2025 홍천고등학교 여름방학 과학캠프

Al를활용한



춘천고등학교 장원재



01 데이터(빅데이터) 데이터의 특성과 가치 02 인공지능(AI) 인공지능정의 및 능력

03 인공지능(AI) 모델 제작 회귀 및 분류 모델 제작

데이터(빅데이터)

데이터의특성과가치

빅데이터란?

0000의 재료!



- ❖ 빅데이터! (5V)
 - ✓ 크기가 크다!!! (크기가 TiB급 Volume)
 - ✓ 종류가 다양하다!!! (Variety)
 - ✓ 속도가 빠르다(실시간, Velocity)
 - ✓ 가치가 있다(Value)
 - ✓ 정확하다(Veracity)
 - 〉 빅데이터는! 문제00의 재료이다!
 - > 가공되지 않은 날것 그대로!



- 다양한 데이터

사고건수	사망자수	중상자수	경상자수	부상신고자수	
12767	445	5999	6664	458	
3628	146	1438	2007	183	
1973	33	563	1330	121	
2226	21	745	1455	131	
14675	333	4887	9139	1062	
7892	205	3539	8837	588	
71687	404	17756	84667	5781	

[수치데이터]

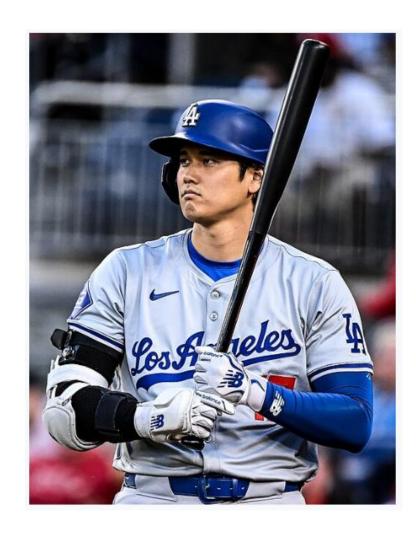
IAM

IVE

다른 문을 열어 따라갈 필요는 없어 넌 너의 길로, 난 나의 길로 음 하루하루마다 색이 달라진 느낌 밝게 빛이 나는 길을 찾아

I'm on my way 넌 그냥 믿으면 돼 I'm on my way 보이는 그대로야

[텍스트 데이터]



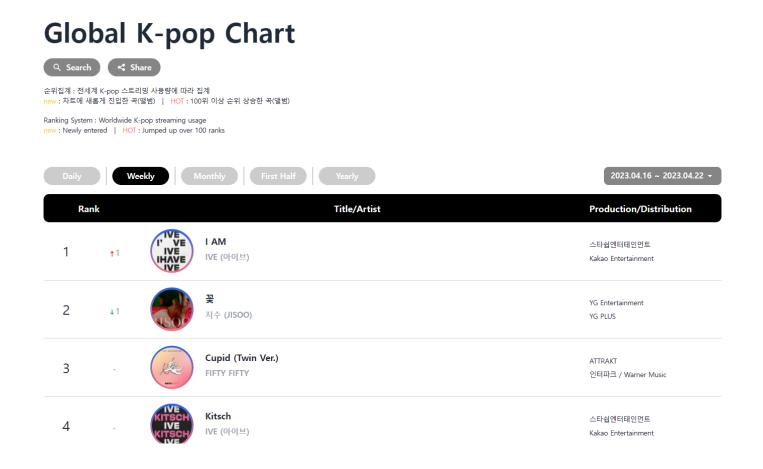
[이미지 데이터]

