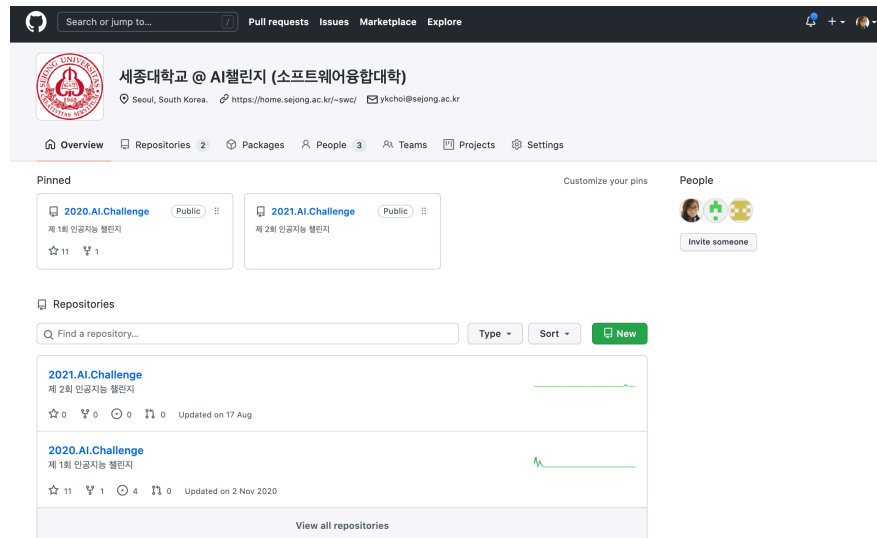


데이터 분석 실습

오렌지3를 통한 AI 챌린지 도전기

AI 챌린지 란?

- 인공지능을 활용하여 주어진 문제를 해결하는 대회
 - 학습 기반 방법론을 활용한 문제 해결 능력을 겨루는 대회
 - 정량적 평가 지표를 활용한 순위 결정 전 (매년 10월 말)
 - 캐글(Kaggle) 이라는 플랫폼을 활용
- 세종 AI 챌린지 운영 깃허브
 - <https://github.com/SejongAI-Challenge>



제1회 AI 챌린지 문제 1번

- 패션 의류 예측 문제

- 문제: <https://www.kaggle.com/t/0621a92d547f4f39adabf258c6973870>

- 데이터 셋 항목

- 본 문제는 패션 의류 10종에 대한 흑백 영상 데이터를 사용한다. 학습 및 테스트 데이터로 28x28 크기의 2d 영상 데이터를 1d 로 가공하여 제공하며, 종류는 0 부터 9 로 라벨링되어 있다. 최종적으로 본 문제는 주어진 데이터를 기반으로 10종에 대한 패션 의류의 종류를 예측하는 것을 목표로 한다.

- 평가 지표

- 28x28 흑백 영상을 기반으로 패션 의류의 종류를 예측하고, Category Accuracy 를 활용하여 예측 모델의 정확도를 측정한다.

제1회 AI 챌린지 문제 1번

- 데이터 파일 설명

- train.csv : 학습 데이터
- test.csv : 테스트 데이터
- submission_sample.csv : 결과 제출 템플릿

- 데이터 설명

- id : 데이터 순번
- pixel 0~784 : 흑백 영상 밝기 값 (0~255로 채워져 있음)
- label : 의류 종류 (0~9로 채워져 있음)

제1회 AI 챌린지 문제 1번: 분류 문제

- 데이터 준비

- 캐글 챌린지 >> Data 탭 이동 >> “Download All” 버튼 클릭 >> 압축 풀기

InClass Prediction Competition

Sejong AI Challenge 문제1

세종대학교 인공지능 챌린지 문제1

a year ago

Overview **Data** Code Discussion Leaderboard Rules Team Host My Submissions **Late Submission** ...

