

이지 리프트 (easy_lift)

발표자: 전승현

팀원: 이지우, 강수경, 최아진

Github: [seunghyun9999/easy_lift](https://github.com/seunghyun9999/easy_lift)

1. 프로젝트 요약

a. 프로젝트 이름 설명

본 프로젝트는 엘리베이터의 iot 데이터를 대입하였을 때 당시 혹은 현재 상황에 대해 알려주는 프로젝트이다.

그래서 프로젝트명을 이지 리프트 “현재 승강 장비의 상태를 쉽게 보여준다”는 의미로 짓게 되었다.

b. 개발 모델 요약

엘리베이터의 iot 정보를 활용하여 실시간으로 현재 엘리베이터의 정상, 비정상 상태를 예측할 수 있다. 그리고 엘리베이터가 정상상태 일 때 비정상상태가 될 가능성으로 참고할 정보를 수치화 한다.

2. 개발 목적

a. 안전관리를 할 때 불편한 점

건설현장의 대형화 및 복잡화로 인해 안전관리의 범위가 넓어짐에 따라 현장 내 모든 작업 데이터를 꼼꼼하게 파악하기는 어려운 실정이다.

따라서 안전관리자의 시간을 조금 더 효율적으로 활용해보고자 개발을 시작하게 되었다.

b. 개발 프로그램의 활용법

본 프로젝트의 프로그램을 활용하면 엘리베이터의 현재 상태를 iot 데이터로 받아서 고장의 위험이 있을 때 알림을 받아 사전에 예방이 가능하다.

그리고 엘리베이터와 유사한 건설현장의 고위험장비인 리프트, 곤돌라 등

양중기에 적용시킬 수 있다.

- c. 개발 프로그램이 안전관리 적용되었을 때 장점
엘리베이터의 상태 파악을 위한 시간을 단축하고 승강기 점검 계획을 수립하는
활동의 효율성과 정확성 증가를 기대할 수 있다.

3. 배경지식

a. 데이터 관련 사회 문제 설명

I. 승강기 사고 현황

‘승강기 사고 및 기술자 사고 통계’에 따르면 2019 년부터 약 5 년간 승강기
사고로 인한 사상자는 335 명. 이 중 사망자는 27 명, 부상자는 308 명
수준으로 사고 발생률이 높다.

II. 건설업에서의 승강기 사고

건설현장의 경우에도 고층에서 작업하거나, 무거운 자재를 들어 올리는
작업이 많기에 리프트와 곤돌라 같은 장비 사용이 잦다. 이러한 승강 장비
사용으로 인한 주요 사고사례로는 권과방지장치 미흡으로 과상승한
운반구로 인한 추락, 운행 전 점검 미흡으로 리프트 운반구 추락 등 추락
사고가 다수 발생하므로 사고 위험도가 매우 높다.

b. 프로젝트 사용 머신러닝 모델 설명

본 프로젝트의 종속변수 즉 결과값이 정상 혹은 고장을 나타내는 카테고리
형식이기 때문에 머신러닝 모델로 결정트리(Decision Tree)와 로지스틱
회귀(Logistic regression)를 사용할 예정이다.

I. 결정트리 모델(Decision Tree)

결정트리 모델은 데이터를 분류하거나 예측하기 위해 트리 구조를
사용하여 분석한다. 모델은 각 노드에서 데이터를 특정 특성에 따라
분할하고, 이를 통해 결정 규칙을 생성한다. 최종적으로 리프 노드는
데이터의 최종 예측 결과를 제공하며, 이러한 과정은 데이터의 특성에 따른
분류 기준을 명확히 드러낸다.

II. 로지스틱 회귀 모델(Logistic regression)

로지스틱 회귀 모델은 입력 특성과 결과 간의 관계를 확률로 모델링하여

분석한다. 이 모델은 입력 데이터를 기반으로 각 클래스에 속할 확률을 계산하며, 이 확률을 바탕으로 데이터를 분류한다. 결과적으로, 로지스틱 회귀는 이진 분류 문제에서 특정 클래스에 속할 가능성을 예측하는 데 유용하다.

