BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC BÀ RỊA - VŨNG TÀU**

**KHOA KỸ THUẬT - CÔNG NGHỆ**

**BÁO CÁO MÔN HỌC  
TRÍ TUỆ NHÂN TẠO**

**Tên đề tài:**

**SỬ DỤNG THUẬT TOÁN HỒI QUY TUYẾN TÍNH DỰ ĐOÁN DỮ LIỆU**

**GVHD:** *TS. Bùi Thị Thu Trang*

**Sinh viên:** *Tạ Quang Khôi*

**Lớp:** *DH20LT*

**Vũng Tàu, tháng 05 năm 2023**

# MỤC LỤC

[MỤC LỤC 2](#_Toc136413671)

[DANH MỤC HÌNH ẢNH 3](#_Toc136413672)

[PHẦN I. TRÍ TUỆ NHÂN TẠO VÀ ỨNG DỤNG 4](#_Toc136413673)

[1. Định nghĩa về trí tuệ nhân tạo 4](#_Toc136413674)

[2. Các ứng dụng của trí tuệ nhân tạo 4](#_Toc136413675)

[PHẦN 2. THUẬT TOÁN HỒI QUY TUYẾN TÍNH VÀ DỰ ĐOÁN DỮ LIỆU 6](#_Toc136413676)

[2.1. Thuật toán Hồi quy tuyến tính 6](#_Toc136413677)

[2.2. Dữ liệu 6](#_Toc136413678)

[2.3. Phân tích dữ liệu 7](#_Toc136413679)

[TÀI LIỆU THAM KHẢO 12](#_Toc136413680)

# DANH MỤC HÌNH ẢNH

[Hình 1: Vài dòng dữ liệu 7](#_Toc136378571)

[Hình 2: Import các thư viện cần thiết để hỗ trợ việc xây dựng model 7](#_Toc136378572)

[Hình 3: Lấy dữ liệu của X và Y ra hai biến riêng 8](#_Toc136378573)

[Hình 4: Lấy ngẫu nhiên hai số để làm chỉ mục của dữ liệu thử nghiệm 8](#_Toc136378574)

[Hình 5: Biến đổi dữ liệu thành mảng hai chiều 9](#_Toc136378575)

[Hình 6: Vẽ biểu đồ bằng thư viện matplotlib.pyplot 9](#_Toc136378576)

[Hình 7: Đồ thị biểu diễn dữ liệu giữa X và Y 10](#_Toc136378577)

[Hình 8: Tạo model hồi quy tuyến tính 10](#_Toc136378578)

[Hình 9: Thử nghiệm model với dữ liệu đã chuẩn bị 11](#_Toc136378579)

# PHẦN I. TRÍ TUỆ NHÂN TẠO VÀ ỨNG DỤNG

## Định nghĩa về trí tuệ nhân tạo

Theo IBM, Trí tuệ nhân tạo đơn giản là sự kết hợp của khoa học máy tính và lưỡng dữ liệu khổng lồ, mục đích để giải quyết một vấn đề nào đó. Nó cũng bao gồm hai lĩnh vực nhỏ bên trong là máy học và học sâu.

## Các ứng dụng của trí tuệ nhân tạo

1. Trí tuệ nhân tạo trong Thiên văn học

Trong lĩnh vực thiên văn học, trí tuệ nhân tạo giúp các nhà khoa học giải quyết các vấn đề phức tạp của vũ trụ. Công nghệ AI giúp tìm hiểu về cách vũ trụ hoạt động, nguồn gốc của nó.[[1]](#endnote-2) Với sự phát triển của công nghệ kính thiên văn, lượng dữ liệu mà các nhà thiên văn học thu thập được từ vũ trụ càng ngày cang nhiều. AI chính là công cụ được dùng để xử lý lượng dữ liệu lớn này. AI giúp các nhà thiên văn học nhận dạng các thiên hà, các ngoại hành tinh.

1. Trí tuệ nhân tạo trong Ngành công nghiệp game

Các nhà phát triển game sử dụng AI để chế tạo ra các nhân vật. AI được dùng để hỗ trợ người chơi tạo ra các trang phục trong game. Một nhà phát triển game khác sử dụng AI để tạo nên các câu trả lời tùy thuộc vào những gì người dùng nói.[[2]](#endnote-3)

1. Trí tuệ nhân tạo trong Ngành tài chính

Trong ngành tài chính, AI được dùng để tìm ra các dấu hiệu lừa đảo trong các giao dịch. Bên cạnh đó, các chatbot được tích hợp AI sẽ giúp đưa ra những lời khuyên về tài chính mang tính cá nhân hóa hơn, trà lởi các câu hỏi của khách hàng, hoặc tự động mở tài khoản mới và cập nhật thông tin khách hàng.[[3]](#endnote-4)

1. Trí tuệ nhân tạo trong An ninh mạng

AI được dùng để tự động phát hiện các mối đe dọa, và phần ứng với các mối đe dọa hiệu quả hơn phần mềm truyền thống. AI sẽ tự động cập nhật dữ liệu về các xu hướng tấn công của hacker để giúp đưa ra quyết định thích hợp. AI giúp dự đoán các rủi ro có thể xảy ra trong tương lai.[[4]](#endnote-5)

1. Trí tuệ nhân tạo trong Giáo dục

AI có thể được sử dụng để kiểm tra sự có mặt và điểm danh học sinh, sau đó thông báo cho phụ huynh nếu có sự vắng mặt hoăc đi trễ. Giáo viên còn có thể sử dụng AI để kiểm tra qua các bài tiểu luận của học sinh với các sai sót về chính tả, cấu trúc câu và đạo văn. Điều này giúp tiết kiệm thời gian cho giáo viên, để họ có thể dành thời gian cho các tác vụ quan trong hơn, hoặc chỉ dạy một đối một cho từng học sinh.[[5]](#endnote-6)

# PHẦN 2. THUẬT TOÁN HỒI QUY TUYẾN TÍNH VÀ DỰ ĐOÁN DỮ LIỆU

## 2.1. Thuật toán Hồi quy tuyến tính

Hồi quy tuyến tình là một thuật toán supervised learning trong lĩnh vực máy học. Nó là phép phân tích thống kê dùng để tìm ra mối liên hệ giữa hai biến. Hàm sẽ đưa ra giả định một mối quan hệ tuyến tính giữa biến độc lập và biến phụ thuộc. Mục tiêu là tìm ra đường thẳng phù hợp nhất để mô ta mối liên hệ.

