**TRÍCH XUẤT THÔNG TIN CỦA CV**

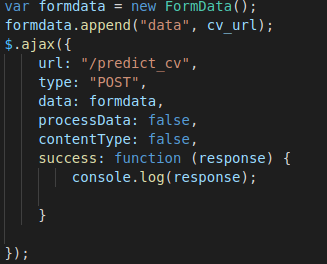
**I. THÔNG TIN CÁC API**

**1. API1: Trả về kết quả trích xuất thông tin của CV**

- Link API: <http://13.212.197.189:8000/predict_cv>

- *Input*: đường dẫn file CV, với định dạng theo FormData

VD cách gọi theo JS:

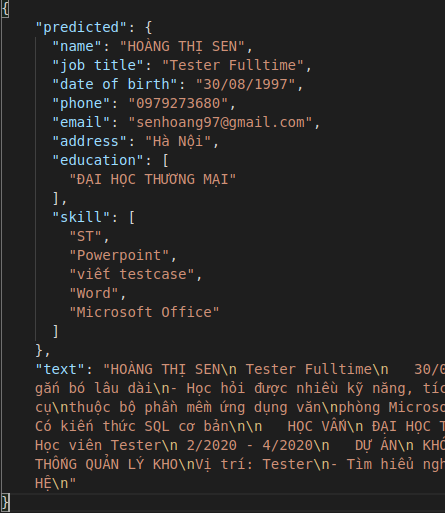


- *Output*: {“predicted”: , “text”: }

Trong đó, “predicted” là kết quả trả về của mô hình

“text” chứa các ký tự từ file pdf hoặc word

VD:



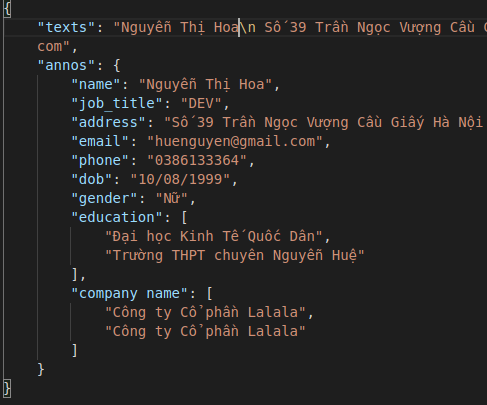
**2. API2: Update thông tin CV**

- Link API: <http://13.212.197.189:8000/update_cv>

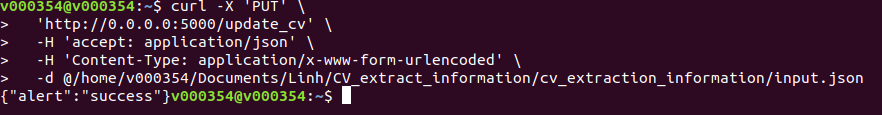
- *Input*: file json với định dạng như sau:

+ texts chứa các ký tự từ file pdf hoặc word (được trả về từ API1)

+ annos chứa các thông tin sau khi chỉnh sửa, các trường "education", "company\_name" là dạng mảng, các trường còn lại là string.



VD cách gọi: truyền file "input.json" như định dạng ở trên



- *Output*: {‘alert’: ‘success’} - thông báo đã chỉnh sửa thành công.

**3. Trang web đánh nhãn**

- Link: <http://13.212.197.189:5000/>

- Tài khoản:

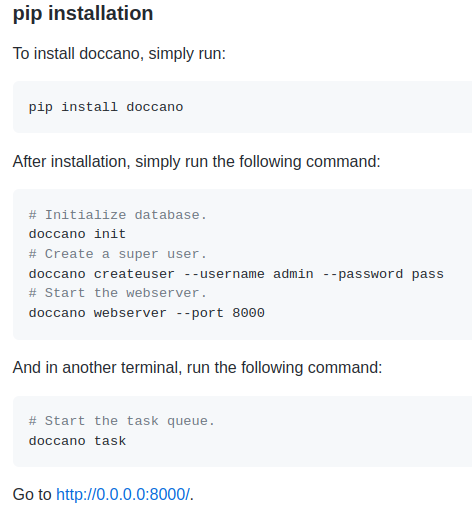
+ User: admin

+ Pass: Vti@1234

- Cách chạy trang web đánh nhãn trên server mới:

+ Tham khảo theo hướng dẫn: <https://github.com/doccano/doccano>

+ Nên sử dụng pip3 (pip) sẽ đơn giản để cài đặt hơn.

