

AI

NEWSLETTER

**K-water
AI Research Lab.**

25 OCT. 2022

AI News

TEXT TO IMAGE

우주 비행사와 관련된 글을 입력했을 때 AI가 언어모델과 생성모델을 활용하여 그려낸 그림들입니다.
(DALL-E2, MidJourney, Stable Diffusion 등)



NVIDIA 신제품 발표

GPU 시장의 강자 NVIDIA에서 지난 9월 21일 RTX 4090, RTX 4080을 발표했습니다. RTX 4090은 이전 세대 RTX 3090 Ti보다 1.5배 높은 성능을 나타낸다고 합니다.

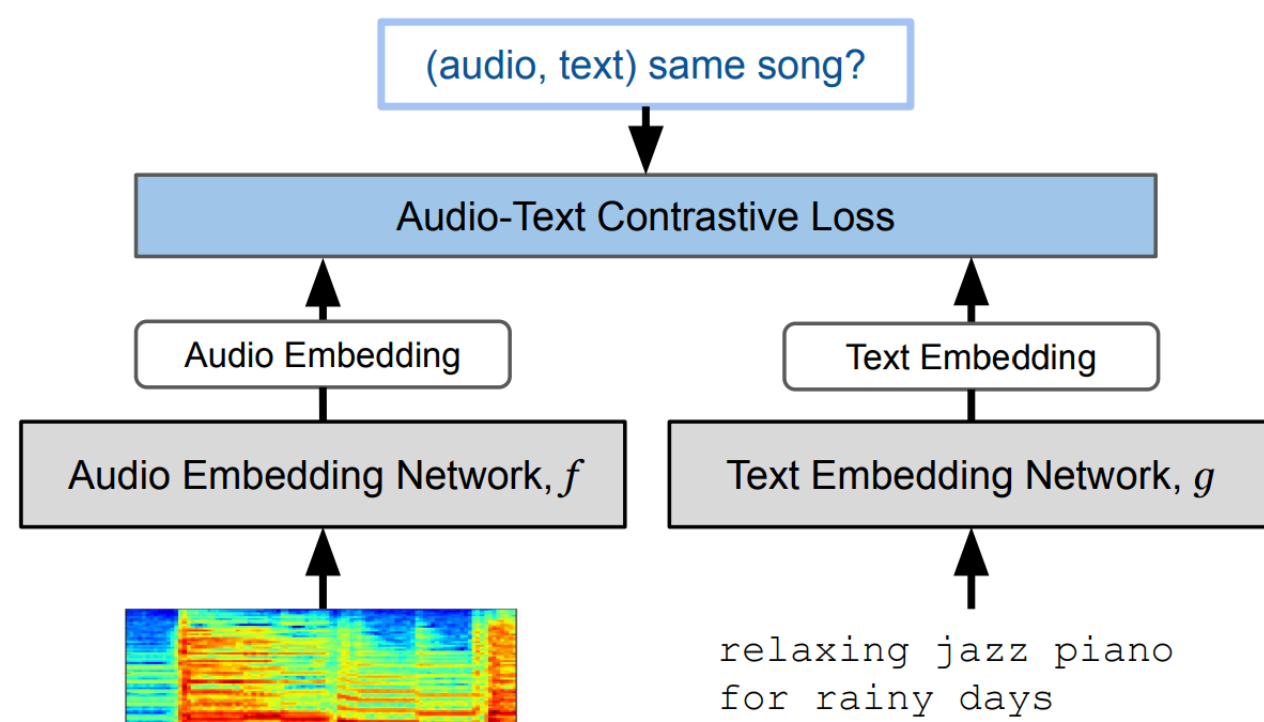
하지만 환율 및 반도체 웨이퍼 생산 단가 상승으로 4090은 263만원, 4080은 192만원(16GB) 부터 시작합니다. 지난 세대인 3000번대에 비하면 가격 및 전력소모가 50%가까이 올랐습니다.



