# 

TRƯỜNG ĐẠI HỌC THỦY LỢI

**KHOA CÔNG NGHỆ THÔNG TIN**

-----🙡🕮🙣-----



**PHAN MINH TRƯỜNG**

**ỨNG DỤNG HỌC MÁY TRONG VIỆC NHẬN DẠNG KHUÔN MẶT**

**ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

**HÀ NỘI, NĂM 2018**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TRƯỜNG ĐẠI HỌC THỦY LỢI  **KHOA CÔNG NGHỆ THÔNG TIN**  -----🙡🕮🙣-----   |  | | --- | | ĐỒ ÁNTỐT NGHIỆP ĐẠI HỌC **Ngành Công nghệ thông tin**  **Đề tài:** **ỨNG DỤNG HỌC MÁY TRONG VIỆC NHẬN DẠNG KHUÔN MẶT** |  |  |  | | --- | --- | | Sinh viên thực hiện: | **PHAN MINH TRƯỜNG** | |  | Lớp: K55TH1; Hệ: Chính quy | | Giáo viên hướng dẫn: | **TS. LẠI HIỀN PHƯƠNG** |  |  | | --- | | ***Hà Nội, năm 2018*** | |

# Lời cảm ơn

Sau hơn 4 năm học tập và nghiên cứu tại khoa Công Nghệ Thông Tin - trường Đại học Thủy Lợi, em đã được trải nghiệm trong môi trường đào tạo tốt và nhận được sự chỉ dạy nhiệt tình của các thầy, các cô trong khoa.

Trước hết, em xin được bày tỏ lòng biết ơn và gửi lời cám ơn chân thành đến **TS. Lại Hiền Phương** đã dạy cho em có được những kiến thức vững chắc, tận tình chỉ bảo và hướng dẫn em trong suốt quá trình làm đồ án tốt nghiệp…

Ngoài ra, em cũng xin được bày tỏ lời cảm ơn sâu sắc đến những thầy cô giáo đã giảng dạy em, những kiến thức mà em nhận được sẽ là hành trang giúp em hoàn thành tốt luâ tốt nghiệp này.

Trong quá trình nghiên cứu của mình, mặc dù được sự hướng dẫn rất nhiệt tình, nghiêm túc của TS. Lại Hiền Phương cùng với sự nỗ lực của cá nhân nhưng cũng không thể tránh được những thiếu sót. Em rất mong nhận được những ý kiến đóng góp từ quý Thầy, Cô và các bạn bè.

Em xin chân thành cảm ơn!

Sinh viên thực hiện

Phan Minh Trường

# Mục lục

[Lời cảm ơn ii](#_Toc501915466)

[Mục lục iii](#_Toc501915467)

[Danh mục hình vẽ v](#_Toc501915468)

[Danh mục bảng biểu vi](#_Toc501915469)

[MỞ ĐẦU vii](#_Toc501915470)

[Chương 1. TỔNG QUAN VỀ HỌC MÁY VÀ NHẬN DIỆN KHUÔN MẶT 1](#_Toc501915471)

[1.1 Giới thiệu về học máy 1](#_Toc501915472)

[1.1.1 Học máy (Meachine learning) là gì? 1](#_Toc501915473)

[1.1.2 Các ứng dụng của học máy 2](#_Toc501915474)

[1.2 Tổng quan về bài toán nhận dạng khuôn mặt 3](#_Toc501915476)

[1.2.1 Định nghĩa 3](#_Toc501915477)

[1.2.2 Một số công trình nghiên cứu ở Việt Nam. 6](#_Toc501915478)

[1.2.3 Những nghiên cứu của Viện Xã hội học 7](#_Toc501915479)

[Chương 2. QUY TRÌNH NHẬN DẠNG KHUÔN MẶT 12](#_Toc501915483)

[2.1 Tổng quan về mô hình hồi quy 12](#_Toc501915484)

[2.1.1 Hồi quy tuyến tính 12](#_Toc501915485)

[2.1.2 Mô hình hồi quy 12](#_Toc501915486)

