
기계학습 4조 중간발표

21011910 박지호

22011401 정세은

정재석

목차

1. 세부 주제 설명
2. 선정한 Dataset과 특징
3. 세부 주제와 Dataset의 적합성
4. 앞으로의 계획

세부주제

- 산업 현장에서 기계학습 기반 예지보전(Predictive Maintenance) 데이터 분석 및 모델 개발
- 예지보전(Predictive Maintenance)은 센서 데이터와 기계의 운전 정보를 바탕으로 고장이나 이상 징후를 사전에 예측하여 불필요한 정비를 줄이고 생산 효율을 높이는 대표적인 산업 현장 적용 사례
- 기계학습 기반 예지보전은 실제 산업 설비의 유지 보수 비용 절감, 생산성 향상, 안전사고 예방 등 실질적인 비즈니스 가치를 창출할 수 있어, 산업 데이터 분석 및 모델 개발의 대표적인 활용 분야

선정한 Dataset

- Machine Predictive Maintenance Classification
([Machine Predictive Maintenance Classification](#))
- 기계 고장 여부와 고장 유형을 예측하는 dataset
 - 목적 : 센서 및 공정 데이터를 바탕으로 기계 고장 예측 및 원인 분석
 - 샘플 수 : 10,000개

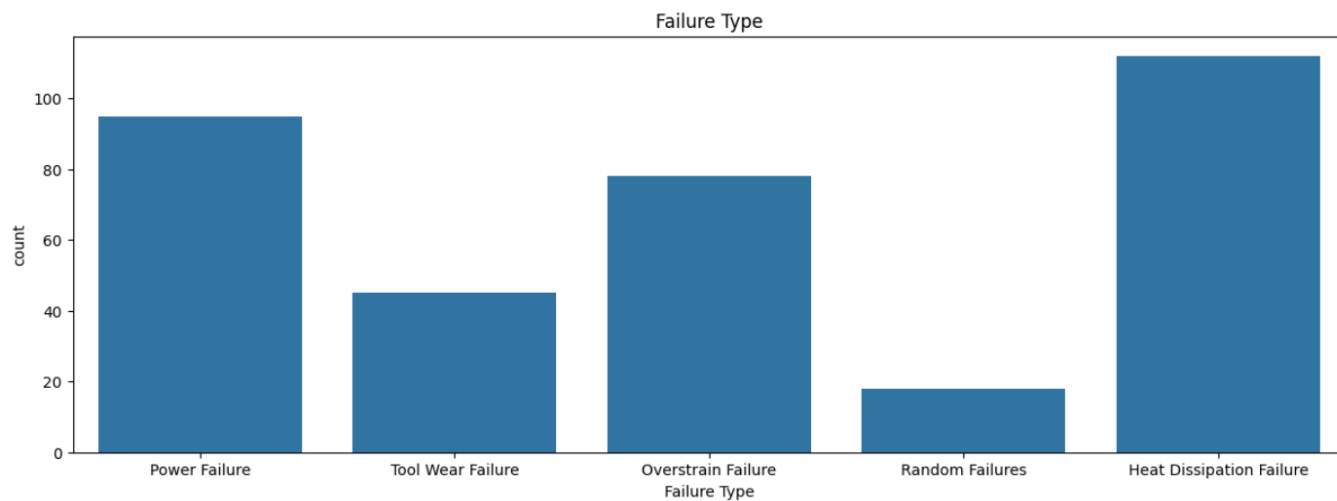
데이터셋의 description, 특징

- 주요 변수

변수명	설명	특징/범위
UID	데이터 행 고유 번호	1~10,000
productID	제품 고유 식별자	예: M14860, L47181 등
Type	제품 등급 (L/M/H)	L(50%), M(30%), H(20%)
air temperature [K]	기계 작동 환경 온도	평균 300K, 표준편차 2K
process temperature [K]	공정 내 온도	air temp+10K, std 1K
rotational speed [rpm]	회전속도	평균 1500~1700, 노이즈
torque [Nm]	회전축 토크	평균 40, std 10, 음수X
tool wear [min]	공구 누적 마모 시간	등급별 누적 증가

데이터셋의 description, 특징

- 타겟 변수
 - Target (0 - 정상, 1 - 고장)
 - Failure Type : (고장 유형, 다중 클래스)
 - 5가지로 분류
 - 고장 유형 중 하나라도 발생 시 고장 (Target = 1)



세부 주제와 Dataset의 적합성

- 예지보전은 실제 산업 현장에서 활용되고 있는 중요한 기술
 - 예지보전(Predictive Maintenance)은 센서 데이터를 이용해 기계가 언제 고장 날지를 미리 예측해 불필요한 정비를 줄이고, 생산 효율을 높일 수 있는 대표적인 산업 응용 분야
- 수업 시간에 배운 모델들을 적용하고 비교하기 좋은 데이터셋
 - 이 데이터는 장비에서 측정한 다양한 센서 값과 고장 여부(Target)가 포함되어 있어, 로지스틱 회귀, KNN, 판별 분석, 결정 트리 등 다양한 모델로 고장 여부를 예측하는 데 사용 가능
 - Target이 0과 1로 구분되어 있는 이진 분류 문제이기 때문에, 수업 시간에 배운 내용과 직접 연결 가능
 - 필요하다면, 앙상블 학습 방법 사용 예정 (ex. 랜덤 포레스트 등)
- 실제 문제를 해결하는 과정을 경험할 수 있음
 - 데이터를 나누고(Cross Validation), 여러 모델을 비교해보고, 어떤 모델이 정확도가 높은지 확인해보면서 실제 현장에서 어떤 모델이 더 좋은지를 판단하는 과정을 직접 경험할 수 있음

프로젝트 방향성

- 데이터 분석 및 전처리

- 데이터셋(predictive_maintenance.csv) 분석
- 결측치, 이상값, 클래스 비율 확인 및 처리
- '고장 여부'를 Target으로 이진 분류 문제 정의

- 모델 학습 및 비교 실험

- 수업에서 배운 기법들 적용:
 - Logistic Regression
 - K-Nearest Neighbors (KNN)
 - Linear/Quadratic Discriminant Analysis (LDA/QDA)
 - Decision Tree
- 각 모델에 대해 Cross Validation을 통한 성능 비교

프로젝트 방향성

- 실전 문제 고려 요소 반영
 - 클래스 불균형 문제 대응 (예: 샘플링, 평가 지표 활용 등)
 - 모델 성능 평가 (정확도뿐 아니라 Recall, F1-score 등도 고려)
- 결과 해석 및 산업 적용 가능성 논의
 - 어떤 모델이 고장을 가장 잘 예측하는지 분석
 - 산업 현장에서의 적용 시나리오 제시