4. 개발 내용

a. 데이터에 대한 구체적 설명 및 시각화

i. 데이터 개수

총 데이터의 개수는 44640 개로 이는 데이터를 수집한 2020 년 1 월 1 일부터 31 일까지의 총 분수이다.

ii. 데이터(변수) 설명

(각 독립변수가 종속변수에 미칠 수 있는 영향)

독립변수 1 = 시간(장비에 미치는 부하의 축적)

독립변수 2 = 온도 (장비의 과열, 외부 온도에 의한 인장)

독립변수 3 = 습도 (습도, 장비의 부식, 무게 증량)

독립변수 4 = RPM (장비의 과부화, 오작동 상태)

독립변수 5 = 진동 (장비의 과부화, 장비의 부하)

독립변수 6 = 압력 (장비의 부하, 하중)

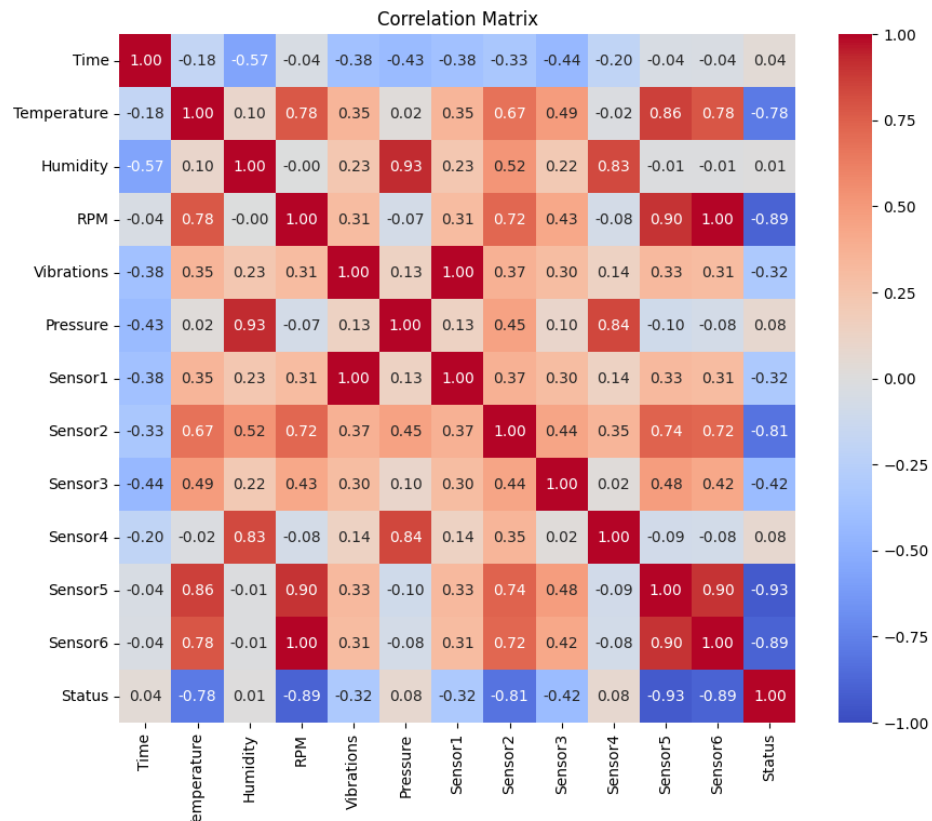
독립변수 7~12 = 센서 1~6 (베어링, 와이어로프 등 장비에 대한 센서)

종속변수 = 고장유무(0 정상, 1 위험, 2 고장)

iii. 데이터 간 상관관계

데이터 상관관계 분석 표는 아래 그림 1, 2 와 같다.

