



# GIỚI THIỆU NHÓM DSAI@OU

ThS Nguyễn Thị Phương Trang

# NỘI DUNG

---

1. Giới thiệu nhóm DSAI@OU
2. Mục đích nhóm
3. Giới thiệu lĩnh vực nghiên cứu
4. Giới thiệu đề tài
5. Quy tắc nhóm

# GIỚI THIỆU NHÓM

---

- DSAI@OU (Data Science & Artificial Intelligence) là nhóm nghiên cứu về Khoa học dữ liệu và Trí tuệ nhân tạo.
- Đối tượng nghiên cứu: Dữ liệu tiếng Việt (ưu tiên text)
- Lĩnh vực nghiên cứu: Các công nghệ giúp máy tính xử lý và hiểu ngôn ngữ của con người, đặc biệt là tiếng Việt, để xây dựng các hệ thống phân tích dữ liệu, hệ thống khuyến nghị và hỗ trợ ra quyết định dựa trên dữ liệu tiếng Việt.

# MỤC ĐÍCH NHÓM

---

- Truyền đạt kiến thức liên quan để giúp sinh viên có kiến thức cơ bản về khoa học dữ liệu và AI.
- Hỗ trợ, hướng dẫn sinh viên nghiên cứu khoa học, làm luận văn tốt nghiệp

# LĨNH VỰC NGHIÊN CỨU

---

- Phân tích mạng xã hội
  - Xác định tác nhân theo vùng miền
  - Xác định tác nhân theo giới tính, độ tuổi
- Recommend trong thương mại điện tử
  - Sử dụng Association Rule, collaboration user-based hoặc collaboration user-based để recommend
- Machine learning, Deep learning
  - Tìm hiểu về FCA (Phân tích khái niệm hình thức (Formal Concept Analysis - FCA)) Xây dựng cấu trúc ngữ nghĩa cho dữ liệu huấn luyện trong ứng dụng máy học.
  - Tìm hiểu về RMB (Restricted Boltzmann Machine) - Ứng dụng vào bài toán dự đoán bệnh.
  - Tìm hiểu về web ngữ nghĩa và xây dựng website tra cứu bệnh thông minh.
  - Tìm hiểu về Ontology và xây dựng Ontology bệnh dựa trên phương pháp cộng tác.

# Data science là gì?

---

- Là hướng nghiên cứu đề cập đến việc **xác định, trình bày và trích xuất** các **thông tin có ý nghĩa** từ các nguồn dữ liệu, được sử dụng cho mục đích nào đó.
- Ví dụ:
  - Dựa trên việc thu thập thông tin từ các post/comment/status trên mạng xã hội, facebook... Data Scientist có thể tìm ra được các thông tin như: người đó là người vùng miền nào? Các mặt hàng đang được quan tâm là gì, ...
  - Dựa vào phân tích dữ liệu bán hàng, có thể biết được thời gian sắp tới khách hàng sẽ mua những mặt hàng nào ...

# Các công ty khoa học dữ liệu hàng đầu

---

The Google logo, featuring the word "Google" in its characteristic multi-colored font: blue 'G', red 'o', yellow 'o', blue 'g', green 'l', and red 'e'.The Amazon logo, consisting of the word "amazon" in a bold, black, lowercase sans-serif font, with a curved orange arrow underneath it pointing from the 'a' to the 'z'.

# Bạn có thể làm gì với Data Science?

---

- Đưa ra được những kết luận hoặc tạo ra được những câu hỏi mở.
- Trích xuất khối lượng lớn dữ liệu từ các nguồn bên ngoài và nội bộ.
- Triển khai thống kê, học máy và phương pháp phân tích.
- Làm tinh gọn dữ liệu, giúp dữ liệu có thể sẵn sàng cho việc xử lý và phân tích.
- Có cái nhìn đa chiều từ các góc độ khác nhau để xác định các mẫu, quan hệ và xu hướng ẩn mình trong khối lượng dữ liệu khổng lồ đó.
- Sử dụng kết hợp các công cụ thuật toán và công cụ tự động hóa.
- Thiết kế lại quy trình, hệ thống, sử dụng phương pháp tiếp cận theo hướng dữ liệu.



