2023
제조 빅데이터 분석 경진대회

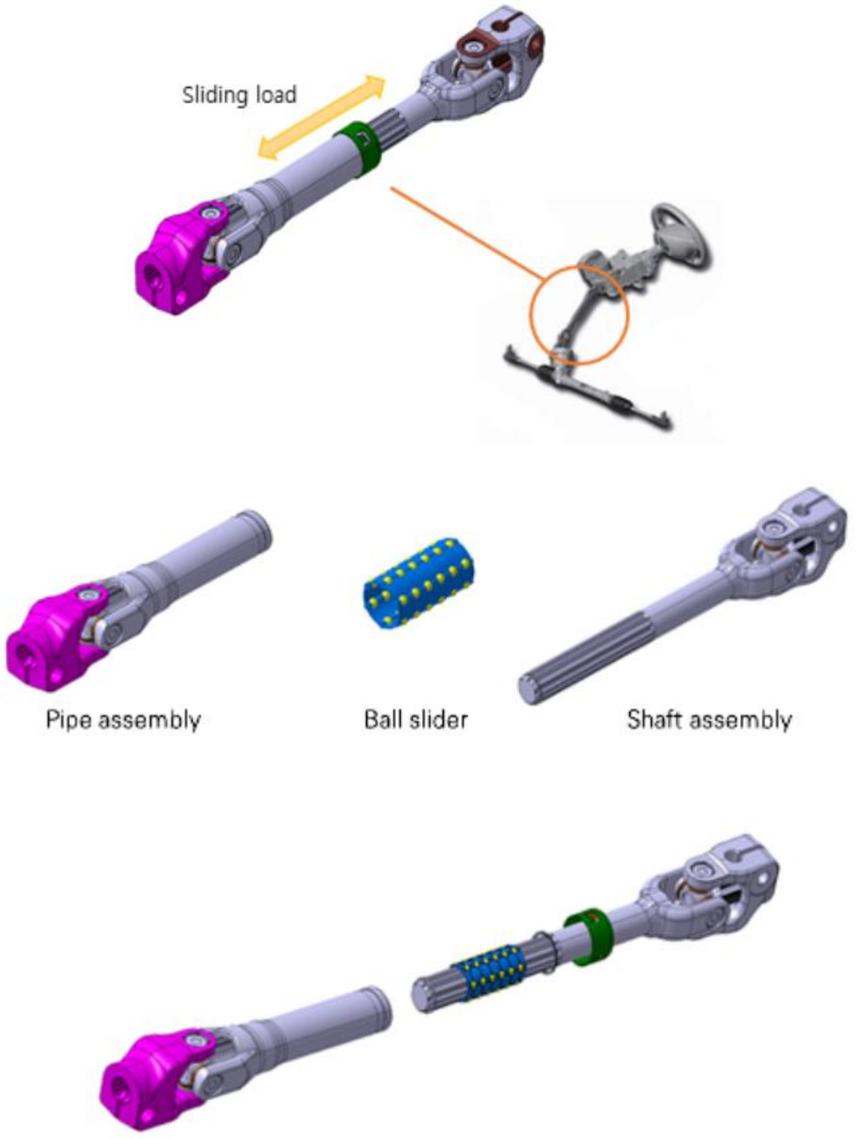
팀 일구공삼

목표 설정

핵심적인 변수를 선정, 해당 변수가 Ball size에
얼마나 영향을 주는지 파악

Table of Contents

목표 설정



" 주어진 조향 장치의 데이터를 활용하여 최적의 Ball size 찾기 "

TUBE

Shaft

Ball size

데이터를 보고 설정한 가설들

1) 영향력이 클 것 같은 범위 가정

2) 외경과 내경의 차이인 Gap의 영향력

3) 변수 간의 상관성 분석의 필요성

데이터 분석

전처리

"TD3", "SC1", "SC2", "SC3"

67%

Grade 매칭시도

주어진 7개

11개

상관계수 분석

Ball Size에 영향 주는 정도

0.07

파생변수 생성

기존 18개

58개

원인 변수의 분산이 결과 변수의 분산을 얼마나 잘 설명하는지?

