

# AI EDUTOM. ● Service Manual

서비스 매뉴얼

[06 AI CAN DO]

**이메일** [aiedutom.theimc.co.kr](mailto:aiedutom.theimc.co.kr)

**주소** 대구광역시 수성구 알파시티1로35길 17

**홈페이지** [aiedutom.co.kr](http://aiedutom.co.kr)



# 06

# AI CAN DO.

정형 데이터셋을 활용한 AI 서비스

01	Intro. 인트로	03
02	내 리스트	06
03	[STEP 1] 데이터셋 선정	14
04	[STEP 2] 데이터 선택	32
05	[STEP 3] 모델 학습	38
06	포스터 만들기	62

# AI CAN DO.

정형 데이터셋을 활용한 AI 서비스

01	Intro. 인트로	03
02	내 리스트	06
03	[STEP 1] 데이터셋 선정	14
04	[STEP 2] 데이터 선택	32
05	[STEP 3] 모델 학습	38
06	포스터 만들기	62



## - 학습 단계 (4단계) -

STEP 1.



### 데이터셋 선정

활용할 데이터셋을 선정하는 단계

- 1) 제공되는 데이터셋 중 선택
- 2) 보유 중인 데이터셋 업로드

STEP 2.



### 데이터 선택

1) 분류 / 회귀 모델 중 학습시킬

모델 유형을 선택하고,

2) 변수로 사용할 칼럼을 선택

STEP 3.



### 모델 학습

기계학습 기법을 선택하여

데이터를 학습하고,

학습 결과를 통해 모델을 평가

STEP 4.



### 포스터 만들기

나의 인공지능 모델을 소개하는

포스터를 만들며, 제작 과정을

복습하고 활용 방안을 생각

정형 데이터셋을 학습시켜 '나만의 AI'를 만들고 데이터를 예측할 수 있습니다.

내가 선택한 모델 유형과 입력 데이터에 따라 달라지는 학습 결과를 확인해 보세요!



STEP  
01

데이터셋 선정

어떤 모델을 만들 것인지를 바탕으로 데이터셋을 선택해 주세요.  
선택한 데이터셋에 따라 생성되는 인공지능이 달라집니다!

STEP  
02

데이터 선택

결과 데이터에 따라 분류 모델과 회귀 모델 중 모델 유형을 선택하고,  
데이터셋에서 결과변수와 입력변수로 사용할 칼럼을 선택해 주세요!

STEP  
03

모델 학습

선택한 모델 유형에 적합한 기계학습 기법을 선택합니다.  
입력한 데이터 칼럼에 따라 사용자가 값을 입력하고,  
그에 따른 학습된 모델의 학습 결과를 확인할 수 있습니다.

STEP  
04

포스터 만들기

내가 만든 인공지능 모델을 소개하는 포스터를 만들며, 모델 제작  
과정을 복습하고 활용할 수 있는 방안에 대해 생각해 볼 수 있어요!

# AI CAN DO.

정형 데이터셋을 활용한 AI 서비스

01	Intro. 인트로	03
02	내 리스트	06
03	[STEP 1] 데이터셋 선정	14
04	[STEP 2] 데이터 선택	32
05	[STEP 3] 모델 학습	38
06	포스터 만들기	62



The screenshot shows the AI EDUTOM website interface. At the top, there is a navigation bar with tabs: Feel, Learn, Do (highlighted with a red box and a red circle with the number 1), Edu, and Communication. Below the navigation bar, there is a main banner with the text "이제는 실전! 내가 만들어보는 나만의 AI" and "나도 AI를 만들어볼 수 있을까?". Below the banner, there is a section titled "다양한 데이터셋을 활용하여 모델을 구축하며 나만의 인공지능을 만들어 볼 수 있습니다." and a sub-section "알고리즘에 대한 이해력만 있다면, 코딩기술 없이 다양하고 새로운 인공지능을 만들어낼 수 있습니다." Below this, there are two service cards: "AI CAN DO" and "I CAN CLASSIFY". The "AI CAN DO" card has a red box around its "시작하기" button, and a hand cursor is pointing at it, with a red circle and the number 2 next to it.

1. 메인페이지 상단에 위치한 [Do] 버튼을 눌러, Do 단계로 이동해주세요.
2. AI CAN DO 서비스의 [시작하기] 버튼을 눌러 입장해 주세요.



1. AI CAN DO 서비스 메인페이지에서 [내 리스트] 버튼을 클릭해 주세요.

나의 '업로드 데이터'와 완성한 '학습 모델' 및 '포스터'를 관리할 수 있는 페이지로 이동합니다.





1. [업로드 데이터] 버튼을 클릭해 주세요.
2. 직접 보유하고 있는 데이터 중 업로드 한 데이터 리스트를 확인할 수 있습니다.  
( 파일명을 변경하거나, 데이터를 선택하여 삭제할 수 있고, 제목을 입력해 검색할 수도 있습니다.)

\* 데이터 업로드 방법은 목차 : [\[STEP 1\] 데이터 선정 - 보유 데이터셋 업로드](#)에서 확인해 보세요.



내 리스트

나의 업로드 데이터와 완성한 학습모델 및 포스터를 관리하세요.

업로드 데이터    **학습 모델**    포스터

전체 4건     전체선택/해제     삭제    제목을 입력하세요.    🔍 검색

<input type="checkbox"/> [AI EDUTOM] 위라벨 정수_테스트 STEP3 ▶ 모델 학습 완료   2024-05-31	<input type="checkbox"/> [AI EDUTOM] 아이리스 데이터셋_테스트 STEP1 ▶ 데이터셋 선정 완료   2024-05-27	<input type="checkbox"/> [AI EDUTOM] 타이타닉 생존자_테스트 STEP1 ▶ 데이터셋 선정 완료   2024-05-27
--	---	--

1. [학습 모델] 버튼을 클릭해 주세요.
2. 내가 만든 완성된 학습 모델 리스트를 확인할 수 있습니다.  
( 파일명을 변경하거나, 학습 모델을 선택하여 삭제할 수 있고, 제목을 입력해 검색할 수도 있습니다.)
3. [STEP 3 : 모델 학습] 단계까지 모두 진행이 완료된 학습 모델을 선택할 경우, 미리보기 화면을 통해 학습 모델의 내용을 확인할 수 있습니다.



전체 4건  전체선택/해제  삭제

[AI EDUTOM] 워라벨 점수\_테스트

STEP3 >>> 모델 학습 완료 | 2024-05-31

**1**

미리보기

결과 성능 지표	
	Errors
평균 절대 오차	24.6482
평균 백분율 오차 (%)	3.7085
평균 계급 오차	950.8501
평균 계급근 오차	30.8359

시각화 자료

① 평균 절대 오차: 오차(실제값-예측값)의 절대값을 모두 더하여 데이터의 수로 나눈 값. 절대값을 취하기 때문에 가장 직관적으로 오차를 알 수 있는 지표입니다.

② 평균 백분율 오차: 오차의 백분율을 더하여 데이터의 수로 나누고 다시 100을 곱하여 백분율로 나타낸 값. 오차가 어느 정도 차이나는지 퍼센트로 알 수 있습니다.

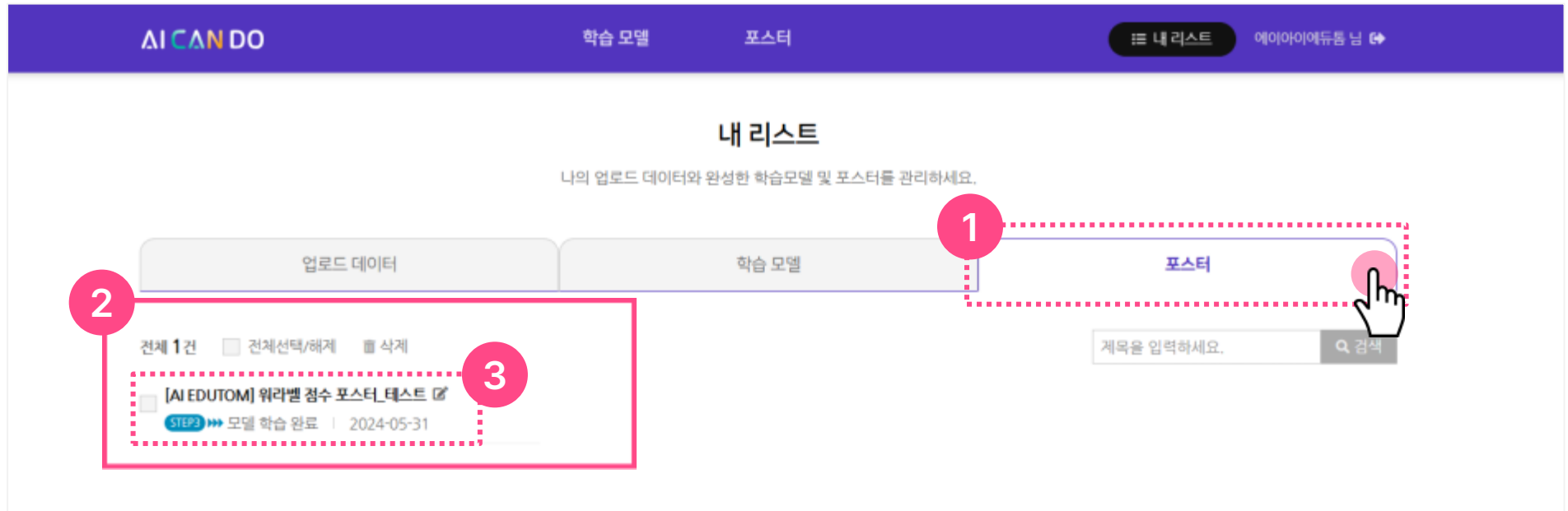
③ 평균 계급 오차: 오차의 계급들의 합을 데이터수로 나눈 값. 직관적으로 어느 정도 오차가 발생하는지 확인 하기 어렵고, 이상치가 있을 경우 값이 매우 커질 수 있습니다.

④ 평균 계급근 오차: 평균계급오차에 무드를 씌운 값. 이 방법을 통해 실제값과 유사한 단위 값으로 확인 가능하며, 이상치가 있을 경우 값이 매우 커질 수 있습니다.

능력증명회수	하루_과일채소섭취량	연령	주간_회년회수	성별	하루_결혼	워라벨점수	예측
1	5	3	21~35	여자	5	651.6	653.5
2	8	3	21~35	여자	5	650.5	687.2
3	2	1	21~35	남자	6	623.5	639.3
4	7	2	21~35	남자	4	602.5	651.0

⑤ 테스트 데이터셋의 결과 실제값과 모델이 예측한 예측값을 2차원 공간에 표현 한 것입니다. 모델이 잘 예측 할수록 초록색 점이 보라색 직선(y=x) 가까이 분포되며, 어떤 구간에서 예측을 잘 하고 못하였는지 확인 할 수 있습니다.

1. 미리보기 화면을 통해, 학습 모델의 내용을 확인해 보세요.



1. [포스터] 버튼을 클릭해 주세요.
2. 내가 만든 포스터 리스트를 확인할 수 있습니다.  
( 포스터 제목을 변경하거나, 포스터를 선택하여 삭제할 수 있고, 제목을 입력해 검색할 수도 있습니다.)
3. 포스터 제목을 클릭하면, 미리보기 화면을 통해 포스터 내용을 확인할 수 있습니다.



**1** 미리보기

[AI EDUTOM] 워라벨 점수\_테스트 / 2024-05-31

[AI EDUTOM] 워라벨 점수 포스터\_테스트

소속 : 더아이엠씨 (매뉴얼제작 용) 1학년 | 작성자 : 9jh1225

**01. 나의 AI는?**

[AI EDUTOM] 워라벨 점수 예측 AI

**02. 데이터셋**

- 데이터명 : wal.csv
- 데이터 정보 : 행 : 5270행, 열 : 23열
- 입력 데이터 : 능력공명횟수, 하루\_과일채소섭취량, 연령, 주간\_화난횟수, 성별, 하루\_걸음수, 침단위
- 결과 데이터 : 워라벨점수
- 라벨 정보 : 최소값 : 480, 최대값 : 804.3, 평균 : 667.63256166983

**04. 모델 결과**

• 시각화

• 결과 성능 지표

	평균 오차 값
평균 절대 오차	24.6482
평균 백분율 오차 (%)	3.7085
평균 계급 오차	950.8501
평균 계급근 오차	30.8359

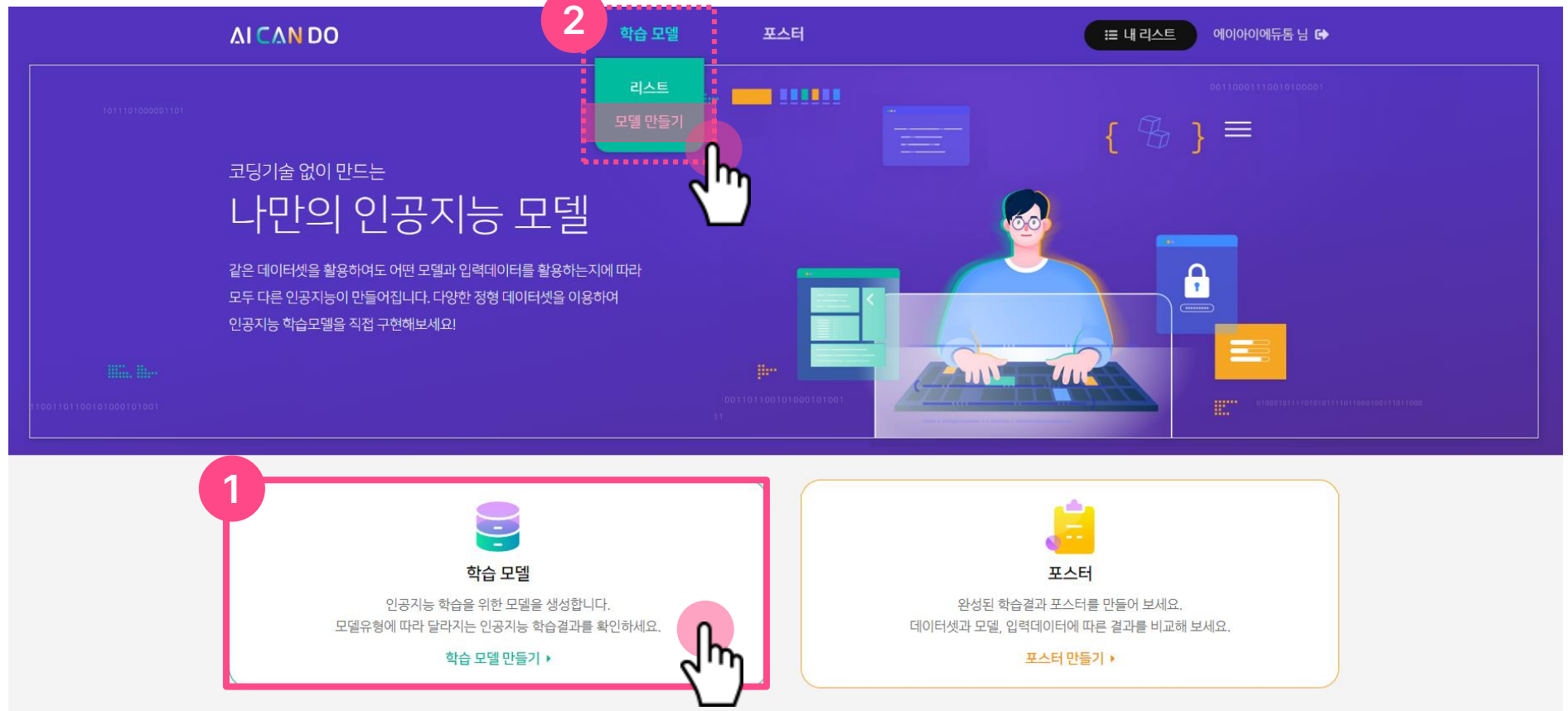
**05. 느낀점 및 활용방안**

1. 미리보기 화면을 통해, 내가 만든 포스터의 내용을 확인할 수 있습니다.

