TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP

KHOA ĐIỆN TỬ

A circular logo with a star and a blue star with a red and yellow circle

Description automatically generated with medium confidence

**BÀI TẬP LỚN**

**KHOA HỌC DỮ LIỆU**

**BỘ MÔN: CÔNG NGHỆ THÔNG TIN**

**HỆ : ĐẠI HỌC CHÍNH QUY**

ĐỀ TÀI: XÂY DỰNG CHƯƠNG TRÌNH NHẬN DẠNG KHUÔN MẶT ỨNG DỤNG VÀO BÀI TOÁN ĐIỂM DANH

THÁI NGUYÊN - 2024

TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP

KHOA ĐIỆN TỬ

A circular logo with a blue star and a yellow and red circle

AI-generated content may be incorrect.

**BÀI TẬP KHOA HỌC DỮ LIỆU**

**BỘ MÔN : CÔNG NGHỆ THÔNG TIN**

ĐỀ TÀI: XÂY DỰNG CHƯƠNG TRÌNH NHẬN DẠNG KHUÔN MẶT ỨNG DỤNG VÀO BÀI TOÁN ĐIỂM DANH

|  |  |  |
| --- | --- | --- |
| **Giảng viên hướng dẫn** | **:** | **TS. Nguyễn Văn Huy** |
| **Họ và tên sinh viên** | **:** | **La Đức Thắng** |
| **Ngành học** | **:** | **Kỹ thuật Máy tính** |
| **MSSV** | **:** | **K215480106120** |
| **Lớp** | **:** | **K57KMT.01** |

THÁI NGUYÊN - 2024

|  |  |
| --- | --- |
| TRƯỜNG ĐHKTCN | CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM |
| KHOA ĐIỆN TỬ | Độc lập - Tự do - Hạnh phúc |

BÀI TẬP LỚN

KHOA HỌC DỮ LIỆU

**Sinh viên:** La Đức Thắng

**Lớp:** K57KMT **Khoá:** 2021-2025

**Bộ môn:** Công Nghệ Thông Tin

**Giáo viên hướng dẫn:** TS. Nguyễn Văn Huy

**Tên đề tài** : Xây dựng mô hình nhận diện khuôn mặt ứng dụng vào bài toán điểm danh

Yêu cầu :

* Phân tích yêu cầu, dữ liệu
* Trình bày phương pháp học máy được sử dụng
* Huấn luyện và kiểm thử mô hình
* Giao diện chương trình ứng dụng: Web, android, Windows app

|  |  |  |
| --- | --- | --- |
| BCN KHOA | TRƯỞNG BỘ MÔN | GIÁO VIÊN HƯỚNG DẪN |
| (Ký và ghi rõ họ tên) | (Ký và ghi rõ họ tên) | (Ký và ghi rõ họ tên) |

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

Xếp loại:……………… Điểm :……

Thái Nguyên, ngày….tháng…..năm 20....

GIÁO VIÊN HƯỚNG DẪN

(Ký và ghi rõ họ tên)

MỤC LỤC

[NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN 4](#_Toc199153849)

[DANH MỤC TỪ VIẾT TẮT 7](#_Toc199153850)

[LỜI NÓI ĐẦU 8](#_Toc199153851)

[LỜI CẢM ƠN 9](#_Toc199153852)

[CHƯƠNG I: KHÁI QUÁT VỀ BÀI TOÁN 10](#_Toc199153853)

[1.1. Đặt vấn đề 10](#_Toc199153854)

[1.2. Hướng giải quyết 10](#_Toc199153855)

[CHƯƠNG II: CƠ SỞ LÝ THUYẾT 11](#_Toc199153856)

[2.1. Pandas 11](#_Toc199153857)

[2.2. Matplotlib 11](#_Toc199153858)

[CHƯƠNG III: THIẾT KẾ VÀ XÂY DỰNG CHƯƠNG TRÌNH 13](#_Toc199153859)

[3.1. Sơ đồ khối hệ thống 13](#_Toc199153860)

[3.2. Sơ đồ khối các thuật toán chính 13](#_Toc199153861)

