[2.1 Xử lí dữ liệu 3](#_Toc137195515)

[2.1.1 Đọc dữ liệu 3](#_Toc137195516)

[2.1.2 Lọc dữ liệu 3](#_Toc137195517)

[2.2 Gộp nhóm và thống kê cơ bản 3](#_Toc137195518)

[2.2.1 Gộp nhóm 3](#_Toc137195519)

[2.2.2 Thống kê cơ bản 3](#_Toc137195520)

[2.3 Trực quan hóa dữ liệu 3](#_Toc137195521)

[2.3.1 Biểu đồ đường 3](#_Toc137195522)

[2.3.2 Biểu đồ cột 3](#_Toc137195523)

[2.3.3 Biểu đồ tròn 3](#_Toc137195524)

[PHẦN 3: BÀI TOÁN TƯƠNG QUAN, ƯỚC LƯỢNG, KIỂM ĐỊNH 4](#_Toc137195525)

[3.1 Tương quan 4](#_Toc137195526)

[3.2 Ước lượng 4](#_Toc137195527)

[3.3 Kiểm định 4](#_Toc137195528)

# PHẦN 1: KHÁI QUÁT DỮ LIỆU CỦA NHÓM

## Giới thiệu dữ liệu

Trong bài báo cáo này nhóm sử dụng dữ liệu từ file "ds\_salaries.xlsx" thu thập trên internet là một tập hợp các thông tin về mức lương của nhân viên trong lĩnh vực khoa học dữ liệu và máy học. File này chứa các thuộc tính như năm làm việc, mức độ kinh nghiệm, loại hình làm việc, chức danh công việc, mức lương, đơn vị tiền tệ, mức lương chuyển đổi sang đơn vị USD, quốc gia cư trú của nhân viên, tỷ lệ làm việc từ xa, địa điểm của công ty và quy mô công ty.

Thông qua dữ liệu này, chúng ta có thể khám phá và phân tích các yếu tố quan trọng liên quan đến mức lương trong lĩnh vực này. Chẳng hạn, chúng ta có thể tìm hiểu sự ảnh hưởng của mức độ kinh nghiệm, chức danh công việc và địa điểm làm việc đến mức lương của nhân viên. Ngoài ra, cũng có thể khám phá sự tương quan giữa mức lương và các yếu tố như loại hình làm việc, quốc gia cư trú, tỷ lệ làm việc từ xa và quy mô công ty.

Dữ liệu trong file "ds\_salaries.xlsx" cung cấp thông tin đa dạng và chi tiết, cho phép chúng ta thực hiện các phân tích và so sánh để hiểu rõ hơn về thị trường lương trong ngành khoa học dữ liệu và máy học.

## Mô tả dữ liệu

### 1.2.1 Thực thể

Trong tập dữ liệu của nhóm có hai thực thể chính được đề cập:

1. Nhân viên (Employee): Đại diện cho các cá nhân làm việc trong lĩnh vực khoa học dữ liệu và máy học. Thông tin về nhân viên bao gồm mức độ kinh nghiệm, loại hình làm việc, chức danh công việc, mức lương, đơn vị tiền tệ, mức lương chuyển đổi sang USD, quốc gia cư trú và tỷ lệ làm việc từ xa.
2. Công ty (Company): Đại diện cho các công ty hoạt động trong lĩnh vực khoa học dữ liệu và máy học. Thông tin về công ty bao gồm địa điểm, quy mô và vị trí của công ty.

Các thuộc tính trong file đề cập đến các thông tin của các thực thể này và cho phép phân tích và so sánh các yếu tố liên quan đến mức lương và các yếu tố khác trong lĩnh vực này.

### 1.2.2 Thuộc tính

Dữ liệu cần phân tích chứa thông tin về mức lương của các nhân viên trong lĩnh vực khoa học dữ liệu và máy học. Các thuộc tính trong dữ liệu được mô tả như sau:

1. work\_year: Năm làm việc, đại diện cho năm mà dữ liệu lương được thu thập.
2. experience\_level: Mức độ kinh nghiệm của nhân viên, được biểu thị bằng các mã viết tắt, ví dụ SE (Senior Engineer) hoặc MI (Mid-level).
3. employment\_type: Loại hình làm việc, có thể là FT (toàn thời gian) hoặc CT (hợp đồng).
4. job\_title: Chức danh công việc của nhân viên, ví dụ Principal Data Scientist, ML Engineer, Applied Scientist, Data Scientist,...
5. salary: Mức lương của nhân viên, được ghi bằng đơn vị tiền tệ tương ứng với thuộc tính salary\_currency.
6. salary\_currency: Đơn vị tiền tệ được sử dụng để biểu thị mức lương, ví dụ EUR (Euro) hoặc USD (US Dollar).
7. salary\_in\_usd: Mức lương chuyển đổi sang đơn vị USD (US Dollar), cho phép so sánh trực tiếp các mức lương dựa trên cùng một đơn vị.
8. employee\_residence: Quốc gia hoặc khu vực cư trú của nhân viên.
9. remote\_ratio: Tỷ lệ làm việc từ xa (remote) của nhân viên, được biểu thị dưới dạng phần trăm (%).
10. company\_location: Địa điểm của công ty, tương ứng với quốc gia hoặc khu vực.
11. company\_size: Quy mô của công ty, có thể là L (lớn) hoặc M (vừa).

Đây là các thuộc tính mô tả thông tin quan trọng về mức lương, chức danh công việc, đơn vị tiền tệ, kinh nghiệm, quốc gia cư trú và các yếu tố liên quan đến công ty và loại hình làm việc.

# PHẦN 2: THAO TÁC VỚI DỮ LIỆU BẰNG PYTHON

## 2.1 Xử lí dữ liệu

### 2.1.1 Đọc dữ liệu

### 2.1.2 Lọc dữ liệu

## 2.2 Gộp nhóm và thống kê cơ bản

### 2.2.1 Gộp nhóm

### 2.2.2 Thống kê cơ bản

## 2.3 Trực quan hóa dữ liệu

### 2.3.1 Biểu đồ đường

### 2.3.2 Biểu đồ cột

### 2.3.3 Biểu đồ tròn

# PHẦN 3: BÀI TOÁN TƯƠNG QUAN, ƯỚC LƯỢNG, KIỂM ĐỊNH

### 3.1 Tương quan

### 3.2 Ước lượng

### 3.3 Kiểm định