# AI CAN DO.

정형 데이터셋을 활용한 AI 서비스

01	Intro. 인트로	03
02	내 리스트	06
03	[STEP 1] 데이터셋 선정	<b>14</b>
04	[STEP 2] 데이터 선택	32
05	[STEP 3] 모델 학습	38
06	포스터 만들기	62



1. 서비스 메인페이지에서 [학습 모델 만들기] 버튼을 클릭해 다음 페이지로 이동해 주세요.
2. 또는 상단의 [학습 모델 → 모델 만들기] 카테고리를 통해 이동할 수도 있습니다.

1. 내가 제작할 학습 모델의 제목을 입력해 주세요.
2. [정형데이터 선택]에서 AI EDUTOM이 제공하는 4가지 정형 데이터셋 중 하나를 선택해 주세요.





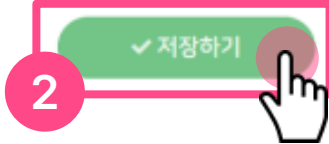
정형데이터란?

정형 데이터는 데이터베이스의 정해진 규칙(Rule)에 맞게 수치만으로 의미 파악이 쉬운 데이터를 의미합니다.  
 데이터는 미리 만들어진 형식 또는 틀에 저장되며 그 값이 의미를 파악하기 쉽고, 규칙적인 값으로 데이터들을 표의 형태로 확인 할 수 있습니다.

1

데이터 미리보기

	승객번호	좌석등급	성별	나이	동승자_형제/자매/배우자	동승자_부모/자식	승객요금	탑승지	생존여부
1	1	3	남자	22	1	0	7.25	사우스햄프턴	0
2	2	1	여자	38	1	0	71.2833	세르부르	1
3	3	3	여자	26	0	0	7.925	사우스햄프턴	1
4	4	1	여자	35	1	0	53.1	사우스햄프턴	1
5	5	3	남자	35	0	0	8.05	사우스햄프턴	0
6	6	3	남자		0	0	8.4583	퀸즈타운	0
7	7	1	남자	54	0	0	51.8625	사우스햄프턴	0
8	8	3	남자	2	3	1	21.075	사우스햄프턴	0
9	9	3	여자	27	0	2	11.1333	사우스햄프턴	1
10	10	2	여자	14	1	0	30.0708	세르부르	1
11	11	3	여자	4	1	1	16.7	사우스햄프턴	1
12	12	1	여자	58	0	0	26.55	사우스햄프턴	1







1. 데이터 미리보기를 통해 선택한 데이터셋의 내용을 확인할 수 있습니다.  
 데이터셋에 따라 구성하고 있는 칼럼의 요소가 다릅니다. ( ex : 타이타닉 생존자 → 좌석 등급, 성별 등 )
2. [저장하기] 버튼을 눌러주세요.  
 작성한 모델의 이름과 데이터셋이 저장되며, 다음 단계로 이동합니다.



## 어떤 정형 데이터셋을 활용할 수 있나요?

### 정형데이터 선택

1	 타이타닉 생존자	분류
2	 동물 종	분류
3	 워라벨 점수	회귀
4	 인도 중고차 가격	회귀

1. 타이타닉 생존자 (분류) : 타이타닉호 탑승자의 개별정보를 통해 생존 여부를 예측할 수 있습니다.
2. 동물종 (분류) : 동물의 특징을 통해 동물의 종을 예측할 수 있습니다.
3. 워라벨 점수 (회귀) : 설문조사 결과를 통해 참여자의 워라벨 점수를 예측할 수 있습니다.
4. 인도 중고차 가격 (회귀) : 인도에서 판매되는 중고차 정보를 통해 중고차 가격을 예측할 수 있습니다.



데이터 리스트 (칼럼)

승객번호	좌석등급	성별	나이	동승자_형제/자매/배우자	동승자_부모/자식	승객요금	탑승지	생존여부
1	3	남자	22	1	0	7.25	사우샘프턴	0
2	1	여자	38	1	0	71.2833	세르부르	1
3	3	여자	26	0	0	7.925	사우샘프턴	1
4	1	여자	35	1	0	53.1	사우샘프턴	1
5	3	남자	35	0	0	8.05	사우샘프턴	0
6	3	남자		0	0	8.4583	퀸즈타운	0
7	1	남자	54	0	0	51.8625	사우샘프턴	0
8	3	남자	2	3	1	21.075	사우샘프턴	0
9	3	여자	27	0	2	11.1333	사우샘프턴	1
10	10	여자	14	1	0	30.0708	세르부르	1
11	11	여자	4	1	1	16.7	사우샘프턴	1
12	12	여자	58	0	0	26.55	사우샘프턴	1
13	13	남자	20	0	0	8.05	사우샘프턴	0
14	14	남자	39	1	5	31.275	사우샘프턴	0

분류

타이타닉 생존자

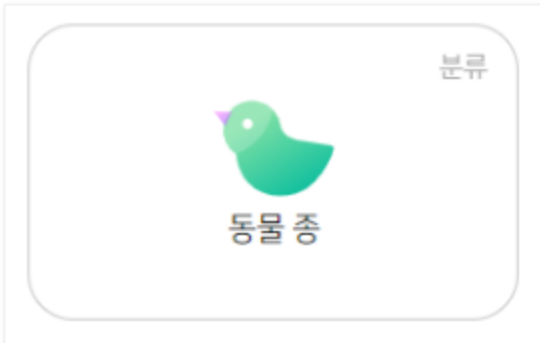
① 타이타닉 생존자

[타이타닉 생존자] 데이터셋을 구성하고 있는 칼럼 요소를 확인해 보세요!

승객번호, 좌석등급, 성별, 동승자\_형제/자매/배우자, 동승자\_부모/자식, 승객 요금, 탑승지, 생존여부, 나이



데이터 리스트 (칼럼)



② 동물종

	동물이름	털	깃털	알	모유	비행	이빨	척추	호흡	다리수	꼬리	애완동물	동물종
	땅돼지	1	0	0	1	0	1	1	1	4	0	0	포유류
	명양	1	0	0	1	0	1	1	1	4	1	0	포유류
	배스	0	0	1	0	0	1	1	0	0	1	0	어류
	곰	1	0	0	1	0	1	1	1	4	0	0	포유류
	수탉	1	0	0	1	0	1	1	1	4	1	0	포유류
	물소	1	0	0	1	0	1	1	1	4	1	0	포유류
7	송아지	1	0	0	1	0	1	1	1	4	1	1	포유류
	잉어	0	0	1	0	0	1	1	0	0	1	1	어류
	메기	0	0	1	0	0	1	1	0	0	1	0	어류
10	기니피그	1	0	0	1	0	1	1	1	4	0	1	포유류
11	치타	1	0	0	1	0	1	1	1	4	1	0	포유류
12	닭	0	1	1	0	1	0	1	1	2	1	1	조류
13	처브	0	0	1	0	0	1	1	0	0	1	0	어류
14	대합	0	0	1	0	0	0	0	0	0	0	0	무척추동물


**[동물종] 데이터셋을 구성하고 있는 칼럼 요소를 확인해 보세요!**

동물 이름, 털, 깃털, 알, 모유, 비행, 수중, 육식, 이빨, 척추, 호흡, 독, 지느러미, 다리 수, 꼬리, 애완동물, 동물종



데이터 리스트 (칼럼)

성별	연령	능력증명횟수	친한사람수	미래계획년수	비만여부	수입만족여부	하루_대화시간	하루_수면시간	하루_스트레스횟수	
			4	10	1	1	3	6	1	
			6	3	0	1	8	6	4	
			8	8	1	1	8	9	5	
			4	10	1	1	10	8	2	
			5	1	0	1	10	7	4	
			4	4	1	1	3	6	2	
			3	3	0	1	8	6	4	
	여자	51이상	5	10	5	1	1	8	6	3
			5	0	0	0	10	8	5	
			0	0	1	0	0	1	5	
11	여자	20이하	2	6	4	0	1	9	8	2
12	여자	36~50	3	10	10	1	1	10	7	3
13	남자	36~50	10	7	10	1	1	5	6	3
14	여자	51이상	2	8	0	0	1	2	8	1



회귀

워라벨 점수


③ 워라벨 점수

**[워라벨 점수] 데이터셋을 구성하고 있는 칼럼 요소를 확인해 보세요!**

성별, 연령, 능력증명횟수, 친한사람수, 미래계획년수, 비만여부, 수입 만족 여부, 걸음수, 과일채소섭취량, 대화시간, 수면시간, 스트레스 횟수, 자기 투자시간, 집중시간, 화낸 횟수, 휴식활동시간, 목표성취빈도, 기부횟수, 새로운장소방문횟수, 성공경험, 성공을도와준수, 휴가반납수, 워라벨점수



데이터 리스트 (칼럼)



회귀

인도 중고차 가격

④ 인도 중고차 가격

	제조사	지역	생산년도	운전거리	연료구분	구동방식	소유자구분	주행연비	엔진_CC	마력	좌석수	중고차가격
	바이		2010	72000	천연가스	수동	첫번째	26.6	998	58.16	5	1.75
	네		2015	41000	경유	수동	첫번째	19.67	1582	126.2	5	12.5
	나이		2011	46000	휘발유	수동	첫번째	18.2	1199	88.7	5	4.5
	나이		2012	87000	경유	수동	첫번째	20.77	1248	88.76	7	6
	임바토르		2013	40670	경유	자동	두번째	15.2	1968	140.8	5	17.74
	이데라바드		2012	75000	LPG	수동	첫번째	21.1	814	55.2	5	2.35
	이푸르		2013	86999	경유	수동	첫번째	23.08	1461	63.1	5	3.5
	바이		2016	36000	경유	자동	첫번째	11.36	2755	171.5	8	17.5
	네		2013	64430	경유	수동	첫번째	20.54	1598	103.6	5	5.2
10	타타	첸나이	2012	65932	경유	수동	두번째	22.3	1248	74	5	1.95
11	마루티	코치	2018	25692	휘발유	수동	첫번째	21.56	1462	103.25	5	9.95
12	혼다	콜카타	2012	60000	휘발유	자동	첫번째	16.8	1497	116.3	5	4.49
13	마루티	자이푸르	2015	64424	경유	수동	첫번째	25.2	1248	74	5	5.6
14	랜드로버	델리	2014	72000	경유	자동	첫번째	12.7	2179	187.7	5	27

[인도 중고차 가격] 데이터셋을 구성하고 있는 칼럼 요소를 확인해 보세요!

제조사, 지역, 생산년도, 운전 거리, 연료 구분, 구동 방식,  
소유자 구분, 주행 연비, 엔진\_CC, 마력, 좌석 수, 중고차 가격



학습 모델 만들기

다양한 경험데이터를 이용하여 나만의 AI 모델을 만들어 보세요.

제작 할 학습 모델의 제목을 입력해 주세요.

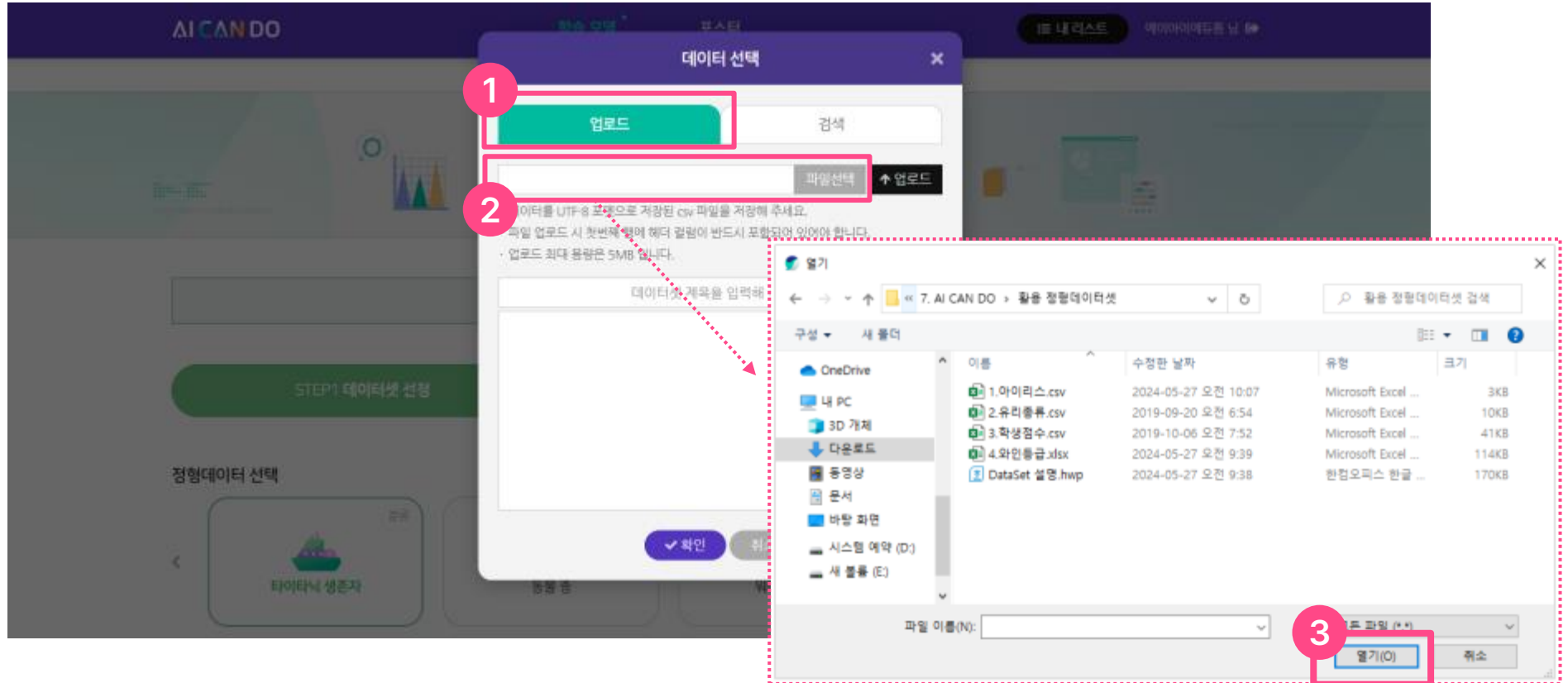
STEP1 데이터셋 선정    STEP2 데이터 선택    STEP3 모델 학습