Summary

- 3 files
- 1573 columns

Download All

+ New Version

76	1
79	1
102	1
116	1
119	1
120	1
124	1
126	1
133	1

sample_submission.csv

test.csv

train.csv

리더보드 제출 포맷

테스트 데이터

학습 데이터

제1회 AI 챌린지 문제 1번: 분류 문제

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

■ 학습 데이터 열기

- ① Data 탭 클릭 >> ② File 드래그&드랍 >> ③ File 위젯 더블클릭
- ④ File 경로 선택 >> ⑤ 데이터 Role 변경 >> ⑥ Apply 버튼 클릭 >> ⑦ Dialog 닫기

① Data 탭 클릭 >> ② File 드래그&드랍 >> ③ File 위젯 더블클릭

④ File 경로 선택 >> ⑤ 데이터 Role 변경 >> ⑥ Apply 버튼 클릭 >> ⑦ Dialog 닫기

sample_submission.csv 리더보드 제출 포맷
test.csv 테스트 데이터
train.csv 학습 데이터

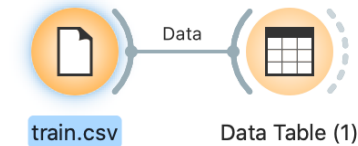
Name	Type	Role	Values
id	N numeric	skip	
pixel780	N numeric	feature	
pixel781	N numeric	feature	
pixel782	N numeric	feature	
pixel783	N numeric	feature	
pixel784	N numeric	feature	
label	C categorical	target	0, 1, 3, 4, 5, 6, 7, 8, 9, NAN, missing

제1회 AI 챌린지 문제 1번: 분류 문제

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

■ 학습 데이터 전처리

- ① Data Table 위젯 드래그&드랍
- ② File 위젯에 연결하여 데이터 이상 유무 확인



Info

33840 instances (no missing data)
784 features
Target with 11 values
No meta attributes

Variables

- ☒ Show variable labels (if present)
- ☐ Visualize numeric values
