

발표를 시작하겠습니다.

P0(표지). 안녕하세요. 팀 '일단'입니다. 저희는 카테고리 파생변수를 활용한 분류모델 생성과 데이터 기반 작업자 안전 관리 서비스 'HD솔루션'에 대해 발표하겠습니다. 목차입니다.

P1. 먼저 알고리즘 코드 리뷰입니다. 저희는 EDA를 통해 데이터 특성을 파악하고, 전처리 수행 후 모델을 선택하여 성능을 확인했습니다. 시간관계상 핵심만을 짚어 발표하겠습니다.

우선 이번 데이터의 큰 특징은, train데이터와 test데이터의 컬럼이 다르다는 것이었습니다. 왼쪽 그림처럼 train데이터에는 test데이터에 없는 컬럼들이 포함되어 있어서 이 괴리를 어떻게 해결해야할지 고민했습니다. 컬럼 수 차이에 의미가 있다고 생각했고, 주어진 데이터를 최대한 활용하는 관점으로 접근했습니다. 저희는 train에만 존재하는 컬럼들을 train전용 컬럼, 그리고 나머지를 test컬럼으로 명명했습니다.

P2. 원소 컬럼의 분포를 확인했습니다. 왼쪽그림에서 보시다시피 대부분의 원소는 0값이 전체 비율의 대부분을 차지하고 있습니다. 정규분포를 이루는 원소는 거의 존재하지 않아서 저희는 원소 컬럼이 0값이 많다는 사실에 집중했습니다.

P3. 전처리 부분으로 넘어가겠습니다. 앞서 언급한 train과 test의 컬럼 괴리 때문에 train전용 컬럼들 중 의미가 없다고 판단하는 컬럼들은 과감하게 삭제했습니다. train전용 컬럼은 test컬럼과는 다르게 null 값이 매우 많았고 null이 있는 19개의 컬럼들을 제거했습니다.

P4. train전용 컬럼 중 0의 비율이 높은 컬럼들도 삭제했습니다. 0이 많다는 것이 의미가 있는지 없는지는 알 수 없습니다. 하지만 test에 없으면서 0의 비율이 높다는 점은 학습에 도움이 되지 않을 것이라 판단하기에 충분했습니다.

P5. 저희는 특히 원소 컬럼의 전처리에 집중했는데요. 우선 정상과 비정상의 분포에 차이가 없는 원소를 통계적 검정인 ks검정을 통해 확인하고 삭제했습니다.

P6. 다음은 카테고리 변수를 생성하는 과정입니다. 정상과 비정상 별 컬럼의 Outlier수를 확인하고 최소 outlier를 구했습니다. 그리고 비정상에서 outlier가 더 많은 컬럼들을 선별했습니다. 그 컬럼들마다 각각의 최소 outlier보다 클 경우 yes로 하여 파생변수를 생성했습니다.

P7. '특정 원소가 오일에 포함되어 있다 없다'로 해석을 하여 수치 값 대신 카테고리 변수로 변경했습니다.

P8. 앞서 전환한 컬럼들을 2개, 3개, 4개로 조합하여 정상과 비정상의 차이가 있는 경우 파생변수를 생성했습니다. 예를 들어 코발트와 크로뮴이 둘 다 no일 경우 'no', 하나라도 yes일 경우 yes로 값을 주었습니다. 특정 원소들이 동시에 조금이라도 존재할 경우 오일이 비정상일 확률이 높을 수 있다고 판단했습니다.

P9. 그렇게 최종적으로 1차 분류 모델에 사용될 train data1을 완성했습니다.

P10. 모델링 부분입니다. 저희는 주어진 데이터를 최대한 활용하고자 하는 목적으로 증류 기법을 사용했습니다. 이 과정에서 catboost 분류, 회귀 모델을 사용했는데요. 그 이유는 전처리한 데이터가 범주형 데이터가 많았기 때문입니다.

P11. 증류란 '미리 잘 학습된 큰 네트워크의 지식을 실제로 사용하고자 하는 작은 네트워크에게 전달하는 것'을 말하는데요. 연구용과는 다르게 현장 자원의 제한 속에서 배포용 모델로 사용되기 때문에 지식 증류 기법을 선택했습니다. 전체 학습 과정은 크게 세 단계로 이루어집니다.

첫번째는 이진분류 y_label 을 확률로 바꾸기 위한 분류 모델을 생성합니다.

두번째 바뀐 y_label 데이터를 예측하는 회귀 모델 생성합니다.

세번째 회귀 모델의 예측 값을 다시 이진분류 합니다.

P12. 첫 과정에서는 train data1을 사용합니다. 즉, test data에 없는 컬럼을 포함한 데이터로 학습을 하는 것이고, predict_proba 함수를 통해 두번째 모델에 사용할 y_label2 를 만들었습니다. 오른쪽은 y_label2 의 모습입니다.

P13. 그 후, train 전용 컬럼들을 모두 제거하고 train data2를 만듭니다.

catboost 회귀 모델에 train2를 입력하여 y_label2 를 예측하는 모델을 만듭니다.

