

한국타이어와 함께하는
제 5 회 KAIST-POSTECH-UNIST
데이터 사이언스 경진대회

공동주최



POSTECH

UNIST

주관

KAIST

후원사

HANKOOK
driving emotion



Lunit® THYROSCOPE

PuzzleData

Fast campus

제 5 회 데이터 사이언스 경진대회 개요

{ “대회 주최 및 주관” }

- ✓ KAIST, POSTECH, UNIST 공동 주최
- ✓ 이번 대회는 KAIST 주관 (세 학교가 돌아가며 주관)

{ “지난 대회들” }

[제 1 회] UNIST 산업공학과 주관 <ul style="list-style-type: none">✓ 한국석유공사 메인 후원✓ 세일가스 생산량 예측 및 생산정 설치 의사결정 문제	[제 3 회] POSTECH 산업경영공학과 주관 <ul style="list-style-type: none">✓ LG전자 메인 후원✓ 부품 수요 예측 및 주문 의사결정 문제
[제 2 회] KAIST 산업및시스템공학과 주관 <ul style="list-style-type: none">✓ 하나은행 메인 후원✓ 하나은행 앱 접속 시기 예측 및 마케팅 의사결정 문제	[제 4 회] UNIST 산업공학과 주관 <ul style="list-style-type: none">✓ KB증권 메인 후원✓ 기업 보고서를 고려한 재무 질의 답변 및 서비스 기획 문제

