

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
NGÀNH TOÁN-TIN HỌC**



**PYTHON CHO KHOA HỌC DỮ LIỆU**

**FINAL SEMESTER PROJECT  
REALTIME SENTIMENT ANALYSIS DASHBOARD**

**Giảng viên phụ trách  
Hà Văn Thảo**

## MỤC LỤC

### I. [INTRODUCE](#)

- A. Thông tin nhóm
- B. Nguyên tắc hoạt động nhóm
- C. Mục tiêu project
- D. Data source

### II. [ARCHITECTURE](#)

- A. Dictionary tree
- B. EDA and models overview
- C. Deploy model
  - 1. Setting up data source
  - 2. Transform and load
  - 3. Backend
  - 4. Frontend
  - 5. Docker build file

### III. [SETUP](#)

- A. All you have to do is docker compose up

Tài liệu tham khảo

### **Lưu ý báo cáo**

Thưa thầy, vì do file code của nhóm em không chỉ có mỗi file ibnpy mà còn có code của file py trên vscode chúng em không thể chụp lại được vì nó quá dài. Nên thầy thông cảm giúp nhóm chúng em khi phải paste code trên doc ạ. Tụi em xin chịu mọi trách nhiệm nếu code trong doc khác với code trong file code của tụi em ạ. Em cảm ơn thầy

Thầy có thể xem file code tổng của tụi em ở link github dưới đây ạ

[https://github.com/Amature123/new\\_new\\_sentiment](https://github.com/Amature123/new_new_sentiment)

## INTRODUCE

### 1.1 Thông tin nhóm

- Thành viên:
  - 22110191 - Nguyễn Thành Tài
  - 22110212 - Nguyễn Quang Thịnh
  - 22110068 - Lê Kim Hùng

### 1.2 Nguyên tắc hoạt động nhóm

- Ý tưởng được đóng góp từ tất cả thành viên của nhóm.
- Mọi người tôn trọng ý kiến của nhau.
- Giao và làm việc đúng thời hạn đưa ra. Nếu trễ thì sẽ chịu trách nhiệm về phần của mình.

### 1.3 Mục tiêu project

Với mục tiêu làm một pipeline đầy đủ về các chức năng cũng như vai trò của data scientist, nhóm em đã thực hiện project Sentiment Analysys dashboard. Project tập trung về quá trình train model và cũng như cách để deploy model lấy tiếp dữ liệu từ trang web voz.vn để có thể visualize và analyze data.

### 1.4 Data source

- Vietnamese sentiment dataset: Chứa dữ liệu của đánh giá khách hàng, số sao đánh giá, và tình trạng của câu nói  
<https://www.kaggle.com/datasets/linhlpv/vietnamese-sentiment-analyst>
- Voz.vn: Nơi scrap các data về để chạy thử mô hình  
<https://voz.vn/whats-new/>

## Architecture

### 2.1 Dictionary tree

Ở phần này chúng em chỉ liệt kê các thành phần chính trong code. Các phần khác của code đã tự động tạo khi khai triển 1 dự án của fullstack .

- **Vite\_project**: Nơi chứa các code để làm frontend
  - "node\_modules": File dependencies của Nodejs
  - "src": Code chính của frontend
    - App.jsx: File chính tạo backend
  - "Dockerfile": tạo container cho frontend
- **Voz\_neww**: Backend của project
  - "model": file model để deploy dự án
  - "postgres": file tạo database schema mẫu
  - "Voz\_neww": Khung sườn của backend
    - pipeline.py: Ống sự kiện, nơi quản lý luồng dữ liệu
    - spiders/demospider.py: Code scraping data
  - "Dockerfile": tạo container cho backend
  - "main.py": file code khởi tạo backend
  - "output.json": file data mẫu do Voz\_neww scrap về
  - "requirements.txt": file các thư viện cần thiết cho python
- **Docker-compose.yaml**: Xây dựng các container cho docker và chạy code tự động
- **Các version pipeline**: sự thay đổi của code trong Voz\_neww cho các phiên bản của model khác nhau
- **test.py**: Mock code