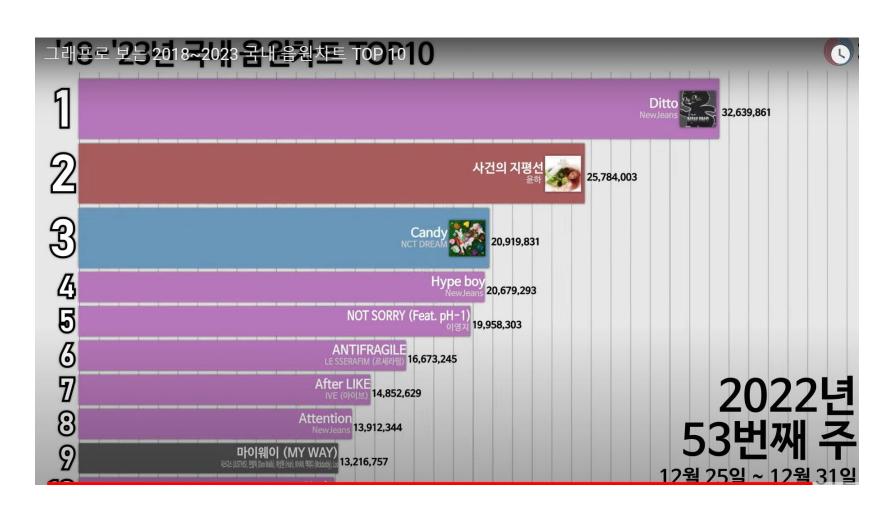
이외에도 다양하고 많은 양의 데이터가 현실에 존재한다!



데이터의 분석?

- 자료를 정보로! (표 데이터를 애니메이션 시각화로 바꿨을 때 이점은?)



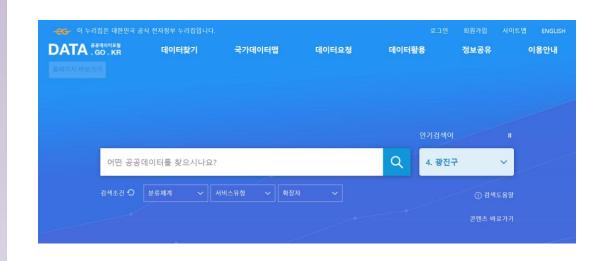


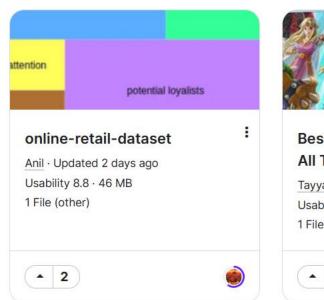
https://www.youtube.com/channel/UCs-fnRl8JjhbMax7wRb9jwA (데이터 다람쥐 Youtube 채널)

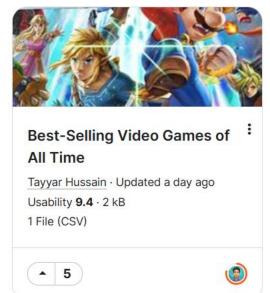


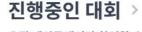
데이터를 어디서 얻을 수 있을까?

- 데이터를 얻을 수 있는 다양한 온라인 사이트



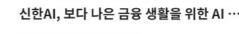






오직 데이콘에서만 참여할 수 있어요





	D-24
알고리즘 비전 분류 MLOps	

2023.04	D-17				
아이디어	금융	AI	서비스	플랫폼	
아고리즈	저서피	17L			

[공공데이터 포털]

: 대한민국 정부, 지방자치단체, 공기업, 기관 등에서 조사, 생산한 데이터를 얻을 수 있는 포털

[Kaggle(캐글)]

: 국제적인 데이터, AI 관련 사이트 다양한 주제, 다양한 형태의 데이터를 얻을 수 있으며 이를 이용한 AI 콘테스트가 열림

[데이콘]

: 한국의 캐글!

데이터와 인공지능(AI)은 어떤 관계일까?

학생과 책-학용품-선생님의 관계!