음악을 학습하는 AI 모델

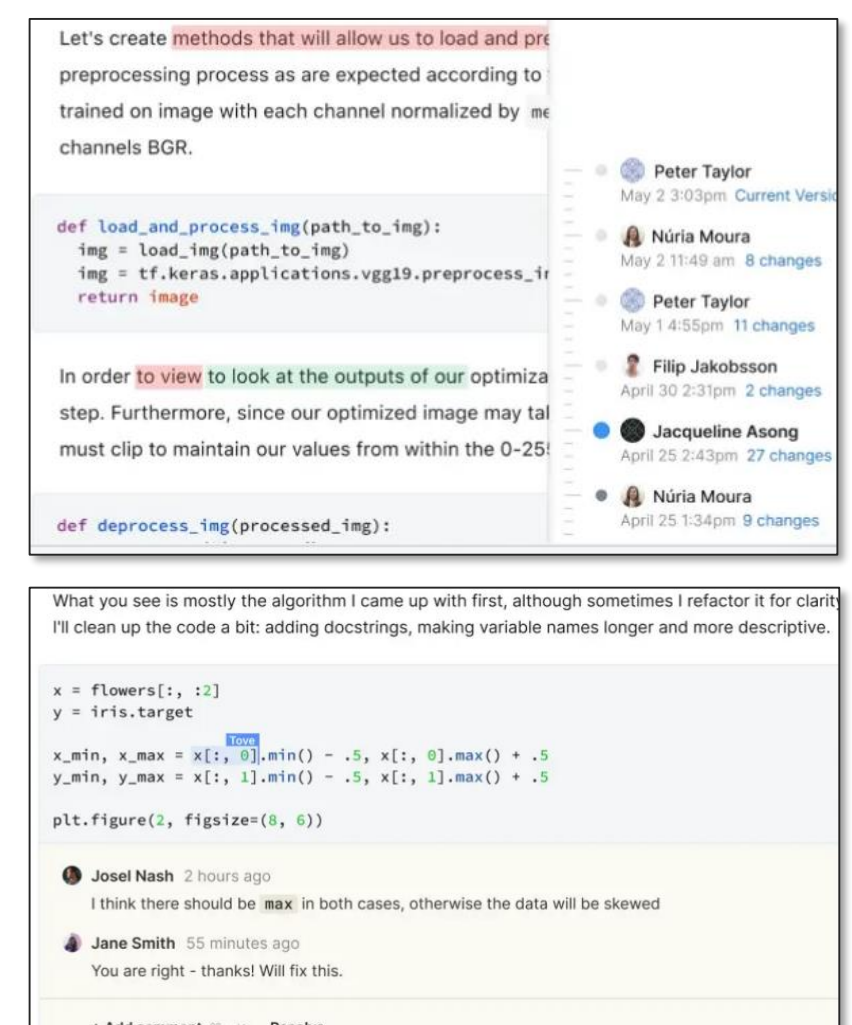
오디오(음악)와 자연어를 연결하는 새로운 임베딩 모델에 대한 논문이 공개되었습니다.

*<https://arxiv.org/pdf/2208.12415.pdf>



DeepNote

DeepNote는 Python 노트북을 공유하여 활용할 수 있는 플랫폼입니다. 구글의 Colab과 유사하기도 하고 요금이 비싸다는 단점이 있지만, 동일한 프로젝트에 대해 공동으로 작업할 경우 장점이 많습니다.





Hands-On AI Project



먹을 수 있는 물인지 예측

1. 데이터를 준비 준비합니다 (Kaggle)

```
!kaggle datasets download -d adityakadiwal/water-potability
data = pd.read_csv('water_potabiliby.csv')
```

2. 데이터에 뭐가 들어있는지 봅니다

- 수질관련 다양한 자료(ph, 경도, 고형물, 전도도 등)와
마지막에는 음용가능여부(Potability)가 있네요

```
data.info()
```

```
RangeIndex: 3276 entries, 0 to 3275
```