# Hướng nghiên cứu của Data science

---

- Analysis: Phân tích dữ liệu tìm ra các insight giá trị.
- Building: xử lý/lưu trữ dữ liệu -> Hệ thống khuyến nghị (recommendation system)
  - Tạo sản phẩm là phần mềm (data mining)
    - Người dùng đã mua những món đồ gì, có đặc điểm gì, những món đồ nên mua kèm, những món đồ mà người dùng khác có hành vi tương tự đã mua ... (Amazon)
  - Tạo model cho các hệ thống máy học
    - Machine learning là lĩnh vực của trí tuệ nhân tạo, trong đó các thuật toán được sử dụng để tự học dựa trên dữ liệu đưa vào mà không phải lập trình cụ thể

# Data Science quan trọng đối với những lĩnh vực nào?

---

- Data Science
- Machine Learning
- Data Mining
- Big Data
- Deep learning

# Kiến thức cần có đối với Data Scientist

---

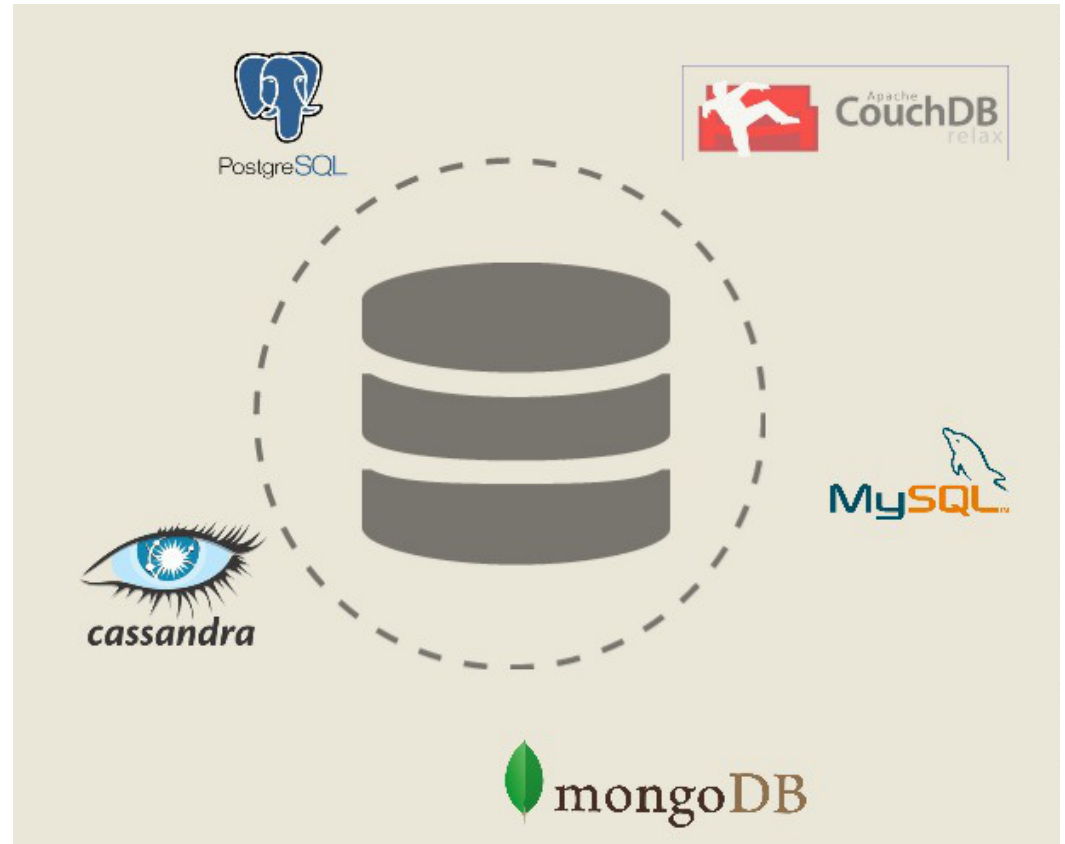
- Xác suất và thống kê
- Đại số tuyến tính
- Machine Learning  
(<https://machinelearningcoban.com>)

Một số khóa học online có thể tham khảo:

- [Học toán](#), Khan Academy.
- [Đại số tuyến tính \(Linear algebra\)](#), MIT Open Course Ware.
- [Giới thiệu về thống kê \(Statistics\)](#), Udacity.
- [Học máy \(Machine learning\)](#), khóa học rất nổi tiếng do giáo sư Andrew Ng (ĐH Stanford) dạy trên Coursera.

