**㈜PCN**

**산업 데이터 처리**

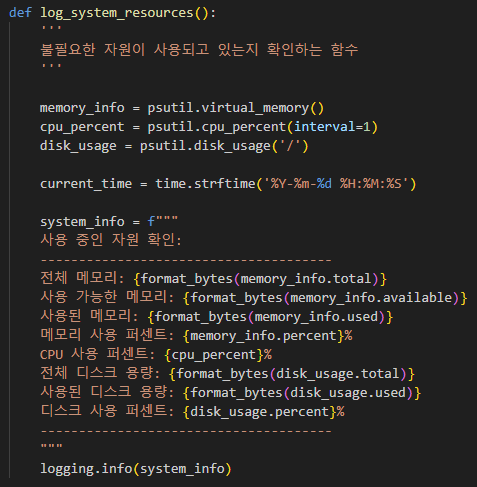
**기술문서**

**Ver. 1.2**

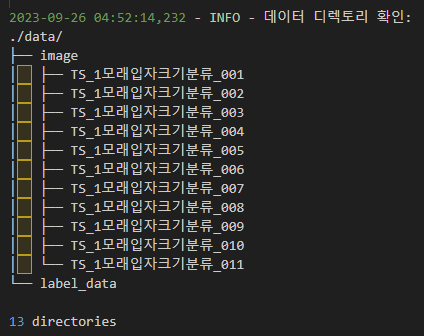
**목 차**

1. **개발환경**
   1. 개발언어
   2. IDE
   3. 라이브러리
2. **시스템 리소스 및 데이터 디렉토리 확인**
   1. 데이터 처리 전 사용 가능한 자원 확인
   2. 저장된 데이터 경로 확인
3. **데이터 로드(Load) 및 메타 데이터 추출** 
   1. 데이터 로드(Load)
   2. 메타 데이터 추출
   3. 테이블 생성
4. **데이터 처리 결과 확인 및 데이터 베이스 저장**
   1. 데이터 처리 결과 확인
   2. 데이터베이스 저장
5. **메타 데이터 추출 결과 조회**
   1. 지도학습 기반 분류 모델 결과 확인

1. **개발환경**
   1. **개발언어**
      1. python 3.10
   2. **IDE**
      1. Vscode
   3. **라이브러리**
      1. os
      2. glob
      3. time
      4. json
      5. psutil
      6. sqlite3
      7. logging
      8. subporcess
      9. pandas
      10. tqdm
      11. pathlib
2. **시스템 리소스 및 데이터 디렉토리 확인**
   1. **데이터 처리 전 사용 가능한 자원 확인**

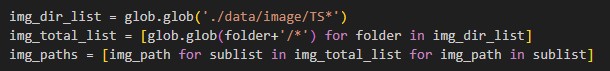


* + 1. 메모리, CPU, 디스크 등 사용 가능한 용량 확인
    2. 메모리, CPU, 디스크 등 사용된 용량 확인
  1. **저장된 데이터 경로 확인**

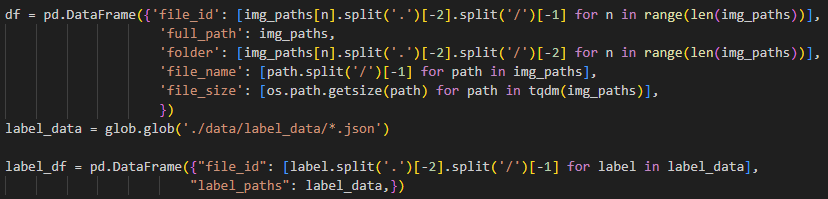


* + 1. 데이터 디렉토리가 정상적으로 존재한다면 데이터 로드 및 추출 시작

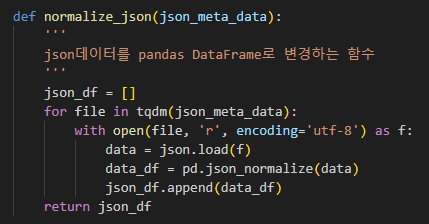
1. **메타 데이터 추출 및 테이블 생성**
   1. **데이터 로드(Load)**