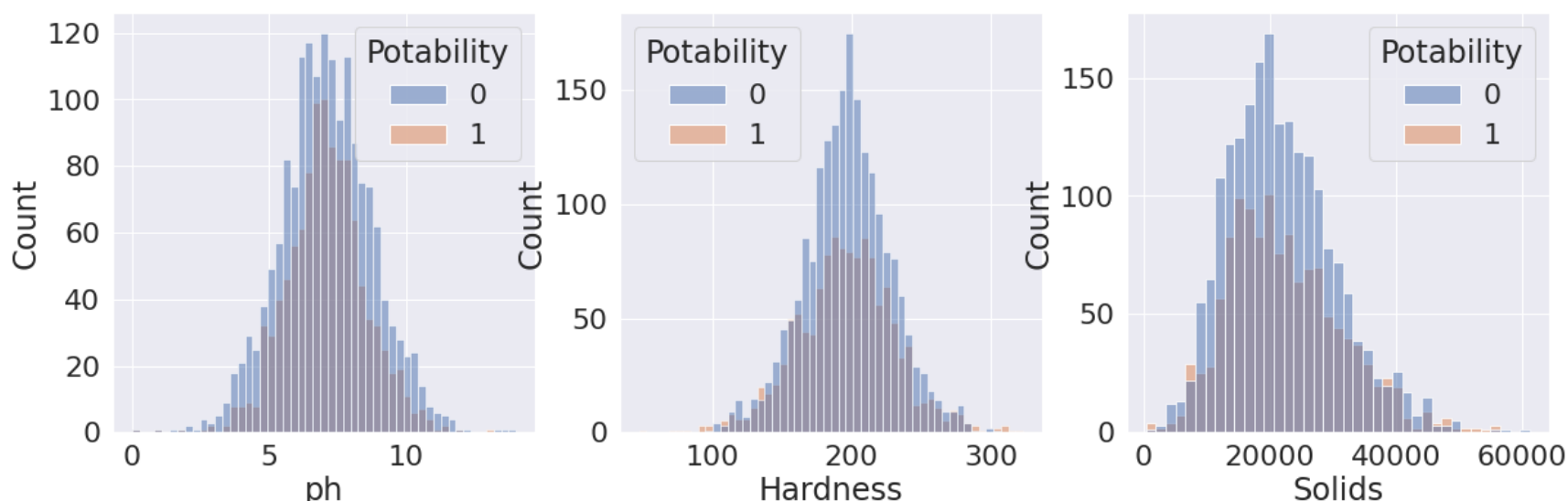
```
Data columns (total 10 columns):
```

#	Column	Non-Null Count	Dtype
0	ph	2785 non-null	float64
1	Hardness	3276 non-null	float64
2	Solids	3276 non-null	float64
3	Chloramines	3276 non-null	float64
4	Sulfate	2495 non-null	float64
5	Conductivity	3276 non-null	float64
6	Organic_carbon	3276 non-null	float64
7	Trihalomethanes	3114 non-null	float64
8	Turbidity	3276 non-null	float64
9	Potability	3276 non-null	int64

3. 음용가능 여부에 따라 데이터가 다를까요?

- 크게 차이가 나는지는 잘 모르겠습니다..

```
sns.histplot(data=data, x=col, hue='Potability', ax=axes[idx])
```



4. 간단하게 Random Forest를 돌려봅니다

- 0.67정도의 정확도를 보여줍니다

```
rf_clf = RandomForestClassifier(random_state=42)
```

```
rf_clf.fit(X_train, y_train)
```

```
rf_clf.score(X_test, y_test)
```

```
0.6738
```

가짜뉴스 판단하기

1. 데이터를 준비 준비합니다 (Github)

```
!wget https://github.com/muttinenisairohith/FakeNewsDe ....
```

2. 핵심 라이브러리들을 불러옵니다

```
from keras models import Model, Sequential
from keras layers import Input, Dense, Dropout, Embedding
from transformers import AutoTokenizer, TFBertModel
```

3. 데이터를 불러오고 합쳐줍니다

```
df_fake = pd.read_csv("Fake.csv")
df_true = pd.read_csv("True.csv")
df_fake['Label'] = 'Fake'
df_true['Label'] = 'True'
df = pd.concat([df_fake, df_true])
df = df.sample(frac=1).reset_index(drop=True)
```

