**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN HÀ NỘI**

**KHOA TOÁN – CƠ – TIN HỌC**

**----\*\*\*----**



**BÀI BÁO CÁO**

***Đề tài: Tạo web phỏng đoán tin tức thật giả bằng Python***

***Giảng viên hướng dẫn: PGS.TS. Lê Trọng Vĩnh***

***Nhóm sinh viên thực hiện : Nhóm 5***

***Lớp học phần: Lập trình nâng cao***

***Hà Nội, tháng 11/2022.***

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN HÀ NỘI**

**KHOA TOÁN – CƠ – TIN HỌC**

**----\*\*\*----**



**BÀI BÁO CÁO**

***Đề tài: Tạo web phỏng đoán tin tức thật giả bằng Python***

**Thành viên nhóm 5:**

1. Đỗ Thị Hường – 20001929
2. Ngô Thị Thu Hằng – 20001913
3. Nguyễn Thị Thu Hiền – 20001914

# 1. Mô phỏng bài toán và quy trình làm việc:

## 1.1 Mô phỏng bài toán:

Web sẽ giúp người dùng phỏng đoán tin tức được nhập vào theo 2 cách là: Nhập url của trang web chứa tin tức hoặc nhập tin tức thủ công bằng Input Box có trên trang web. Bài toán sẽ được chia nhỏ thành 2 bài toán con: Xử lý dữ liệu để đưa ra phỏng đoán về thông tin được truyền vào và tạo lập trang web với PyScript.

**Phần 1:** Xử lý dữ liệu để đưa ra phỏng đoán về thông tin được truyền vào: Xử lý dữ liệu traning và lưu lại kết quả, từ những traning có được sẽ phỏng đoán dữ liệu chứ tin tức mới nhập được từ người dùng là thật hay giả.

Tập dữ liệu trainning của bài toán được lấy từ <https://www.kaggle.com/c/fake-news/data?select=train.csv>. **train.csv** là một tập dữ liệu training đầy đủ với các thuộc tính sau:

id: id duy nhất cho mỗi tin tức

title: tiêu đề của tin tức

author: tác giả của tin tức

text: văn bản của bài viết, có thể không đầy đủ

label: nhãn đánh dấu bài viết có khả năng không đáng tin cậy

1: không đáng tin cậy

0: đáng tin cậy

**Phần 2:** Xây dựng trang web với PyScript

**<Viết thêm vào>**

## 1.2 Quy trình làm việc:

- Đầu tiên: Chuẩn hóa dữ liệu đầu vào, để tối ưu hóa phân tích dữ liệu.

- Chuyển đổi dữ liệu văn bản thành các vector để máy tính có thể hiểu được.

- Ghi các kết quả bước vừa rồi vào model để phỏng đoán.

- Lưu trữ lại các kết quả training được.

- Tạo lập web để nhận các dữ liệu cần phỏng đoán.

- Chuẩn hóa dữ liệu và chuyển nó thành các vector

- Truyền nó vào model đã tạo ở các bước trên để model đưa ra phỏng đoán.

# 2. Chi tiết thực hiện:

## 2.1 Phần 1: Xử lý dữ liệu để đưa ra phỏng đoán về thông tin được truyền vào:

## 2.1.1: Giới thiệu các thư viện sử dụng:

Với một tập dữ liệu training như đã trình bày ở phần trên, ta sẽ chuẩn hóa nó bằng hàm