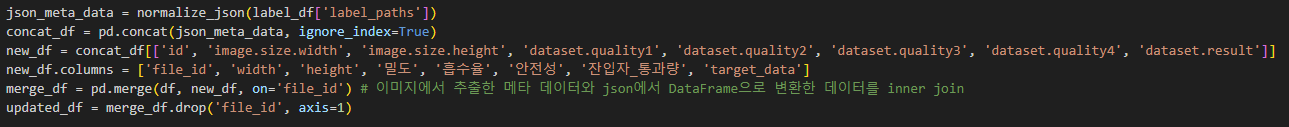
* + 1. 디렉토리에 저장되어 있는 이미지 데이터의 경로를 하나의 리스트 변수에 할당.
  1. **메타 데이터 추출**



* + 1. 리스트에 할당한 전체 데이터 경로를 읽으며 id, path, folder, name, file size 등 추출
    2. 이미지 데이터에 대한 Label 정보를 담고 있는 전체 json 파일의 경로를 하나의 리스트 변수에 할당
  1. **테이블 생성**

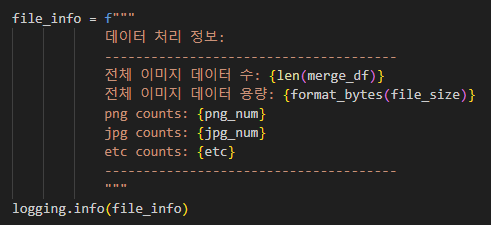


* + 1. 여러 디렉토리에 있는 json파일을 pandas DataFrame로 변환하며 데이터 로드(Load)

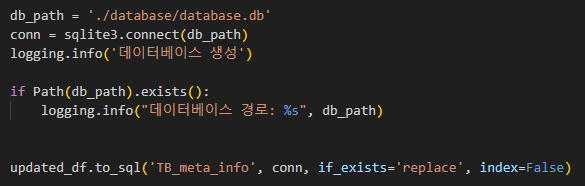


* + 1. 이미지 데이터에서 추출한 메타 데이터와 json에서 Label 정보를 추출해서 생성한 DataFrame을 inner join하여 단일 테이블 생성