4. 훈련/테스트 데이터를 분리하고 토큰화합니다

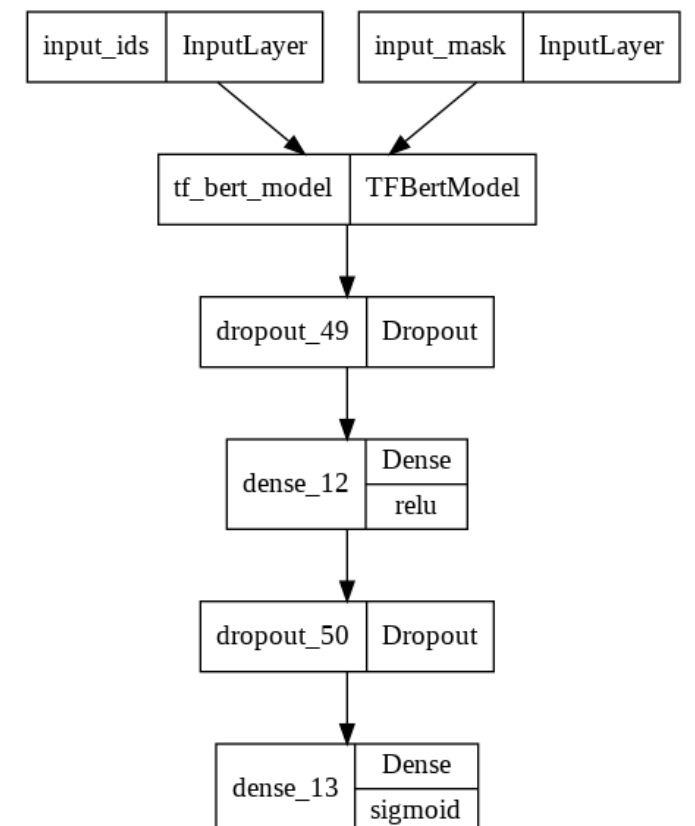
```
X_train, X_test, y_train, y_test = train_test_split(X, y)
X_train_tokens = tokenize(X_train)
X_test_tokens = tokenize(X_test)
```

5. 분류모델의 구조는 아래와 같습니다.

```
# Layer
Embedding Layer →
Dense Layer (64) →
Dense Layer (1)

# MODEL
Dropout Rate: 0.2
Optimizer: Adam
Loss: binary_crossentropy
Metric: accuracy

# Compile
Validation split: 0.2
Batch size: 32
Callbacks: Early Stopping
```



6. 테스트 데이터로 훈련이 잘 됐는지 확인합니다

```
y_pred = model.predict({'input_ids':X_test_tokens[ ...
Y_pred = np.where(y_pred >= 0.5, 1, 0)
classification_report(y_test, y_pred)
```

7. 100%의 정확도가 가짜뉴스와 진짜뉴스를 구분하는 모델이 만들어졌습니다.