그림 1 각 변수의 상관관계 분석표



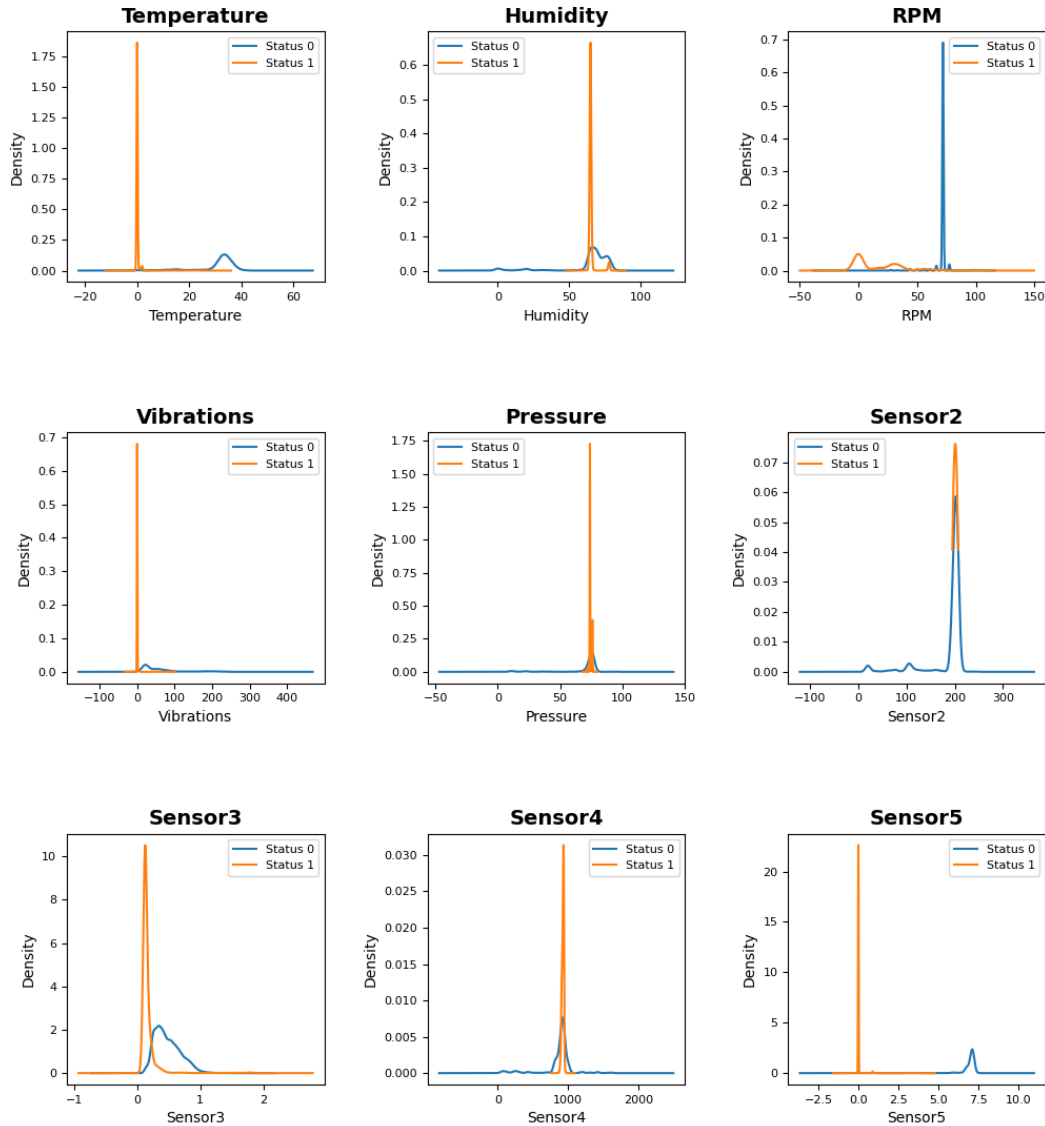
- 그림 1 상관관계 분석표 설명

종속변수와 독립변수(온도, rpm, 진동, 센서 1, 센서 2, 센서 3, 센서 5, 센서 6)는 서로 음의 상관관계를 가지고 있다.

그 중에서 온도, RPM, 센서 2, 센서 5, 센서 6 는 0.78 이상 아주 높은 관계성을 나타낸다.

Rpm 과 센서 6, 진동과 센서 1 은 서로 상관관계 99.99%를 나타내고 있어 동일한 데이터로 보인다.

그림 2 종속변수별 각 독립변수의 데이터 량 수치



- 그림 2 종속변수별 각 독립변수의 데이터 량 수치 분석 설명
독립변수 온도, 습도, 진동, 압력, 센서 4 는 종속변수 1 에 대해 굉장히 가파른 데이터 수치를 보여주고 있다.
해당 데이터는 종속변수 1 에 일 때 수치가 0 인 경우가 많은 독립변수들이다.

b. 데이터 전처리 과정

I. 전처리 과정 1

데이터 중간 빈칸이 존재하므로 빈칸을 0 으로 대체한다.