****

+ Thay đổi username hoặc password và cổng chạy server.

**II. HƯỚNG DẪN HUẤN LUYỆN MÔ HÌNH**

**1. Các bước huấn luyện mô hình**

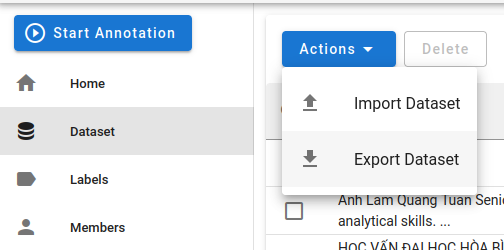
**B1: Cài đặt các thư viện**

Tại cửa sổ terminal của project API, sử dụng câu lệnh sau:

pip3 install -r requirements.txt

**B2: Chuẩn bị dữ liệu**

- Tại cửa sổ "Dataset" của trang web đánh nhãn, click vào "Actions" và chọn "Export Dataset" => chọn JSONL => click vào "Export" để tải dữ liệu về theo format JSONL như hình dưới đây:





- Unzip file vừa tải về, chọn file *admin.jsonl*

- Tại terminal hiện tại, cd vào folder utils: cd utils

- Chạy file *convert\_label.py* theo câu lệnh sau:

python3 convert\_label.py -input đường dẫn file *admin.jsonl* -output đường dẫn file output

+ Có thể bỏ -output vì đã để mặc định file output là file Resumes.json trong folder data.

- Quay lại thư mục project hiện tại: cd ..

**B3: Huấn luyện mô hình**

- Chạy file *train.py* theo câu lệnh sau:

python3 train.py -epoch (số lượng epoch) -input (đường dẫn file output sau khi chạy file *convert\_label.py)* -output (đường dẫn thư mục chứa output của model) -config (đường dẫn file config)

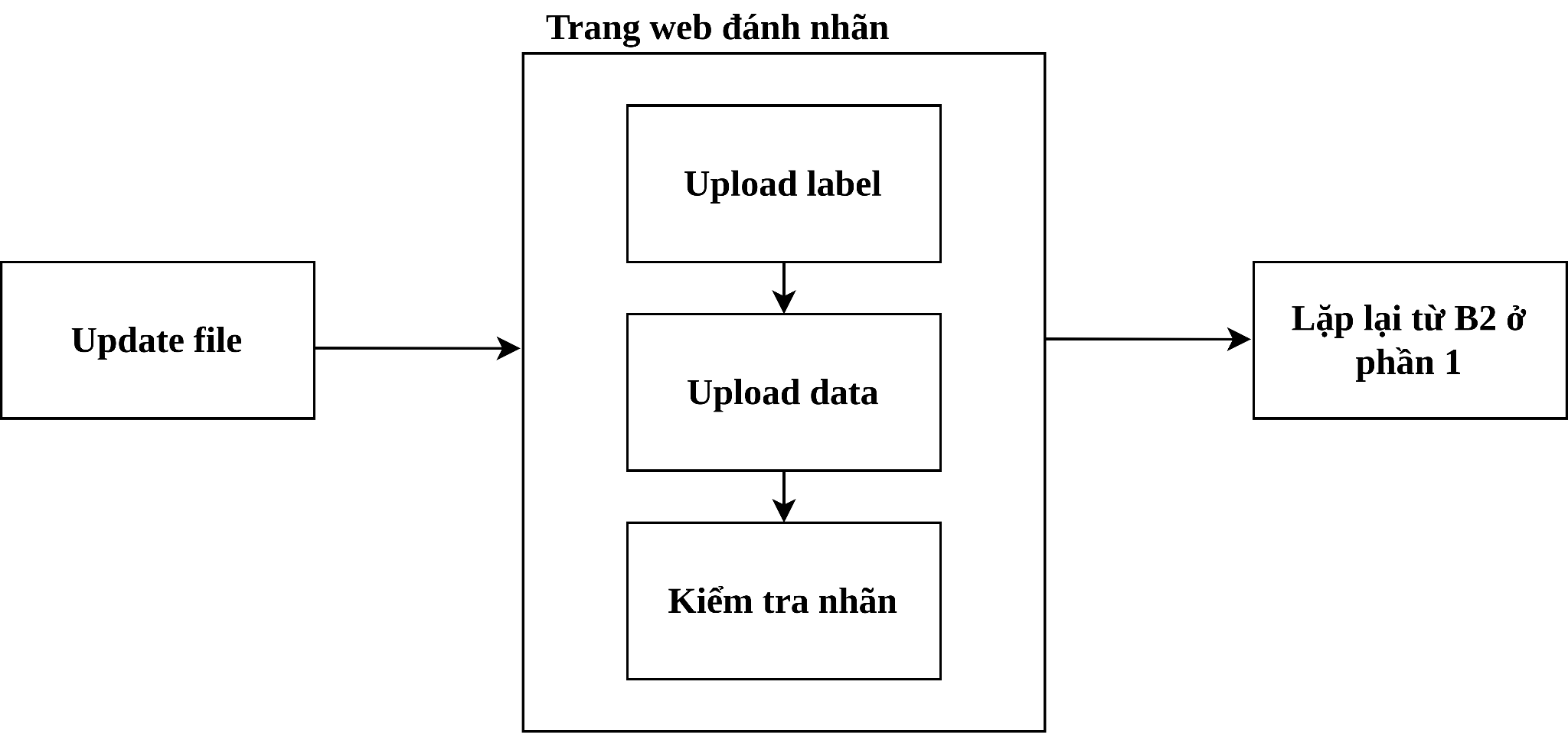
+ Khi chạy, có thể bỏ qua *-output* và *-config*. Vì mặc định thư mục chứa output của model sẽ là thư mục *weights*, và file config mặc định là *configs/config.yaml*