정형데이터 선택

타이타닉 생존자    동물 종    워라벨 점수    인도 중고차 가격    + 업로드 데이터

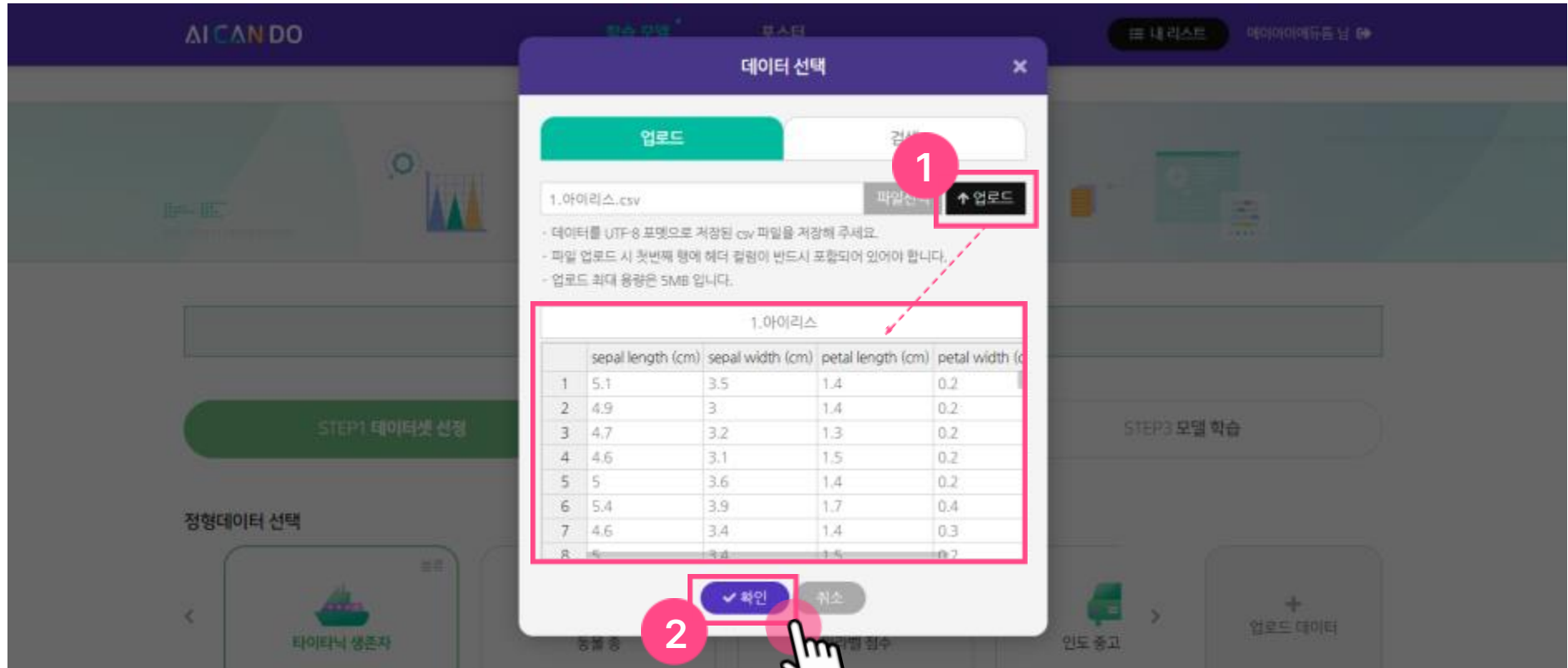
AI EDUTOM에서 제공하는 데이터셋 외에도, 사용자가 보유하고 있는 데이터셋을 활용할 수 있습니다.

1. [업로드 데이터] 버튼을 클릭해 주세요.



1. 생성된 팝업창에서 [업로드]를 눌러주세요.
2. [파일 선택] 버튼을 클릭하면, 파일을 가져올 수 있는 창이 생성됩니다.
3. 보유하고 있는 데이터셋을 [열기] 버튼을 눌러 불러와 주세요.





1. [업로드] 버튼 클릭해 주세요. 선택한 데이터 셋의 데이터(칼럼) 내용을 확인할 수 있습니다. 이때, 제목은 업로드한 파일 제목과 동일하게 설정됩니다.
2. [확인] 버튼을 누르면, 보유데이터 업로드가 완료됩니다.



타이타닉 생존자
동물 중
워라벨 점수
인도 중고

**1** 정형데이터란?  
 정형 데이터는 데이터베이스의 정해진 규칙(Rule)에 맞게 수치만으로 의미 파악이 쉬운 데이터를 의미합니다.  
 데이터는 미리 만들어진 형식 또는 틀에 저장되며 그 값의 의미를 파악하기 쉽고, 규칙적인 값으로 데이터들을 표의 형태로 확인 할 수 있습니다.

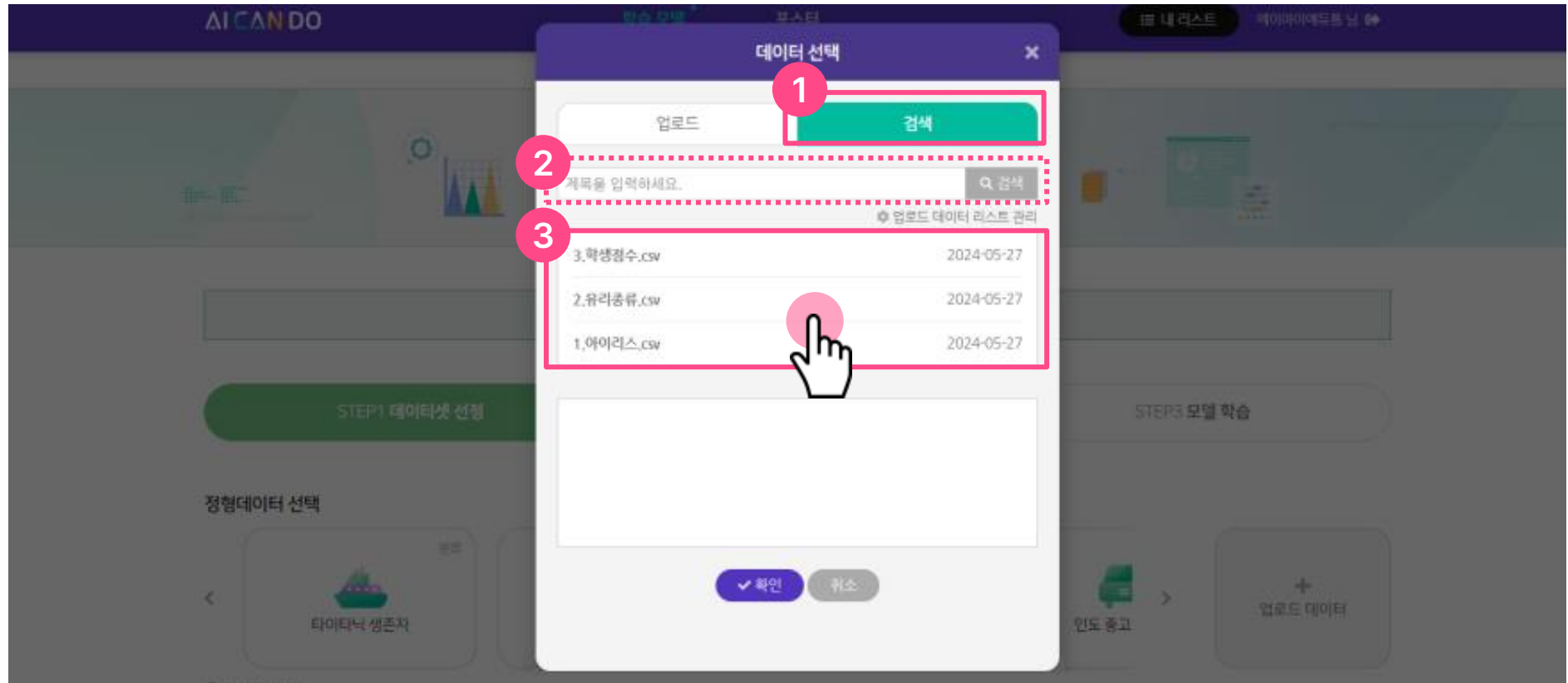
**1** 데이터 미리보기

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
1	5.1	3.5	1.4	0.2	0
2	4.9	3	1.4	0.2	0
3	4.7	3.2	1.3	0.2	0
4	4.6	3.1	1.5	0.2	0
5	5	3.6	1.4	0.2	0
6	5.4	3.9	1.7	0.4	0
7	4.6	3.4	1.4	0.3	0
8	5	3.4	1.5	0.2	0
9	4.4	2.9	1.4	0.2	0
10	4.9	3.1	1.5	0.1	0
11	5.4	3.7	1.5	0.2	0

✓ 저장하기

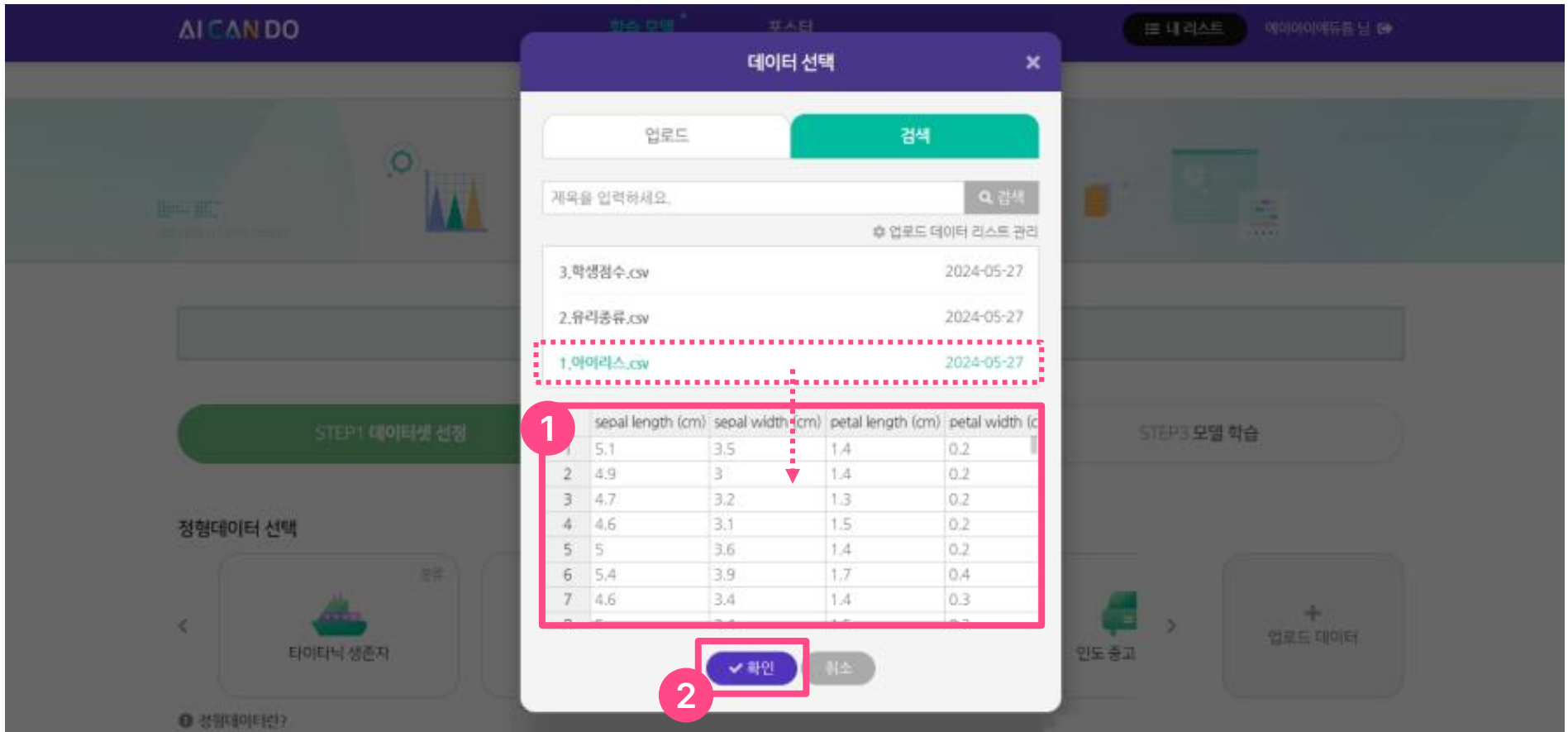
**2**

1. 데이터 미리보기를 통해 업로드한 보유 데이터셋의 내용을 확인할 수 있습니다.
2. [저장하기] 버튼을 눌러주세요.  
업로드한 데이터셋이 저장되며, 다음 단계로 이동합니다.