[2.2 Hồi quy tuyến tính nhiều biến 14](#_Toc501915487)

[2.2.1 Xây dựng mô hình hồi quy tuyến tính nhiều biến 14](#_Toc501915488)

[2.2.2 Ứng dụng hồi quy để dự đoán 15](#_Toc501915489)

[2.3 Tiêu chuẩn đánh giá 16](#_Toc501915490)

[Chương 3. XÂY DỰNG CHƯƠNG TRÌNH NHẬN DẠNG KHUÔN MẶT 17](#_Toc501915491)

[3.1 Vị trí địa lý, tự nhiên, quy mô dân số và kinh tế xã hội ở Tân Ninh 17](#_Toc501915492)

[3.1.1 Vị trí địa lý, đặc điểm khí hậu 17](#_Toc501915493)

[3.1.2 Dân số và tình hình phát triển kinh tế- xã hội 18](#_Toc501915494)

[3.2 Công cụ và công nghệ 19](#_Toc501915495)

[3.2.1 Giới thiệu về ngôn ngữ R 19](#_Toc501915496)

[3.2.2 Tải R và cài đặt cho máy 21](#_Toc501915497)

[3.2.3 Khởi động và ngưng chạy R 21](#_Toc501915498)

[3.3 Xử lý dữ liệu 22](#_Toc501915499)

[3.4 Mô tả dữ liệu 24](#_Toc501915500)

[3.5 Công cụ sử dụng chạy thử nghiệm 26](#_Toc501915501)

[3.6 Kết quả chạy thử nghiệm 27](#_Toc501915502)

[3.6.1 Dự đoán số trẻ em sinh năm sau 27](#_Toc501915503)

[3.6.1.1 Phương án 1 27](#_Toc501915504)

[3.6.1.2 Phương án 2 29](#_Toc501915505)

[3.6.1.3 Phương án 3 30](#_Toc501915506)

[3.6.1.4 Phương án 4 32](#_Toc501915507)

[3.6.2 Dự đoán số người mất năm sau 34](#_Toc501915508)

[3.6.2.1 Phương án 1 34](#_Toc501915509)

[3.6.2.2 Phương án 2 36](#_Toc501915510)

[3.6.2.3 Phương án 3 38](#_Toc501915511)

[3.6.2.4 Phương án 4 40](#_Toc501915512)

[3.6.2.5 Phương án 5 42](#_Toc501915513)

[3.7 Kết luận chương 44](#_Toc501915514)

[Chương 4. Đánh giá kết quả thực nghiệm 17](#_Toc501915491)

[3.1 Vị trí địa lý, tự nhiên, quy mô dân số và kinh tế xã hội ở Tân Ninh 17](#_Toc501915492)

[3.1.1 Vị trí địa lý, đặc điểm khí hậu 17](#_Toc501915493)

[3.1.2 Dân số và tình hình phát triển kinh tế- xã hội 18](#_Toc501915494)

[3.2 Công cụ và công nghệ 19](#_Toc501915495)

[3.2.1 Giới thiệu về ngôn ngữ R 19](#_Toc501915496)

[3.2.2 Tải R và cài đặt cho máy 21](#_Toc501915497)

[3.2.3 Khởi động và ngưng chạy R 21](#_Toc501915498)

[KẾT LUẬN 46](#_Toc501915515)

[TÀI LIỆU THAM KHẢO 48](#_Toc501915516)

[PHỤ LỤC 49](#_Toc501915517)

# MỞ ĐẦU

Cùng với sự phát triển ngày càng mạnh mẽ của khoa học kỹ thuật trong những thập kỷ gần đây, lĩnh vực xử lý ảnh vẫn còn là một ngành khoa học rất mới mẻ so với các ngành khoa học khác nhưng nó đã là một lĩnh vực thu hút rất đông đảo nhà khoa học quan tâm và phát triển, thúc đẩy các lĩnh vực nghiên cứu chuyên sâu về lĩnh vực này.