+ Số lượng epoch là số nguyên thường nằm trong khoảng 50-100.

- Sau khi kết thúc training, ta sẽ có được model.

**2. Các bước sau khi update thông tin của CV**

**Tổng quan**



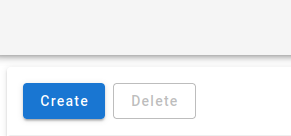
**B1: Update file**

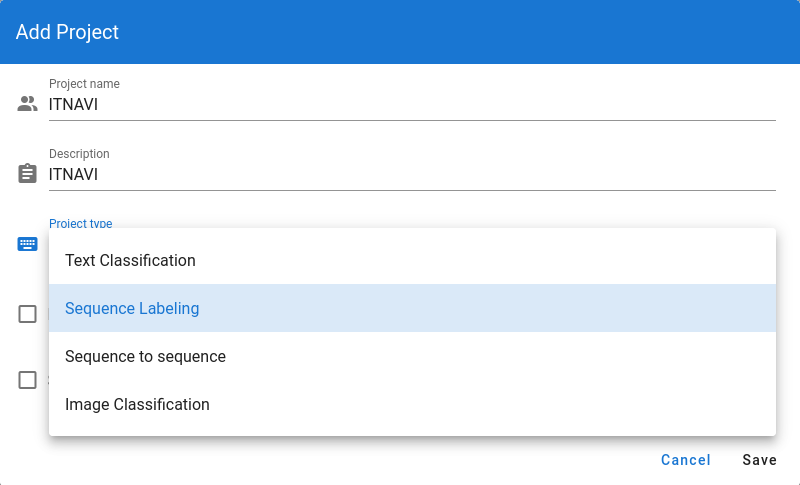
- Update file là file chứa các trường thông tin sau khi update (mặc định: *data/update.txt*.) Nếu muốn đổi tên file hoặc đường dẫn, bạn hãy vào file *configs/config.yaml* tìm đến dòng chứa "update\_file" để sửa đường dẫn.



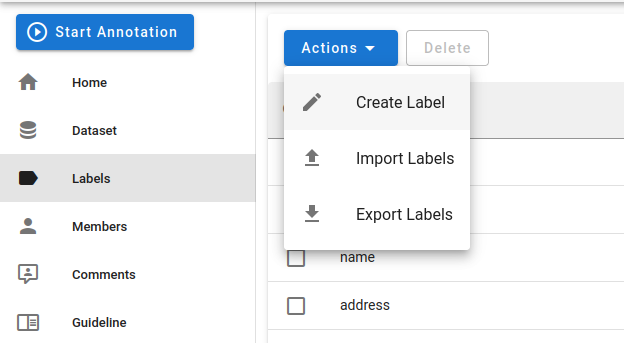
**B2: Kiểm tra nhãn**

- Upload label: Tại trang projects của trang web đánh nhãn, Click *Create* để tạo một project đánh nhãn mới => Điền các thông tin như *Project name, Description* => Ở trường *Project type* chọn *Sequence Labeling* => Click *Save*