[3.3. Cấu trúc dữ liệu 13](#_Toc199153862)

[3.4. Chương trình 13](#_Toc199153863)

[CHƯƠNG IV: THỰC NGHIỆM VÀ KẾT LUẬN 14](#_Toc199153864)

[4.1. Thiết kế hệ thống 14](#_Toc199153865)

[4.1.1. Phân tích chương trình 14](#_Toc199153866)

[4.1.2. Phân tích hệ thống 15](#_Toc199153867)

[4.1.3. Cơ sở dữ liệu 15](#_Toc199153868)

[4.1.4. Giao diện chương trình 16](#_Toc199153869)

[4.2. Kiểm thử chương trình 17](#_Toc199153870)

[CHƯƠNG V: TỔNG KẾT 20](#_Toc199153871)

[5.1. Kết quả đã đạt được 20](#_Toc199153872)

[5.2. Hướng phát triển của đề tài 20](#_Toc199153873)

[KẾT LUẬN 21](#_Toc199153874)

[TÀI LIỆU THAM KHẢO 22](#_Toc199153875)

# DANH MỤC TỪ VIẾT TẮT

1. CSDL: Cơ sở dữ liệu.
2. CNTT: Công nghệ thông tin

LỜI NÓI ĐẦU

Trong quá trình thực hiện đề tài, em đã cố gắng tìm hiểu, nghiên cứu và áp dụng những kiến thức, kỹ năng đã học để xây dựng một chương trình hoàn chỉnh, có khả năng hoạt động ổn định và đáp ứng được yêu cầu thực tế. Tuy nhiên, do kiến thức và kinh nghiệm còn hạn chế, chắc chắn sẽ không tránh khỏi những thiếu sót. Em rất mong nhận được sự góp ý, đánh giá từ quý thầy cô và các bạn để đề tài được hoàn thiện hơn.

LỜI CẢM ƠN

Trong suốt quá trình học tập và thực hiện bài tập lớn, em đã nhận được sự giúp đỡ tận tình của thầy giáo TS. Nguyễn Văn Huy trong bộ môn Tin học công nghiệp – Khoa Điện tử - Trường Đại học Kỹ thuật Công Nghiệp – Đại học Thái Nguyên. Em bày tỏ lòng biết ơn thầy đã tận tình giúp đỡ, hướng dẫn em trong thời gian thực hiện đề tài này.

Mặc dù đã cố gắng hết sức, song do điều kiện thời gian và kinh nghiệm thực tế còn ít, cho nên đề tài không thể tránh khỏi thiếu sót. Vì vậy, em rất mong nhận được sự đóng góp ý kiến của các thầy giáo, cô giáo và các bạn.

Em xin chân thành cảm ơn!

CHƯƠNG I: KHÁI QUÁT VỀ BÀI TOÁN

## Đặt vấn đề

## Hướng giải quyết

CHƯƠNG II: CƠ SỞ LÝ THUYẾT

## 2.1. Pandas

Pandas là một thư viện mã nguồn mở mạnh mẽ của Python, được thiết kế đặc biệt để xử lý và phân tích dữ liệu dễ dàng và hiệu quả. Tên "Pandas" bắt nguồn từ "Panel Data" (dữ liệu bảng) hoặc "Python Data Analysis".

Cấu trúc dữ liệu chính:

* Series: Giống như một mảng 1 chiều (có nhãn index)
* DataFrame: Cấu trúc 2 chiều giống bảng tính Excel hoặc bảng SQL

Tính năng chính:

* Đọc/ghi dữ liệu từ nhiều định dạng (CSV, Excel, SQL, JSON,...)
* Lọc, sắp xếp, nhóm dữ liệu
* Xử lý dữ liệu thiếu (NaN)
* Thống kê cơ bản và nâng cao
* Hợp nhất và biến đổi dữ liệu

Pandas có thể tích hợp tốt với Matplotlib/Seaborn để trực quan hóa. Phù hợp với các công việc liên quan tới khoa học dữ liệu hoặc phân tích dữ liệu.