1. **데이터 처리 결과 확인 및 데이터베이스 저장**
   1. **데이터 처리 결과 확인**

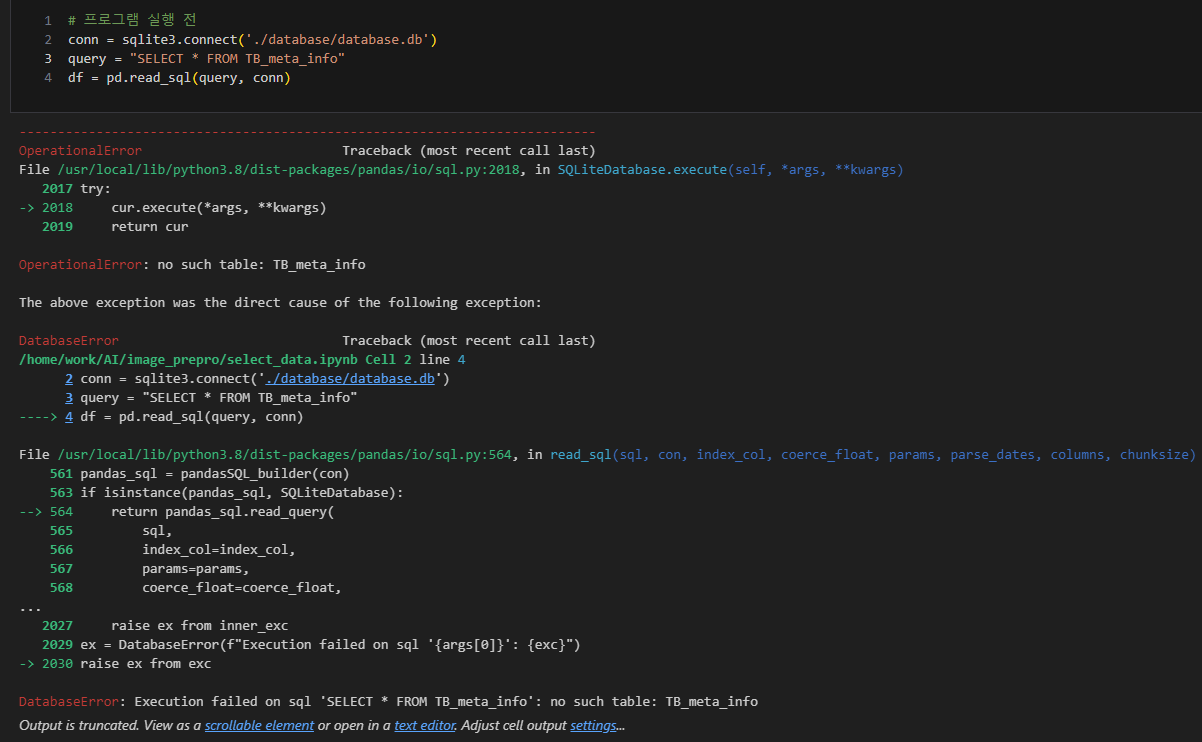


* + 1. 처리한 전체 데이터 수, 용량 로그 저장
  1. **데이터베이스 저장**

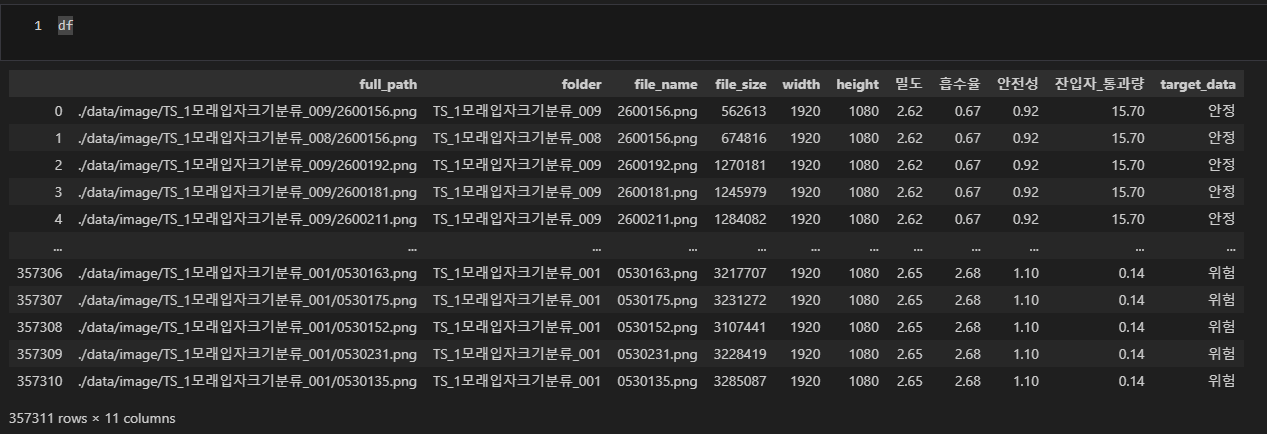


* + 1. SQLite3 데이터베이스 생성 및 추출한 메타 데이터 저장

1. **메타 데이터 추출 결과 조회**
   1. 지도학습 기반 분류 모델 결과 확인

****

* + 1. 데이터베이스에 데이터가 비었는지 확인
    2. 데이터베이스가 비었다면 DatabaseError 출력



* + 1. 데이터베이스에 데이터가 정상적으로 저장되었다면 테이블 조회 가능