[학생 : 인공지능]

: 여러가지 문제를 해결하기 위해 읽기, 쓰기, 말하기, 암기, 실험, 체험하기 등 다양한 활동을 수행함



[책, 학습자료 : 데이터]

: 학생이 학습을 하기 위해서는 책, 수업 프린트 등 다양한 자료가 필요하다. 이것이 바로 데이터이다. 또한 이를 기반으로 수업하기 위해서는 연필, 교사 등 다양한 지원이 필요하다.



[학생 : 인공지능]

: 학습을 마치면 학습한 내용을 토대로 다양한 문제를 해결할 수 있다. 또한 성공적으로 학습이 이루어졌다면 비슷한 유형의 다른 문제(접하지 않은 데이터)를 해결할 수 있다!

인공지능(AI)

인공지능의정의와능력

인공지능이란?

사람이 인공적으로 구현한 지능!



- ❖ 지능(Artificial Intelligence)
- 사람이 문제에 대해 합리적으로 사고하고 해결하는 학습 능력과 인지능력
- 학습: 새로운 정보나 방법을 배우고 이를 외부 도움 없이 재생하거나 활용할 수 있는 능력
- 인지능력 : 문제를 해결하기 위한 최선의 방법을 판단하고 예측할 수 있는 능력
- ❖ 인공지능(Artificial Intellignce)
- 인간이 가진 지능의 일부 또는 전체를 인공적으로 구현한 것

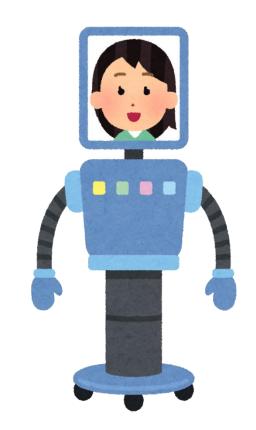
인공지능의 발전 단계

약인공지능 - 강인공지능 - 초인공지능



[약인공지능]

: 인간이 가진 여러 가지 지적 능력 중 일부만을 수행할 수 있는 인공지능



[강인공지능]

: 인간을 완벽하게 모방하여 인간이 할 수 있는 모든 일을 수행할 수 있는 인공지능 범용인공지능, 일반인공지능(Artificial General Intelligence(AGI)이라 하기도 함



[초인공지능]

: 인간이 가진 능력을 초월(인간에게 무한한 시간과 자원을 준다 하더라도 수행할 수 없는 능력을 수행할 수 있는)하여 일을 수행할 수 있는 인공지능

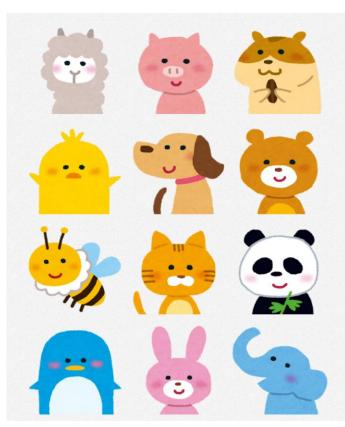
인공지능이할수있는것

회귀, 분류, 군집



[회귀모델]

: 외부의 입력(독립변수)을 받으면 사용자가 원하는 형태의 결과값을 출력해주는 모델: 주거조건 입력 - 〉 주택 예상 가격 / ChatGPT(문장 입력 이에 대한 답변) / Dall-E2 등



[분류모델]

: 입력을 받으면 사전에 학습한 방법에 따라 값의 종류를 분류해주는 모델 : 꿀벌 -〉 곤충 / 고양이 -〉 식육목 / 햄스터 -〉설치목



[군집모델]

: 입력값이 가진 특징에 따라 사용자가 지시한 방법대로(2가지, 3가지, N가지) 데이터를 나누어주는 모델: 위 4개의 사진을 2종류로, 4종류로 나누어봐라



인공지능의 학습방법

지도학습, 비지도학습, 강화학습



[지도학습]