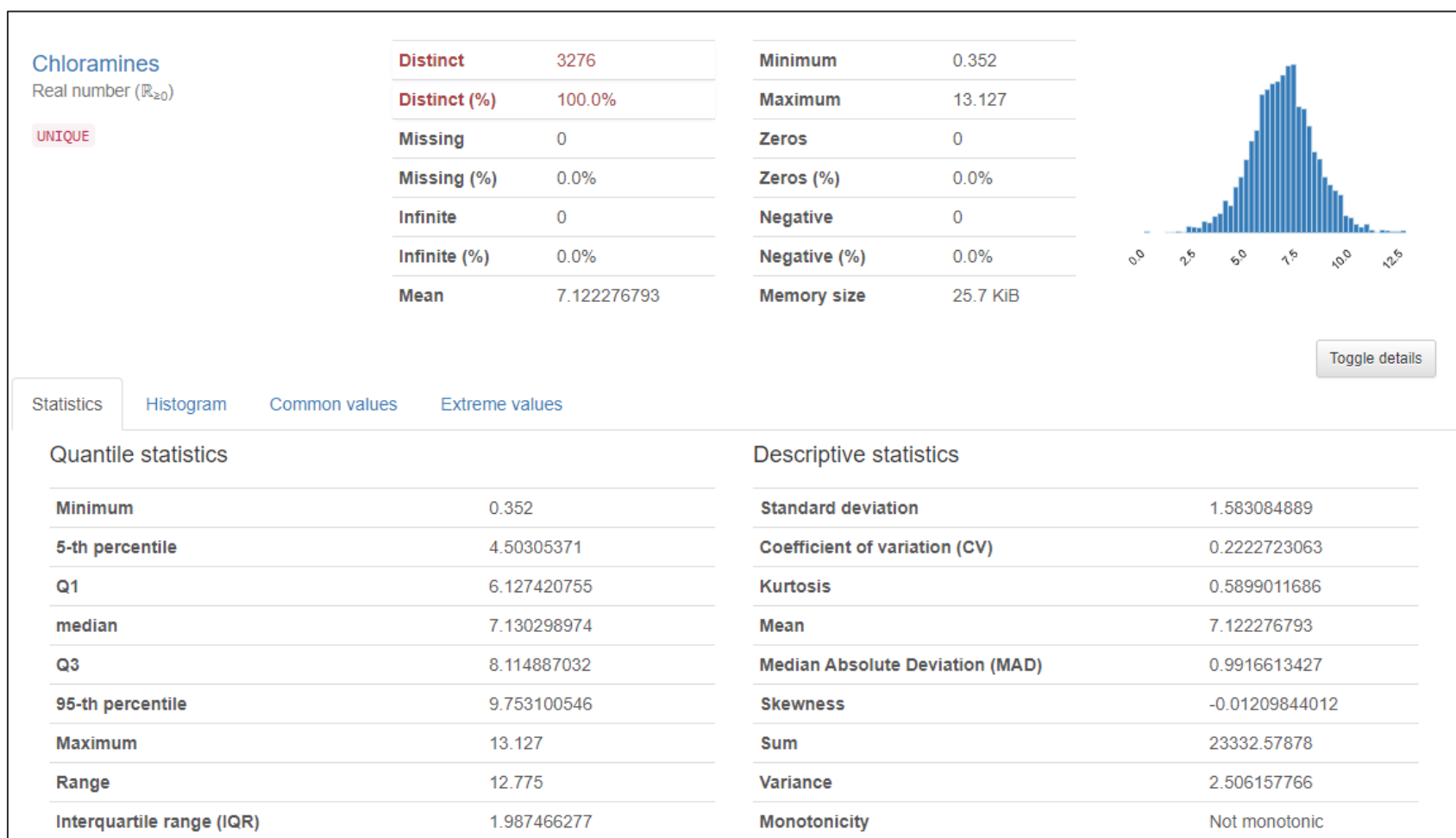
TIPS

Pandas.profile

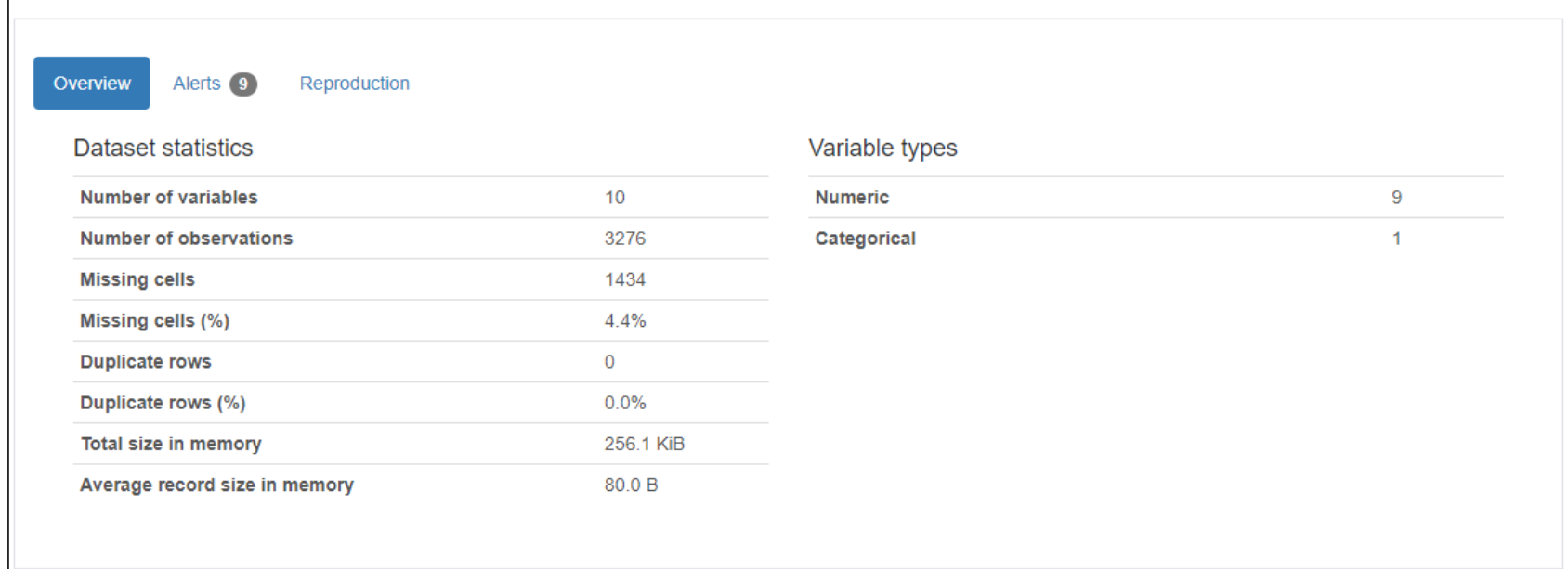
Python의 영원한 친구 Pandas 애드온입니다.

```
!pip install pandas_profiling
df = pd.read_csv('my_file.csv')
profile = df.profile_report()
profile
```