- ☒ Color by instance classes

Selection

- ☒ Select full rows

Restore Original Order

☒ Send Automatically

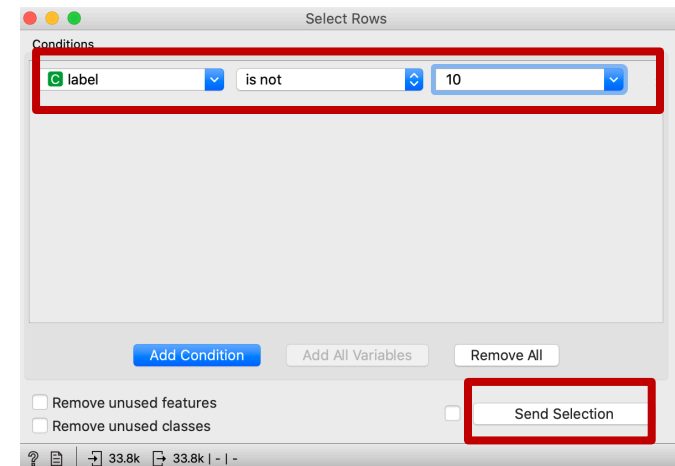
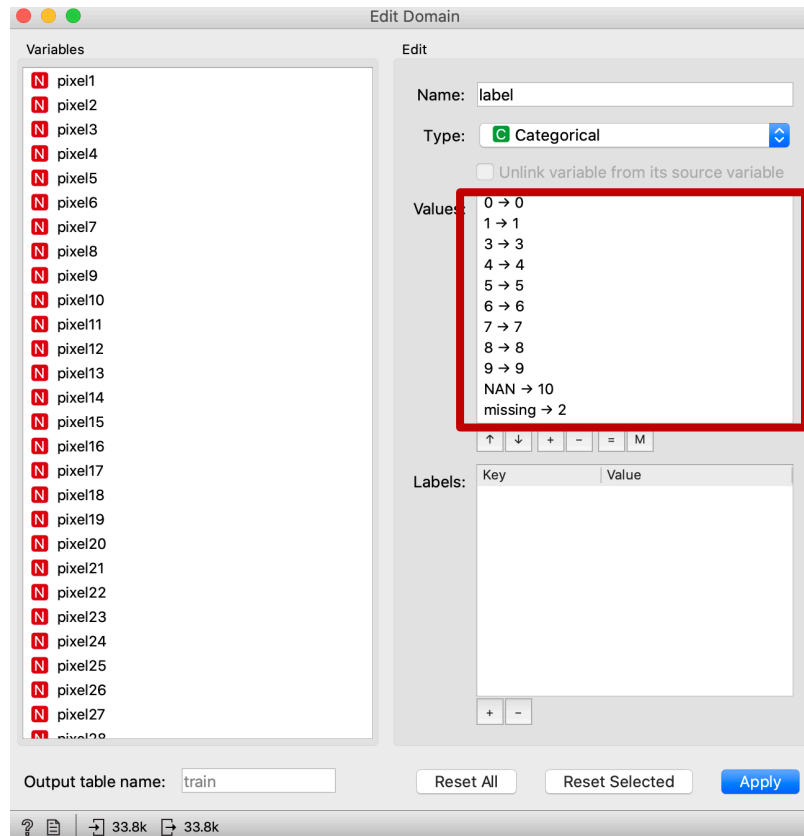
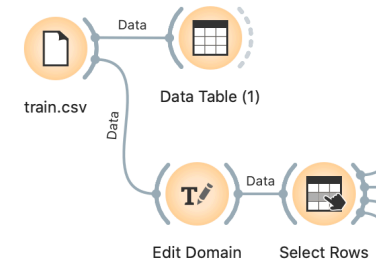
	label	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	pixel9	pixel10
1	missing	0	0	0	0	0	0	0	0	0	0
2	9	0	0	0	0	0	0	0	0	0	0
3	6	0	0	0	0	0	0	0	5	0	0
4	NAN	0	0	0	0	0	0	0	0	0	0
5	3	0	0	0	0	0	0	0	0	0	0
6	4	0	0	0	5	4	5	5	3	5	6
7	4	0	0	0	0	0	0	0	0	0	0
8	5	0	0	0	0	0	0	0	0	0	0
9	4	0	0	0	0	0	0	3	2	0	0
10	8	0	0	0	0	0	0	0	0	0	0
11	NAN	0	0	0	0	1	0	0	0	0	41
12	9	0	0	0	0	0	0	0	0	0	0
13	NAN	0	0	0	0	0	0	0	0	0	0
14	missing	0	0	0	0	1	1	0	0	0	0
15	9	0	0	0	0	0	0	0	0	0	0
16	3	0	0	0	0	0	0	0	0	0	38
17	3	0	0	0	0	0	0	0	0	0	0
18	8	0	0	0	0	0	0	0	0	0	0
19	4	0	0	0	0	0	1	0	1	0	0
20	4	0	0	0	0	2	0	0	0	1	0
21	4	0	0	0	0	0	0	0	1	0	0
22	8	0	0	0	0	0	0	0	1	0	0
23	7	0	0	0	0	0	0	0	0	0	0
24	1	0	0	0	0	0	0	0	0	0	171
25	5	0	0	0	0	0	0	0	0	0	0
26	5	0	0	0	0	0	0	0	0	0	0
27	3	0	0	0	0	0	0	0	0	0	0
28	missing	0	0	0	0	0	0	0	0	0	0
29	7	0	0	0	0	0	0	0	0	0	0
30	4	0	0	0	0	0	0	0	0	0	13
31	missing	0	0	0	0	0	0	0	0	0	0
32	6	0	0	0	0	0	0	0	0	0	2
33	9	0	0	0	0	0	0	0	0	0	0
34	6	0	0	0	0	0	0	0	0	0	4
35	NAN	0	0	0	0	0	1	1	0	0	0