: 선생님과 함께 공부하는 것처럼 입력값(문제)에 대한 결과값(정답)을 알려주면서 인공지능 모델을 학습시키는 방법: 이미지 분류 / 문장 생성 등 회귀모델에서 많이 사용



[비지도학습]

: 지도학습과는 달리 정답을 알려주지 않고 주어진 기준에 따라 모델이 스스로 정답을 찾아 나가는 방식으로 이루어지는 학습: 유튜브 추천 알고리즘 / 스포츠 1군 2군 선수 능력치 구분



[강화학습]

점수라는 목표치와 이를 달성하기 위해 수정가능한 입력값(독립변수)을 제시한 후 모델이높은 점수를 얻기 위해 입력값을 바꿔가며 최적의 방법을 찾는 방식 : 게임 몬스터인공지능 / 게임 클리어 인공지능

인공지능 활용 및 프로그래밍

인공지능모델제작및프로그래밍



인공지능 모델이란?

특정한 기능을 수행하는 인공지능을 의미한다!

[Al 모델의 이름]



- 왼쪽 이미지는 치와와와 머핀 사진을 모아 놓은 것이다.
- 사람은 이 사진을 보고 어떤 것이 치와와인지 머핀인지 쉽게 분류할 수 있다.
- 사람의 이와 같은 능력을 인공적으로 구현하고 사용할 수
 있다면 이를 인공지능 모델이라고 한다.
- 인공지능 모델을 설명할 때 보통 데이터의 종류-Al 모델의 수행기능 - 모델의 학습방법 순으로 나타낸다.
- 가령 치와와-머핀 사진을 분류하는 모델은
- 이미지 분류 머신러닝-딥러닝 모델이라 나타낸다.
- 이런 내용을 토대로 Python을 활용한 이미지 분류 딥러닝 모델 실습을 진행해보자!

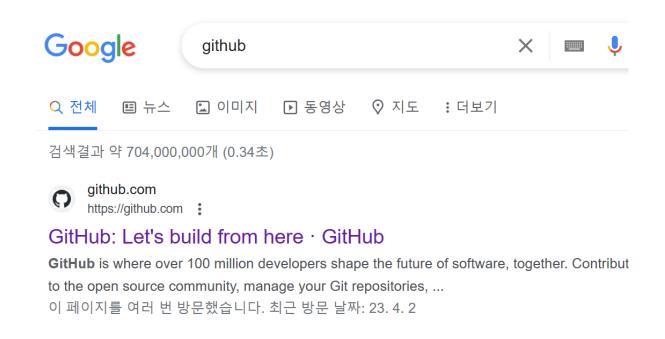


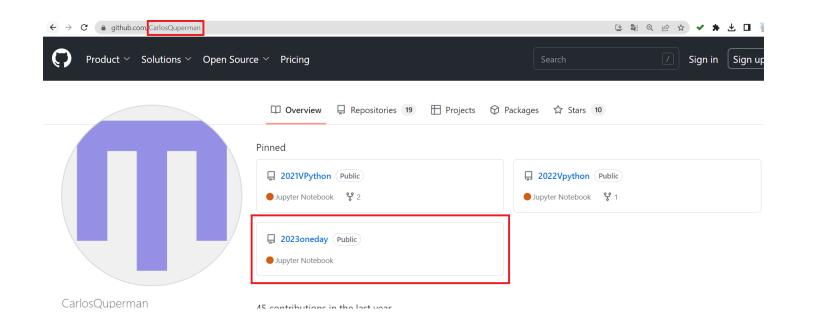
Python을 활용한 이미지 분류 모델 제작 프로그래밍

1. Github 접속 후 실습 관련 페이지 내용 확인하기 (구글 Colab 접속 후)

[실습관련 Github 페이지 접속]

- 1. https://github.com/CarlosQuperman/2023oneday 접속
- 2. 실습관련 데이터와 프로그래밍 실습 코랩 파일을 모아놓은 페이지 살펴보기
- 3. Github 검색 후 뒤에 /를 붙이고 CarlosQuperman을 적어 접속 후 2023oneday를 선택하면 쉽게 접근가능