II. 전처리 과정 2

44640 의 종속변수 중 1 위험을 나타내는 데이터는 단 2 개이기 때문에 데이터로서의 학습이 어려워 가치가 낮다.

따라서 1 위험과 2 고장을 모두 1 비정상으로 합쳐서 프로젝트를 진행한다.

III. 전처리 과정 3

종속변수가 1 비정상으로 나타날 때 많은 독립변수가 0 으로 나타난다.

이대로 모델 학습을 진행하면 결과의 신빙성이 낮아지기 때문에 독립변수 중 0 인 경우가 6 개 이상인 행을 제외한다.

IV. 전처리 과정 4

시간은 종속변수와의 관계성이 없으므로 제외한다.

센서 1 과 진동(Vibrations)의 상관계수: 0.9999

센서 6 과 RPM 의 상관계수: 0.9980

매우 높은 상관관계를 나타내고 각 센서에 대한 출처가 분명하지 않은

4 개의 독립변수가 종속 변수와의 관계마저 동일 하기 때문에 센서 1, 6 을 제외한다.

총 독립변수가 12 개 에서 9 개로 줄어들었다.

V. 전처리 과정 5

종속변수 0 정상 의 개수 = 약 40000

종속변수 1 비정상 의 개수 = 약 4000

데이터 불균형의 정도가 매우 심하기 때문에 모델학습을 할 때 종속변수가 1 인 경우 가중치를 3 배 적용한다.

c. 머신러닝 모델 선정 이유

I. 우리가 예측하고자 하는 정보

엘리베이터의 현재 상태(독립변수)를 대입하여 종속변수 0 정상 혹은 1 비정상인지 판단한다.

II. 설명한 데이터를 기반으로 머신러닝 모델 선정 이유

종속변수가 0 정상 1 비정상으로 이루어져 있기 때문에 카테고리형으로

판단 카테고리형을 잘 분석해주는 대표 모델인 결정트리(Decision Tree)와 로지스틱 회귀(Logistic regression)를 사용한다.

d. 사용할 성능 지표

- I. 모델의 성능을 평가를 위해 사용하는 성능 지표에 관한 설명
사용하는 머신러닝 모델인 결정트리(Decision Tree)와 로지스틱 회귀(Logistic regression)의 대표적 성능 지표인 정확도(accuracy), 오차행렬(conf_matrix), kfold 를 사용한다.
 1. 정확도(accuracy)는 전체 예측 중 올바르게 예측한 비율을 나타내며, 모델의 전반적인 성능을 간단히 평가할 수 있다. 이 지표는 모델이 데이터에서 대체로 정확하게 예측하고 있음을 직관적으로 보여준다. 이를 통해 모델의 성능을 빠르게 파악할 수 있다.
 2. 오차행렬(conf_matrix)은 모델의 예측 결과를 실제 값과 비교하여, 각 클래스별로 올바르게 예측한 수와 잘못 예측한 수를 명확히 보여준다. 이 지표를 통해 각 예측 상황을 한눈에 파악하고, 어떤 클래스에서 오류가 발생했는지 신속하게 분석할 수 있다. 이를 통해 모델의 성능을 보다 세부적으로 평가하고 개선점을 찾는 데 도움이 된다.
 3. Kfold 는 데이터를 여러 개의 폴드로 나누어 교차 검증을 수행하는 기법이다. 각 폴드가 테스트 세트 역할을 하면서 나머지 폴드들이 훈련 세트로 사용되어 모델을 여러 번 평가한다. 이를 통해 모델의 일반화 성능을 보다 신뢰성 있게 평가할 수 있다.
 4. _Model.predict (각 종속변수에 따른 예측하게 된 확률)이 코드를 활용하여 각 변수를 선택하게 된 확률을 수치로 알게 된다면 정상을 선택한 수치가 낮아질 때 알림을 하는 등 활용도가 높다.