제1회 AI 챌린지 문제 1번: 분류 문제

■ 학습 데이터 전처리

- ③ Edit Domain 위젯 사용 - 노이즈 데이터 편집
- ④ Select Rows 위젯 사용 - 10번 (NaN) 클래스 삭제

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

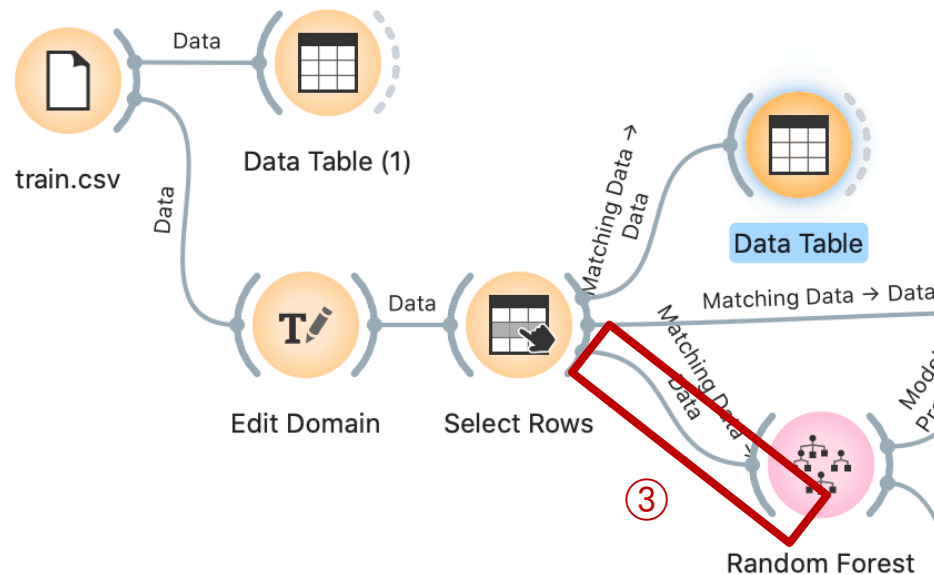
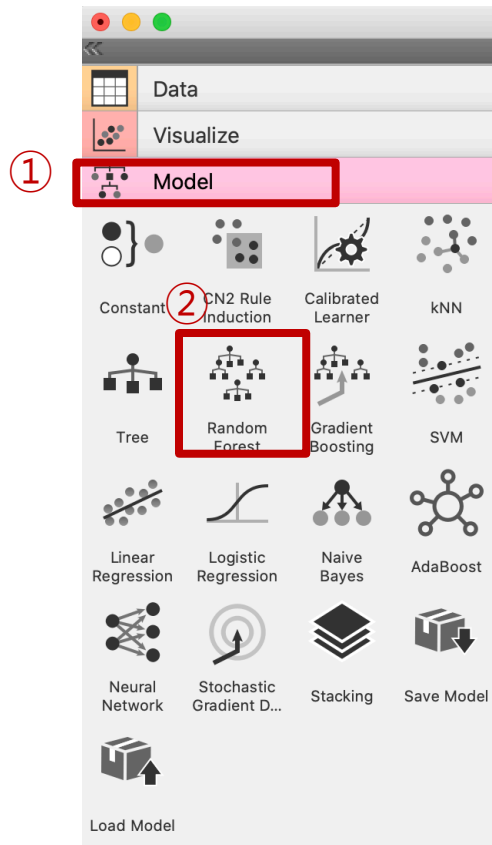


제1회 AI 챌린지 문제 1번: 분류 문제

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

■ 학습 데이터로 모델 학습

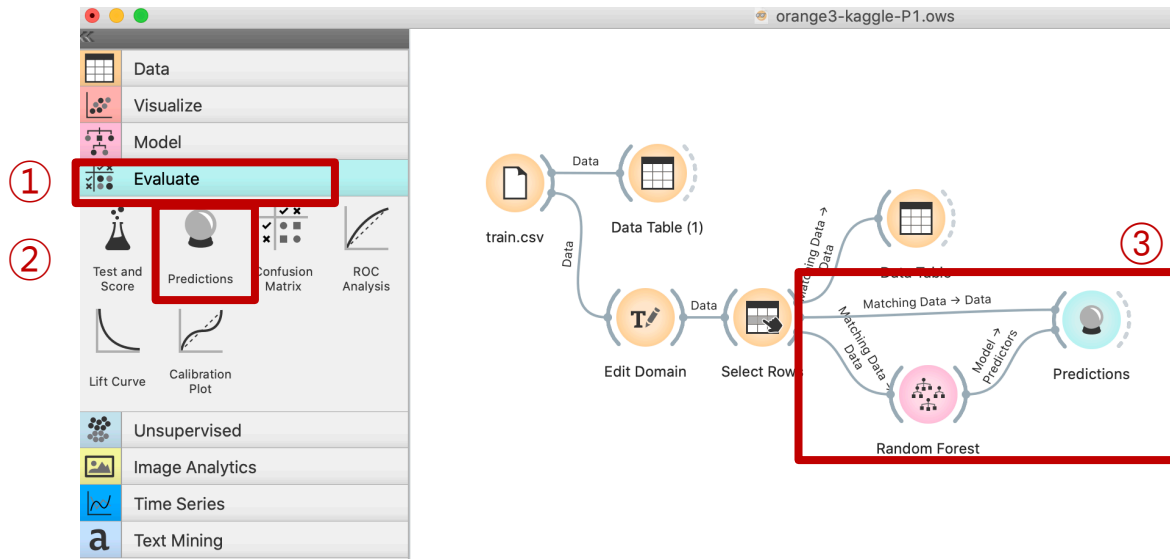
- ① Model 탭 클릭 >> ② RF 드래그 & 드랍 >> ③ 위젯 사이 연결