제 5 회 데이터 사이언스 경진대회 개요

{ “대회 시상 및 상금” }



대상 (1 팀) 300만원



금상 (1 팀) 200만원



은상 (1 팀) 100만원



동상 (2 팀) 50만원

제 5 회 데이터 사이언스 경진대회 개요

{ “대회 일정” }

① 문제 설명회

11월 10일 월요일
19시 00분

② 참가 신청

11월 11일 화요일
00시 00분
~ 11월 16일 일요일
23시 59분

③ 예선 시작

11월 11일 화요일
00시 00분

④ 예선 제출 마감

12월 12일 금요일
23시 59분

제 5 회 데이터 사이언스 경진대회 개요

{ “대회 일정” }

⑤ 예선 결과 발표

12월 19일 금요일
13시 00분

⑥ 본선 발표 평가

2026년 1월 5일 월요일

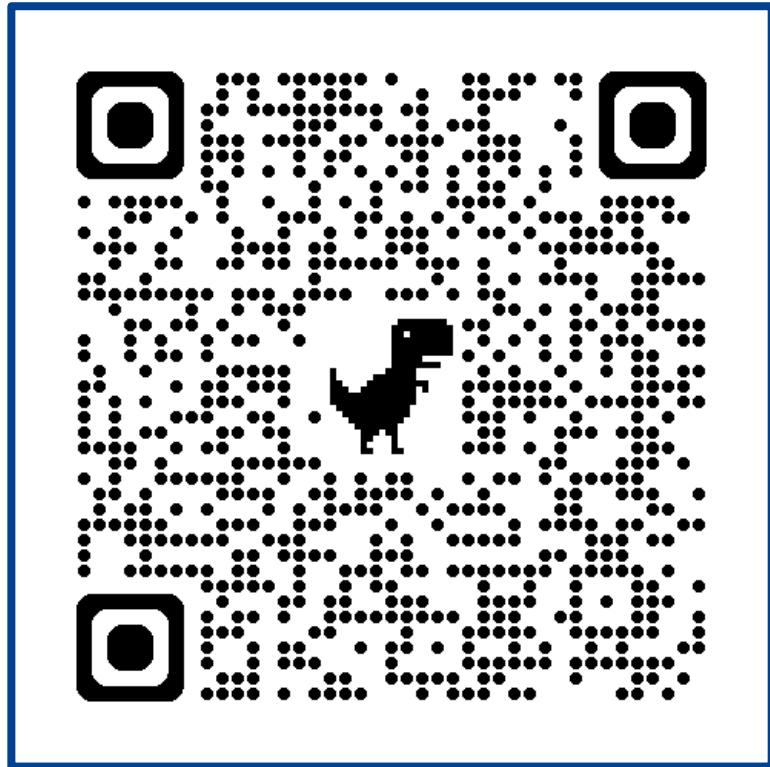
제 5 회 데이터 사이언스 경진대회 문제 설명

{ “참가자 자격” }

- ✓ 참가자 자격: 대한민국 학사과정 학생
- ✓ 팀 구성 조건
 - ✓ 산업공학 전공(복수 전공 포함) 학생을 팀 리더로 하여 2인 이상 4인 이하로 구성
 - ✓ 타 전공 학생 포함 가능, 서로 다른 학교의 학생들로 팀 구성 가능
- ✓ 기타
 - ✓ 팀 구성과 관련한 특이 사항은 주최측에 문의 요망(카카오톡 오픈채팅방, hyangsuk.min@kaist.ac.kr)

제 5 회 데이터 사이언스 경진대회

{ “참가 신청” }



제 5회 KAIST - POSTECH - UNIST AI & 데이터사이언스 경진대회

제 5회 AI / 데이터사이언스 경진대회

안녕하세요.

2025년도 제5회 KAIST-POSTECH-UNIST AI-데이터사이언스 경진대회에 관심을 가져주셔서 감사합니다.
신청 전, 홈페이지 공지사항의 신청요강을 반드시 확인해 주시기 바랍니다.
신청서를 모두 작성하신 후에는 ‘제출’ 버튼을 꼭 눌러주시기 바랍니다.

“해당 경진대회 신청서는 팀 1개로 팀장만 제출하면 됩니다.”

문의사항이 있으신 경우 hyangsuk.min@kaist.ac.kr 로 연락해 주시면 안내드리겠습니다.

• 일정: 2025년 11월 11일 (화) ~ 12월 12일 (금) (본선 진출 팀 발표평가 26년 01월 05일)
• 주제: 유한요소법(FEM) 기반 시뮬레이션 결과를 활용한 **데이터 불량률 예측 및 시험 생산 의사결정**을 수행하라!
• 참가 대상: 산업공학 전공(복수 전공 포함) 학생을 팀 리더로 하여 2인 이상 4인 이하로 구성
* 타 전공 학생 포함 가능, 서로 다른 학교의 학생들로 팀 구성 가능

팀 이름 / 휴대전화 / 캐글 참여 이메일 (위에 기입한 이메일과 달라도 무방합니다.)
팀장 이름, 연락처, Kaggle 참여용 이메일을 아래 형식에 따라 작성해 주세요.

* 본 경진대회는 Kaggle에서 진행되며, 아래에 기입한 이메일로만 대회 페이지 접속 및 리더보드 제출이 가능합니다. 입력 후에는 변경이 불가하니 신중히 작성해 주세요.

Ex)
홍길동 / 010-1234-5678 / gildong.hong@gmail.com

Team Leader's Information (Name / Mobile Phone / Kaggle Email Address)
This may differ from the email provided above.

Please provide the team leader's name, mobile number, and the email address to be used for Kaggle participation, following the format below.

* The competition will be held on Kaggle, and only the email submitted here will be authorized for accessing the competition page and submitting to the leaderboard.

No changes can be made after submission, so please enter it carefully.

Example:
1) Dong Hong / KAIST Department
gildong.hong@kaist.ac.kr
2) Soo Kim / POSTECH Department
kim@postech.ac.kr
3) Yu Han / UNIST Department o
* Each team must consist of 2 to 4 members.