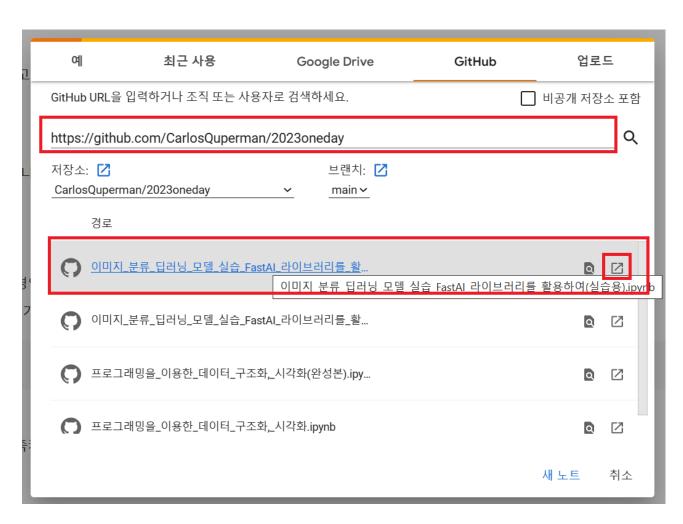




* 구글에서 Github 검색하여 접속 후 주소창에 /CarlosQuperman 추가 후 엔터 빨간색으로 표시된 부분 클릭

Python을 활용한 이미지 분류 모델 제작 프로그래밍

2. Google Colab에서 Github 페이지 불러온 후 구글 드라이브에 복사하기



[Github 저장소 파일 불러오기]

- 1. https://github.com/CarlosQuperman/2023oneday
 주소 복사하여 붙여넣기 후 돋보기 클릭
 - 2. 파일 오른쪽에 있는 새창에서 열기 아이콘 클릭



[Drive로 복사 클릭]

- 1. Drive로 복사 버튼 클릭하여 실습 기본 Colab 파일 자신의 구글 계정으로 복사
 - 2. 이후 선생님의 지시를 받아 실습 수행!

Python을 활용한 이미지 분류 모델 제작 프로그래밍

3. 모델 제작을 위한 라이브러리 설치, 불러오기, 구글 드라이브 권한 설정



[프로그래밍 기본 설정]

- Colab에서 기본 설치되지 않은 라이브러리 (프로그래밍 도구집)라면 사용 전 설치가 필요하기에 fastai2 라이브러리와 fastbook을 설치해준다.
- 우리는 구글 Colab에서 프로그래밍을 진행하고 있고, 그러면 학습 데이터를 스스로의 구글 드라이브에서 저장하고 관리하는 것이 편하기에 우리 코드가 구글 드라이브에 접근할 수 있도록 권한을 허락해준다.
- 도구집을 사더라도 프로그래밍의 책상 위에 올리지 못하면 사용하지 못하기에 도구를 import 해준다.