# Ngôn ngữ cần có đối với Data Scientist

---



# Giới thiệu các đề tài

---

1. Tìm hiểu về Blockchain, đưa ra quy trình quản lý bệnh án bệnh nhân liên quốc gia.
2. Dự đoán hành vi người dùng trên mạng xã hội
3. Hệ thống khuyến nghị trong mô hình thương mại điện tử
4. Xây dựng ứng dụng y tế thông minh
5. Tiếp cận deep learning qua giải thuật RMB (Restricted Boltzmann Machine) - Ứng dụng vào bài toán dự đoán hành vi người dùng.
6. Khuyến nghị danh sách từ vựng cần học cho sv thi TOEIC

# Giới thiệu các đề tài

---

1. Tích hợp Google Maps APIs trong website chỉ đường ASP.Net MVC
4. Xây dựng website thương mại điện tử sử dụng công nghệ ASP.Net MVC – thanh toán tích hợp Paypal. (Có thể chọn loại website khác như: Bán thiết bị điện tử, Bán VLXD, website thời trang, ...)

# Giới thiệu các đề tài

---

1. Tìm hiểu về Blockchain, đưa ra quy trình quản lý bệnh án bệnh nhân liên quốc gia.

*Xem file 1. Blockchain*

# Giới thiệu các đề tài

---

## 2. Dự đoán hành vi người dùng trên mạng xã hội



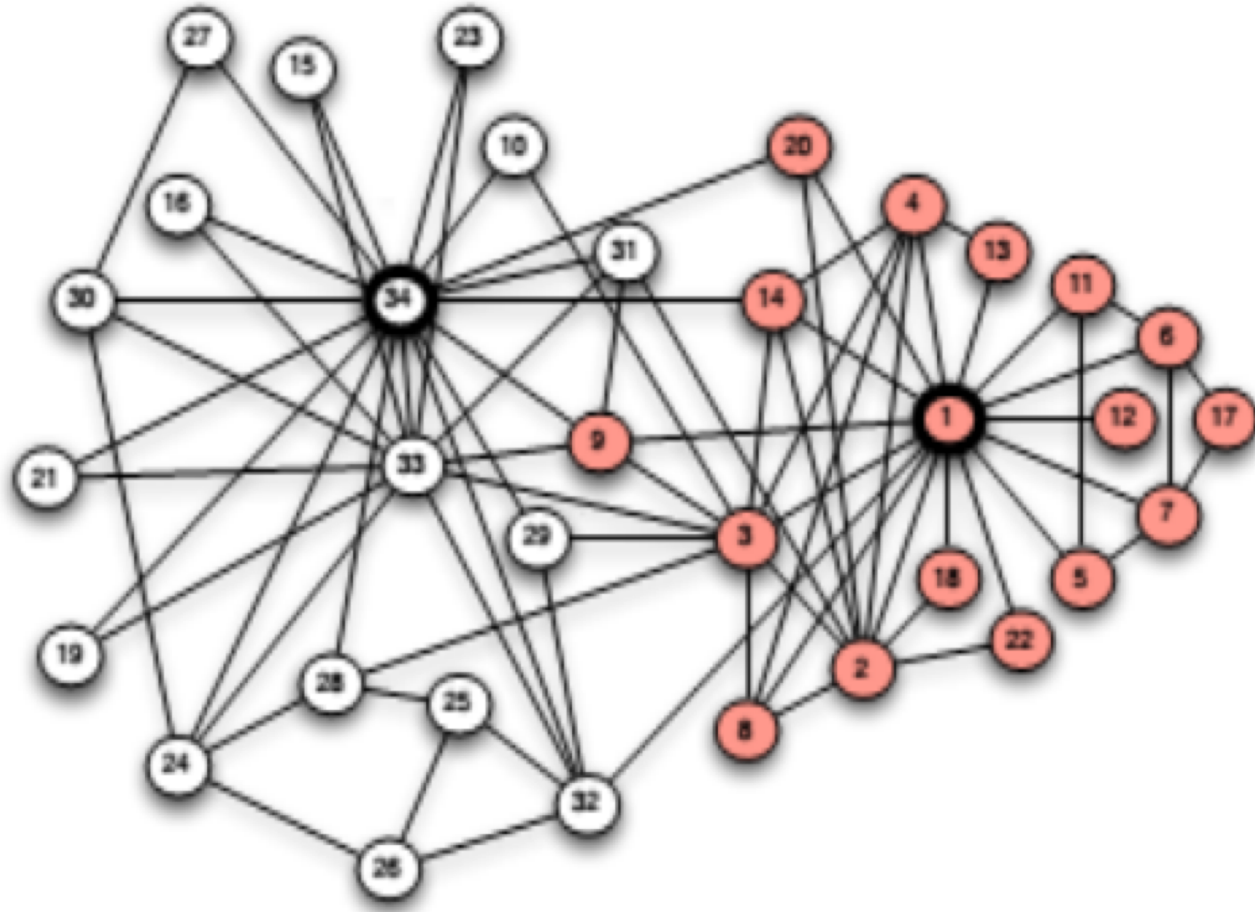
# Mạng xã hội - social network

---

- Mạng xã hội: nút là tác nhân xã hội và cạnh là quan hệ/tương tác giữa các tác nhân đó.
- Là một cấu trúc xã hội bao gồm các cá nhân hay tổ chức, thường được biểu diễn bằng các nút, cùng với các quan hệ xã hội, tương ứng với các liên kết giữa các nút.

# Mạng xã hội - social network

---



[Easley10] David Easley and Jon Kleinberg (2010). Networks, Crowds, and Markets: Reasoning about a Highly Connected World, Cambridge University Press, 2010

# Giới thiệu các đề tài

---

3. Hệ thống khuyến nghị trong mô hình thương mại điện tử

Xem file 3. recommender\_SV

# Giới thiệu các đề tài

---