모델 적용 및 변수 수정

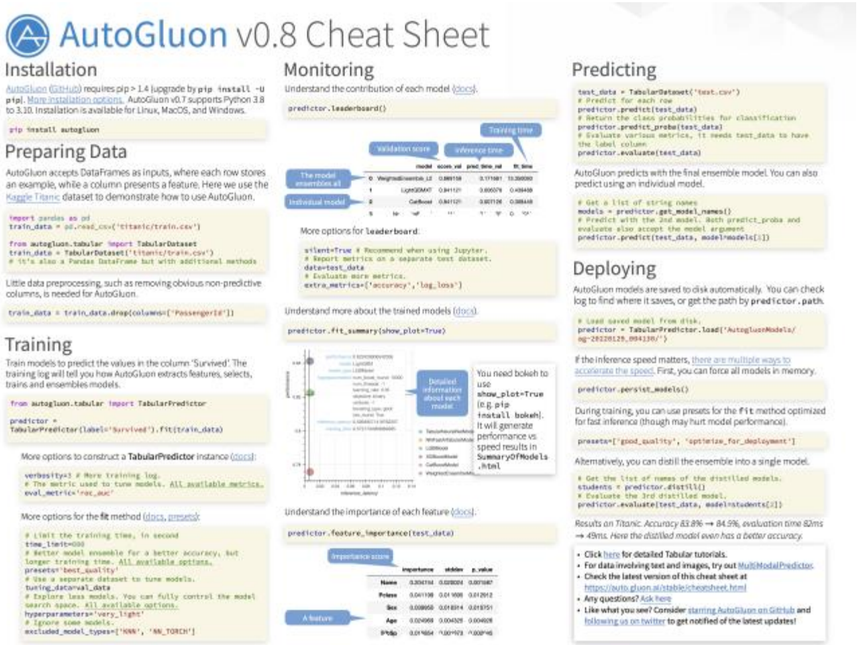
Auto Gluon

R-squared 3.3% 개선

AutoGluon-Tabular(Auto ML)

Framework	Wins	Losses	Failures	Champion	Avg. Rank	Avg. Rescaled Loss	Avg. Time (min)
AutoGluon	-	-	**1**	**23**	**1.8438**	**0.1385**	201
H2O AutoML	4	26	8	2	3.1250	0.2447	220
TPOT	6	27	5	5	3.3750	0.2034	235
GCP-Tables	5	20	14	4	3.7500	0.3336	**195**
auto-sklearn	6	27	6	3	3.8125	0.3197	240
Auto-WEKA	4	28	6	1	5.0938	0.8001	244

Framework	Wins	Losses	Failures	Champion	Avg. Rank	Avg. Percentile	Avg. Time (min)
AutoGluon	-	-	**0**	**7**	**1.7143**	**0.7041**	**202**
GCP-Tables	3	7	1	3	2.2857	0.6281	222
H2O AutoML	1	7	3	0	3.4286	0.5129	227
TPOT	1	9	1	0	3.7143	0.4711	380
auto-sklearn	3	8	**0**	1	3.8571	0.4819	240
Auto-WEKA	0	10	1	0	6.0000	0.2056	221



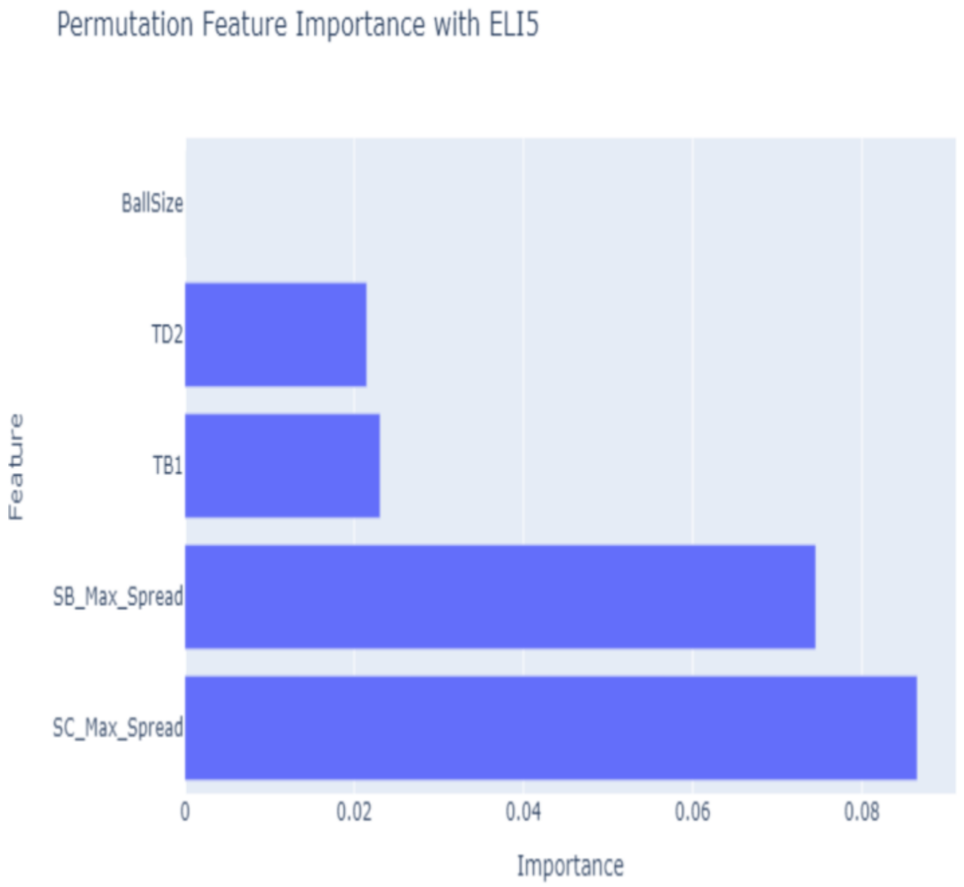
딥러닝 모델을 제작할 뿐만 아니라 최적의 하이퍼파라미터 세트로 튜닝까지 해주는 AutoML라이브러리