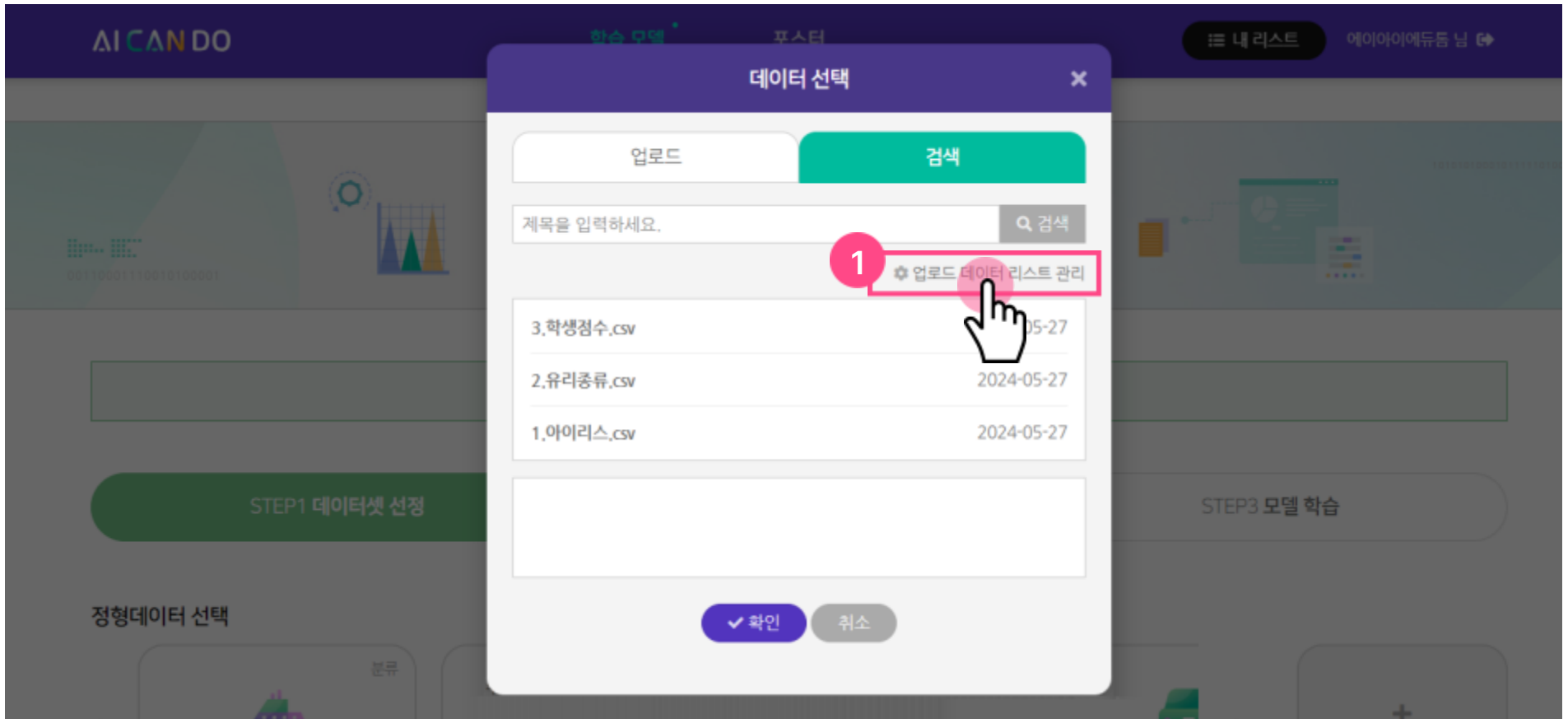


'데이터 선택 - 업로드' 단계에서 업로드해 둔 데이터 리스트를 검색하여 불러올 수 있습니다.

1. 생성된 팝업창에서 [검색]을 눌러주세요.
2. 검색하여 데이터셋을 불러오하고자 할 경우, 검색창에 제목 중 일부를 입력해 주세요.
3. 혹은, 리스트에 있는 데이터셋을 바로 선택할 수도 있습니다.



1. 리스트에 있는 데이터셋의 파일명을 클릭하면, 선택한 데이터셋의 내용을 데이터 미리보기 영역에서 확인할 수 있습니다.
2. [확인] 버튼을 누르면, 보유데이터 업로드가 완료됩니다.



1. [업로드 데이터 리스트 관리] 버튼을 클릭하면 '나의 리스트 관리' 페이지로 이동하며, 업로드해둔 데이터 중 불필요한 데이터를 삭제할 수 있습니다.



**AI CAN DO**      학습 모델      포스터      **내 리스트**      에이아이에듀동 님

## 내 리스트

나의 업로드 데이터와 완성한 학습모델 및 포스터를 관리하세요.

업로드 데이터
학습 모델
포스터

전체 3건     전체선택/해     삭제

3.학생점수.csv    2.유리종류.csv    1.아이리스.csv

undefined | 2024-05-27    undefined | 2024-05-27    undefined | 2024-05-27

제목을 입력하세요.    🔍 검색

1. 삭제하고자 하는 데이터를 선택해 주세요.
2. [삭제] 버튼을 클릭하면, 업로드 데이터가 삭제됩니다.



## [STEP 2] 단계로 넘어가기 전, 꼭 확인해 보세요!

### 놓치면 안 되는 꿀팁은?

#### [STEP 1] 데이터셋 선정 :

만들고자 하는 모델의 학습 데이터셋을 선정하는 단계입니다.

학습하고자 하는 데이터셋의 각 변수(칼럼)의 구성요소를

미리 확인해 본 후, 그다음 단계인 STEP 2로 넘어가 주세요!



# AI CAN DO.

정형 데이터셋을 활용한 AI 서비스

01	Intro. 인트로	03
02	내 리스트	06
03	[STEP 1] 데이터셋 선정	14
04	[STEP 2] 데이터 선택	<b>32</b>
05	[STEP 3] 모델 학습	38
06	포스터 만들기	62



**학습 모델 만들기**  
다양한 경험데이터를 이용하여 나만의 AI 모델을 만들어 보세요.

[AI EDUTOM] 동물 종\_테스트

STEP1 데이터셋 선정    **STEP2 데이터 선택**    STEP3 모델 학습

**2** 모델 유형

**분류모델**  
선택한 데이터가 어떤 그룹에 속하는지 예측하는 모델입니다.

**회귀모델**  
선택한 데이터의 결과값에 가까운 실수값을 예측하는 모델입니다.

① 분류모델: 결과데이터가 그룹을 나타내는 값일 경우 선택하면 됩니다. 예시) 동물종, 생존여부  
회귀모델: 결과데이터가 연속형 숫자형 데이터(실수값)일 경우 선택하면 됩니다. 예시) 자동차 가격, 워라벨 점수 등

**데이터 미리보기**

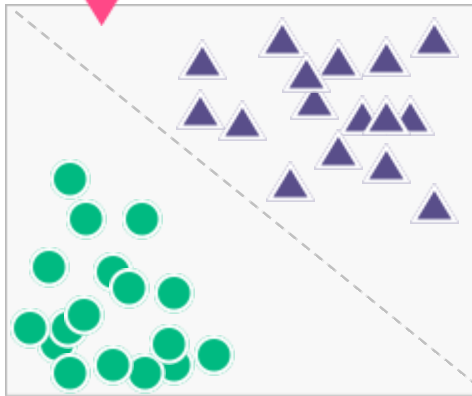
	동물이름	털	깃털	알	모유	비행	수중	육식	이빨	척추	호흡	독	지느러미	다리수	고리	애완동물	동물종
1	당대지	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	포유류

1. 'STEP 2 : 데이터 선택' 단계입니다.
2. 이전 단계에서 선택한 데이터셋에 맞는 모델 유형을 [분류 모델 / 회귀 모델] 중에서 선택해 주세요.
  - ex) 타이타닉 생존자, 동물종 → [분류 모델] 선택
  - ex) 워라벨 점수, 인도 중고차 가격 → [회귀 모델] 선택



# 분류 모델 / 회귀 모델은 어떤 차이가 있나요?

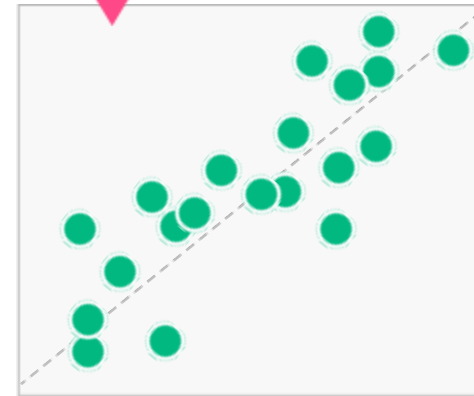
분류 모델



선택한 데이터가 어떤 그룹에 속하는지 예측하는 모델입니다.

결과 데이터가 '그룹에 대한 값'을 나타내는 경우

회귀 모델



선택한 데이터의 결괏값에 가까운 실숫값을 예측하는 모델입니다.

결과 데이터가 '연속 숫자형 값'을 나타내는 경우

↶ 📄 ☰

### 데이터 컬럼 선택

**1**

입력 데이터

비행

깃털 ✕ + 모유 ✕ + 꼬리 ✕ + 알 ✕ +  
 독 ✕ + 비행 ✕ +

**2**

결과 데이터

동물종

동물종

선택 데이터 미리보기

	깃털	모유	꼬리	알	독	비행	동물종
1	0	1	0	0	0	0	포유류
2	0	1	1	0	0	0	포유류
3	0	0	1	1	0	0	어류
4	0	1	0	0	0	0	포유류
5	0	1	1	0	0	0	포유류
6	0	1	1	0	0	0	포유류
7	0	1	1	0	0	0	포유류
8	0	0	1	1	0	0	어류
9	0	0	1	1	0	0	어류
10	0	1	0	0	0	0	포유류

✓ 저장하기

1. 데이터셋 중에서 '입력 데이터'로 사용할 컬럼을 선택해 주세요.  
 입력 데이터는 여러 개의 컬럼을 선택할 수 있으며, 최소 2개 이상 선택해 주세요.  
**\* 선택한 컬럼의 수에 따라 모델의 성능이 달라집니다. 컬럼을 많이 선택하면 성능이 높아질 수 있습니다.**
2. 데이터셋 중 '결과 데이터'로 사용할 컬럼을 선택해 주세요.  
 결과 데이터는 한가지 컬럼만 선택할 수 있습니다.

데이터 칼럼 선택

입력 데이터: 비행 (선택됨), 깃털 ✕ +, 모유 ✕ +, 꼬리 ✕ +, 알 ✕ +, 독 ✕ +, 비행 ✕ +

결과 데이터: 동물종 (선택됨), 동물종

1 선택 데이터 미리보기

	깃털	모유	꼬리	알	독	비행	동물종
1	0	1	0	0	0	0	포유류
2	0	1	1	0	0	0	포유류
3	0	0	1	1	0	0	어류
4	0	1	0	0	0	0	포유류
5	0	1	1	0	0	0	포유류
6	0	1	1	0	0	0	포유류
7	0	1	1	0	0	0	포유류
8	0	0	1	1	0	0	어류
9	0	0	1	1	0	0	어류
10	0	1	0	0	0	0	포유류

2 저장하기

1. '선택 데이터 미리보기'를 통해 선택한 칼럼 데이터로 구성된 데이터셋을 미리 확인할 수 있습니다. 잘못 선택한 칼럼은 입력데이터 란에서 삭제하여 재설정할 수 있습니다.

2. [저장하기] 버튼을 눌러주세요.

선택한 칼럼들로 구성된 데이터셋이 저장되며, 다음 단계로 이동합니다.

\* 모델 유형 및 선택 칼럼이 올바르지 않을 경우, 다음 단계로 넘어갈 수 없습니다.



## [STEP 3] 단계로 넘어가기 전, 꼭 확인해 보세요!

놓치면 안 되는 꿀팁은?

### [STEP 2] 데이터 선택 :

모델 유형을 선택하고, 입력/결과 데이터를 선택하는 단계입니다.

데이터셋의 결과 데이터에 맞는 모델 유형을 선택해 주세요!

어떤 칼럼을 선택하는가에 따라 모델의 성능이 달라질 수 있습니다.



# AI CAN DO.

정형 데이터셋을 활용한 AI 서비스

01	Intro. 인트로	03
02	내 리스트	06
03	[STEP 1] 데이터셋 선정	14
04	[STEP 2] 데이터 선택	32
05	[STEP 3] 모델 학습	38
06	포스터 만들기	62

**학습 모델 만들기**  
다양한 정형데이터를 이용하여 나만의 AI 모델을 만들어 보세요.

[AI EDUTOM] 동물 종\_테스트

STEP1 데이터셋 선정    STEP2 데이터 선택    **STEP3 모델 학습**

**2 기계 학습 기법**

- 선형 회귀**  
입력변수와 결과변수가 선형 상관관계를 가지는 모델로 연속형 숫자 데이터를 예측하는 회귀모델로 적합합니다.
- 서포터벡터머신**  
데이터 그룹의 분류를 위한 경계를 표현하며 특히 이진(참, 거짓) 분류모델로 적합합니다.
- 로지스틱**  
변수들 간의 선형 결합을 가지고 사건의 발생 가능성(확률)을 예측하며 범주형 데이터를 분류하는데 적합합니다.
- 의사결정나무**  
변수들 사이에 존재하는 특징을 예측 가능한 규칙의 조합으로 나타내며 그 모양이 나뭇가지와 비슷합니다.
- 신경망**  
분류, 회귀  
신경세포인 뉴런을 모방한 모델로 네트워크를 생성하고 가중치를 스스로 지도학습 합니다.

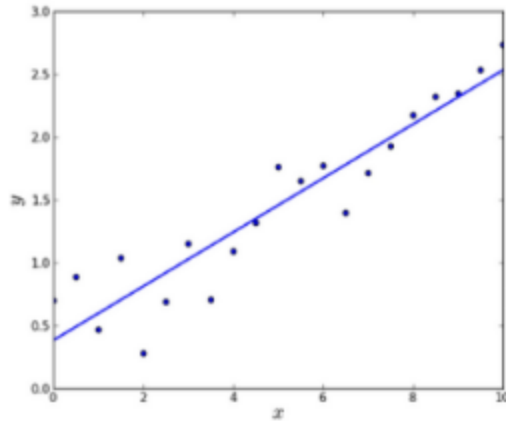
**모델 학습하기 >**

1. 'STEP 3 : 모델 학습' 단계입니다.
2. 모델을 학습할 [기계 학습 기법] 중 한 가지를 선택해 주세요.  
이전 단계에서 선택한 모델 유형(분류/회귀)을 바탕으로 '기계 학습 기법'이 활성화됩니다.
3. [모델 학습하기] 버튼을 클릭해 주세요. 인공지능 모델이 생성되며 학습을 시작합니다.



# 어떤 기계 학습 기법이 있나요??

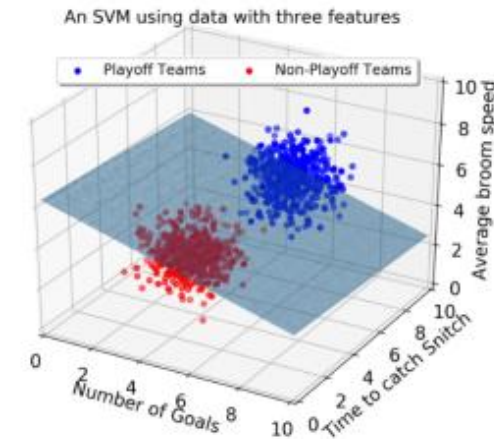
## ① 선형 (선형 회귀)



입력변수와 결과변수가  
선형 상관관계를 가지는 모델입니다.

연속성 숫자값을 예측하는 **회귀 모델**에 적합

## ② 서포트 벡터 머신



데이터 그룹의 분류를 위한  
기준선을 표현하는 모델입니다.