제1회 AI 챌린지 문제 1번: 분류 문제

■ 학습 데이터로 모델 평가

- ① Evaluate 탭 클릭 >> ② Predictions 드래그 & 드랍 >> ③ 위젯 사이 연결



- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

Predictions						
	Random Forest	pixel1	pixel2	pixel3	pixel	
1	0.00 → 2	0	0	0	0	
2	0.00 → 9	0	0	0	0	
3	0.00 → 6	0	0	0	0	
4	0.00 → 3	0	0	0	0	
5	0.00 → 4	0	0	0	5	
6	0.00 → 4	0	0	0	0	
7	0.00 → 5	0	0	0	0	
8	0.00 → 4	0	0	0	0	
9	0.00 → 8	0	0	0	0	
10	0.00 → 9	0	0	0	0	
11	0.00 → 2	0	0	0	0	
12	0.00 → 9	0	0	0	0	
13	0.00 → 0	0	0	0	0	
14	0.00 → 3	0	0	0	0	
Model Performance						
Model	AUC	CA	F1	Precision	Recall	
Random Forest	1.000	0.997	0.997	0.997	0.997	

33.8k | 33.8k | 1x33753

제1회 AI 챌린지 문제 1번: 분류 문제

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

■ 테스트 데이터 열기

- ① Data 탭 클릭 >> ② File 드래그&드랍 >> ③ File 위젯 더블클릭
- ④ File 경로 선택 >> ⑤ Dialog 닫기

① Data 탭 클릭 >> ② File 드래그&드랍 >> ③ File 위젯 더블클릭 >> ④ File 경로 선택 >> ⑤ Dialog 닫기

sample_submission.csv 리더보드 제출 포맷
test.csv 테스트 데이터
train.csv 학습 데이터

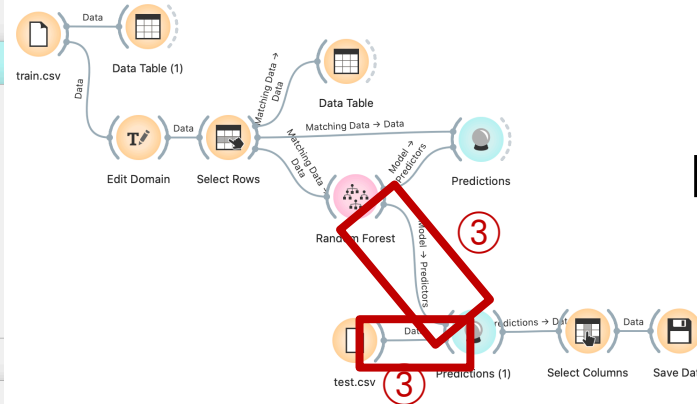
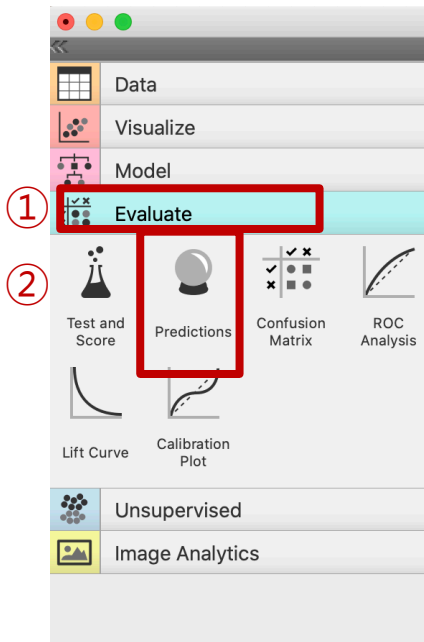
Columns (Double click to edit)	Name	Type	Role	Values
1	id	N numeric	skip	
2	pixel1	C categorical	feature	0
3	pixel2	N numeric	feature	
4	pixel3	N numeric	feature	
5	pixel4	N numeric	feature	
6	pixel5	N numeric	feature	
7	pixel6	N numeric	feature	

제1회 AI 챌린지 문제 1번: 분류 문제

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

■ 테스트 데이터로 모델 평가

- ① Evaluate 탭 클릭 >> ② Predictions 드래그 & 드랍 >> ③ 위젯 사이 연결



테스트데이터를 이용한 모델 예측 결과

A screenshot of the 'Predictions' widget showing test data results. The table has three columns: an index, 'Random Forest', and 'pixel1'. The data is as follows:

