**TRƯỜNG ĐẠI HỌC SÀI GÒN**

KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO BÀI TẬP

**Khảo sát đề tài**

**Fraud detection using machine learning**

**HỌC PHẦN:** PHƯƠNG PHÁP NGHIÊN CỨU

**THUỘC NHÓM NGÀNH:** CÔNG NGHỆ THÔNG TIN

**GIẢNG VIÊN:** ĐỖ NHƯ TÀI

THÀNH VIÊN NHÓM 77:

3123410024 - NGÔ THƯỢNG BẢO

3123410242 - NGUYỄN TRỌNG NHÂN

3123410174 - NGUYỄN PHẠM TUẤN KHÔI

3123410387 - NGUYỄN HỮU TRI

**Mục lục**

[I. Introduction 1](#_Toc194093708)

[II. Aim of the Study and Scope of Research 2](#_Toc194093709)

[III. Approach 3](#_Toc194093710)

[1) Naïve Bayes 4](#_Toc194093711)

[2) SVM (Support Vector Machine) 4](#_Toc194093712)

[3) Decision Tree 5](#_Toc194093713)

[4) K-nearest neighbor 6](#_Toc194093714)

[IV. References 6](#_Toc194093715)

# Introduction

Với bối cảnh tăng trưởng người dùng sử dụng Internet, các mail, tin nhắn rác cũng ngày một gia tăng với các thủ đoạn ngày một tinh vi. Các mail, tin nhắn này được sử dụng cho các mục đích xấu như lừa đảo hay chiếm đạt thông tin. Người gửi các văn bản rác có nhiều thủ đoạn khác nhau để phục vụ mục đích nói trên, như gắn mã độc hay cố gắng dụ dỗ người đọc nhập thông tin cá nhân của họ vào một đường link lạ. Các đối tượng trên thường chủ yếu nhắm đến những cá nhân không biết về mánh khóe của chúng. Vì vậy mà việc xác định và nhận diện các mail spam, mail lừa đảo này là cần thiết, chính vì thế mà nhóm quyết định áp dụng máy học vào giải quyết vấn đề này.

# Aim of the Study and Scope of Research

Như đã nói ở phần trước, mục tiêu của đề tài/dự án này là giải quyết được vấn đề nhận diện, phân loại các loại mail xem đâu là mail lừa đảo và đâu là mail bình thường. Việc phân loại được các mail, tin nhắn rác sẽ giúp tiết kiệm được thời gian, tiền, và không gian kho lưu trữ. Đề tài cũng hi vọng sẽ phần nào giúp biến Internet trở thành một nơi an toàn hơn cho tất cả mọi người và nâng cao tính cảnh giác đối với các loại thư văn bản được gửi tới người đọc.

Phạm vi nghiên cứu của đề tài trài dài từ việc phân tích ngôn ngữ, văn bản cho đến các loại thủ đoạn, xu hướng mà các đối tượng gửi mail, tin nhắn rác thường sử dụng. Vì việc cách mà mail và tin nhắn rác được soạn và gửi đi luôn liên tục thay đổi theo thời gian nên phạm vi nghiên cứu của đề tài này sẽ không cố định mà được mở rộng theo thời gian.

# Approach

Trước đây hầu hết các tin nhắn, mail rác đều có thể bị chặn do số lượng người gửi còn ít và dễ nhận diện thông qua tên hoặc địa chỉ email, số điện thoại. Nhưng trong các năm gần đây thì mánh khóe của các đối tượng trên lại càng trở nên tinh vi, sử dụng các tài khoản khác nhau, email “ăn liền” để phục vụ việc gửi đi các tin nhắn và mail spam này. Chính vì thế mà mô hình máy học được đưa vào sử dụng để phát hiện và phân loại các loại mail và tin nhắn nói trên.

Đa số các phương pháp nhận diện và phân loại không theo mô hình máy học đều dựa vào các yếu tố như, phân tích văn bản, blacklist và whitelist các nguồn, địa chỉ gửi thư và văn bản, cũng như dựa vào sự hỗ trợ của cộng đồng. Tuy nhiên thì các giải pháp cũng có phần kém hiệu quả do sự lạc hậu về công nghệ cũng như các quy trình quá mức khuôn khổ.

A diagram of a machine learning model

AI-generated content may be incorrect.

Việc áp dụng mô hình máy học được đánh giá là sẽ hiệu quả hơn, một bộ dữ liệu sẽ được sử dụng nhằm phục vụ việc đào tạo, các dữ liệu này bao gồm các mail, văn bản được phân loại sẵn thành các danh mục khác nhau. Về thuật toán sử dụng, mô hình sử dụng các thuật toán như:

## Naïve Bayes

Thuật toán Naïve Bayes là thuật toán tính xác suất của từng class và chọn class có xác suất cao nhất làm kết quả cuối cùng. Khi áp dụng vào phân tích văn bản, nếu có bất cứ từ nào xuất hiện trong bộ dữ liệu tin nhắn, mail rác mà không xuất hiện trong bộ dữ liệu các tin nhắn, mail bình thường thì mail đó là mail rác.

## SVM (Support Vector Machine)

Trong thuật toán này, chúng ta vẽ đồ thị dữ liệu là các điểm trong n chiều ( ở đây n là số lượng các tính năng bạn có), trong trường hợp này 2 chiều tương đường với spam và bình thường, với giá trị của mỗi tính năng sẽ là một phần liên kết. Sau đó thực hiện tìm "đường bay" (hyper-plane) phân chia các lớp, mô hình sẽ dựa vào đường bay để đưa ra quyết định mail, tin nhắn là spam hay bình thường.

A diagram of a support line

AI-generated content may be incorrect.

## Decision Tree

Là thuật toán phân loại/dự đoán dựa trên kết quả lấy được sau khi học bộ dữ liệu được đưa vào, sau đó tùy theo các điều kiện mà đưa ra kết quả cuối cùng.

A diagram of a person's age

AI-generated content may be incorrect.

ví dụ của một cây quyết định

## K-nearest neighbor

Là thuật toán phân loại sử dụng các dữ liệu trong bộ dữ liệu training để so sánh với dữ liệu đầu vào mới, từ đó đưa ra kết quả cuối cùng. Khi có một văn bản tin nhắn hay mail vào, thuật toán sẽ so sánh với các văn bản có sẵn trong bộ dữ liệu training, nếu kết quả so sánh bao gồm các văn bản thuộc phân loại spam thì thư, tin nhắn sẽ được đánh dấu spam, và ngược lại.

Và còn nhiều thuật toán khác nữa…

# References

Các thông tin, văn bản trên được tham khảo từ cái bài báo sau:

Kumar, N., Sonowal, S., and Nishant (2020) ‘Email Spam Detection Using Machine Learning Algorithms’, in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India: IEEE, pp. 108–113. Available at: https://doi.org/10.1109/ICIRCA48905.2020.9183098.

Email Spam Detection  
<https://www.kaggle.com/code/zabihullah18/email-spam-detection>

Enron1 Spam Mail Dataset

<https://www.kaggle.com/datasets/venky73/spam-mails-dataset>