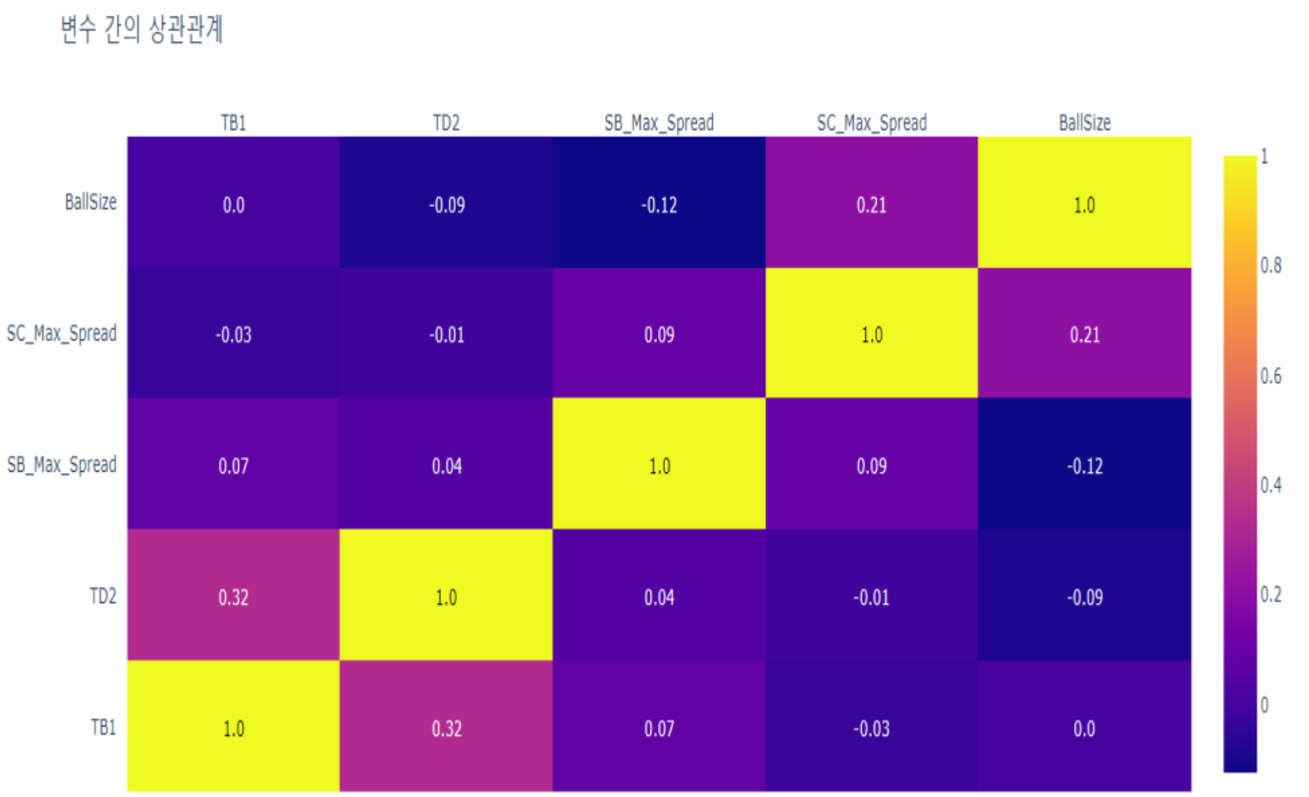
쉽고 빠르게 모델을 생성을 도와주는 기능으로 모델 개발 프로세스의 효율성을 향상시켜줌