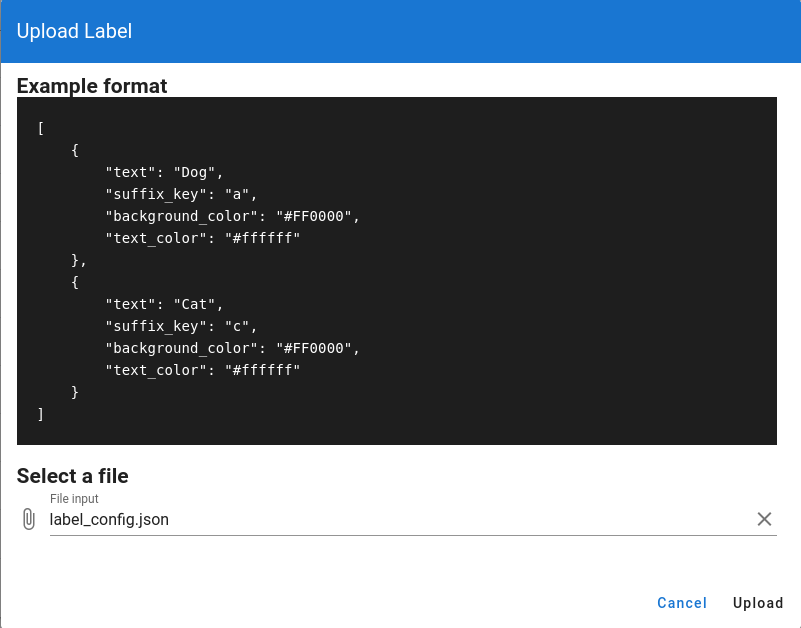




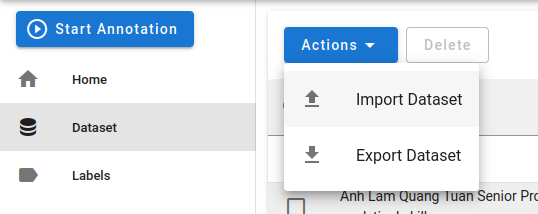
+ Chọn project vừa tạo, ở bên trái cửa sổ chọn *Labels* => Click *Actions* => Chọn *Import Label* nếu chưa có.



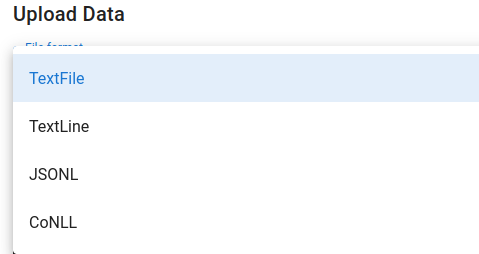
+ Tại cửa sổ Upload, chọn file *configs/label\_config.json* trong project hiện tại để upload lên => Click *Upload*



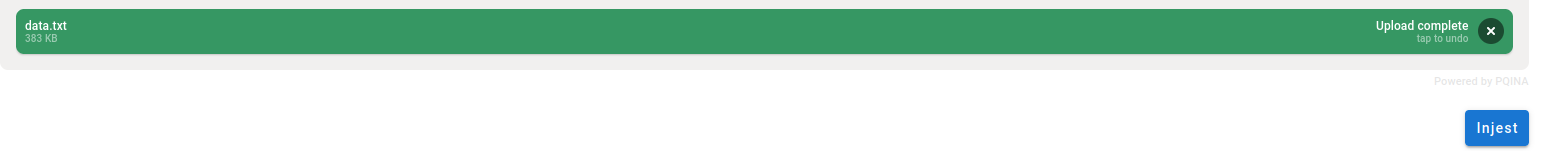
- Upload data: Sau khi upload label thành công, chọn vào *Dataset* phía bên trái cửa sổ => Click *Actions* => Chọn *Import Dataset*



+ Tại cửa sổ *Upload data*, Click *File format* => Chọn *TextFile =>* Click *Drop files here* để tải file *data/update.txt* ở B1 lên.



+ Đợi khi Upload thành công Click *Injest*



+ Chọn lại vào Dataset phía bên trái cửa sổ để tiến hành kiểm tra lại nhãn. Đánh nhãn là bước quan trọng để mô hình có thể học được tốt hơn.

**B3: Huấn luyện tiếp mô hình**

Sau khi đã kiểm tra lại nhãn, lặp lại từ B2 ở phần 1 trên để có thể tiếp tục huấn luyện mô hình.

\* ***Note***: File *configs/skill.txt* chứa các kỹ năng. Nếu update thêm các trường skill, update trong file này. Hoặc đổi đường dẫn ở file *configs/config.yaml* ở dòng skill