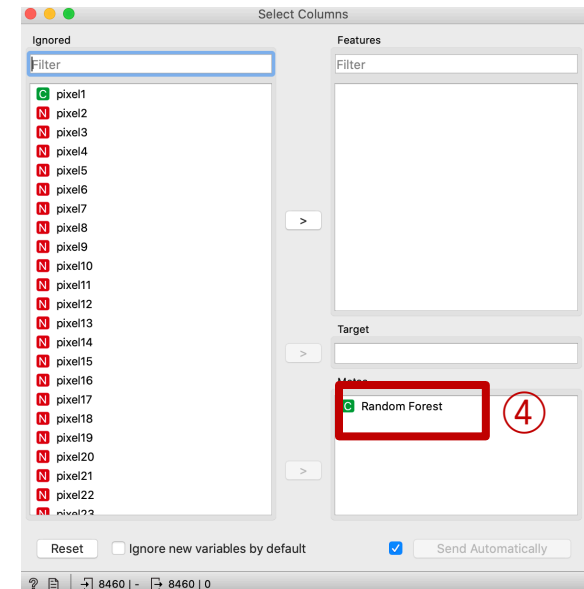
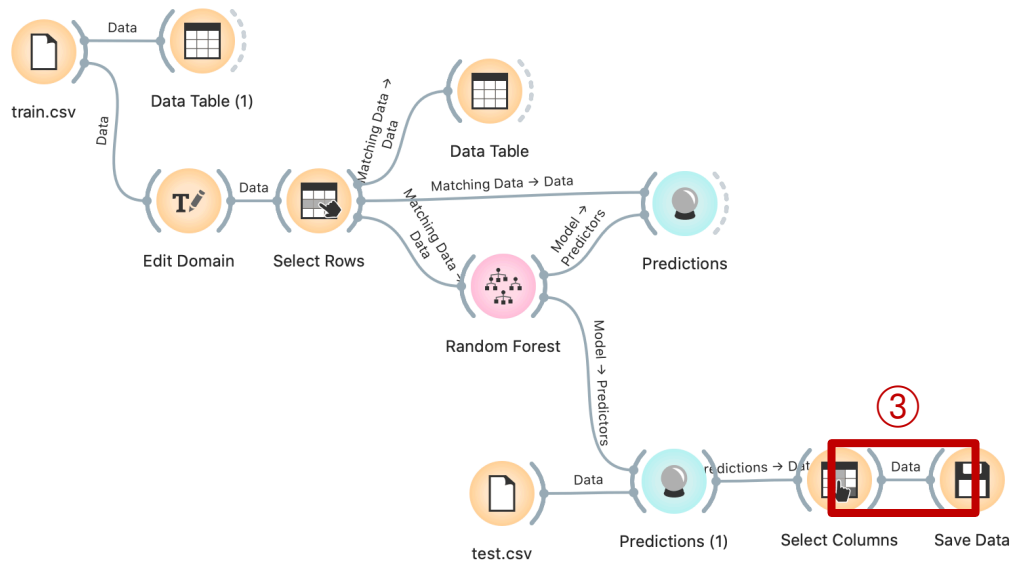
	Random Forest	pixel1
1	6	0
2	2	0
3	4	0
4	7	0
5	4	0
6	0	0
7	0	0
8	6	0
9	1	0
10	0	0
11	5	0
12	1	0
13	4	0
14	5	0
15	6	0
16	6	0
17	1	0

제1회 AI 챌린지 문제 1번: 분류 문제

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

■ 테스트 데이터 예측 결과 저장

- ① Data 탭 클릭 >> ② Select Columns 드래그 & 드랍 >> ③ 위젯 사이 연결
- ④ 모델 예측(Random Forest) 데이터만 선택