기본 변수 기준 중요도 파악

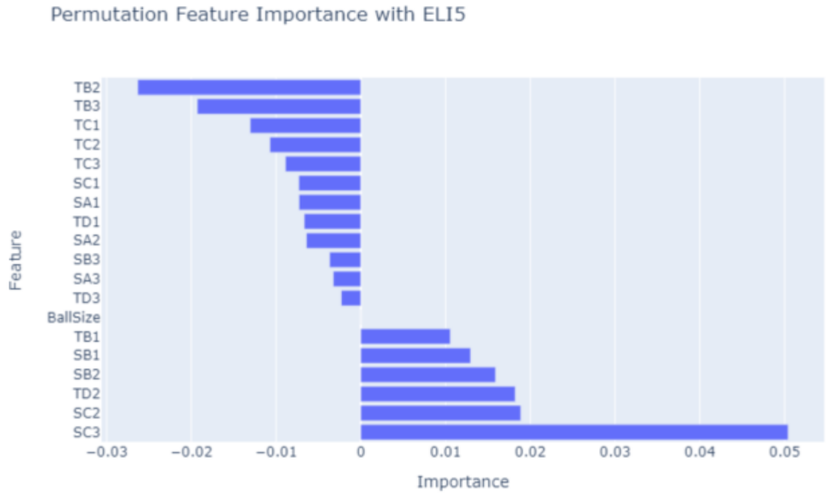
변수 간의 상관관계

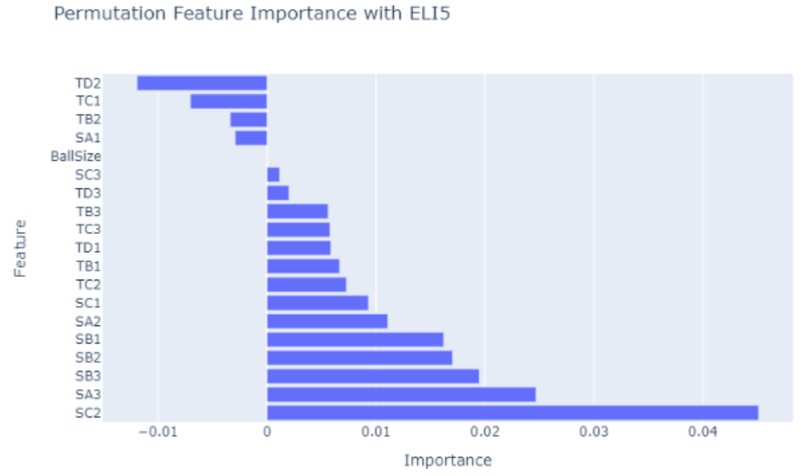
기본 변수 기준 중요도 파악





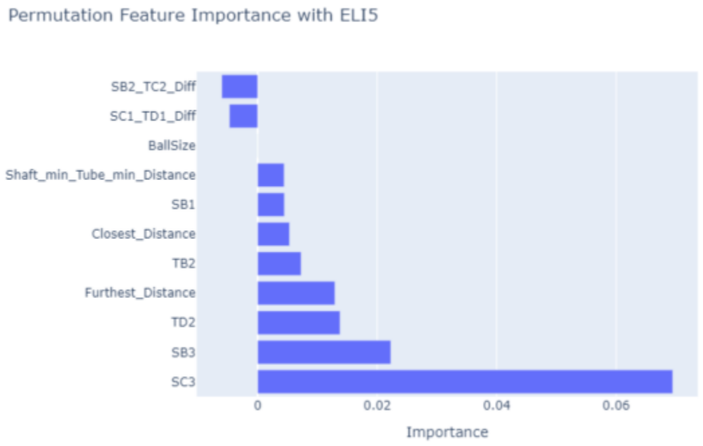
변수 중요도 파악

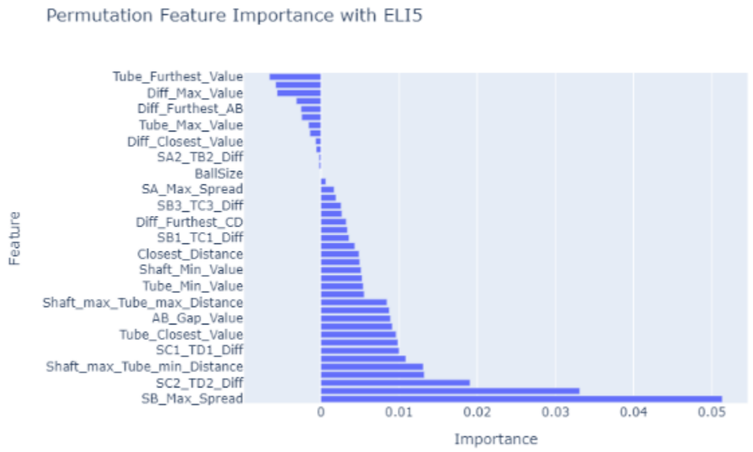


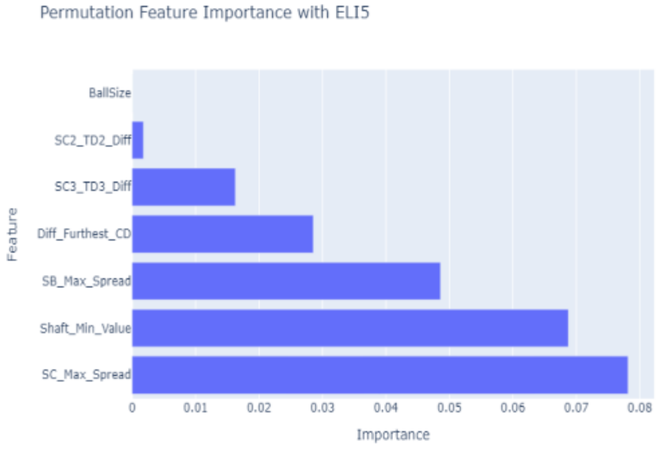


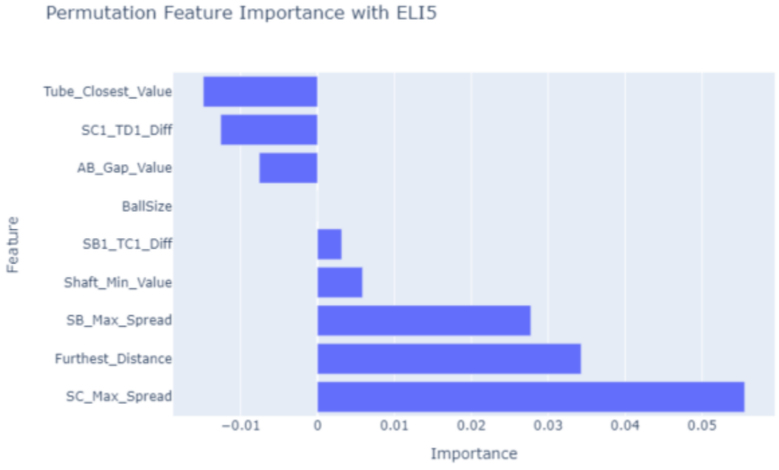
'TB1', 'TB2', 'TB3', 'TC1', 'TC2', 'TC3', 'TD1', 'TD2', 'TD3', 'SA1',
'SA2', 'SA3', 'SB1', 'SB2', 'SB3', 'SC1', 'SC2', 'SC3',
'SA1_TB1_Diff', 'SA2_TB2_Diff', 'SA3_TB3_Diff', 'SB1_TC1_Diff',
'SB2_TC2_Diff', 'SB3_TC3_Diff', 'SC1_TD1_Diff', 'SC2_TD2_Diff',
'SC3_TD3_Diff', 'Diff_Furthest_AB', 'Diff_Furthest_BC',
'Diff_Furthest_CD', 'AB_Gap_Value', 'BC_Gap_Value', 'CD_Gap_Value',
'inner_Average_Selected', 'outer_Average_Selected',
'Tube_Furthest_Value', 'Shaft_Furthest_Value', 'Diff_Furthest_Value',
'Tube_Closest_Value', 'Shaft_Closest_Value', 'Diff_Closest_Value',
'Shaft_Max_Value', 'Shaft_Min_Value', 'Tube_Max_Value',
'Tube_Min_Value', 'Diff_Max_Value', 'Diff_Min_Value',
'Furthest_Distance', 'Closest_Distance', 'Both_big_Distance',
'Both_small_Distance', 'Shaft_max_Tube_min_Distance',
'Shaft_max_Tube_max_Distance', 'Shaft_min_Tube_max_Distance',
'Shaft_min_Tube_min_Distance', 'SA_Max_Spread', 'SB_Max_Spread',
'SC_Max_Spread' ,'BallSize'

변수 중요도 파악(R-squared)