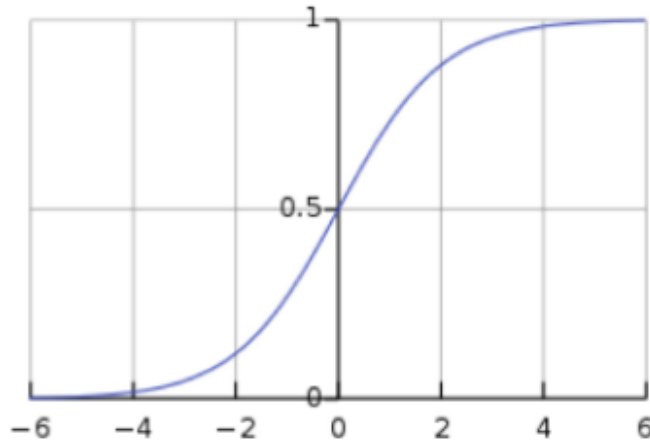
이항형 (참/거짓)의 **분류 모델**에 적합  
경우에 따라 분류 / 회귀 모델 모두 적용될 수 있음





# 어떤 기계 학습 기법이 있나요??

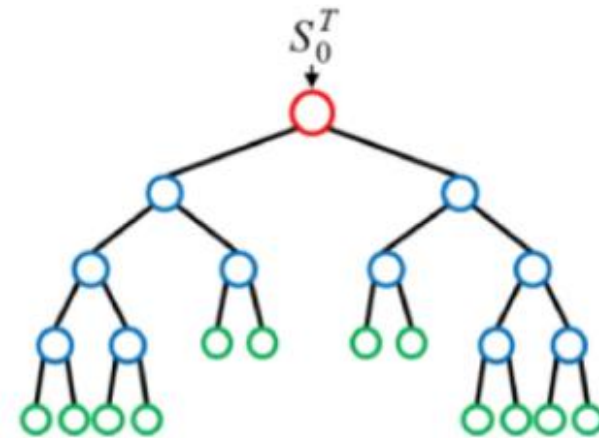
## ③ 로지스틱



변수 간의 선형 결합을 가지고 사건의 발생 가능성(확률)을 예측하는 모델입니다.

범주형 데이터를 분류 예측하는 **분류 모델**에 적합

## ④ 의사결정나무



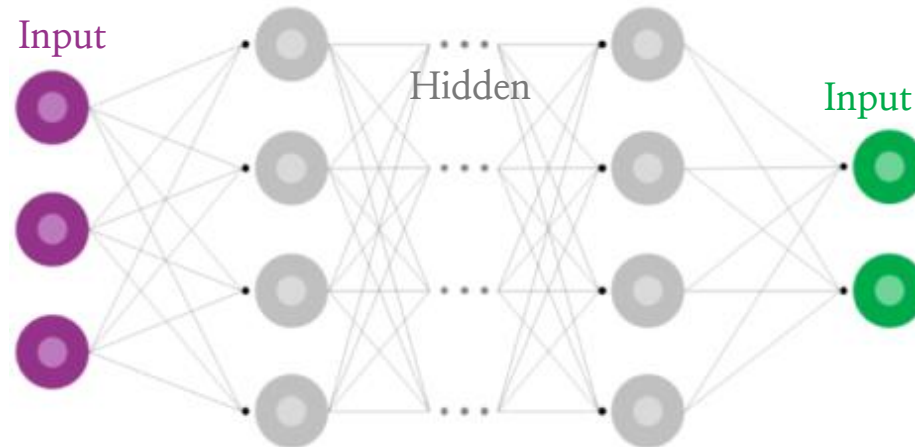
변수들 사이에 존재하는 특징을 예측 가능한 규칙의 조합으로 나타냅니다.

**분류 / 회귀 모델**에 모두 사용 가능



# 어떤 기계 학습 기법이 있나요??

## ⑤ 신경망



신경세포인 뉴런을 모방한 모델로,  
네트워크를 생성하고 가중치를 스스로 지도학습합니다.

**분류 / 회귀 모델에 모두 사용 가능**



## 분류 모델

타이타닉, 동물종 → 서포터벡터머신 / 로지스틱 / 의사결정나무 / 신경망 선택 시



### 학습 결과

1

테스트 입력 데이터셋

	깃털	모유	꼬리	털	독	비행
1	0	1	1	1	0	0
2	0	1	1	1	0	0
3	1	0	1	0	0	1
4	0	0	1	0	0	0
5	0	0	0	0	1	0
6	1	0	1	0	0	0
7	0	1	0	1	0	0
8	0	1	1	1	0	0
9	0	0	1	0	1	0
10	0	0	0	0	0	0
11	0	1	1	1	0	0
12	0	0	0	0	0	0
13	1	0	1	0	0	1

결과 비교

	동물종	예측값
1	포유류	포유류
2	포유류	포유류
3	조류	조류
4	어류	어류
5	무척추동물	무척추동물
6	조류	조류
7	포유류	포유류
8	포유류	포유류
9	파충류	파충류
10	무척추동물	무척추동물
11	포유류	포유류
12	무척추동물	무척추동물
13	조류	조류

① 선택한 컬럼들로 구성된 입력 데이터셋을 학습과 테스트의 비율을 7 : 3으로 데이터셋은 분리하여 학습에 사용되지 않은 테스트 데이터셋의 입력 데이터 컬럼을 나타낸 데이터 테이블입니다.

② 테스트 입력 데이터셋의 실제 결과 데이터와 인공지능 모델이 예측 값을 비교하는 테이블입니다.

1. 테스트 입력 데이터셋은 이전 단계에서 선택한 입력 데이터셋의 테스트를 위한 값입니다.

\* 입력 데이터셋은 7 : 3 비율로 모델 학습과 테스트에 활용되며, 랜덤으로 분리하여 사용됩니다.

입력 데이터셋	70%	'모델 학습'에 사용
	30%	'테스트 데이터'로 사용 (완성된 모델을 테스트하는 용도)



## 분류 모델

타이타닉, 동물종 → 서포터벡터머신 / 로지스틱 / 의사결정나무 / 신경망 선택 시



테스트 입력 데이터셋

	깃털	모유	꼬리	털	독	비행
1	0	1	1	1	0	0
2	0	1	1	1	0	0
3	1	0	1	0	0	1
4	0	0	1	0	0	0
5	0	0	0	0	1	0
6	1	0	1	0	0	0
7	0	1	0	1	0	0
8	0	1	1	1	0	0
9	0	0	1	0	1	0
10	0	0	0	0	0	0
11	0	1	1	1	0	0
12	0	0	0	0	0	0
13	1	0	1	0	0	1

① 선택한 컬럼들로 구성된 입력 데이터셋을 학습과 테스트의 비율을 7 : 3으로 데이터셋은 분리하여 학습에 사용되지 않은 테스트 데이터셋의 입력 데이터 컬럼을 나타낸 데이터 테이블입니다.

결과 비교

	1 동물종	2 예측값
1	포유류	포유류
2	포유류	포유류
3	조류	조류
4	어류	어류
5	무척추동물	무척추동물
6	조류	조류
7	포유류	포유류
8	포유류	포유류
9	파충류	파충류
10	무척추동물	무척추동물
11	포유류	포유류
12	무척추동물	무척추동물
13	조류	조류

① 테스트 입력 데이터셋의 실제 결과 데이터와 인공지능 모델이 예측 값을 비교하는 테이블입니다.

**결과 비교**는 이전 단계에서 선택한 결과 데이터셋과 인공지능의 예측값을 비교하는 테이블입니다.

1. [동물종]은 이전 단계(STEP 2)에서 내가 선택한 결과 데이터셋 컬럼입니다.
2. [예측값]은 테스트 입력 데이터셋을 기준으로 인공지능이 예측한 값을 나타냅니다.

\* 분류 모델의 결과 데이터값은 숫자 형태가 아닌, '문자의 그룹' 형태를 씁니다. (ex : 포유류, 조류, 어류 등)



여기서 잠깐!

## 학습 결과 데이터는 어떻게 읽을까요?

테스트 입력 데이터셋

	깃털	모유	꼬리	털	독	비행
1	0	1	1	1	0	0
2	0	1	1	1	0	0
3	1	0	1	0	0	1
4	0	0	1	0	0	0
5	0	0	0	0	1	0
6	1	0	1	0	0	0
7	0	1	0	1	0	0
8	0	1	1	1	0	0

결과 비교

	동물종	예측값
1	포유류	포유류
2	포유류	포유류
3	조류	조류
4	어류	어류
5	무척추동물	무척추동물
6	조류	조류
7	포유류	포유류
8	포유류	포유류

예시) ▶



데이터 읽는 방법

'테스트 입력 데이터셋' → '결과 비교 : 예측값' 순서대로 데이터값을 읽어주세요!

\* 예시)

3번째 행에 위치한 동물종은 다음과 같은 특징이 있습니다.

[ 깃털과 꼬리가 있고 (1) / 비행하며 (1) / 모유를 먹지 않고 (0) / 털과 독이 없음 (0) ]

→ 위 데이터셋으로 예측한 동물종은 '조류'입니다.



## 분류 모델

타이타닉, 동물종 → 서포터벡터머신 / 로지스틱 / 의사결정나무 / 신경망 선택 시



1

시각화 자료

실제값 \ 예측값	0	1	2	3	4	5	6
0	1	1	0	0	0	0	0
1	0	3	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0	3	0	1	0
4	0	0	0	0	6	0	0
5	0	0	0	1	0	1	0
6	0	0	0	0	0	0	12

결과 성능 지표

라벨번호	라벨	정답	전체
0	근종류	1	2
1	무척추동물	3	3
2	양서류	0	1
3	어류	3	4
4	조류	6	6
5	파충류	1	2
6	포유류	12	12
합계		26	30

- ① 라벨번호 : 라벨 값을 대신하는 번호입니다.
- 라벨 : 해당 그룹의 유일 값을 나타냅니다.
- 정답 : 테스트 데이터를 학습된 모델이 예측하여 해당 라벨을 맞춘 수입니다.
- 전체 : 각 라벨 별 테스트 데이터의 수입니다.

1. 시각화 자료는 테스트 데이터 셋의 모델 예측 결과를 혼합 행렬로 표현한 것입니다.  
가로축은 '모델이 예측한 라벨의 개수' / 세로축은 '실제 테스트 데이터 라벨의 개수'를 나타냅니다.



## 분류 모델

타이타닉, 동물종 → 서포터벡터머신 / 로지스틱 / 의사결정나무 / 신경망 선택 시



시각화 자료

실제값 \ 예측값	0	1	2	3	4	5	6
0	1	1	0	0	0	0	0
1	0	3	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0	3	0	1	0
4	0	0	0	0	6	0	0
5	0	0	0	1	0	1	0
6	0	0	0	0	0	0	12

결과 지표	1	2	3	4
라벨번호	라벨	정답	전체	
0	곤충류	1	2	
1	무척추동물	3	3	
2	양서류	0	1	
3	어류	3	4	
4	조류	6	6	
5	파충류	1	2	
6	포유류	12	12	
합계		26	30	

- ① 라벨번호 : 라벨 값을 대신하는 번호입니다.
- 라벨 : 해당 그룹의 유일 값을 나타냅니다.
- 정답 : 테스트 데이터를 학습된 모델이 예측하여 해당 라벨을 맞춘 수입니다.
- 전체 : 각 라벨 별 테스트 데이터의 수입니다.

결과 성능 지표는 모델이 학습을 완료한 후, 테스트 데이터를 활용해 성능 점검을 진행한 결과를 나타냅니다.

1. 라벨 번호 : 라벨값을 대신하는 번호입니다.
2. 라벨 : 해당 그룹의 유일 값을 나타냅니다.
3. 정답 : 테스트 데이터를 학습된 모델이 예측하여 해당 라벨을 맞춘 수입니다.
4. 전체 : 각각의 라벨별 테스트 데이터의 수입니다.



## 분류 모델

타이타닉, 동물종 → 서포터벡터머신 / 로지스틱 / 의사결정나무 / 신경망 선택 시



시각화 자료

실제값	0	1	2	3	4	5	6
0	1	1	0	0	0	0	0
1	0	3	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0	3	0	1	0
4	0	0	0	0	6	0	0
5	0	0	0	1	0	1	0
6	0	0	0	0	0	0	12
	0	1	2	3	4	5	6

예측값

결과 성능 지표

1 라벨번호	라벨	2 정답	전체
0	근충류	1	2
1	무척추동물	3	3
2	양서류	0	1
3	어류	3	4
4	조류	6	6
5	파충류	1	2
6	포유류	12	12
합계		26	30

- 1 라벨번호 : 라벨 값을 대신하는 번호입니다.
- 라벨 : 해당 그룹의 유일 값을 나타냅니다.
- 정답 : 테스트 데이터를 학습된 모델이 예측하여 해당 라벨을 맞춘 수입니다.
- 전체 : 각 라벨 별 테스트 데이터의 수입니다.

1. [라벨 번호]는 결과 데이터(동물종)의 칼럼의 고유 번호를 의미합니다.

2. [라벨]은 결과 데이터(동물종)의 칼럼의 유일값을 나타냅니다.

( \* 유일값 : 중복된 값들을 제외한 값으로, 특정 값이 여러 번 나오더라도 한 번만 기입합니다. )

\* 이때 결과 데이터는 [STEP 2 데이터 선택] 단계에서 선택한 결과 데이터를 기준으로 값이 나타납니다.

\* '라벨 번호'와 '시각화 자료의 라벨 번호'는 일치합니다.





## 분류 모델

타이타닉, 동물종 → 서포터벡터머신 / 로지스틱 / 의사결정나무 / 신경망 선택 시



시각화 자료

실제값	0	1	2	3	4	5	6
0	1	1	0	0	0	0	0
1	0	3	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0	3	0	1	0
4	0	0	0	0	6	0	0
5	0	0	0	1	0	1	0
6	0	0	0	0	0	0	12
	0	1	2	3	4	5	6

예측값

결과 성능 지표

라벨번호	라벨	정답	전체
0	곤충류	1	2
1	무척추동물	3	3
2	양서류	0	1
3	어류	3	4
4	조류	6	6
5	파충류	1	2
6	포유류	12	12
합계		26	30

- ① 라벨번호 : 라벨 값을 대신하는 번호입니다.
- 라벨 : 해당 그룹의 유일 값을 나타냅니다.
- 정답 : 테스트 데이터를 학습된 모델이 예측하여 해당 라벨을 맞춘 수입니다.
- 전체 : 각 라벨 별 테스트 데이터의 수입니다.

1. [전체]는 앞서 선택한 데이터셋에서 테스트 데이터의 전체 수를 의미합니다.
2. [정답]은 예측에 대한 통계값(결괏값)으로, [전체]에서 맞춘 정답의 수를 의미합니다.  
학습된 모델이 테스트 데이터를 예측하여 라벨을 맞춘 수를 나타냅니다.

\* 동물종 테스트 데이터값 중에서 결과적으로 곤충류가 몇 개 등장했는지, 양서류가 몇 개 등장했는지에 대한 실제 값입니다.



여기서 잠깐!