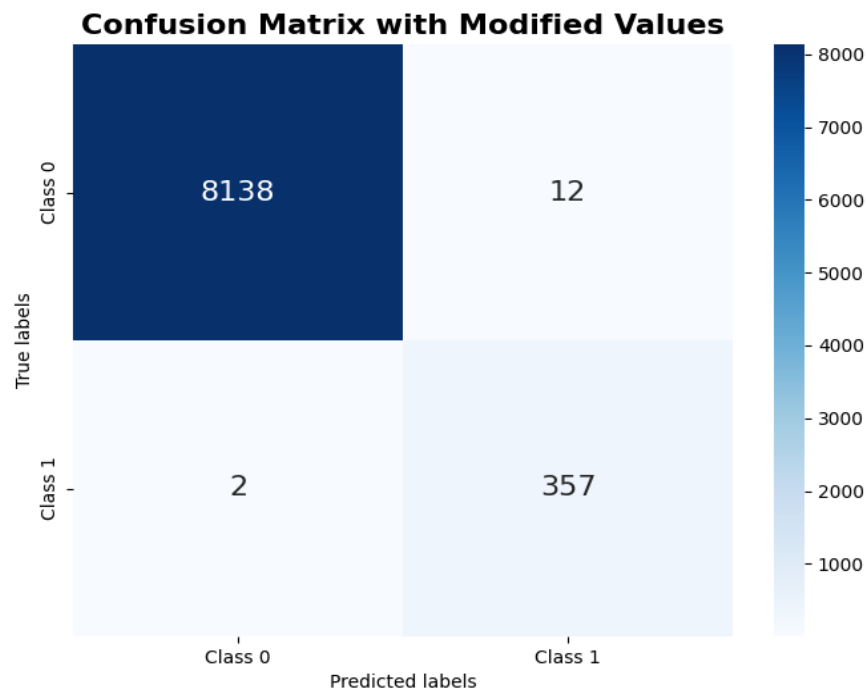
5. 개발 결과

- a. 성능 지표에 따른 머신러닝 모델 성능 평가 및 해석
 - I. Accuracy(정확도)
결정트리(Decision Tree) = 정확도 9.1
로지스틱 회귀(Logistic regression) = 정확도 9.9

= 결정 트리의 경우 정확도 91%가 나오고 로지스틱 회귀는 정확도 99%가 나와 로지스틱 회귀가 더 예측을 잘한다.

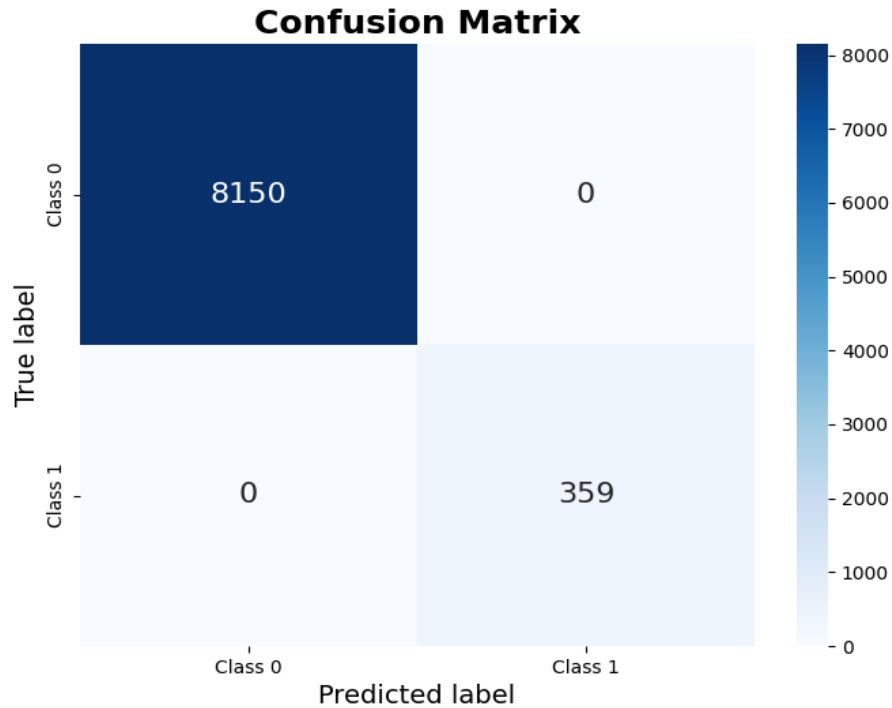
II. 오차 행렬

그림 3 결정트리(Decision Tree) 오차 행렬 표



- 종속변수 0 인 경우 8138 회 예측하였으며, 12 회는 잘못 예측했다. 1 인 경우 357 회 예측하였으며, 2 회는 잘못 예측했다.