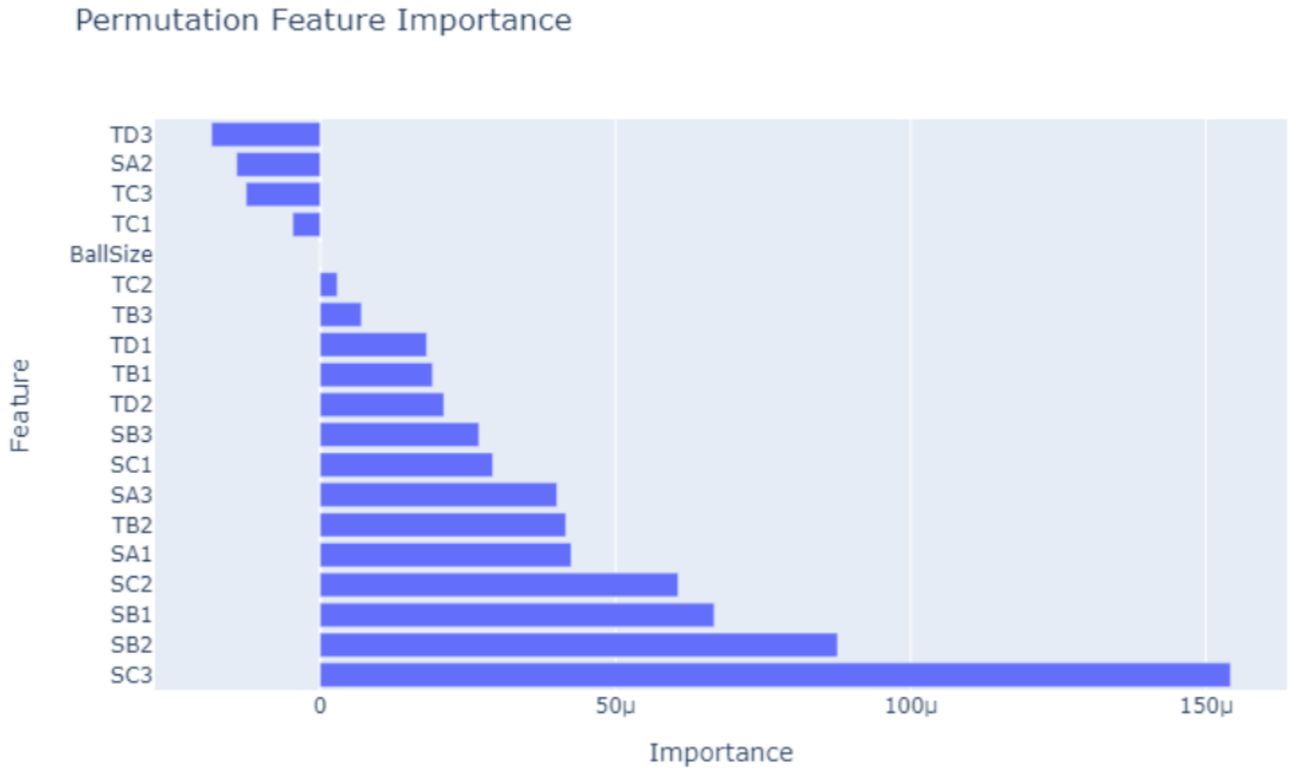




주요변수	변수 중요도 값
SC_Max_Spread	0.066031
Shaft_min_Tube_max_Distance	0.013028
Diff_Furthest_BC	0.008383
SC2_TD2_Diff	0.003732
Shaft_Min_Value	0.002039
SB_Max_Spread	0.001343

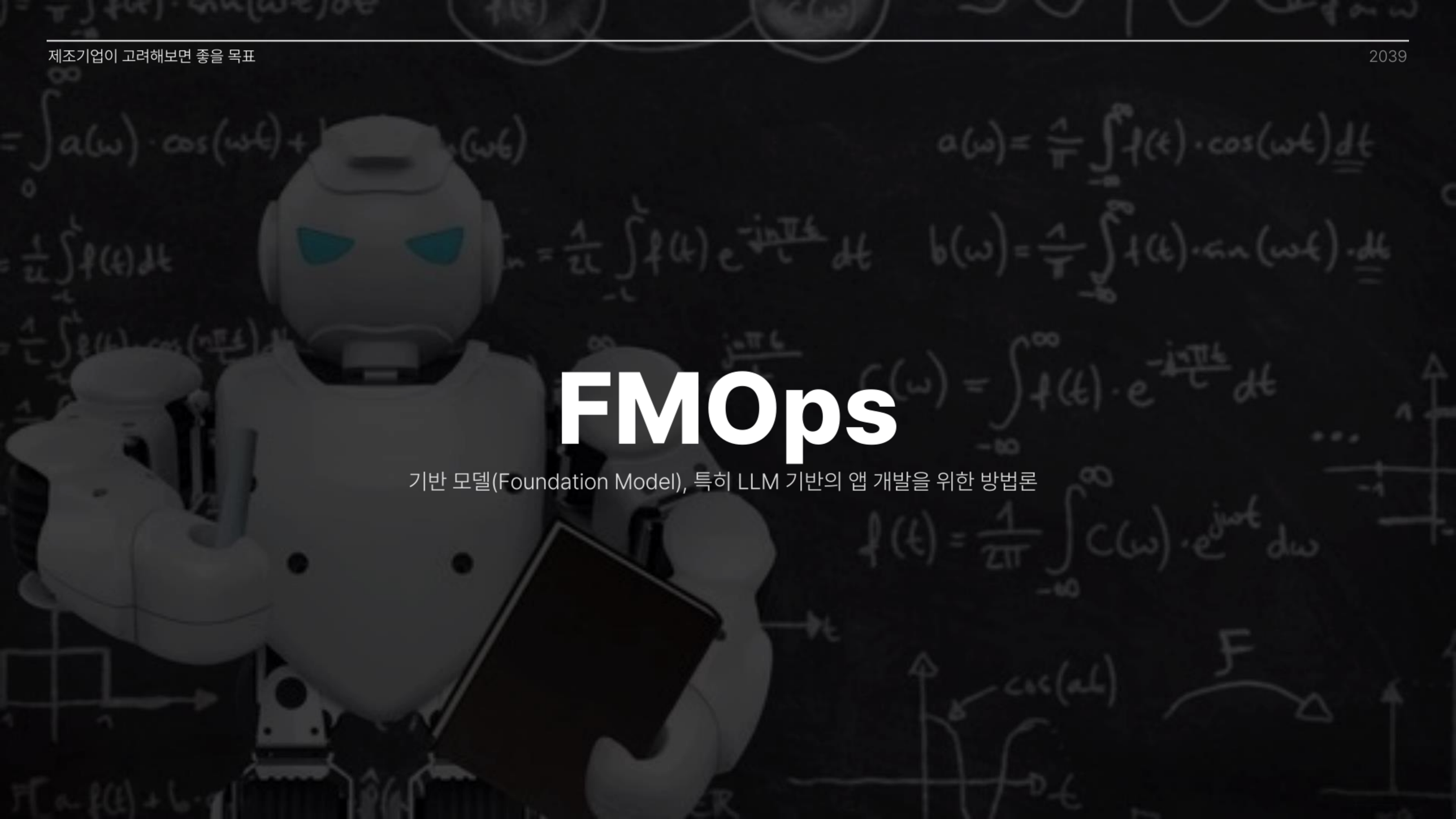
변수 중요도 파악(RMSE)