Python을 활용한 이미지 분류 모델 제작 프로그래밍

4. 키워드 입력 후 이를 사용자의 구글 드라이브에 저장하기



```
1 # 분류할 type을 key_types에 문자열로 저장한다.
 2 key_types = 'grizzly', 'panda', 'polar bear
 3 # 드라이브에 사진을 저장할 기본 폴더를 만든 후 이를 경로로 지정한다
 4 path = Path('<u>/content/drive/MyDrive</u>/(2023) 원데이 클래스 AI 이미지 분류')
 5 # 만약 path에 저장된 경로가 없으면 경로를 만들어준다
 6 if not path.exists():
 7 | path.mkdir()
 1 #o라는 이름으로 key_type을 돌면서
 2 for o in key_types:
 3 print(path, o) # 확인용
 4 dest = (path/o) #각각의 키워드에 따라 사진을 저장할 폴더 경로
 5 print(dest) # 확인용
 6 dest.mkdir(exist_ok = True) #폴더가 없으면 폴더를 만들고
   results = search_images_ddg(o,max_images=100) #키워드에 해당하는 이미지 주소를 가져오고
 8│ download_images(dest, urls = results) #설정한 경로에 내용을 파일로 저장한다.
/content/drive/MyDrive/(2023) 원데이 클래스 AI 이미지 분류 grizzly
/content/drive/MyDrive/(2023) 원데이 클래스 AI 이미지 분류/grizzly
/content/drive/MyDrive/(2023) 원데이 클래스 AI 이미지 분류 panda
/content/drive/MyDrive/(2023) 원데이 클래스 AI 이미지 분류/panda
/content/drive/MyDrive/(2023) 원데이 클래스 AI 이미지 분류 polar bear
/content/drive/MyDrive/(2023) 원데이 클래스 AI 이미지 분류/polar bear
```

[이미지 분류 모델 제작을 위한 이미지 수집]

- 1. Result를 활용하여 이미지 주소 불러오기 및 이미지 다운을 통해 이미지가 잘 저장되는지 확인 후
- 2. key_types에 분류 모델이 분류할 키워드를 설정 후 이미지를 사용자의 구글 드라이브에 다운로드 받는다.
- 3. 이후 이미지 다운에 실패한 정리코드를 실행해준다.

Python을 활용한 이미지 분류 모델 제작 프로그래밍

5. 다운로드 된 이미지 확인

내 드라이브 > (2023) 원데이 클래스 AI 이미지 분류 > grizzly ▼

내 드라이브 → (2023) 원데이 클래스 AI 이미지 분류 → panda ▼

Oe1fdc92-2ca1-4... Oc650779-6cf5-4... Ob06b5eb-adce-... Oc711a7a-c90f-4... Odcba702-7ff7-4... Of2324c5-8388-4... 1c2bcc46-928e-4... If6c75c5-5023-4... 4befd4e7-6df0-4... 5b46e9cb-45ba-... 1afaa516-1e61-4d... 3d33c922-9dac-4... 3acf0625-2793-4... 1af7ebaf-a6aa-47... 1e1746cc-42a5-4... 3ad3db1e-db4c-4...

[다운로드 된 이미지 확인]

- 1. 여러분들의 구글 드라이브에 접속 후 경로를 확인하면 이미지 다운이 성공적으로 됨을 확인할 수 있다.
- 2. 이미지를 확인 후 학습에 적합하지 않은 이미지는 삭제를 통해 걸러준다. (예: 자이언트 팬더 사이에 섞인 랫서팬더)



Python을 활용한 이미지 분류 모델 제작 프로그래밍

6. 데이터 전처리







2 #예) 50분 공부하고 10분 쉬고 암기로 공부하는 방법을 -> 영어단어에 적용한다.

4 dls = typetools.dataloaders(path) #공부방법.dataloaders(공부과목)

[데이터 나누기]

- 1. 우리가 인공지능 모델을 만들고 활용하는 이유는 학습한 문제 뿐 아니라 학습하지 않은 문제 역시 잘 해결하기 위해서이다.
- 2. 이를 위해서는 데이터를 학습데이터와 검증데이터로 나눈 후 학습이 끝나면 검증 데이터를 활용해 검증하는 것이 필요하다.
- 3. 관련된 코드를 실행해주자!

Python을 활용한 이미지 분류 모델 제작 프로그래밍

7. 데이터 증강



[데이터 증강]

- 1. 인공지능 모델을 학습하기 위해서는 양질의 많은 데이터가 필요하다.
- 2. 그런데 현실에서는 이런 데이터를 수집하는 데에 어려움이 따른다.
- 3. 데이터 증강은 이를 해결하는 데 도움을 줄 수 있는 수단이다.
- 4. 데이터 증강은 동일 이미지를 다르게 표현함으로써 학습 데이터를 증가하는 기법이다.
- 5. 네 가지 데이터 증강기법을 코드와 실습을 통해 확인해보자!