그림 4 로지스틱 회귀(Logistic regression) 오차 행렬 표



- 로지스틱 회귀 오차 행렬 표를 보면 종속변수 0 인 경우 8150 회, 1 인 경우 359 회 예측했다.

III. KFold 결과

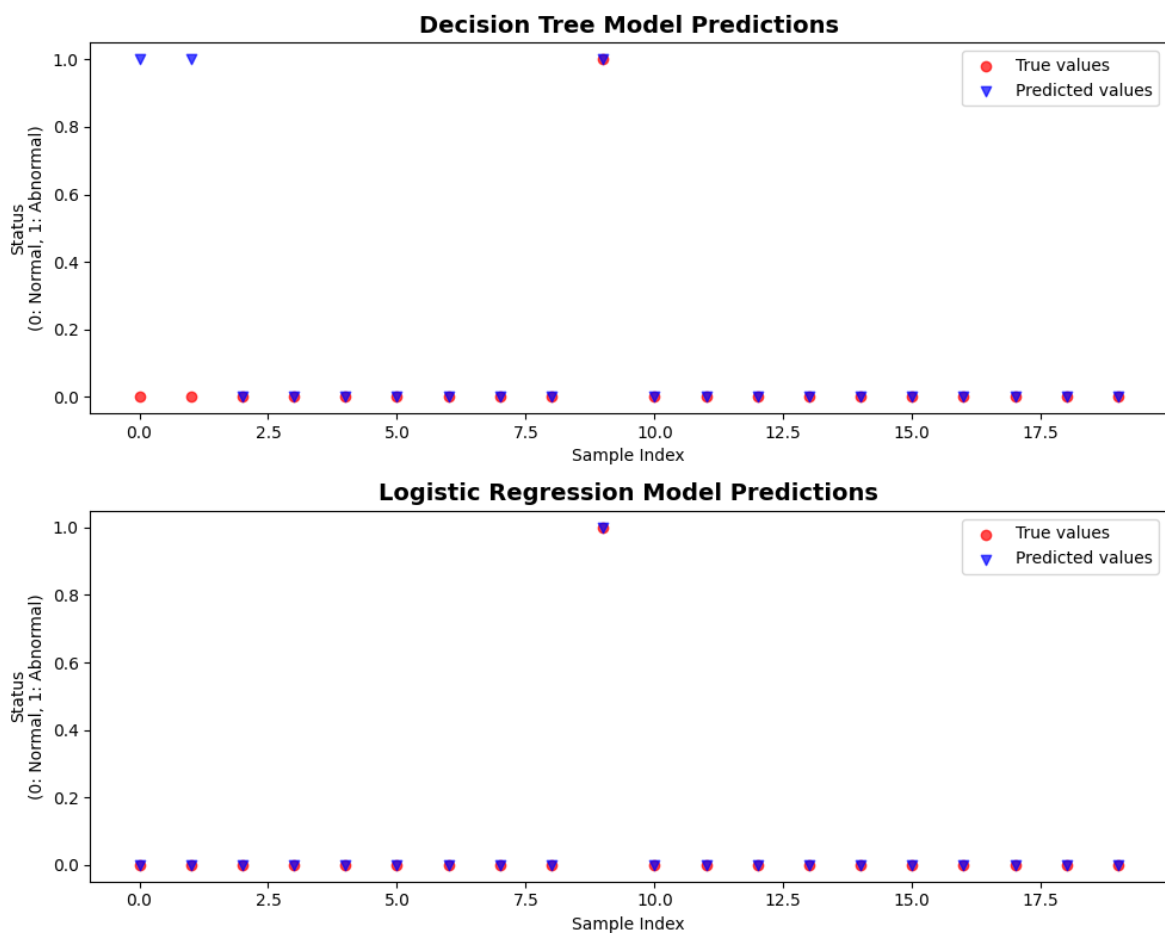
결정트리(Decision Tree) = K-Fold 교차검증 정확도: 0.919 ± 0.081

로지스틱 회귀(Logistic regression) = K-Fold 교차검증 정확도: 0.999 ± 0.001

= 랜덤 테스트 데이터 42 개를 가지고 5 회 테스트 진행했을 때 정확도가 결정트리의 경우 91% 로지스틱 회귀의 경우 99%가 나와 로지스틱 회귀의 신뢰성이 높다.