주요변수
SC_Max_Spread
SB_Max_Spread
Shaft_min_Tube_max_Distance
CD_Gap_Value
Shaft_Min_Value
SA1_TB1_Diff
Shaft_Closest_Value



AutoML의 AutoGluon 모델을 통한 결과값 도출

SC위치에서 값들이 Ball Size를 결정하는데 있어서 중요한 요소로 작용



FMOps

기반 모델(Foundation Model), 특히 LLM 기반의 앱 개발을 위한 방법론

기업 데이터를 활용한 AI 인프라

Task	MLOps	FMOPS
목적	모델을 새로 개발하는 것	이미 학습이 완료된 기반 모델과 자신의 데이터를 활용항 커스텀한 AI앱을 만드는 것
주로 다루는 모델의 유형	예측성 모델분류, 랭킹, 회귀	생성형 모델데이터셋을 기반으로 새로운 콘텐츠를 생성
모델 접근 방식	기업이 모델과 모든 파이프라인을 직접 개발인하우스에서 학습시키기 위함	모델을 API를 통해 접근
모델 성능 향상	Fine tuniong적정양의 정형 및 비정형 데이터로 구성된 기반 모델의 파라미터를 변경하여 반복적으로 재학습	프롬프트 엔지니어링프롬프트 체이닝데이터 임베팅과 파인튜닝
최종적으로 배포하는 것	LLM을 API로 배포	LLM과 사용자의 커스텀 데이터 기반의 AI앱을 배포이때, 앱의 형태는 대개 대화형 챗봇, 어시스턴트이다.

기업 데이터를 활용한 AI 인프라

인이지, "제조업 혁신을 위해서도 LLM 필요"

"세계 제조업의 혁신 공장에서도 AI를 도입하고 있습니다. 에너지 효율적인 사용이나 인구구조 변화에 대응하기 위해서 제조업의 AI 활용은 필수가 될 겁니다"

최재식 인이지 대표는 'AI를 통한 제조의 변화, 생성형 AI·대화형언어모델의 제조업 적용 방안'를 주제로 발표했다. 인이지는 국내 대표적인 산업 현장 AI 전문 스타트업 중 하나다. 최재식 KAIST 교수가 2019냔 창업한 인이지는 제조업체 대상 공정 최적화 AI를 만든다. 포스코의 스마트고로에 AI 솔루션을 적용하며 이름을 알렸다. 용광로 쇳물 온도의 예측 오차를 줄여 연간 647억원 상당의 연료비를 줄였다.

한경 **한국 AI 스타트업, '초거대 언어 모델'로 무엇까지 할 수 있을까 [긱스]**

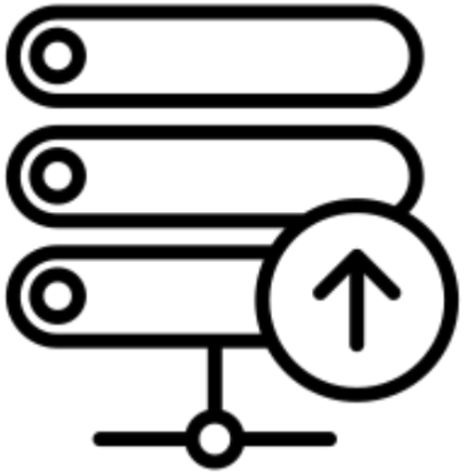
최 대표는 국내 제조업에서 AI 도입이 더 필요하다고 주장했다. 그는 "최근 한전의 적자 규모가 급격히 커지고 에너지 비용 상승으로 제조업의 원가 부담도 늘었다"라며 "AI로 불필요한 에너지 소비를 줄이고 생산성도 높일 수 있다"고 말했다. 저출산고령화에 따른 젊은 인력 감소에 대응하는 방법도 AI 도입이다.