Python을 활용한 이미지 분류 모델 제작 프로그래밍

8. 학습방법 결정 및 학습

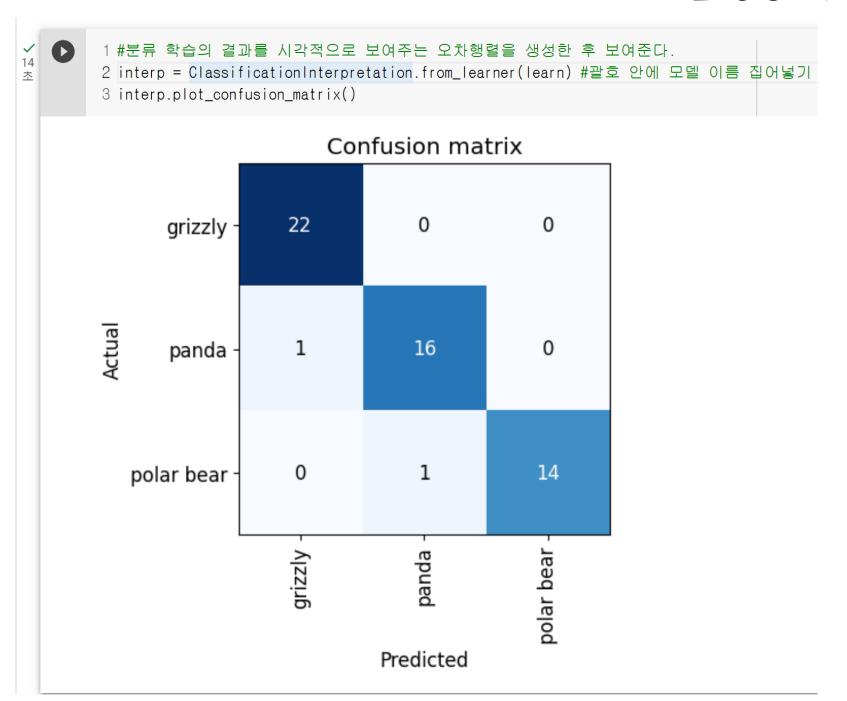


[학습모델 및 횟수 결정]

- 1. FastAi 라이브러리에서는 다양한 인공지능 모델 학습방법을 제공한다.
- 2. 제시된 코드에서 우리는 이미지 분류용으로 학습된 다양한 CNN 모델을 활용할 수 있다.
- 3. CNN 모델은 딥러닝 기반 모델로 이미지를 처리하는데 특화된 모델이다.
- 4. 이미지 분류 시 딥러닝 뿐 아니라 머신러닝 모델을 활용할 수도 있지만 딥러닝이 더 강점을 보이기에 여기서는 딥러닝 모델을 활용한 실습을 진행한다.
- 5. 활용 모델 종류와 학습횟수는 수정이 가능하다.

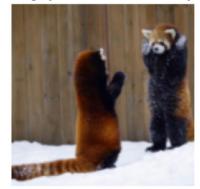
Python을 활용한 이미지 분류 모델 제작 프로그래밍

9. 모델 성능 확인 및 이미지 확인



```
(31] 1 #가장 오차가 큰 값 20개를 보여줌
2 interp.plot_top_losses(20, nrows=5) #n개의 값을 n개의 줄에 표시하기
3
4 #예측값/실제값/오차/가능성
```

grizzly/panda / 2.95 / ÞBBda/polar bear / 0.91 / @





polar bear/polar bear / 0.02 of Oz9B/arizzly / 0.02 of Olas

[오차행렬 및 오차 데이터 확인]

- 1. 오차행렬은 학습된 모델에 검증 데이터를 집어넣었을 때 결과를 확인해주는 인공지능 모델 성능 평가 방법이다.
- 2. 또 코드를 통해 오차가 제일 큰 데이터의 모습을 살펴볼 수도 있다.
- 3. 이후 학습 데이터와 검증 데이터를 적절하게 수정할 수 있다.