P14. 그렇게 만들어진 두번째 모델에 test 데이터를 넣어서 나온 결과입니다. 출력 값이 분포되어 있음을 알 수 있는데요. 이진 분류를 하려면 임계치가 필요했기에, train data를 활용했습니다.

두번째 모델에 사용된 train data2의 출력 값을 임의의 임계치를 기준으로 0과 1로 변환시켰고, 처음 주어진 y_label1 과 비교했습니다. for문을 돌아 macro-f1-score가 가장 높은 값을 임계치로 정했으며, 이진 분류 결과를 도출했습니다.

이상으로 알고리즘 부문을 마치며 기획 부문을 시작하겠습니다.

P15.

중장비 관리와 작업자 안전을 생성형 ai로. 더 똑똑하게 !!

데이터 기반 작업자 안전 관리 서비스 hd솔루션을 소개합니다.

P16. '중대재해처벌법'이 2022년부터 시행되었으나 끊임없이 발생하는 안전사고는 현장과 기업 모두에 지속적인 위험 요소로 자리잡고 있습니다. 특히 건설업에서 53%의 비율로 사망자 수가 가장 많았습니다. 변수가 많은 건설 현장의 특성상, 계획 단계에서 미래의 위험 요소를 고려하지 않고 관성적으로 넘어가는 경우가 많습니다. 또한 다수의 협력 업체와 근로자가 같이 일하기 때문에 사고 예방을 위해 명확한 소통과 공유, 작업 상황에 대한 체계적이고 신속한 관리의 습관화가 필요합니다.

P17. 건설업 근로자와 관리자를 해당 서비스의 페르소나로 설정했습니다. 과거 중장비 기사들은 중장비의 위험 상태를 즉각적으로 인지하지 못 했는데요. 복잡한 건설현장의 환경, 장비에 대한 체계적인 안내와 교육을 통해 안전한 작업 환경을 제공하고자 합니다.

또한 현장 관리자들은 복잡한 건설현장을 끊임없이 모니터링하는데 어려움을 겪습니다. 일정 변

경을 공유하고 관리하는 기능을 통해 효율적인 관리 체계를 제공하고자 합니다.

P18. 해당 서비스에서 제공하는 기능입니다. 중장비 이상 감지를 통한 신속한 정비, 이상 감지 알람으로 위험 상황 즉각 대응, 안전 수칙 및 작업에 대한 맞춤형 교육, 그리고 캘린더를 통한 일정 관리입니다. 이를 통해 더욱 안전하고 효과적인 작업환경을 기대할 수 있습니다.

P19. 중장비 관리와 작업자의 안전한 근무환경을 생성형 ai가 혁신적으로 도와준다고 정리할 수 있습니다. 저희는 생성형 ai를 도입한 hd솔루션을 제안하려고 합니다.

P20.

첫번째 메인 기능은 센서데이터를 기반으로 한 중장비 실시간 관리 기능입니다. 중장비 부품의 세부 컴포넌트를 모니터링하고, 실시간 센서 데이터를 통해 위험 감지가 가능합니다.

부품 이상 알람 발생 시, 생성형 Ai기능을 도입한 HD 솔루션이 중장비 부품 고장 솔루션을 도출하여 즉각적인 조치를 가능하게 합니다.

그리고 알람 보내기를 통해 사고 발생 지역의 작업자들에게 실시간 알람을 보내 빠른 대피 또한 가능하게 합니다.

P21.

두번째 메인 기능은 중장비 사용 공유 캘린더입니다. 중장비 사용 일정이 표시된 캘린더를 통해, 관리자와 중장비 기사 모두 세부 일정을 파악해 공사 전반의 과정을 이해한 후 작업에 참여할 수 있게 하였습니다 .

또한 관리자가 중장비 기사 초대하기 기능을 통해 해당 일정의 중장비 기사에게 다시 한번 공사에 대한 인지를 시킬 수 있습니다.

P22.

초대된 중장비 기사에게 hd솔루션 속 캘린더를 통해 3중 안전체크를 제공합니다.

첫째, '중장비 맞춤 Ai안전지침'은 공사 일정에 기록된 중장비 종류에 맞게 생성형 Ai가 안전 예방 지침을 제공합니다.

p23.

둘째, 'Ai안전체크리스트'는 앞선 Ai 안전 지침을 다시 한번 요약합니다. 그리고 이를 중장비 기사가 직접 체크하는 인터랙션을 통해 이중 안전 지침을 제공합니다.

셋째, '시뮬레이션 사고 예방 퀴즈 ' 는 수집된 센서데이터를 기반으로 사용 중장비에 맞는 Ai 시뮬레이션 사고예방 퀴즈를 직접 풀며 최종적인 안전 지침 숙지를 도와줍니다.

P24. 다음과 같은 기능을 HD 솔루션에 도입하여 효율적인 작업과 안전한 근무 환경을 제공할 수 있습니다. HD 솔루션의 도입으로 사고율이 감소하고, 작업 효율성이 향상되어 생산성이 증가할 것입니다. 또한, 교육 비용이 절감되어 기업의 비용 효율성이 향상될 것이며, 작업자들의 만족도와 함께 기업 이미지도 높아질 것입니다. 서비스 도입으로 보다 안전하고, 효율적인 근무 환경이 되길 희망합니다. 감사합니다.