Xử lý ảnh đóng vai trò quan trọng trong nhiều ứng dụng thực tế và khoa học kỹ thuật như: giám sát an ninh, nhận dạng đối tượng, nhận dạng khuôn mặt, phát hiện chuyển động, theo dõi chuyển động, nhận dạng các khối u trong y học, hiệu chỉnh các ảnh và video,…

Một trong những bài toán quan trọng trong xử lý ảnh đó là "Nhận diện khuôn mặt".

Công nghệ nhận diện khuôn mặt (Face Detection) là một kỹ thuật từ một ảnh đầu vào, thông qua các thuật toán phân tích đặc trưng của ảnh, máy tính sẽ phát hiện được đó là khuôn mặt của ai hoặc tìm kiếm những ảnh tương đồng.

Đề tài "Ứng dụng học máy trong việc nhận diện khuôn mặt" nhằm khảo sát, phân tích một số phương pháp xử lý ảnh để tìm kiếm cũng như phân lớp ảnh. Nội dung của đồ án gồm 4 phần như sau:

**Chương I**: Giới thiệu học máy và một số ứng dụng thực tế

**Chương II**: Quy trình nhận dạng khuôn mặt

**Chương III**: Xây dựng chương trình nhận dạng khuôn mặt bằng phương pháp KNN

**Chương IV**:Đánh giá kết quả

# Chương 1. TỔNG QUAN VỀ HỌC MÁY VÀ ỨNG DỤNG

## Học máy (Meachine learning) là gì?

Là một lĩnh vực của trí tuệ nhân tạo bao gồm quá trình đúc rút tri thức từ các quan sát, trải nghiệm thực tiễn bằng cách xây dựng các mô hình từ dữ liệu; các phương pháp học và nhận dạng tự động các mẫu phức tạp (complex patterns) từ dữ liệu.

Cụ thể hơn, học máy là một phương pháp để tạo ra các chương trình máy tính bằng việc phân tích các tập dữ liệu. Học máy là lĩnh vực liên quan mật thiết đến thống kê, do cả hai lĩnh vực đều nghiên cứu việc phân tích dữ liệu. Tuy nhiên học máy có sự khác biệt với thống kê, học máy tập trung vào nghiên cứu sự phức tạp của các giải thuật trong việc thực thi tính toán, xử lý dữ liệu. Trên thực tế, có nhiều bài toán suy luận được xếp loại là bài toán nhị phân khó, vì thế một phần của học máy là nghiên cứu sự phát triển của giải thuật suy luận xấp xỉ để có thể xử lý được lớp các bài toán nhị phân một cách tổng quát nhất.

Học máy được xem là phương pháp tạo ra các chương trình máy tính sử dụng kinh nghiệm, quan sát hoặc dữ liệu trong quá khứ để cải thiện công việc của mình trong tương lai.

## Các giải thuật học máy

Học có giám sát (Supervised Learning): Huấn luyện cho các giải thuật học máy xây dựng mô hình từ các mối quan hệ trong dữ liệu, dựa trên các cặp đầu vào – đầu ra của các quan sát.

Học không giám sát (Unsupervised Learning): Ngược với học máy có giám sát, ở  thuật toán này chúng ta không biết được outcome hay nhãn mà chỉ có dữ liệu đầu vào.

Học không giám sát sẽ dựa vào cấu trúc của dữ liệu để thực hiện một công việc nào đó, ví dụ như phân nhóm (clustering) hoặc giảm số chiều của dữ liệu (dimension reduction) để thuận tiện trong việc lưu trữ và tính toán.

Đây là 2 giải thuật cơ bản của học máy. Ngoài ra còn có một số giai thuật khác như: Học tăng cường, Học máy bán giám sát, Học củng cố…

## Thuật toán K-nearest neighbor:

K-nearest neighbor là một trong những giải thuật Supervised Learning đơn giản nhất (mà hiệu quả trong một vài trường hợp). Khi training, thuật toán này không học một điều gì từ dữ liệu training, mọi tính toán được thực hiện khi nó cần dự đoán kết quả của dữ liệu mới. KNN có thể được áp dụng vào cả hai loại của bài toán Classification và Regression