IV. 모델의 예측결과 오차 시각화 자료

그림 5 각 모델 별 실제 값과 예측 값을 시각화한 자료



- 실제 값과 예측 값 차이 시각화 설명

로지스틱 회귀의 경우 예측을 100% 성공한 반면 디시전 트리의 경우 20 번의 예측 중 2 개의 예측이 빗나간 모습이 나타났다.

V. Model.predict(각 종속변수에 따른 모델이 예측하게 된 확률)

결정트리(Decision Tree) =[1. 0.]

로지스틱 회귀(Logistic regression) =[9.952e-01 4.8000e-03]

= 결정트리의 경우 각각의 종속변수를 결정함에 확률이 100%지만, 로지스틱 회귀의 경우 각각의 종속변수를 결정함에 다른 변수의 고려도 하는 모습을 보인다.

이 수치를 활용하여 정상일 때 비정상을 고려하는 수치가 높게 나오면 위험한 상황이라는 것을 예측할 수 있다.

6. 결론

a. 머신러닝 모델 개발에 관한 간략한 요약 및 결과 설명

i. 모델 개발 요약

엘리베이터의 내.외부의 데이터를 수집하여 현재 엘리베이터의 상태를 정상, 비정상으로 나타내는 모델을 개발 모델의 예측 정확도는 99.99%로 현재 엘리베이터 상태에 대해서 정확한 판단을 내린다.

ii. 개발 모델의 장, 단점

개발 모델의 장점은 매우 높은 정확도를 가지고 있다는 점이다.

하지만 단점은 데이터의 불균형이 높고 신빙성이 낮아 모델이 높은 정확도를 가지고 있더라도 결과를 활용하기 어렵다는 점이다.

iii. 현재 개발 모델의 활용

현재 프로젝트로 진행한 모델도 바로 안전활동에 활용이 가능하다.

5-a-v 에서 나온 Model.predict 값 중에서 종속변수가 0 일 때 종속변수 1 을 고려한 확률 수치가 높을 때 안전점검을 권장하는 프로그램으로 만들어 적절한 순간에 점검을 실행할 수 있다.

또 데이터 중에 상관관계가 높은 데이터가 급하게 변화했을 때도 알림을 통해 위험을 예측할 수 있다.

b. 개발 의의 등

건설현장의 안전관리자는 업무가 굉장히 방대하고 양이 많다. 따라서 현장을 점검할 때 놓치는 부분이 생기기 마련이다. 이때 리프트의 안전점검을 조금 더 효율적으로 진행할 수 있게 만들어주는 시스템을 구축하려고 한다. 본 프로젝트의 모델에 리프트 안전 장치인 권과방지장치, 과부하방지장치, 완충장치, 방호울 등을 감지하는 센서를 추가한다면 더 안전활동에 도움이 되는 모델이 될 것이다.

c. 머신러닝 모델의 한계 및 앞으로 개발 가능성

i. 머신러닝 모델의 아쉬운 점

현재 진행한 두 모델 모두 데이터의 순서에 상관없이 모든 데이터를 분석해서 테스트로 넣은 독립변수의 상태에 따라 현재 종속변수를

찾아주는 모델이다. 따라서 미래를 예측하는 능력이 떨어져 이미 사고가 일어난 후에 엘리베이터의 고장을 결과로 말한다.

ii. 머신러닝 모델 개선 방법

데이터를 학습하는 모델 중 시계열 즉 시간의 순서에 따라 데이터를 분류하고 그 데이터를 일정 범위단위로 지정하여 미래를 예측할 수 있는 모델이 있다.

새로운 모델을 활용하면 종속변수가 0 일 때 1로 변화하는 과정 자체를 학습한다. 따라서 종속변수 0 일 때 가중치가 훨씬 높아지기 때문에 현재 앞에 **6-a-ii, 6-c-i** 에서 말한 개발 모델의 단점도 커버가 가능하다.