## 2.2. Matplotlib

Matplotlib là thư viện đồ họa mạnh mẽ và phổ biến nhất của Python, dùng để trực quan hóa dữ liệu dưới dạng các biểu đồ 2D và 3D chuyên nghiệp. Nó được thiết kế để mô phỏng cách vẽ đồ thị trong MATLAB nhưng với cú pháp Python.

`pyplot` là module thường dùng nhất, cung cấp cách tiếp cận giống MATLAB – thuận tiện khi viết code nhanh.

2.3. Scikit-Learn

Scikit-Learn là một thư viện mã nguồn mở nổi bật trong Python dùng để triển khai thuật toán ML, cung cấp:

- Các mô hình học máy chuẩn hóa: classification, regression, clustering,...

- Giao diện API thống nhất, dễ học, dễ dùng

- Tích hợp với NumPy, Pandas, Matplotlib

2.4. Flask

2.5. Hồi quy tuyến tính

2.6. Một số thư viện khác

CHƯƠNG III: THIẾT KẾ VÀ XÂY DỰNG CHƯƠNG TRÌNH

## 3.1. Sơ đồ khối hệ thống

## 3.2. Sơ đồ khối các thuật toán chính

## 3.3. Cấu trúc dữ liệu

## 3.4. Chương trình

CHƯƠNG IV: THỰC NGHIỆM VÀ KẾT LUẬN

# 4.1. Thiết kế hệ thống

### 4.1.1. Phân tích chương trình

Cách thức hoạt động của chương trình có thể được miêu tả bởi biểu đồ hoạt động sau:

A diagram of a diagram

AI-generated content may be incorrect.

Hình 4.1: Biểu đồ hoạt động của chương trình

Chương trình sẽ nhận hình ảnh, sau đó sử dụng mô hình đã được huấn luyện để phân tích các đặc điểm của khuôn mặt rồi so sánh nó với các khuôn mặt trong dataset. Nếu hình ảnh giống với khuôn mặt nào thì sẽ trả về tên thư mục chứa khuôn mặt đó và gắn contour lên để dễ dàng nhận biết.

Tiếp đó, khi ấn điểm danh, chương trình sẽ:

* Nếu sinh viên chưa điểm danh: Lưu dữ liệu bao gồm MSSV, họ tên, thời gian điểm danh và trạng thái (đúng giờ, đi muộn) vào CSDL điểm danh.
* Nếu sinh viên đã điểm danh: Báo rằng sinh viên đã điểm danh.
* Nếu không nhận diện được khuôn mặt, chương trình sẽ báo lỗi.

### 4.1.2. Phân tích hệ thống

Hệ thống sẽ có 2 tác nhân chính là giảng viên và sinh viên. Chức năng và mối quan hệ giữa các tác nhân được miêu tả bởi biểu đồ ca sử dụng sau:

A diagram of a diagram

AI-generated content may be incorrect.

Hình 4.2: Biểu đồ ca sử dụng của hệ thống

Khi chạy chương trình, cửa sổ điểm danh sẽ hiện ra. Thông tin tài khoản được lưu trong CSDL của hệ thống.

* Tài khoản giảng viên sẽ là người quyết định thời gian vào lớp của tiết học. Cuối buổi học, giảng viên sẽ thông kê lại danh sách điểm danh của tiết học đó.
* Tài khoản sinh viên sẽ điểm danh trên hệ thống và hệ thống sẽ gửi thông tin về CSDL.

### 4.1.3. Cơ sở dữ liệu

Để lưu thông tin điểm danh của sinh viên, chúng ta cần sử dụng CSDL. Hệ thống này sẽ sử dụng SQLite.