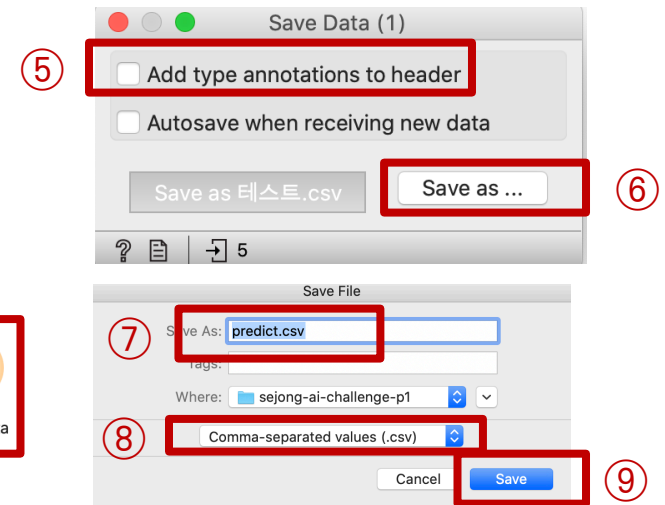
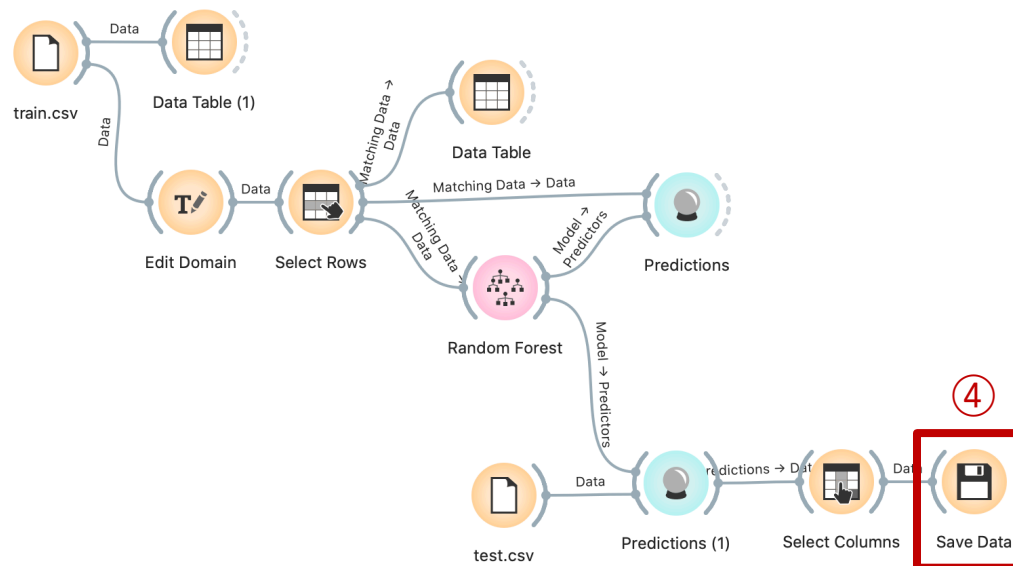


제1회 AI 챌린지 문제 1번: 분류 문제

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

■ 테스트 데이터 예측 결과 저장

- ④ Save Data 더블 클릭 >> ⑤ Add type.. 체크 해제 >> ⑥ Save as 클릭
- ⑦ 파일명 설정 >> ⑧ 확장자 csv 변경 >> ⑨ Save 클릭

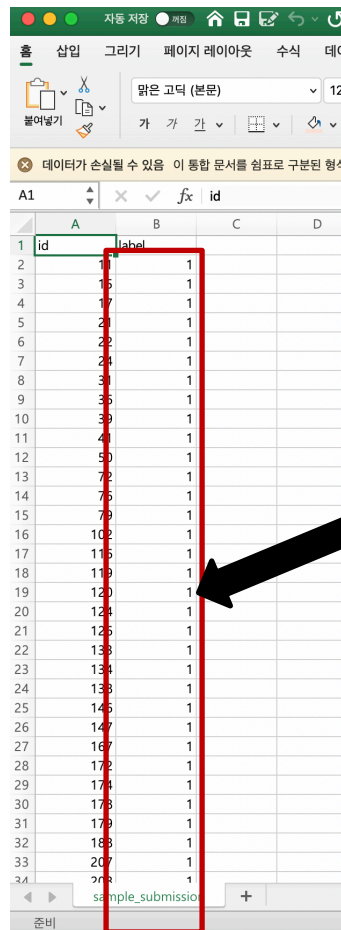


제1회 AI 챌린지 문제 1번: 분류 문제

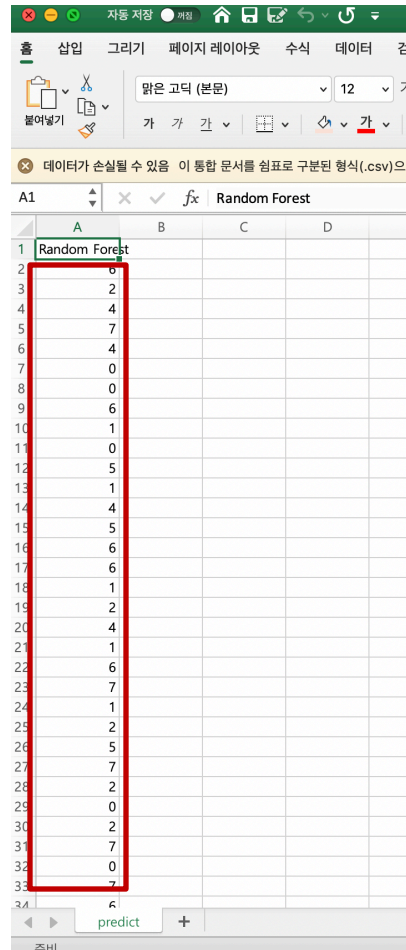
- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