내 답변

<https://forms.gle/N9CG4N8Yrx183p9E7>

제 5 회 데이터 사이언스 경진대회

{ “교육 자료” }

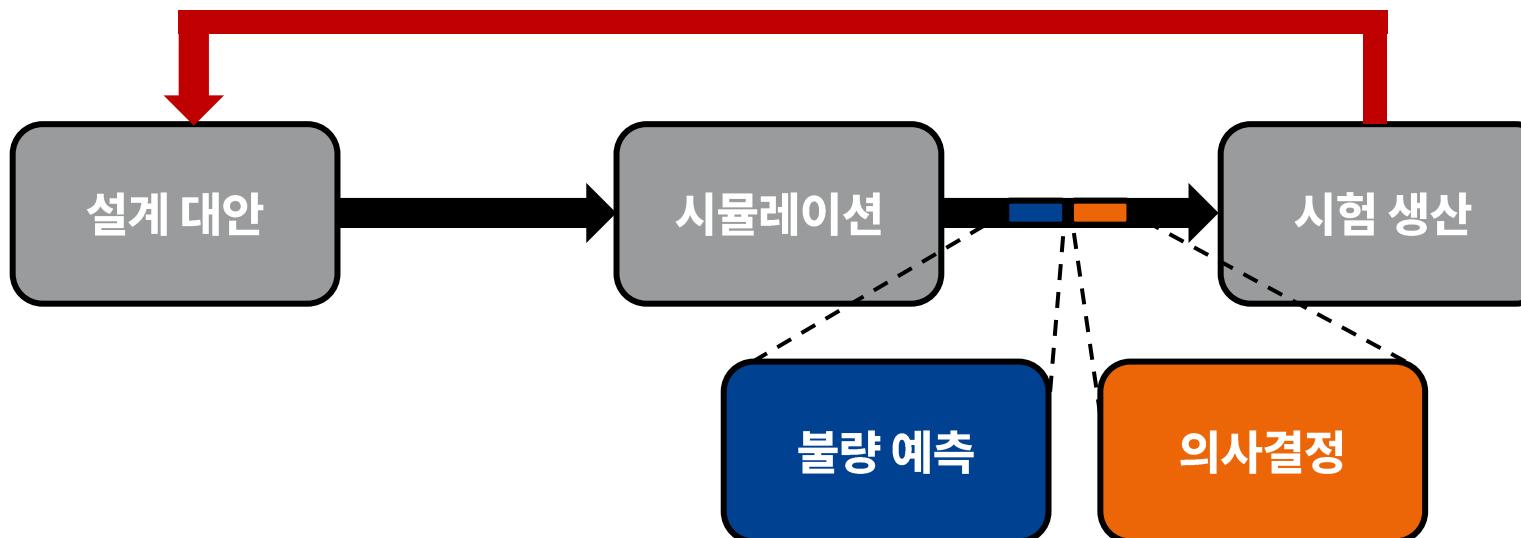


- ✓ 참가 학생들에게 무료로 제공
- ✓ 제공 방법: 참가 신청 후에 각 팀별 안내 예정

제 5 회 데이터 사이언스 경진대회 문제 설명

{ “경진대회 주제” }

유한요소법(FEM) 기반 시뮬레이션 결과를 활용한
타이어 **불량 예측** 및 **시험 생산 의사결정**

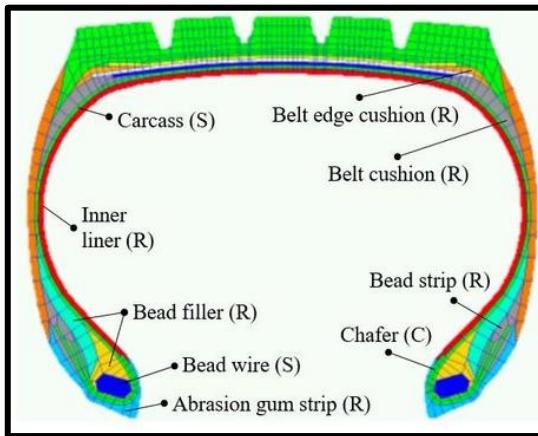


제 5 회 데이터 사이언스 경진대회 문제 설명

{ “과업 1: 불량 예측” }



+



설계 스펙

단면폭, 편평비, 인치,
설계 시점, 생산 공장

시뮬레이션 결과

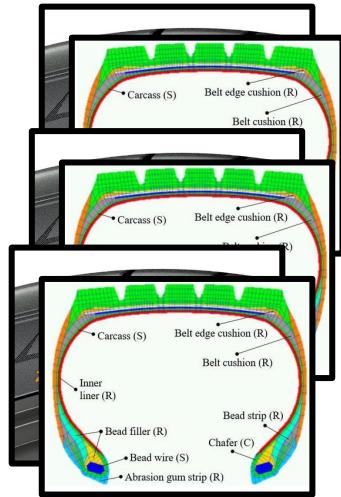
2차원 위치 및 압력 값



시험 생산 불량 추론

제 5 회 데이터 사이언스 경진대회 문제 설명

{ “과업 2: 시험 생산 의사결정” }



시험 생산 후보군 묶음

시험 생산에 대한
추론된 불량 점수들

시험 생산 비용

시험 생산 의사결정

제 5 회 데이터 사이언스 경진대회 문제 설명

{ “경진대회 데이터 – 학습 데이터” }

“train.csv” (720 행, 799 열)

