분석 환경

데이터 개요 및 분석 목표

학습 모델

* 머신러닝 (오토, 개별 5개)
* 딥러닝 (CNN, RNN)

===========================

데이터 분석

머신러닝 학습/예측

딥러닝 학습/예측

데이터 전처리

머신러닝 학습/예측

CNN 학습/예측

최종 결과

[분석 개요]

1. 사용 언어 : Python
2. 분석 환경 : [CPU] Apple M2, [RAM] 8GB, [GPU] Colab T4
3. 분석 데이터 정보
   1. 수집 기관 : Ford 자동차 제조사
   2. 데이터 내용 : 엔진 상태와 관련 있는 센서 500개의 측정 값
   3. 데이터 파일 : Train Dataset 1file, Test Dataset 1file
4. 분석 데이터 구조
   1. Features : 500 columns (Sensor values)
   2. Samples : 4921 raw (Train 3601, Test 1320)
   3. Labels : 1 or 0 (Normal or Abnormal)
5. 학습 및 예측 모델
   1. LogisticRegression
   2. TabularPredictor (Auto ML)
   3. XGBClassifier
   4. LGBMClassifier
   5. CatBoostClassifier
   6. RandomForestClassifier
   7. CNN
   8. RNN

[데이터 분석]

1. 데이터를 불러와서 DataFrame 형식으로 변환

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

1. Train, Test dataset info.

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

* + train\_df : 3601 samples, 501 columns (Include target 1)
  + test\_df : 1320 samples, 501 columns (Include target 1)
  + 500 feature type : float64, target type : object

1. Target value 1과 0으로 변환

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

1. Target value 빈도의 균형이 맞는지 확인

텍스트, 스크린샷, 소프트웨어, 도표이(가) 표시된 사진

자동 생성된 설명

* + train, test dataset 의 target value 빈도는 균형이 맞다.

1. features value 분포 확인

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

1. 0과 1 클래스 각 샘플의 센서 값 분포 비교 (각 3개 샘플 비교)

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

* + 클래스별 센서 값에 차이가 있음 -> 학습 가능할 것으로 보임
  + 전체 샘플의 센서 별 평균도 볼 필요 있음 (클래스 구분)

1. 센서 별 클래스 평균 값 분포

텍스트, 스크린샷, 폰트, 소프트웨어이(가) 표시된 사진

자동 생성된 설명

도표, 그래프, 스크린샷이(가) 표시된 사진

자동 생성된 설명

* + 센서 후반부 (s403 ~) 클래스에 따라 센서 값이 양수와 음수로 나눠짐 -> 후반부 센서만 학습시켜 볼 필요 있음

1. 전체 샘플에 대한 센서 값들의 분포 (클래스별)

텍스트, 스크린샷, 라인, 폰트이(가) 표시된 사진

자동 생성된 설명

* + 각 샘플 값이 시계열 데이터일 경우 시간의 흐름에 따른 패턴은 보이지 않음
  + RNN 모델로 학습 해볼 필요는 있음.
  + 클래스별 평균 값으로 다시 시각화 해볼 필요 있음 (다음 장)

1. 샘플별 500개 센서 값의 평균

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

* + 출력 결과 다음 페이지

스크린샷, 텍스트, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

* + 0클래스에 노이즈 샘플이 있음.
  + 노이즈 샘플 제거 후 학습 해볼 필요 있음.

1. 센서 별 상관관계 분석

텍스트, 스크린샷, 다채로움, 디자인이(가) 표시된 사진

자동 생성된 설명

인접한 센서와 상관계수가 높음

* CNN 으로 학습 가능 할 것으로 보임
* kernel\_size를 3으로하면 학습 효과가 높을 것으로 보임

1. Target에 영향을 주는 센서 확인

텍스트, 스크린샷, 디자인이(가) 표시된 사진

자동 생성된 설명