CSDL gồm 4 bảng:

* Bảng in4\_sv: Lưu thông tin sinh viên

|  |  |
| --- | --- |
| **Tên cột** | **Kiểu dữ liệu** |
| MSSV | TEXT |
| Họ và tên | TEXT |
| Lớp | TEXT |
| Khoa | TEXT |

* Bảng Diem\_danh: Lưu thông tin điểm danh

|  |  |
| --- | --- |
| **Tên cột** | **Kiểu dữ liệu** |
| Id | TEXT |
| MSSV | TEXT |
| Thời gian điểm danh | DATETIME |
| Trạng thái điểm danh | TEXT |

* Bảng Thoigian\_tiet: Lưu thời gian điểm danh của tiết

|  |  |
| --- | --- |
| **Tên cột** | **Kiểu dữ liệu** |
| Id | INTEGER |
| Giờ | INTEGER |
| Phút | INTEGER |

* Bảng Login: Lưu thông tin tài khoản

|  |  |
| --- | --- |
| **Tên cột** | **Kiểu dữ liệu** |
| Id | TEXT |
| Password | TEXT |
| Loại tài khoản | TEXT |

### 4.1.4. Giao diện chương trình

Chương trình sẽ được thiết kế giao diện bằng html. Có những giao diện chính như sau:

* login: Cửa sổ đăng nhập
* index\_gv: Cửa sổ làm việc của tài khoản giảng viên
* index\_sv: Cửa sổ làm việc của tài khoản sinh viên
* diemdanh\_list: Hiển thị danh sách điểm danh

## 4.2. Kiểm thử chương trình

A screenshot of a computer

AI-generated content may be incorrect.

Hình 4.3: Cửa sổ đăng nhập

A screenshot of a computer

AI-generated content may be incorrect.

Hình 4.4: Cửa sổ index\_gv

A screenshot of a computer

AI-generated content may be incorrect.

Hình 4.5: Cửa sổ index\_sv

A screenshot of a computer

AI-generated content may be incorrect.

Hình 4.6: Điểm danh thành công

Link GitHub của chương trình:

A qr code with a dinosaur

AI-generated content may be incorrect.

CHƯƠNG V: TỔNG KẾT

## 5.1. Kết quả đã đạt được

Hệ thống nhận diện khuôn mặt được thử nghiệm với tập dữ liệu gồm 18 sinh viên, mỗi sinh viên có 3-5 ảnh trong thư mục dataset. Kết quả cho thấy: 18/18 sinh viên được nhận diện chính xác trong điều kiện ánh sáng phòng học thông thường, khoảng cách từ camera đến khuôn mặt từ 0.5 đến 1m.

Tuy nhiên, khi góc quay khuôn mặt nghiêng quá 45 độ hoặc ánh sáng yếu, mô hình sẽ gặp khó khăn trong việc khuôn mặt trong khung hình vì hình sẽ bị nhiễm muối, dẫn đến nhãn "Unknown" được gán, hoặc đôi lúc sẽ ra một người hoàn toàn khác.

## 5.2. Hướng phát triển của đề tài

Trong tương lai, em dự kiến sẽ có một số hướng phát triển như sau:

* Nâng cấp mô hình để nâng cao hiệu suất và cải thiện chất lượng nhận dạng.
* Làm cho khoảng cách Ơ-clit trở nên linh hoạt theo môi trường chứ không cố định một ngưỡng, nhằm

KẾT LUẬN

Tin học hoá trong công tác quản lý là đề tài có tính chất thực tế. Nếu đề tài này được hoàn thiện và áp dụng vào thực tế sẽ nâng cao được hiệu quả công tác quản lý, giúp cho cán bộ quản lý giảm bớt khó khăn, tránh được sai sót trong công việc đồng thời nâng cao hiệu quả công việc, giảm bớt được rất nhiều thời gian trong việc lập báo cáo thống kê, tra cứu... Tạo điều kiện phục vụ tốt hơn cho công tác quản lý điểm danh.

Do trình độ của em còn hạn chế nên đề tài của em vẫn chưa thể đáp ứng được nhu cầu thực tế, còn nhiều những thiếu sót hạn chế nhất định. Nhưng qua đó em đã rút ra rất nhiều kinh nghiệm cho bản thân về cách làm đề tài quản lý.

Em xin chân thành cảm ơn thầy Nguyễn Văn Huy đã giúp em hoàn thành đề tài này.

TÀI LIỆU THAM KHẢO

<https://chatgpt.com>

<https://github.com/bariarviv/VGGFace2>

<https://github.com/timesler/facenet-pytorch>

<https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>