- ✓ Mass_Pilot: 설계 시점 (True: 양산 설계, False: 예비 설계)
- ✓ Width, Aspect, Inch: 단면폭, 편평비, 인치
- ✓ Plant: 시험 생산 공장
- ✓ Proc_Param1, ..., Proc_Param11: 공정 파라미터
- ✓ X1, ..., X5, Y1, ..., Y5: 타이어 위치 (X_i, Y_i) 형태의 벡터
- ✓ x0, ..., x255, y0, ..., y255, p0, ..., p255: 시뮬레이션 결과 (x_i, y_i, p_i) 형태의 위치별 압력 값
- ✓ G1, ..., G4: 시뮬레이션 결과 중 변화량 통계
- ✓ Class: 불량 여부 (NG: 불량, Good: 양품) (NG 개수 : Good 개수 = 107 개 : 613 개)

제 5 회 데이터 사이언스 경진대회 문제 설명

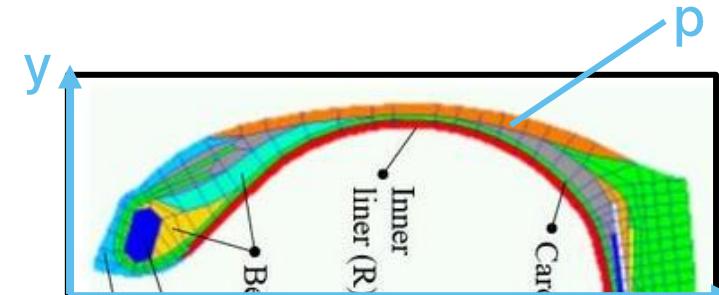
{ “경진대회 데이터 – 학습 데이터” }

	A	B	C	IU	IV
1	x0	x1	x2	x254	x255
2	239.6308	239.8321	240.0334	354.889	355.2151
3	239.2115	239.4113	239.6111	349.836	350.108
4	239.2586	239.4792	239.6998	365.7623	366.1219
5	214.2677	214.4957	214.7238	346.7823	347.1409

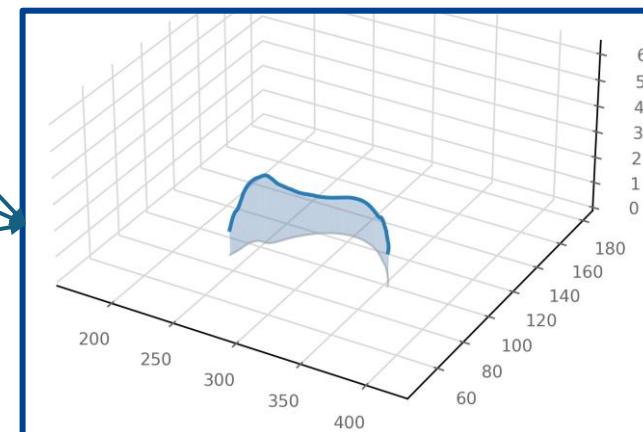
	A	B	C	IU	IV
1	y0	y1	y2	y254	y255
2	87.05644	87.55857	88.06071	97.95088	97.51923
3	92.5168	93.01635	93.5159	100.9907	100.5265
4	92.63725	93.1854	93.73355	105.6461	105.1772
5	87.14947	87.717	88.28452	96.08815	95.59266

	A	B	C	IU	IV
1	p0	p1	p2	p254	p255
2	0.988158	1.110555	1.232952	1.320051	1.318755
3	1.10726	1.241411	1.375563	1.109317	1.113692
4	0.874673	0.983378	1.092084	1.403577	1.402061
5	0.994704	1.114723	1.234743	1.254474	1.25601

시뮬레이션 결과 (x_i, y_i, p_i) 데이터



유한요소법(FEM) 시뮬레이션



제 5 회 데이터 사이언스 경진대회 문제 설명

{ “경진대회 데이터 – 테스트 데이터” }

“test.csv” (466 행, 799 열)

