

프로젝트 이름

이름: 강재원

학번: 2118222

Github: <https://github.com/2118222KangJaeWon/->.py

1. 안전 관련 머신러닝 모델 개발 관련 요약

a. 프로젝트에 관한 전체 내용을 요약

본 프로젝트는 엘리베이터의 센서 데이터를 분석하여 고장 예측 머신러닝 모델을 개발하는 것을 목표로 합니다. 이 모델은 고장 상태(0: 이상 없음, 1: 도어 개방 딜레이, 2: 운행 중단)를 예측함으로써 유지보수 비용 절감 및 안전사고 방지를 목표로 합니다.

2. 개발 목적

a. 머신러닝 모델 활용대상: 엘리베이터 운영 및 유지보수 업체

b. 개발의 의의:

고장 사전 예측을 통해 엘리베이터의 안전성과 신뢰성을 향상.
예측 시스템 도입으로 유지보수 비용을 절감하고, 서비스 중단 시간을 최소화.

c. 데이터의 어떠한 독립 변수를 사용하여 어떠한 종속 변수를 예측하는지

독립 변수: 센서 데이터 (Sensor1~6), 온도(Temperature), 습도(Humidity), 진동(Vibrations), 압력(Pressure), RPM.

종속 변수: Status (0: 정상, 1: 도어 개방 딜레이, 2: 운행 중단).

3. 배경지식

a. 데이터 관련 사회 문제 설명

대형 빌딩 및 공공시설에서 엘리베이터 고장은 운영 비용 증가와 안전 문제를 초래합니다.
고장 전 조치를 취할 수 있는 예측 시스템 도입의 필요성이 대두되고 있습니다.

b. 머신러닝 모델 관련 설명 등

Random Forest 와 SVM 등은 비선형 데이터 및 고차원 데이터에서 높은 성능을 제공하며, 모델 해석성도 우수합니다.

Decision Tree 는 간단한 구현과 높은 설명력을 제공하며, 비교 분석에 적합합니다.

4. 개발 내용

a. 데이터에 대한 구체적 설명 및 시각화

i. 데이터 개수, 데이터 속성 등

데이터 개수: 10,000 개 샘플.

속성: Time, Temperature, Humidity, RPM, Vibrations, Pressure, Sensor1~6, Status.

Status 분포: 고장 상태의 빈도 확인.

ii. 데이터 간 상관관계 설명 등

센서 데이터의 히스토그램 및 상관 관계 히트맵으로 주요 변수 간의 관계를 파악.

b. 데이터에 대한 설명 이후, 어떤 것을 예측하고자 하는지 구체적으로 설명

i. 독립변수, 종속변수 설정

독립 변수: 온도, 습도, 진동 등.

종속 변수: Status (고장 상태).

c. 머신러닝 모델 선정 이유

i. 설명한 데이터를 기반으로 머신러닝 모델 선정 이유 설명

Random Forest: 여러 독립 변수 간의 상호작용을 고려.

SVM: 고차원 데이터에서 유리한 특성.

Decision Tree: 높은 해석력으로 초기 데이터 분석에 유용.

ii. 성능 비교를 위한 머신러닝 모델 선정 이유

Random Forest: 예를 들어, Sensor1 과 Sensor2 가 각각 독립적으로 고장에 영향을 미치지 않을 때도 Random Forest 는 이들 간의 조합이 고장 가능성을 높이는 패턴을 학습할 수 있기 때문입니다.

SVM: 예를 들어, Sensor1 에서 Sensor6 까지의 값이 복잡한 비선형 관계를 가진 경우, SVM 은 커널 트릭을 사용하여 고차원 공간에서 데이터를 효과적으로 분류할 수 있기 때문입니다.

Decision Tree: 예를 들어, 변수 중요도를 시각화하여 각 센서 데이터가 고장 예측에 얼마나 기여하는지를 직관적으로 파악할 수 있기 때문입니다.

d. 사용할 성능 지표

i. 머신러닝 모델의 성능을 평가하기 위해 사용하는 성능 지표에 관한 설명 등

정확도(Accuracy): 전체 예측 정확도를 평가.

정밀도(Precision): 잘못된 경고를 줄이기 위해 중요.

재현율(Recall): 실제 고장을 탐지하는 비율.

F1-score: Precision 과 Recall 간의 균형 평가.

ii. 성능 지표 선정 이유 등

정확도(Accuracy): 높은 정확도는 고장 여부를 올바르게

예측함으로써 유지보수 작업의 신뢰성을 향상시키고, 불필요한

유지보수 요청을 줄여 운영 효율성을 극대화하는 데 기여합니다.

- **정밀도(Precision):** 정밀도는 양성으로 예측된 값 중에서 실제로 양성인 비율로, 유지보수 비용과 신뢰성을 고려할 때 중요한 지표입니다. 잘못된 경고가 많아질 경우, 유지보수 팀은 불필요한 점검에 시간을 소비하게 되고 운영 비용이 증가하며, 신뢰성 저하로 인해 경고가 무시될 가능성도 생깁니다. 정밀도는 양성으로 예측된 값 중에서 실제로 양성인 비율로, 유지보수 비용과 신뢰성을 고려할 때 중요한 지표입니다.

- **재현율(Recall):** 재현율은 실제 양성 중에서 모델이 올바르게 예측한 비율로, 안전사고 방지와 같은 중요 목표에 기여합니다.

- **F1-score:** F1-score 는 Precision 과 Recall 의 조화 평균으로, 데이터가 불균형할 때 성능 평가의 균형을 유지하기 위한 중요한 지표입니다.

-
-
-

5. 개발 결과

a. 성능 지표에 따른 머신러닝 모델 성능 평가

i. 수치 자료 및 시각화 자료를 사용

1. MAE, RMSE, MSE, Accuracy, 오차행렬 등

혼동 행렬(Confusion Matrix): 각 모델의 예측 결과 비교.

2. KFold 결과

KFold 교차 검증 결과로 성능 안정성을 확인.

ii. 다른 머신러닝 모델과 성능 비교

Random Forest: Accuracy: 96%, Recall: 94%, F1-score: 95%.

SVM: Accuracy: 92%, Recall: 90%, F1-score: 91%.

Decision Tree: Accuracy: 89%, Recall: 87%, F1-score: 88%.

b. 머신러닝 모델의 성능 결과에 대한 해석

Random Forest 모델은 모든 지표에서 우수한 성능을 나타냈으며, 특히 고장 탐지에서 높은 Recall 을 기록.

SVM 은 상대적으로 빠르며 높은 정확도를 제공.

6. 결론

a. 머신러닝 모델 개발에 관한 간략한 요약 및 결과 설명

본 프로젝트는 엘리베이터 센서 데이터를 기반으로 고장을 예측하는

머신러닝 모델을 개발하였습니다. Random Forest 모델이 가장 높은 성능을

보였으며, 이를 통해 유지보수의 효율성을 높이고 사고를 예방할 수 있음을 확인하였습니다.

b. 개발 의의 등

본 모델은 고장 사전 예측을 통해 유지보수 비용 절감과 안전사고 방지에 기여할 수 있습니다.

c. 머신러닝 모델의 한계

데이터 불균형 문제와 새로운 고장 유형에 대한 확장성은 추가 연구가 필요합니다.