정말 간단하게 Pandas 데이터프레임에 메소드를 실행시키면 아래와 같은 통계 등이 표출됩니다.



Overview



List Comprehension

간단한 반복문은 list comprehension을 활용

```
# BEFORE
squares = []
for i in range(10):
    squares.append(i^2)

# AFTER
squares = [i^2 for i in range(10)]
```

Variable Swap

변수간 내용을 바꿔주고 싶을 때

```
# BEFORE
tmp = a
a = b
b = tmp
```

```
# AFTER
a, b = b, a
```

AI / 머신러닝 / 딥러닝

○ 인공지능(Artificial Intelligent, AI)

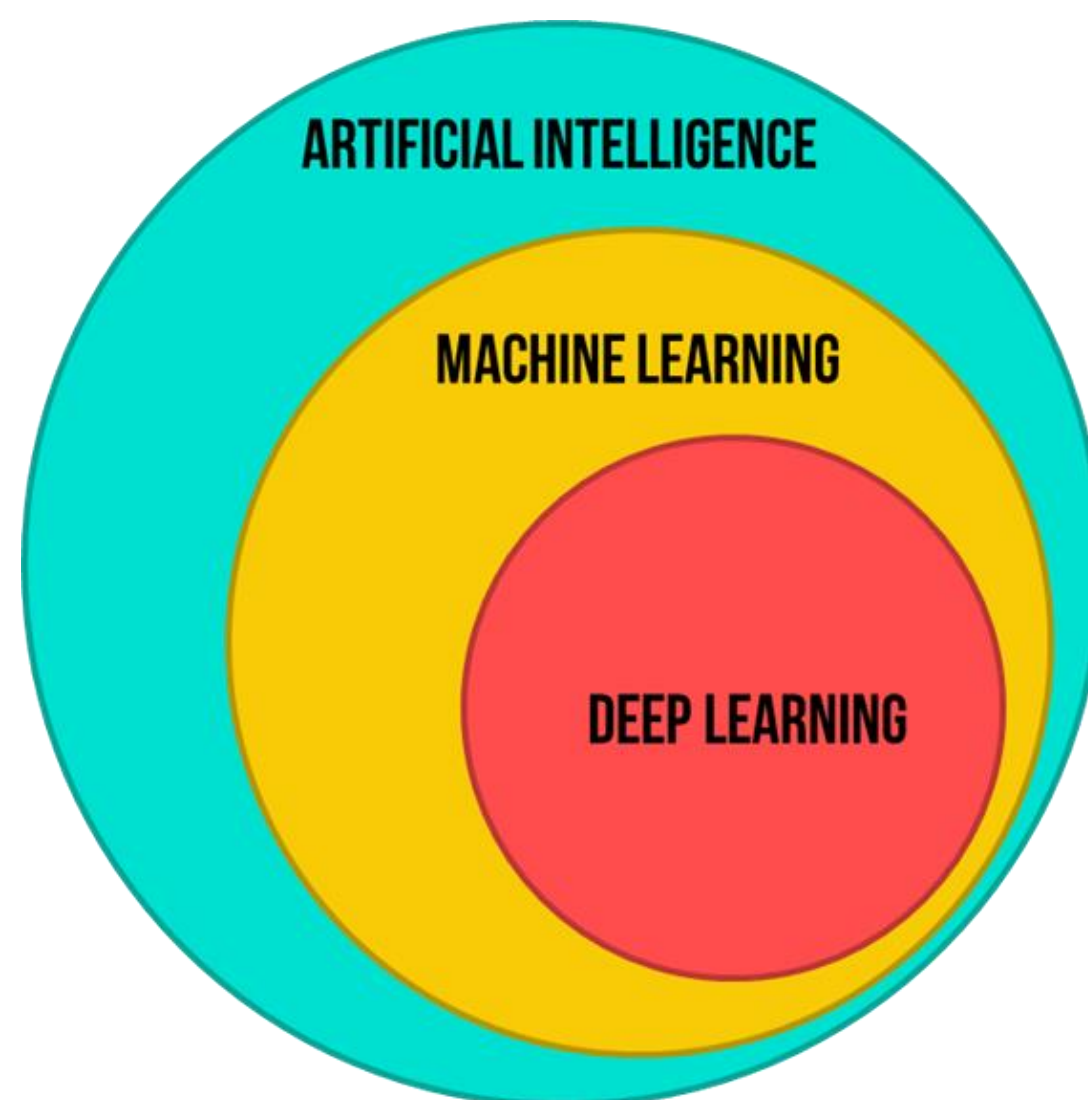
- AI는 인간이 갖고 있는 학습, 추론 및 판단 능력을 컴퓨터로 구현하는 가장 포괄적인 개념입니다