- ✓ ID: 각 행별 고유 식별자
- ✓ ID와 Class를 제외하고 train.csv와 동일하게 열 존재

제 5 회 데이터 사이언스 경진대회 문제 설명

{ “경진대회 규칙 – 제출” }

- ✓ 제출 횟수 제한: 팀당 최대 3회/일 [09시00분 기준]
- ✓ 최종 제출 파일 선택: 팀당 최대 1개 [팀장 선택]
- ✓ 제출 파일 형식: csv 파일 ['ID'(test.csv의 'ID' 변형), 'probability'(0-1 실수형), 'decision'(True/False)]
- ✓ 제약 조건: 'decision'=True 개수는 최대 200개 [200개 초과 시, 과업 2에서 -99,999원 패널티 부여]

주의사항: ID_{N}_L과 ID_{N}_P에는 동일하게 ID_{N}에 대응하는 예측 결과를 제출해야 함

ID	probability	decision	True \leq 200	True \leq 200
ID_0_L	0.000124	True		
ID_1_L	0.991203	False		
ID_2_L	0.121114	False		
:	:	:	=	
ID_465_L	0.015624	True		
submissions.csv				
ID_0_P	0.000124	True		
ID_1_P	0.991203	False		
ID_2_P	0.121114	False		
:	:	:		
ID_465_P	0.015624	True		

제 5 회 데이터 사이언스 경진대회 문제 설명

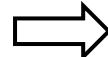
{ “경진대회 규칙 – 평가” }

- ✓ 리더보드 운영: ID_{N}_L 중 50%샘플로 점수 산정 [팀별 제출물 중 가장 높은 점수 기준]
- ✓ 최종 점수: ID_{N}_P 100%로 점수 산정 [선택된 최종 제출 파일 기준]
- ✓ 평가 항목: ROC-AUC Score [과업 1, probability] & Total Net Profit [과업 2, decision]

$$\text{Total Net Profit} = 100 \sum_i I_{\{d_i=True\}} I_{\{y_i=Good\}} - 2,000 \sum_i I_{\{d_i=True\}} I_{\{y_i=NG\}}$$

- ✓ 점수 산정 방식: 과업별 정규화된 점수의 기하평균 [총합 점수 = $\sqrt{\text{과업 1 점수} \times \text{과업 2 점수}}$]

과업	정규화 점수
과업 1	$\frac{\max\{(\text{ROC-AUC Score}) - 0.5, 0\}}{0.5}$
과업 2	$\frac{\max\{(\text{Total Net Profit}), 0\}}{20,000}$



예시) ROC-AUC Score = 0.755 & Total Net Profit: 9,000

$$\rightarrow \sqrt{\frac{\max\{0.755-0.5,0\}}{0.5} \times \frac{\max\{9000,0\}}{20,000}} = 0.479$$

제 5 회 데이터 사이언스 경진대회 문제 설명

{ “경진대회 규칙 – 사용 자원” }

- ✓ 예측 시점 가정: 466개 배치에 대한 설계 스펙 및 시뮬레이션 결과가 모두 주어진 상태
- ✓ 데이터 사용 범위: 모델 학습 과정에 테스트 데이터 사용 허용
- ✓ 외부 리소스 사용: 모든 관련 자원 허용 [외부 데이터셋, 사전학습 모델, 외부 API 및 라이브러리]

제 5 회 데이터 사이언스 경진대회

{ “본선 관련” }

- ✓ 본선 진출 팀 수: 약 10개 팀 선발 예정 [최종 점수 순위 기준]
- ✓ 본선 발표 내용: 분석 “전 과정”에 대한 발표 진행
 - ✓ 데이터 이해 및 분석 전략
 - ✓ 전처리 및 피처 엔지니어링
 - ✓ 모델링 및 학습 방법
 - ✓ 추론 전략 및 자체 검증 방법
- ✓ 평가 방식: 발표 내용에 대한 정성 평가 중심으로 진행 [본선 진출 팀에게 별도 안내 예정]

제 5 회 데이터 사이언스 경진대회

{ “문의 사항” }



<https://open.kakao.com/o/gL7EFLZh>