## 학습 결과 데이터는 어떻게 읽을까요?

시각화 자료



결과 성능 지표

라벨번호	라벨	정답	전체
0	곤충류	1	2
1	무척추동물	3	3
2	양서류	0	1
3	어류	3	4
4	조류	6	6
5	파충류	1	2
6	포유류	12	12
합계		26	30

3번 : 어류

- ① 라벨번호 : 라벨 값을 대신하는 번호입니다.
- 라벨 : 해당 그룹의 유일 값을 나타냅니다.
- 정답 : 테스트 데이터를 학습된 모델이 예측하여 해당 라벨을 맞춘 수입니다.
- 전체 : 각 라벨 별 테스트 데이터의 수입니다.

① [결과 성능 지표]에서 값을 확인한 후 → ② [시각화 자료]를 확인해 주세요.

- [결과 성능 지표]에서 라벨 번호 3번 : 어류 확인 시, 전체 : 4개 / 정답 : 3개 / 오답 : 1개
- [시각화 자료]에서 3번의 x, y축을 확인 시, 겹치는 값(정답) : 3개 / 겹치지 않는 값(오답) : 1개



## 분류 모델

타이타닉, 동물종 → 서포터벡터머신 / 로지스틱 / 의사결정나무 / 신경망 선택 시



“ 원하는 값으로 학습된 모델이 어떻게 결과를 예측하는지 테스트할 수 있어요. ”

모델 테스트

그룹형

사용자 입력

털 <small>그룹형</small>	값을 선택하세요.	척추 <small>그룹형</small>	값을 선택하세요.
호흡 <small>그룹형</small>	값을 선택하세요.	지느러미 <small>그룹형</small>	값을 선택하세요.
꼬리 <small>그룹형</small>	값을 선택하세요.		

① 입력 데이터로 선택한 컬럼 중 모델 학습에 사용 된 컬럼만 표시 됩니다.

그룹형 데이터 중 하나를 선택하세요.

숫자형 최소~최대 값 사이의 실수 값을 입력하세요.

결과보기 ▶

다시하기

**사용자 입력**은 사용자가 입력 데이터로 선택한 컬럼으로, 모델 학습에 사용된 컬럼만 활성화됩니다.

1. 컬럼별로 테스트하고자 하는 값을 입력 혹은 선택해 주세요.  
데이터 형태에 따라 숫자형/그룹형으로 값을 입력하는 방식이 차이가 납니다.

- \* **[그룹형] 데이터** : 데이터셋의 라벨 중 하나를 선택해 입력합니다.
- \* **[숫자형] 데이터** : 데이터셋의 컬럼의 최솟값 ~ 최댓값 사이의 실숫값을 입력합니다.



## 분류 모델

타이타닉, 동물종 → 서포터벡터머신 / 로지스틱 / 의사결정나무 / 신경망 선택 시



꼬리 그림형

① 입력 데이터로 선택한 컬럼 중 모델 학습에 사용 된 컬럼만 표시 됩니다.  
그림형 데이터 중 하나를 선택하세요.  
숫자형 최소~최대 값 사이의 실수 값을 입력하세요.

1 결과보기 ▶ 2 다시하기

3 테스트 결과  
 인공지능 모델이 **동물종**을(를) **어류**로 예측하였습니다.

4 ✔ 완료하기

1. 테스트하고자 하는 값을 입력 혹은 선택한 후, **[결과보기]** 버튼을 눌러주세요.  
아래의 '테스트 결과'를 통해 해당 값을 통해 모델이 예측한 값을 확인할 수 있습니다.
2. **[다시 하기]** 버튼을 클릭하면, 입력한 값을 모두 초기화할 수 있습니다.
3. 사용자가 입력한 값을 통해 인공지능 모델이 예측한 값을 확인할 수 있습니다.
4. **[완료하기]** 버튼을 클릭하면, 모델 만들기가 완료됩니다.



여기서 잠깐!

## 모델 학습 결과가 잘 맞지 않는다면 확인해 보세요!

### 1 최적의 기계학습 기법을 찾아보세요!

다양한 기법들을 적용해 보며, 내 데이터에 더 적합한 기법을 살펴보세요.

(\* 분류의 경우, [서포터벡터머신 / 로지스틱 / 의사결정나무 / 신경망] 4가지 중 택1)

### 2 데이터 칼럼 수를 추가해 보세요!

[STEP 2 데이터 선택]에서 입력 데이터로 선택한 칼럼 수가 부족할 수 있습니다.





## 회귀 모델

위라벨 점수, 인도 중고차 가격 → 선형 / 의사결정나무 / 신경망 선택 시



### 학습 결과

1

테스트 입력 데이터셋

	운전거리	구동방식	마력	좌석수
1	20483	자동	136	5
2	63000	수동	73.9	5
3	62500	자동	174.33	5
4	40496	수동	78.9	5
5	41125	수동	62.1	5
6	120000	수동	76.8	5
7	64806	수동	73.9	5
8	54758	수동	73.75	5
9	89326	수동	68	5
10	68000	수동	102	7
11	27300	수동	117.3	5
12	90000	수동	68	5
13	20604	자동	167.62	5

결과 비교

	중고차가격	예측값	오차값
1	30.3	16.3635	13.9365
2	4.2	2.5616	1.6384
3	19	19.2293	-0.2293
4	2.65	4.4433	-1.7933
5	2.7	2.1611	0.5389
6	2.5	-0.1216	2.6216
7	3.65	2.4643	1.1857
8	6.67	2.9856	3.6844
9	1.55	0.3535	1.1965
10	6.95	8.284	-1.334
11	7	10.2934	-3.2934
12	2.6	0.3172	2.2828
13	39.62	20.5887	19.0313

① 선택한 컬럼들로 구성된 입력 데이터셋을 학습과 테스트의 비율을 7 : 3으로 데이터셋은 분리하여 학습에 사용되지 않은 테스트 데이터셋의 입력 데이터 컬럼을 나타낸 데이터 테이블입니다.

① 테스트 입력 데이터셋의 실제 결과 데이터와 인공지능 모델이 예측 값을 비교하는 테이블입니다.

1. 테스트 입력 데이터셋은 이전 단계에서 선택한 입력 데이터셋의 테스트를 위한 값입니다.

\* 입력 데이터셋은 7 : 3 비율로 모델 학습과 테스트에 활용되며, 랜덤으로 분리하여 사용됩니다.

입력 데이터셋	70%	'모델 학습'에 사용
	30%	'테스트 데이터'로 사용 (완성된 모델을 테스트하는 용도)



## 회귀 모델

위라벨 점수, 인도 중고차 가격 → 선형 / 의사결정나무 / 신경망 선택 시



### 학습 결과

테스트 입력 데이터셋

	운전거리	구동방식	마력	좌석수
1	20483	자동	136	5
2	63000	수동	73.9	5
3	62500	자동	174.33	5
4	40496	수동	78.9	5
5	41125	수동	62.1	5
6	120000	수동	76.8	5
7	64806	수동	73.9	5
8	54758	수동	73.75	5
9	89326	수동	68	5
10	68000	수동	102	7
11	27300	수동	117.3	5
12	90000	수동	68	5
13	20604	자동	167.62	5

결과 비교

	1	2	3
	중고차가격	예측값	오차값
1	30.3	16.3635	13.9365
2	4.2	2.5616	1.6384
3	19	19.2293	-0.2293
4	2.65	4.4433	-1.7933
5	2.7	2.1611	0.5389
6	2.5	-0.1216	2.6216
7	3.65	2.4643	1.1857
8	6.67	2.9856	3.6844
9	1.55	0.3535	1.1965
10	6.95	8.284	-1.334
11	7	10.2934	-3.2934
12	2.6	0.3172	2.2828
13	39.62	20.5887	19.0313

**결과 비교**는 이전 단계에서 선택한 결과 데이터셋과 인공지능의 예측값, 오차값을 비교하는 테이블입니다.

1. [중고차 자격]은 이전 단계(STEP 2)에서 내가 선택한 결과 데이터셋 칼럼입니다.
2. [예측값]은 테스트 입력 데이터셋을 기준으로 인공지능이 예측한 값을 나타냅니다.
3. [오차값]은 결과 데이터의 오차(실제값 - 예측값)를 나타냅니다.

\* 회귀 모델의 결과 데이터값은 문자의 그룹 형태가 아닌, '숫자' 형태를 씁니다. (ex : 30.3, 16.3635 등)



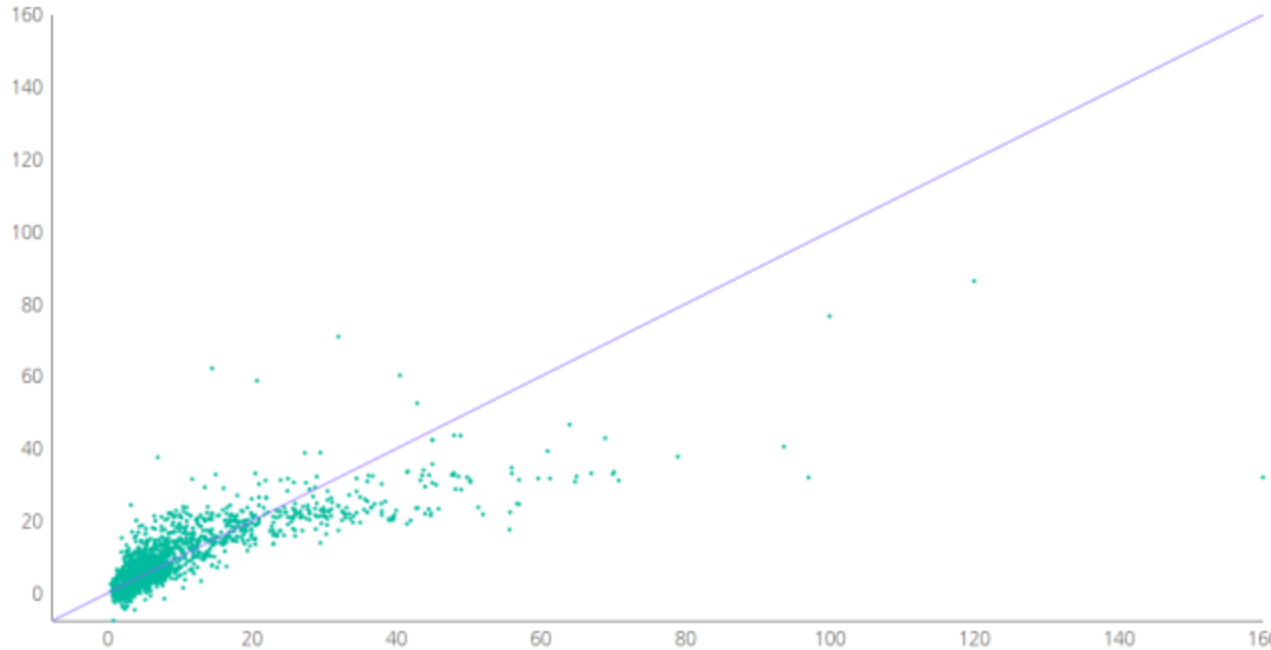
## 회귀 모델

위라벨 점수, 인도 중고차 가격 → 선형 / 의사결정나무 / 신경망 선택 시



1

시각화 자료



① 테스트 데이터셋의 결과 실제 값과 모델이 예측한 예측 값을 2차원 공간에 표현 한 것입니다. 모델이 잘 예측 할 수록 초록색 점이 보라색 직선( $y=x$ ) 가까이에 분포되며, 어떤 구간에서 예측을 잘 하고 못하였는지 확인 할 수 있습니다.

결과 성능 지표

	값
평균 절대 오차	4.088
평균 백분율 오차 (%)	57.9787
평균 제곱 오차	55.0735
평균 제곱근 오차	7.4212

① **평균 절대 오차**: 오차(실제값-예측값)의 절대값을 모두 더하여 데이터의 수로 나눈 값. 절대값을 취하기 때문에 가장 직관적으로 오차를 알 수 있는 지표입니다.

**평균 백분율 오차**: 오차의 백분율을 더하여 데이터의 수로 나누고 다시 100을 곱하여 백분율로 나타낸 값. 오차가 어느정도 차이나는지 퍼센트로 알 수 있습니다.

**평균 제곱 오차**: 오차의 제곱들의 합을 데이터수로 나눈 값. 직관적으로 어느 정도 오차가 발생하는지 확인 하기 어렵고, 이상치가 있을 경우 값이 매우 커질 수 있습니다.

**평균 제곱근 오차**: 평균제곱오차에 루트를 씌운 값. 이 방법을 통해 실제값과 유사한 단위 값으로 확인 가능하며, 이상치가 있을 경우 값이 매우 커질 수 있습니다.

1. **시각화 자료**는 테스트 데이터 셋을 실제 값과 모델이 예측한 값을 2차원 공간에 표현한 것입니다. 모델이 잘 예측할수록 초록색 점이 보라색 직선( $y=x$ ) 가까이에 분포됩니다.

\* 초록색 점 분포를 통해 어떤 구간에서 예측을 잘하고 못하였는지 확인해 보세요.



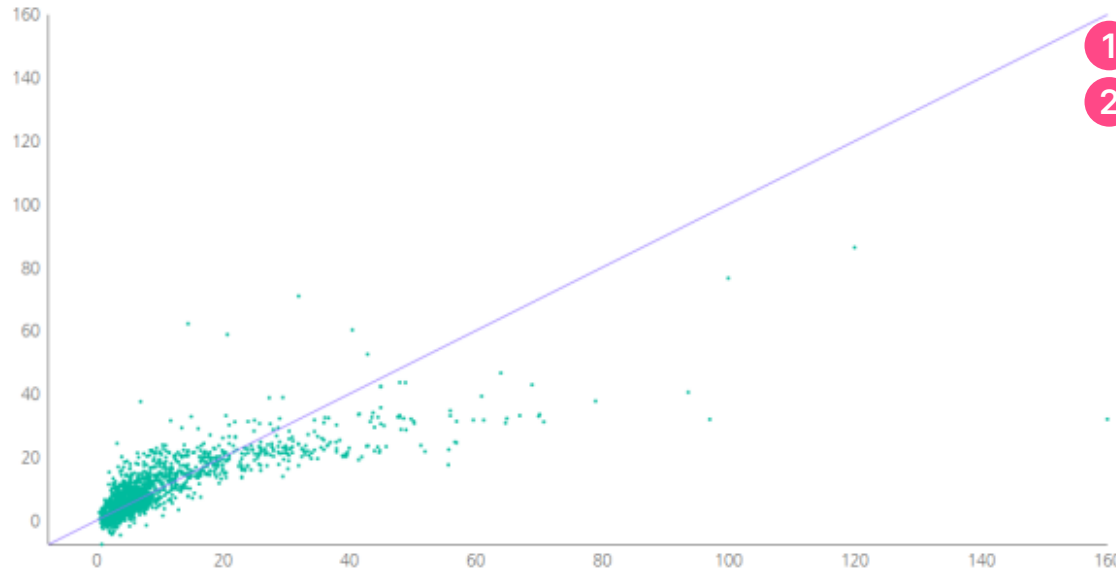


## 회귀 모델

위라벨 점수, 인도 중고차 가격 → 선형 / 의사결정나무 / 신경망 선택 시



시각화 자료



① 테스트 데이터셋의 결과 실제 값과 모델이 예측한 예측 값을 2차원 공간에 표현 한 것입니다. 모델이 잘 예측 할 수록 초록색 점이 보라색 직선( $y=x$ ) 가까이에 분포되며, 어떤 구간에서 예측을 잘 하고 못하였는지 확인 할 수 있습니다.