○ 머신러닝(Machine Learning, ML)

- 데이터의 특성과 패턴을 학습해서 미지의 데이터에 대한 추정치를 계산하는 방법입니다.

○ 딥러닝(Deep Learning, DL)

- 기본적인 신경망(NN)을 여러 층으로 깊게(Deep) 구성해 학습하는 구조와 알고리즘입니다.



K-water AI Lab.

안녕하세요

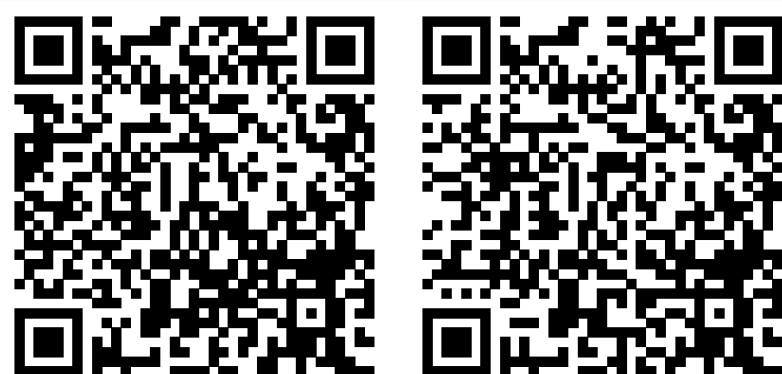
- K-water연구원 AI연구센터입니다.
- 모두가 인식하고 있듯, 최근 AI 기술은 매우 빠른 속도로 발전하고 있으며, 모든 산업군에 빠짐없이 다양한 형태, 목적으로 적용되고 있습니다.
- 그럼에도 불구하고, AI 기술이 중요하다고는 하지만 너무 많은 정보들이 쏟아지고 있어 무슨 정보를 어디서 어떻게 습득해야 할지, 어떤 분야와 업무에 어떤 알고리즘이 적용되고 있는지 파악하기가 쉽지는 않습니다.
- 우리 K-water에서도 현업과 연구/기술 부서를 가리지 않고 많은 적용과 시도가 이루어지고 있는 이 시기에, 최신의 AI관련 S/W, H/W 기법과 동향을 공유함은 물론, 다양한 응용이 가능한 기본기술, 초중급자를 위한 코딩 Tip 등을 담아 공유하고자 합니다

문의

주경원 선임 (7840)

이소령 사원 (7341)

Colab Link



제2회 AI경진대회

- AI연구센터에서는 빠르게 진화하는 AI 분야의 알고리즘 효율적 확보를 위해 AI경진대회를 개최하고 있습니다.
- 11월 초부터 개최예정인 제2회 AI 경진대회에서는 용수 특성이 다른 각각 3개의 배수지에서 향후 2주간 필요한 용수량을 정확하게 예측해야 합니다.

K-water AI Lab. Hub 런칭

- AI연구센터는 깃허브를 통해 AI Lab. 허브를 운영하고 있습니다.
- 저희 AI연구센터와 협업이 필요하시면 언제든지 연락주세요 🙌