Python을 활용한 이미지 분류 모델 제작 프로그래밍

10. 모델 저장

▼ 모델 사용하기

- 위와 같은 과정을 거쳐서 만든 모델을 실제 사용해봅시다
- AI 모델을 사용하는 이유는 학습을 통해서 만든 모델이 학습하지 않은 상황에서도 잘 적용하기 위해서입니다.
- 만든 모델을 파일로 만들고 그 파일을 불러와 학습하지 않은 이미지를 잘 분류하는지 확인해봅시다.



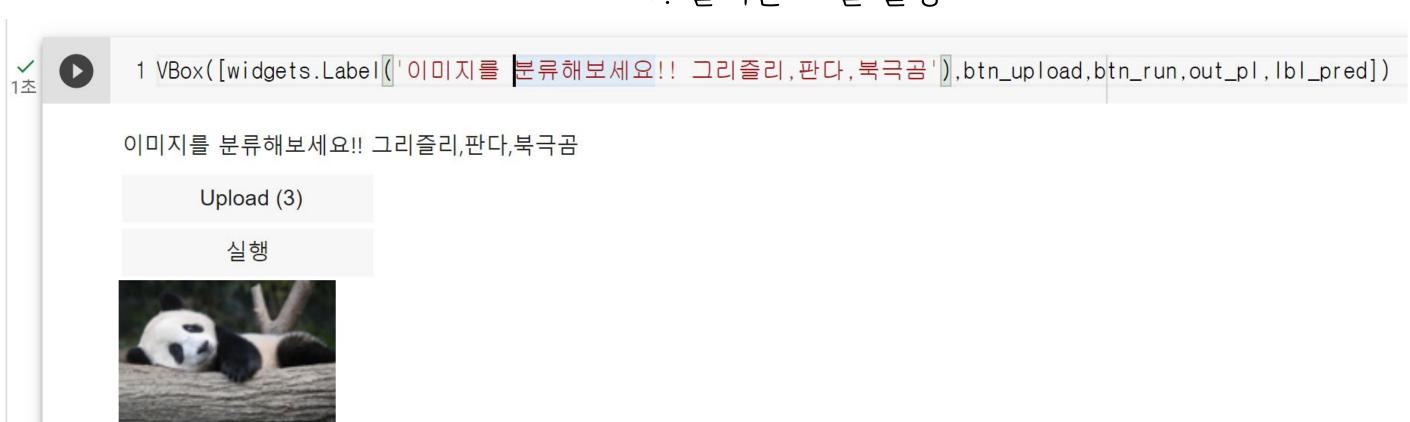
[모델저장]

- 1. 인공지능 모델을 사용할 때마다 학습시키는 것은 매우 비효율적인 방법이다.
- 2. 따라서 학습이 끝난 인공지능 모델은 파일로 저장한 후
- 3. 이 모델을 활용할 때는 단일 파일만 불러와서 활용하면 손쉽게 모델을 활용할 수 있다.
- 4. 우리가 만든 이미지 분류 모델을 드라이브에 파일로 저장하고 이를 불러와보자



Python을 활용한 이미지 분류 모델 제작 프로그래밍

11. 불러온 모델 활용



Predction: panda; Probability: 0.9986

[이미지 업로드 후 모델 활용]

- 1. 제시한 실습 코드를 실행한 후 분류할 모델을 사용자의 컴퓨터에 다운받은 후
- 2. Upload 버튼을 눌러 이미지를 업로드 하고 실행을 누르면
- 3. 모델의 분류 결과와 수치가 출력된다.

2025 홍천고등학교 여름방학 과학캠프

Al를 활용한심화탐구활동