최 대표는 제조업에서 LLM 도입도 늘어날 것으로 전망했다. 그는 "산업 현장에서 중요한 자료 중 하나가 생산 작업 일지인데 이 데이터를 디지털로 전환하고 생성 AI가 학습하면 새로운 기능이 나올 수 있다"고 말했다. 최 대표는 "비가 오면 공정에 많은 영향을 주는데 시멘트 공장에 습도가 올라가면 결과물이 눅눅해질 수 있다"며 "관련 생산일지를 학습해 AI에 '지금 습도가 몇 % 인데 온도를 얼마나 올려야 하나"라고 물어보면 '언제 생산일지를 참고해서 온도를 몇 도 정도 올려라'라는 답을 받을 수도 있다"고 덧붙였다.

공장 설비 메뉴얼을 파악하는 데도 LLM 활용이 가능하다. 최 대표는 "항공기 엔진 같은 큰 설비는 메뉴얼이 상당히 두껍다"라며 "이걸 생성 AI에 학습시켜 설비에 문제가 생길 때 어떻게 대응해야 하는지 신속하게 파악할 수 있을 것"이라고 말했다. 인이지도 관련해 일명 '시계열 예측을 위한 파운데이션 모델'을 개발하고 있다. 제품 수요, 연료 수급 등에 대해 예측하는 AI 솔루션이다.

김주완 기자 kjwan@hankyung.com

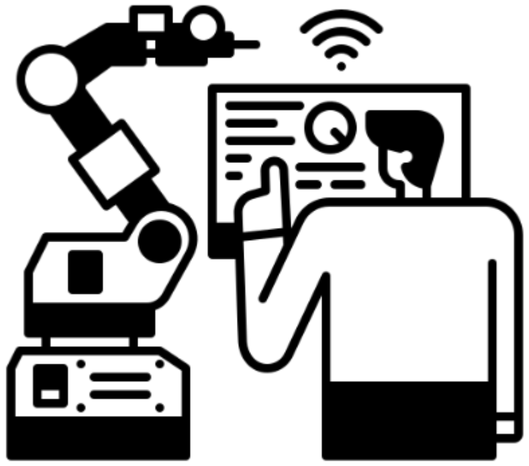
본 대회를 적용할 경우 고려해야 할 점 3가지



1. 데이터의 개수

LLM을 이용하여 증강하여 데이터 Imbalance 문제
> 5번 Grade의 데이터 1개

많은 양의 데이터로 정밀한 치수를 다룰 수 있도록 해결
-> 0.002mm 차이



2. 데이터의 종류

다양한 변수 조합을 활용(파생변수)하여 적절한 변수가 무엇일지 파악

공정상에서 핵심적으로 목표 값에 영향을 주는 추가적인 요인 찾기
-> 압력 데이터 측정

3. 데이터의 품질

치수별로 생성된 데이터의 표준편차값을 줄이기
-> Tube가 Shaft보다 3.07배 크다.
-> 광학식 3D스캐너 측정 방식도 고려

자체 LLM 서비스에 입력하는 예시

Tube의 직경이 16.4cm 이고 Shaft의 직경이 24.5cm 인데 적절한 Ball size 크기를 추천해줘.

조건값을 통해 예상되는 Ball Size 값은 4.82 입니다.

다음은 오답 데이터 10개이고 다음은 정답 데이터 100개야. 이를 활용해서 오답 데이터 100개를 만들어줘. 대신 타겟 변수에 대한 imbalance문제가 발생하지 않도록 만들어줘.

구간에 대해 모두 같은 개수 만큼 나누어지도록 데이터를 생성하고 inference를 하였습니다.

다음은 Ball size를 예측하기 위한 모든 공정에 관한 데이터 1000개야. 이를 통해서 Ball size를 예측한 값을 txt파일로 저장해줘. 예측이 어려운 것에 대해서는 적절한 구간값을 알려줘.

1000개에 대해 생성한 값을 txt 파일로 저장했습니다. 예측이 어려운 것은 0.002mm의 Range로 지정했습니다.

기대 효과

현상에 대한 해석 및 대처 방안을 데이터를 통해 구체적으로 제시하여

공정을 관리하는 사용자들의 의사결정을 더욱 빠르고 효과적으로 할 수 있을 것입니다.


감사합니다.