결과 성능 지표

	값
1 평균 절대 오차	4.088
2 평균 백분율 오차 (%)	57.9787
평균 제곱 오차	55.0735
평균 제곱근 오차	7.4212

① **평균 절대 오차**: 오차(실제값-예측값)의 절대값을 모두 더하여 데이터의 수로 나눈 값. 절대값을 취하기 때문에 가장 직관적으로 오차를 알 수 있는 지표입니다.

**평균 백분율 오차**: 오차의 백분율을 더하여 데이터의 수로 나누고 다시 100을 곱하여 백분율로 나타낸 값. 오차가 어느정도 차이 나는지 퍼센트로 알 수 있습니다.

**평균 제곱 오차**: 오차의 제곱들의 합을 데이터수로 나눈 값. 직관적으로 어느 정도 오차가 발생 하는지 확인 하기 어렵고, 이상치가 있을 경우 값이 매우 커질 수 있습니다.

**평균 제곱근 오차**: 평균제곱오차에 루트를 씌운 값. 이 방법을 통해 실제값과 유사한 단위 값으로 확인 가능하며, 이상치가 있을 경우 값이 매우 커질 수 있습니다.

결과 성능 지표는 데이터값의 오차(실제값-예측값)를 여러 결과 평가 지표로 표현한 결과를 나타냅니다.

1. **평균 절대 오차**: 오차의 절댓값을 모두 더하여 데이터의 수로 나눈 값입니다.

절댓값을 취하기 때문에 가장 직관적으로 오차를 알 수 있는 지표입니다.

2. **평균 백분율 오차(%)**: 오차의 백분율을 더하여 데이터의 수로 나누고 다시 100을 곱한 값입니다.

오차가 어느 정도 차이 나는지 퍼센트로 알 수 있습니다.

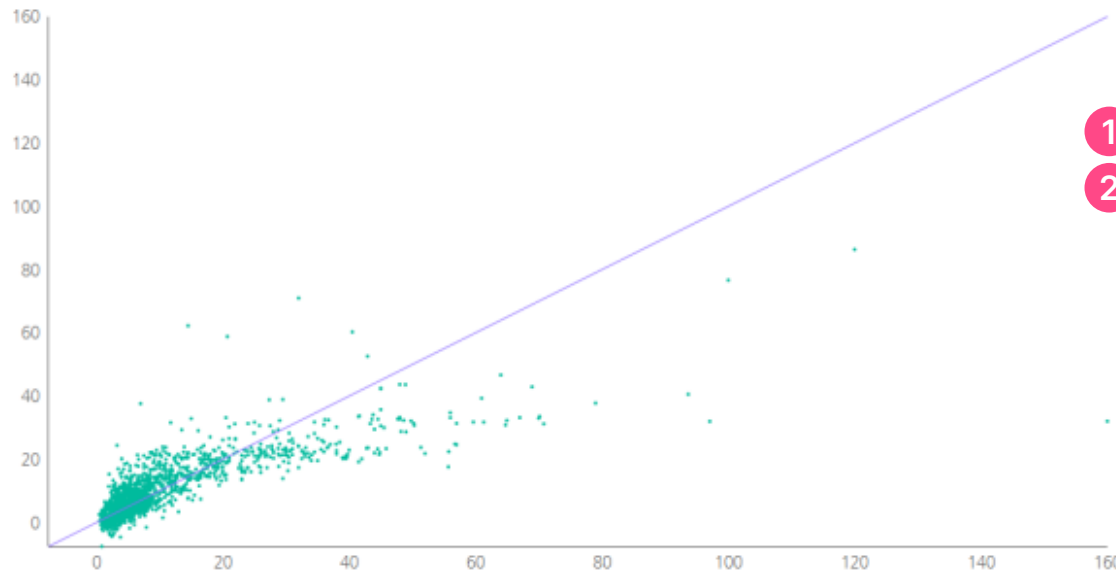


## 회귀 모델

위라벨 점수, 인도 중고차 가격 → 선형 / 의사결정나무 / 신경망 선택 시



시각화 자료



① 테스트 데이터셋의 결과 실제 값과 모델이 예측한 예측 값을 2차원 공간에 표현 한 것입니다. 모델이 잘 예측 할 수록 초록색 점이 보라색 직선( $y=x$ ) 가까이에 분포되며, 어떤 구간에서 예측을 잘 하고 못하였는지 확인 할 수 있습니다.

결과 성능 지표

	값
평균 절대 오차	4.088
평균 백분율 오차 (%)	57.9787
1 평균 제곱 오차	55.0735
2 평균 제곱근 오차	7.4212

① **평균 절대 오차**: 오차(실제값-예측값)의 절대값을 모두 더하여 데이터의 수로 나눈 값. 절대값을 취하기 때문에 가장 직관적으로 오차를 알 수 있는 지표입니다.

**평균 백분율 오차**: 오차의 백분율을 더하여 데이터의 수로 나누고 다시 100을 곱하여 백분율로 나타낸 값. 오차가 어느정도 차이 나는지 퍼센트로 알 수 있습니다.

**평균 제곱 오차**: 오차의 제곱들의 합을 데이터수로 나눈 값. 직관적으로 어느 정도 오차가 발생하는지 확인 하기 어렵고, 이상치가 있을 경우 값이 매우 커질 수 있습니다.

**평균 제곱근 오차**: 평균제곱오차에 루트를 씌운 값. 이 방법을 통해 실제값과 유사한 단위로 확인 가능하며, 이상치가 있을 경우 값이 매우 커질 수 있습니다.

1. **평균 제곱 오차**: 오차의 제곱들의 합을 데이터 수로 나눈 값입니다.

- 직관적으로 어느 정도 오차가 발생하는지 확인 하기 어렵고, 이상치가 있으면 값이 매우 커질 수 있습니다

2. **평균 제곱근 오차**: 평균제곱오차에 루트를 씌운 값입니다.

- 이 방법을 통해 실제값과 유사한 단위로 확인 가능하며, 이상치가 있으면 값이 매우 커질 수 있습니다.



## 회귀 모델

위라벨 점수, 인도 중고차 가격 → 선형 / 의사결정나무 / 신경망 선택 시



“ 원하는 값으로 학습된 모델이 어떻게 결과를 예측하는지 테스트할 수 있어요. ”

모델 테스트  
사용자 입력

숫자형

1

그룹형

1

운전거리 숫자형

171 ~ 300000 사이의 값을 입력하시

구동방식 그룹형

값을 선택하세요.

마력 숫자형

34.2 ~ 560 사이의 값을 입력하세요

좌석수 그룹형

값을 선택하세요.

① 입력 데이터로 선택한 컬럼 중 모델 학습에 사용된 컬럼만 표시 됩니다.

그룹형 데이터 중 하나를 선택하세요.

숫자형 최소~최대 값 사이의 실수 값을 입력하세요.

결과보기 ▶

다시하기

**사용자 입력**은 사용자가 입력 데이터로 선택한 컬럼으로, 모델 학습에 사용된 컬럼만 활성화됩니다.

1. 컬럼별로 테스트하고자 하는 값을 입력 혹은 선택해 주세요.  
데이터 형태에 따라 숫자형/그룹형으로 값을 입력하는 방식이 차이가 납니다.

- \* **[그룹형] 데이터** : 데이터셋의 라벨 중 하나를 선택해 입력합니다.
- \* **[숫자형] 데이터** : 데이터셋의 컬럼의 최솟값 ~ 최댓값 사이의 실숫값을 입력합니다.



## 회귀 모델

위라벨 점수, 인도 중고차 가격 → 선형 / 의사결정나무 / 신경망 선택 시



마력 숫자형

좌석수 그룹형

① 입력 데이터로 선택한 컬럼 중 모델 학습에 사용된 컬럼만 표시 됩니다.

그룹형 데이터 중 하나를 선택하세요.

숫자형 최소~최대 값 사이의 실수 값을 입력하세요.

1

2

결과보기 ▶

다시하기

3 테스트 결과

인공지능 모델이 **중고차가격을(를) 40.3**로 예측하였습니다.

4

✔ 완료하기

1. 테스트하고자 하는 값을 입력 혹은 선택한 후, **[결과보기]** 버튼을 눌러주세요.  
아래의 '테스트 결과'를 통해 해당 값을 통해 모델이 예측한 값을 확인할 수 있습니다.
2. **[다시 하기]** 버튼을 클릭하면, 입력한 값을 모두 초기화할 수 있습니다.
3. 사용자가 입력한 값을 통해 인공지능 모델이 예측한 값을 확인할 수 있습니다.
4. **[완료하기]** 버튼을 클릭하면, 모델 만들기가 완료됩니다.



## [포스터 만들기] 단계로 넘어가기 전, 꼭 확인해 보세요!

### 놓치면 안 되는 꿀팁은?

#### [STEP 3] 모델 학습 :

'분류 / 회귀' 모델 유형에 따라 사용할 기계학습 기법을 선택하여,  
데이터를 모델에 학습시키는 단계입니다.

학습 결과는 2~3단계에 따라 표현 방식 및 성능이 다르게 나타날 수 있습니다.

다양하게 학습을 진행해 본 후 **가장 최적의 모델을 만들어 보세요!**



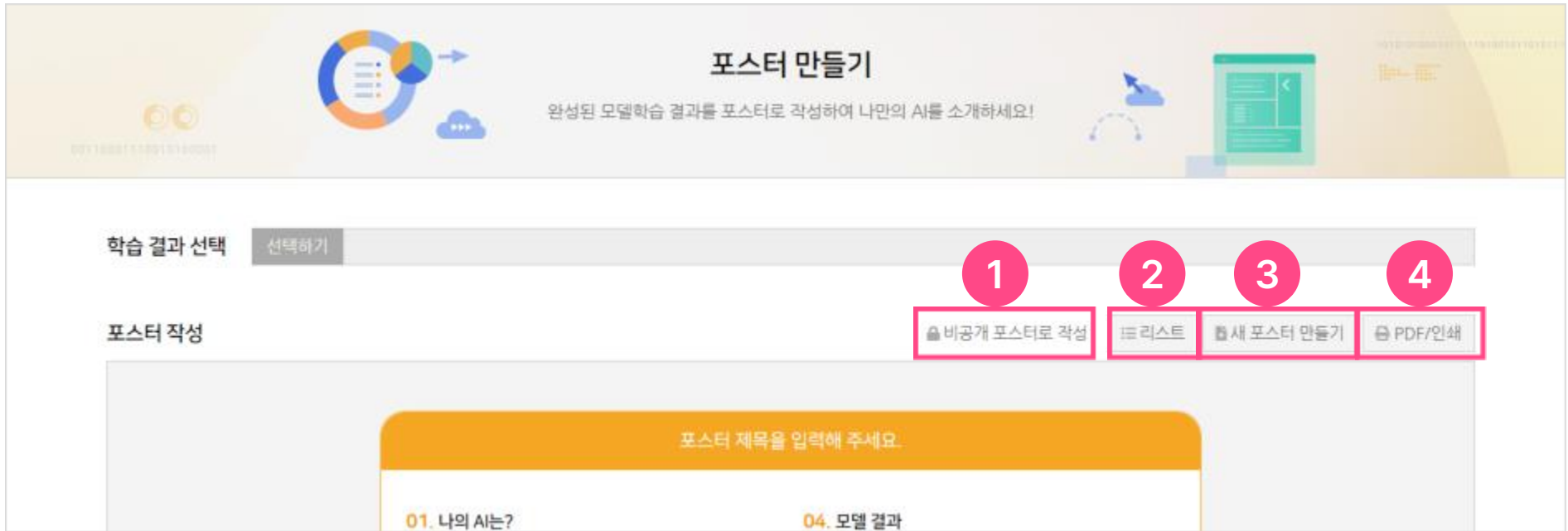
# AI CAN DO.

정형 데이터셋을 활용한 AI 서비스

01	Intro. 인트로	03
02	내 리스트	06
03	[STEP 1] 데이터셋 선정	14
04	[STEP 2] 데이터 선택	32
05	[STEP 3] 모델 학습	38
06	포스터 만들기	62

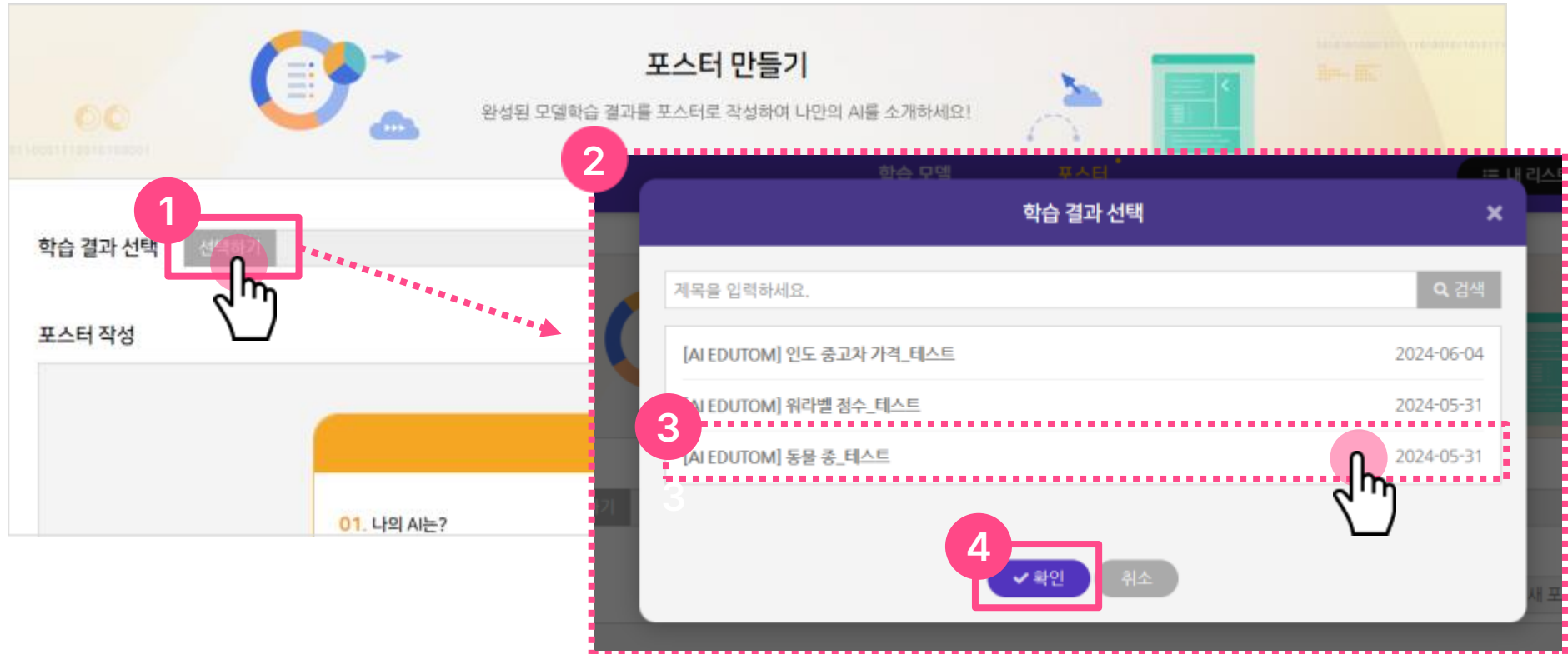


1. 서비스 메인페이지에서 [포스터 만들기] 버튼을 클릭해 다음 페이지로 이동해 주세요.
2. 또는, 상단의 [포스터 → 포스터 만들기] 카테고리를 통해 이동할 수도 있습니다.