■ 캐글 리더보드 제출 파일 만들기

- <sample_submission.csv> 에 <predict.csv> 모델 예측 결과 복사 & 붙여넣기



	A	B	C	D
1	id	label		
2	1	1		
3	15	1		
4	17	1		
5	21	1		
6	22	1		
7	24	1		
8	31	1		
9	35	1		
10	39	1		
11	41	1		
12	50	1		
13	72	1		
14	75	1		
15	79	1		
16	102	1		
17	115	1		
18	119	1		
19	120	1		
20	124	1		
21	125	1		
22	133	1		
23	134	1		
24	138	1		
25	145	1		
26	147	1		
27	167	1		
28	172	1		
29	174	1		
30	178	1		
31	179	1		
32	183	1		
33	207	1		
34	208	1		



	A	B	C	D	E
1	Random Forest				
2	8				
3	2				
4	4				
5	7				
6	4				
7	0				
8	0				
9	6				
10	1				
11	0				
12	5				
13	1				
14	4				
15	5				
16	6				
17	6				
18	1				
19	2				
20	4				
21	1				
22	6				
23	7				
24	1				
25	2				
26	5				
27	7				
28	2				
29	0				
30	2				
31	7				
32	0				
33	7				
34	6				

제1회 AI 챌린지 문제 1번: 분류 문제

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

■ 캐글 리더보드에 답안 제출

- 캐글 챌린지 >> ① Leaderboard 탭 >> ② Late Submission 클릭 >> ③ 결과 파일 제출 >> ④ Make Submission 버튼 클릭

InClass Prediction Competition

Sejong AI Challenge 문제1

세종대학교 인공지능 챌린지 문제1

a year ago

Overview Data Code Discussion **Leaderboard** Rules Team Host My Submissions **Late Submission** ...

```
> kaggle competitions submit -c 2021-ml-p3 -f submission.csv -m "Message"
```

Make a submission for [Yukyung Choi](#)

Step 1
Upload submission file

③

File Format
Your submission should be in CSV format. You can upload this in a zip/gz/rar/7z archive, if you prefer.


Number of Predictions
We expect the solution file to have 550 prediction rows. This file should have a header row. Please see sample submission file on the [data page](#).

④ **Make Submission**

제1회 AI 챌린지 문제 1번: 분류 문제

- 캐글 리더보드에 답안 제출 확인

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

 InClass Prediction Competition

Sejong AI Challenge 문제1

세종대학교 인공지능 챌린지 문제1

a year ago

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [Host](#) [My Submissions](#) [Late Submission](#) [...](#)

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
sample_submission.csv	a few seconds ago	1 seconds	1 seconds	0.85815

Complete

[Jump to your position on the leaderboard ▼](#)

제1회 AI 챌린지 문제 1번: 분류 문제

■ 캐글 리더보드에 답안 제출 확인

- ① 학습 데이터 열기 (필수)
- ② 학습 데이터 전처리 (필수)
- ③ 학습 데이터로 모델 학습 (필수)
- ④ 테스트 데이터 열기 (필수)
- ⑤ 테스트 데이터로 모델 평가 (필수)

InClass Prediction Competition

Sejong AI Challenge 문제1

세종대학교 인공지능 챌린지 문제1

a year ago

Overview Data Code Discussion **Leaderboard** Rules Team Host My Submissions **Late Submission** ...

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
sample_submission.csv	a few seconds ago	1 seconds	1 seconds	0.85815

Complete

[Jump to your position on the leaderboard](#)

Public Leaderboard **Private Leaderboard**

The private leaderboard is calculated over the same rows as the public leaderboard in this competition.
This competition has completed. This leaderboard reflects the final standings.

Refresh

#	Δpub	Team Name	Notebook	Team Members	Score	Entries	Last
1	—	14011139			0.92269	5	1y
2	—	16010213_송성곤			0.90295	7	1y
3	—	18011863			0.83522	6	1y

무려!! 3등!!

