

[Project] Detecting Defective Product(Binary Classification)

주제: 주어진 데이터를 가시화, 분석하여 제품 불량 탐지에 적합한 예측 모델 개발.

목표: 변수의 확률과 통계, 가시화 등 다각도의 방법을 활용하여 모델을 제안하고 검증함.

Machine Learning 모델과 신경망(Neural Network) 구조를 활용함.

내용:

- ✓ PCA, 군집화, 신경망(self-supervised learning) 등을 활용하여 feature selection 혹은 추론
- ✓ 분류 모델 또는 신경망 구조를 활용하여 feature selection 혹은 추론
- ✓ 상기 2가지를 개별적, 혹은 병합하는 모델 개발
- ✓ 분석 결과 발표 및 보고서 작성

보고서:

- ✓ 1. 개요, 2. 방법론 및 구현 방법, 3. 실험 결과 및 분석, 4. 결론 형태로 기술
(표지 포함 15p 이하)
- ✓ 가설을 포함하여 개발 과정의 단계를 기술 (e.g., 가시화, 상관관계 분석 등)
- ✓ 제안한 알고리즘 별 정당성 기술

데이터 및 구현 언어 (language):

- ✓ LMS에 첨부된 파일 참조
 - **train.csv** 을 활용하여 label 변수를 예측함 (0 또는 1; binary classification task)
- ✓ 각 데이터의 명세는 하단을 참고.
 - Num : 제품 번호
 - insp : 검수자 코드
 - date : 검수일자
 - buydate : 구입일자
 - ingr_A : 식품 원료 A 함량
 - ingr_B : 식품 원료 B 함량
 - ingr_C : 식품 원료 C 함량
 - wtcd : 수원지 코드
 - gram : 식품 중량
 - label : 불량 여부(0 : 정상, 1 : 불량)
- ✓ 언어 무관 (Python, MatLab 권장; Tensorflow/Keras, Pytorch)

평가 방법:

- 변수의 확률 관계 분석, 주제에 적합한 지표 선정(20%)
- Feature selection 및 예측 모델을 통한 분석 (20%)
- 제안한 알고리즘 모델의 차별성 및 정당성 발표 (30%)
- 보고서 (20%); 구성원 기여도 보고서 표지에 기재
- 소스코드 및 결과 (10%)
 - Copy & paste 체크
 - 결과의 유효성 체크
 - 참조한 source code가 있을 시, citation 필수

제출 양식 명세:

- ✓ 아래 파일을 zip 파일로 압축하여 제출 (압축 파일명 : DS_프로젝트_#팀.zip)
E.g., “DS_프로젝트_01팀.zip”
 - 보고서 PDF 파일 형식 (e.g., 파일명 : Report_DS_#팀.pdf)
 - 소스코드 단일 파일로 제출 (e.g., 소스코드_DS_#팀.zip)
 - jupyter notebook(.ipynb) 파일 형식으로 제출
 - 모델 save파일 제출 (e.g., **model_학번_이름.pth**)
 - EDA 및 분석에 사용된 소스코드 정리하여 모두 제출
 - 동일한 결과를 산출하기 위한 **seed**고정(재현성 평가)
 - **test_with_labels.csv**는 정확도 평가 지표 계산에만 활용
 - test_*.csv 학습에 활용 일절 금지.(활용시 0점 처리)
 - 기타 참고자료 (supplementary materials) (ppt, pdf)

제출 기한:

2023.12.10 (일) 11:59 PM

발표일:

2023.12.11 (월) 오프라인 (304호 예정)