1. **비공개 포스터로 작성** : 작성하는 포스터를 비공개로 작성할 수 있습니다. 공개 여부를 선택해 주세요.
2. **리스트** : 타인의 포스터를 포함한 전체 포스터 리스트를 확인할 수 있습니다.
3. **새 포스터 만들기** : 새로운 포스터를 만들 수 있습니다.
4. **PDF / 인쇄** : 현재 작성 중인 포스터를 PDF로 저장하거나, 인쇄할 수 있습니다.





1. 학습 결과 선택' 우측의 [선택하기] 버튼을 눌러주세요.
2. STEP 3 : 모델 학습 단계까지 완료된 학습 모델의 결과를 선택할 수 있는 창이 생성됩니다.
3. 활성화된 학습 모델 목록 중 하나를 선택해 주세요.
4. [확인] 버튼을 누르면, 선택한 학습 모델이 포스터에 적용됩니다.



포스터 제목을 입력해 주세요.

소속: 디아이엠씨 1학년 | 작성자: 9hj1225

### 01. 나의 시는?

내용을 입력하세요.

### 04. 모델 결과

• 시각화

6	1	1	0	0	0	0	0
5	0	3	0	0	0	0	0
4	0	0	1	0	0	0	0
3	0	0	0	4	0	0	0
2	0	0	0	0	0	0	0
1	0	0	0	0	2	0	0
0	0	0	0	0	0	0	12
	6	5	4	3	2	1	0

• 결과 성능 지표

라벨번호	라벨	경답	전체
0	곤충류	1	2
1	무척추동물	3	3
2	양서류	1	1
3	어류	4	4
4	조류	6	6
5	파충류	0	2
6	포유류	12	12
합계		27	30

### 02. 데이터셋

- 데이터명 : zoo.csv
- 데이터 정보 : 98행, 17열
- 입력 데이터 : 수증, 알, 모유, 비행, 털, 척추, 호흡, 지느러미, 꼬리
- 결과 데이터 : 동물종
- 라벨 정보 : 곤충류, 무척추동물, 양서류, 어류, 조류, 파충류, 포유류

### 03. 인공지능 모델

분류모델 / 서포터벡터마신

데이터 그룹의 분류를 위한 기준선을 표현하며 특히 이항형(참, 거짓) 분류모델로 적합합니다.

### 05. 느낀점 및 활용방안

내용을 입력하세요.

**AICANDO**

1. 앞에서 선택한 학습 모델이 포스터에 다음과 같이 적용됩니다.

The screenshot shows a poster creation interface for 'AI CAN DO'. It features several sections: a title input field at the top, a section for '01. 나의 AI는?' (My AI?), a data table for '04. 모델 결과' (Model Results), a section for '05. 느낀점 및 활용방안' (Thoughts and Usage Methods), and a section for '02. 데이터셋' (Dataset). Three callouts are present: 1. points to the title input field; 2. points to the '01. 나의 AI는?' section; 3. points to the '05. 느낀점 및 활용방안' section.

**1** 포스터 제목을 입력해 주세요.

**2** 01. 나의 AI는?  
내용을 입력하세요.

**3** 05. 느낀점 및 활용방안  
내용을 입력하세요.

**04. 모델 결과**

시각화

기간	6	1	1	0	0	0	0	0	0	0	0
5	0	3	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0
3	0	0	0	4	0	0	0	0	0	0	0
2	0	0	0	0	5	0	0	0	0	0	0
1	0	0	0	0	0	2	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	12	0
	6	5	4	3	2	1	0	예측값			

결과 성능 지표

라벨번호	라벨	경답	전체
0	곤솔류	1	2
1	무척추동물	3	3
2	양서류	1	1
3	어류	4	4
4	조류	6	6
5	파충류	0	2
6	포유류	12	12
합계		27	30

1. 포스터에 제목을 작성해 주세요.
  2. 나의 AI(인공지능) 모델에 대해 자유롭게 작성해 주세요.
  3. AI를 만들면서 느낀 점 및 해당 AI 모델을 어떻게 활용할지 자유롭게 작성해 주세요.
- \* 인공지능 모델 학습 및 제작의 모든 과정을 다시 생각해 보며 포스터를 작성해 보세요.

① 분류 모델일 경우

1 데이터셋

- 데이터명 : ZOO.CSV
  - 데이터 정보 : 98행, 17열
  - 입력 데이터 : 수중, 알, 모유, 비행, 털, 척추, 호흡, 지느러미, 꼬리
  - 결과 데이터 : 동물종
- 라벨 정보 :**  
곤충류, 무척추동물, 양서류, 어류, 조류, 파충류, 포유류

분류 모델

2 인공지능 모델

분류모델 / 서포터벡터머신

데이터 그룹의 분류를 위한 기준선을 표현하며 특히 이항형(참, 거짓) 분류모델로 적합합니다.

포스터 제목을 입력해 주세요.

소속 : 데이템씨 1학년 | 작성자 : 9기1225

01. 나의 시는?  
내용을 입력하세요.

02. 데이터셋

- 데이터명 : ZOO.CSV
- 데이터 정보 : 98행, 17열
- 입력 데이터 : 수중, 알, 모유, 비행, 털, 척추, 호흡, 지느러미, 꼬리
- 결과 데이터 : 동물종
- 라벨 정보 : 곤충류, 무척추동물, 양서류, 어류, 조류, 파충류, 포유류

03. 인공지능 모델

분류모델 / 서포터벡터머신

데이터 그룹의 분류를 위한 기준선을 표현하며 특히 이항형(참, 거짓) 분류모델로 적합합니다.

04. 모델 결과

시각화

6	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0
	6	5	4	3	2	1	0									

결과 성능 지표

라벨번호	라벨	정답	전체
0	곤충류	1	2
1	무척추동물	3	3
2	양서류	1	1
3	어류	4	4
4	조류	6	6
5	파충류	0	2
6	포유류	12	12
합계		27	30

05. 노란집 및 활용방안

내용을 입력하세요.

AI CAN DO

3

모델 결과

분류 모델



결과 성능 지표

라벨번호	라벨	정답	전체
0	곤충류	1	2
1	무척추동물	3	3
2	양서류	1	1
3	어류	4	4
4	조류	6	6
5	파충류	0	2
6	포유류	12	12
합계		27	30

1. 선택한 데이터셋과 데이터 칼럼에 대한 정보를 확인할 수 있습니다.  
( \* 분류 모델 : 칼럼의 라벨 정보 / 회귀 모델 : 칼럼의 최솟값, 최댓값, 평균값의 결과통계를 나타냅니다. )
2. 선택한 모델 유형과 기계학습 기법에 대한 정보를 확인할 수 있습니다.
3. 분류/회귀 모델에 따른 모델 결과(시각화 자료), 결과 성능지표를 확인할 수 있습니다.

## ② 회귀 모델일 경우

1

### 데이터셋

- 데이터명 : wal.csv
- 데이터 정보 : 5270행, 23열
- 입력 데이터 : 능력증명횟수, 하루\_과일채소섭취량, 연령, 주간\_화면횟수, 성별, 하루\_걸음수\_천단위
- 결과 데이터 : **회귀 모델**   
 위라발점수
- 결과 통계 :   
 최소값 : 480 / 최대값 : 804.3 / 평균 : 667.63

2

### 인공지능 모델

분류모델 / 선형

입력변수와 결과변수가 선형 상관관계를 가지는 모델로 연속형 숫자 데이터를 예측하는 회귀모델로 적합합니다.



3

### 모델 결과



1. 선택한 데이터셋과 데이터 칼럼에 대한 정보를 확인할 수 있습니다.  
( \* 분류 모델 : 칼럼의 라벨 정보 / 회귀 모델 : 칼럼의 최소값, 최대값, 평균값의 결과통계를 나타냅니다. )
2. 선택한 모델 유형과 기계학습 기법에 대한 정보를 확인할 수 있습니다.
3. 분류/회귀 모델에 따른 모델 결과(시각화 자료), 결과 성능지표를 확인할 수 있습니다.



1. 나의 인공지능 모델을 소개하는 포스터를 PDF로 다운받거나 인쇄할 수 있습니다.  
[PDF / 인쇄] 버튼을 눌러주세요.

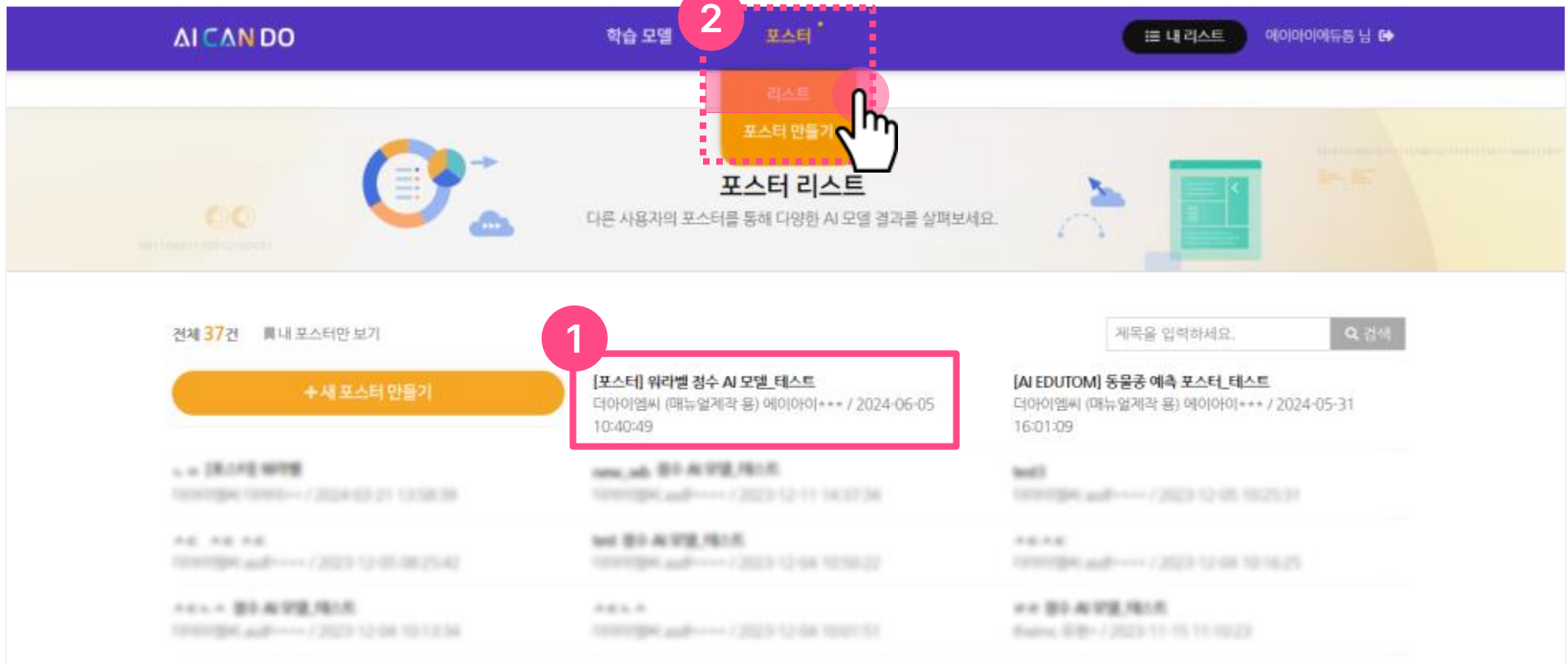


1. 포스터의 내용을 미리 확인해 볼 수 있습니다.
2. [인쇄] 버튼을 눌러 해당 포스터를 인쇄하거나, [저장] 버튼을 눌러 해당 포스터를 PDF 파일 형식으로 다운받을 수 있습니다.



1. [저장하기] 버튼을 누르면, 지금까지 작성한 포스터가 저장됩니다.  
저장한 포스터는 '포스터 리스트'에서 확인할 수 있습니다.





포스터 리스트는 나뿐만 아니라 다른 사용자들이 만든 포스터를 확인할 수 있는 단계입니다.

1. 이전 단계에서 [저장하기] 버튼을 눌러 저장한 포스터를 확인할 수 있습니다.
2. '포스터 리스트'는 상단의 [포스터 → 리스트]를 통해 이동할 수도 있습니다.



AI CAN DO      학습 모델      포스터      ≡ 내 리스트      에이아이에듀툼 님

## 포스터 리스트

다른 사용자의 포스터를 통해 다양한 AI 모델 결과를 살펴보세요.

전체 2건    **내 포스터만 보기**

**+ 새 만들기**

[포스터] 워라벨 점수 AI 모델 테스트 더아이엠씨 (매뉴얼제작 용) 에이아이에듀툼 / 2024-06-05 10:40:49	[AI EDUTOM] 동물종 예측 포스터 테스트 더아이엠씨 (매뉴얼제작 용) 에이아이에듀툼 / 2024-05-31 16:01:09
--	--

제목을 입력하세요.    🔍 검색

1. [내 포스터만 보기] 버튼을 누르면, 내가 만든 포스터만 확인할 수 있습니다.
2. 포스터의 제목 일부를 입력하여 [검색] 버튼을 누르면, 해당 포스터를 검색할 수 있습니다.

# 감사합니다. ●

**이메일** [aiedutom.theimc.co.kr](mailto:aiedutom.theimc.co.kr)

**주소** 대구광역시 수성구 알파시티1로35길 17

**홈페이지** [aiedutom.co.kr](http://aiedutom.co.kr)