## 4. Xây dựng ứng dụng y tế thông minh

Xem file 4. YTeThongMinh.docx

5. Tiếp cận deep learning qua giải thuật RBM (Restricted Boltzmann Machine) - Ứng dụng vào bài toán dự đoán hành vi người dùng.

Xem file RBM.docx

# Giới thiệu các đề tài

---

## 6. Khuyến nghị danh sách từ vựng cần học cho sv thi TOEIC

- Input: Có các đề thi TOEIC ở dạng full-text (file PDF, text). Từ các đề này (sau loại bỏ các stop-word):
- Output:
  1. Danh sách bảng từ và TF, IDF, TF-IDF của nó ứng với D(document) là từng đề, và toàn bộ đề. Ví dụ có 5 đề: sẽ có 6 bảng từ xuất ra (5 bảng ứng 5 đề và 1 bảng ứng toàn bộ)
  2. Xây dựng hệ thống khuyến nghị, dựa trên bộ đề, dựa trên số điểm sv muốn đạt được, hệ thống đưa ra danh sách các từ vựng cần học theo thứ tự ưu tiên

<https://codetudau.com/bag-of-words-tf-idf-xu-ly-ngon-ngu-tu-nhien/index.html>

# Giới thiệu các đề tài

---

1. Tích hợp Google Maps APIs trong website chỉ đường ASP.Net MVC
4. Xây dựng website thương mại điện tử sử dụng công nghệ ASP.Net MVC – thanh toán tích hợp Paypal. (Có thể chọn loại website khác như: Bán thiết bị điện tử, Bán VLXD, website thời trang, ...)

[https://www.w3schools.com/graphics/google\\_maps\\_intro.asp](https://www.w3schools.com/graphics/google_maps_intro.asp)

<http://developers.google.com/maps>

<https://www.youtube.com/watch?v=TscC1HoANzg&list=PL33lvabfss1xKBpkOW4O-afIdwjgpxssz>

Xem thêm tài liệu hướng dẫn đính kèm

# Quy tắc nhóm

---

1. Tham gia các buổi họp nhóm.
2. Ngoại trừ các buổi đào tạo kiến thức, thành viên sẽ được giao nhiệm vụ nghiên cứu và trình bày nội dung nghiên cứu trước cả nhóm.
3. Phân công trình bày nghiên cứu: Xoay vòng theo thứ tự thành viên nhóm.

---

Thank you for your attention