## 2.2. Dữ liệu

**Dữ liệu 1**

Dữ liệu có tên: Gray Kangaroos

Link dữ liệu: https://college.cengage.com/mathematics/brase/understandable\_statistics/7e/students/datasets/slr/frames/slr07.html

Dữ liệu gồm 2 cột X và Y:

X = chiều dài mũi (mm ¥ 10)

Y = chiều rộng mũi (mm ¥ 10)

của một con kangaroo xám đực từ một mẫu ngẫu nhiên của những con vật đó.

Tham khảo: Tạp chí Động vật học Úc, Vol. 28, tr607-613

Tổng dữ liệu có 45 dòng

A screenshot of a computer

Description automatically generated with medium confidence

Hình 1: Vài dòng dữ liệu

## 2.3. Phân tích dữ liệu

Các thư viện đã đã import:

A screenshot of a computer screen

Description automatically generated with low confidence

Hình 2: Import các thư viện cần thiết để hỗ trợ việc xây dựng model

Thư viện pandas được dùng trong báo cáo để đọc và lấy dữ liệu từ tệp excel. Thư viện numpy tạo ra hai chỉ mục ngẫu nhiên dùng để lấy dự liệu thử nghiệm, dữ liệu không đưa vào model. Thư viện matplotlib.pyplot để vẽ và minh họa biểu đồ. Thư viện sklearn dùng để xây dựng thuật toán Hồi quy tuyến tính.

Features được chọn là X, chiều dài mũi của con kangaroo xám đực. Với Target sẽ là Y, chiều rộng mũi của chúng.

A picture containing text, font, white, screenshot

Description automatically generated

Hình 3: Lấy dữ liệu của X và Y ra hai biến riêng

Hai dòng dữ liệu đã được chọn ngẫu nhiên. Bằng cách dùng thư viện numpy lấy hai số bất kỳ để làm chỉ mục.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 4: Lấy ngẫu nhiên hai số để làm chỉ mục của dữ liệu thử nghiệm

Bên cạnh đó, hai giá trị X không nằm trong bộ dữ liệu là 700 và 800 sẽ được chọn để kiểm tra thêm về độ chính xác của model.

Bắt đầu xử lý dữ liệu từ hai biến (x\_col, y\_col) đã có. Chuyển chúng thành mảng hai chiều để tương thích với đầu vào của model.

A screenshot of a computer program

Description automatically generated with low confidence

Hình 5: Biến đổi dữ liệu thành mảng hai chiều

Tiếp theo là sử dụng thư viện matplotlib.pyplot để vẽ biểu đồ minh họa cho dữ liệu.

A screenshot of a computer program

Description automatically generated with low confidence

Hình 6: Vẽ biểu đồ bằng thư viện matplotlib.pyplot

A picture containing screenshot, text, diagram, plot

Description automatically generated

Hình 7: Đồ thị biểu diễn dữ liệu giữa X và Y

Từ đồ thị biểu diễn, có kết luận rằng dự liệu nằm trên một đường thẳng. Nên có thể sử dụng thuật toán Hồi quy tuyến tính để xây dựng mô hình dự đoán.

Tạo biến chứa model Hồi quy tuyến tính với hàm LinearRegression từ lớp linear\_model của thư viện sklearn.

A screenshot of a computer error

Description automatically generated with low confidence

Hình 8: Tạo model hồi quy tuyến tính

Sau khi có model, bắt đầu thử nghiệm nó với các dữ liệu đã có sẵn. Hai dữ liệu là lấy từ bộ dữ liệu gốc mà không bị đưa vào model.

A screenshot of a computer program

Description automatically generated with low confidence

Hình 9: Thử nghiệm model với dữ liệu đã chuẩn bị

Từ kết quả thử nghiệm đặt được với hai dòng dữ liệu được tách riêng.

X là 609, có Y gốc là 241, kết quả dự đoán X từ model là 264, sai số là 23, sai số khá lớn

X là 675, có Y gốc là 217, kết quả dự đoán X từ model là 188, sai số là 29, sai số khá lớn.

**Kết luận**: Thuật toán Hồi quy tuyến tính dễ bị nhiễu dữ liệu, làm kết quả dự đoán có sai số lớn.

# TÀI LIỆU THAM KHẢO

1. 1 What is artificial intelligence (AI)? - https://www.ibm.com/topics/artificial-intelligence [↑](#endnote-ref-2)
2. https://venturebeat.com/games/how-generative-ai-is-changing-game-development-and-ugc/ [↑](#endnote-ref-3)
3. https://cointelegraph.com/news/9-examples-of-artificial-intelligence-in-finance [↑](#endnote-ref-4)
4. https://www.balbix.com/insights/artificial-intelligence-in-cybersecurity/ [↑](#endnote-ref-5)
5. https://www.intel.com/content/www/us/en/education/transforming-education/ai-in-education.html [↑](#endnote-ref-6)