

팀명:성향차이



MBTI란?







Introduction

Naïve Bayes

KNN

Travel

CNN

MBTI(Myers-Briggs Type Indicator)는 마이어스(Myers)와 브릭스(Briggs)가 스위스의 정신분석학자인 카를 융(Carl Jung)의 심리 유형론을 토대로 고안한 자기 보고식 성격 유형검사 도구이며 시행이 쉽고 간편하여 학교, 직장, 군대 등에서 광범위하게 사용되고 있다.

〈유형별 외모 가꾸기〉







프로젝트 기획 의도







Introduction

Naïve Bayes

KNN

Travel



자소서 기반 MBTI 예측

관상을 통해 MBTI 유추





MBTI프로젝트 진행 과정







Introduction

Naïve Bayes

KNN

Travel

주 파트	서브 파트	차수									
		1회	2회	3회	4회	5회	6회	7회	8회	9회	10회
웹 크롤링	MBTI										
	자소서										
	여행 패키지										
	CNN 이미지										
데이터 취합	MBTI										
	자소서										
정제 작업	MBTI										
	자소서										
	CNN 이미지										
나이브 베이즈	MBTI 예측										
KNN	직무 예측										
CNN	MBTI 예측										



자료 수집







Introduction

Naïve Bayes

KNN

Travel

CNN



MBTI 자료 수집시 사용 패키지

- from selenium import webdriver
- from selenium.webdriver.common.by import By
- from bs4 import BeautifulSoup as bs



자기소개서 자료 수집시 사용 패키지

- from selenium import webdriver
- from selenium.webdriver.common.by import By
- import time
- from bs4 import BeautifulSoup as bs
- from urllib.error import URLError, HTTPError

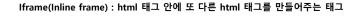


자료 수집 시 문제 해결 방법



브라우저에는 보이는데 엘레먼트를 찾을 수 없다는 내용만 나올 때

- iframe이라는 테그가 있는지 확인 : html안에 다른 html이 있다고 생각하면 된다.
- iframe 안에 html코드를 보려면 프레임 이동을 시켜줘야함
- 작업 끝나면 빠져나오기





Introduction

Naïve Bayes

KNN

Travel

CNN

프레임 id 가져오기

element = driver.find_element_by_id('mainFrame')

프레임 id를 이용하여 프레임 안으로 이동(프레임 안으로 이동하면 안에 있는 태그들에 접근할 수 있다.)

driver.switch_to.frame(element)

MBTI 추출

driver.find_element(By.CSS_SELECTOR,'div#body > div#whole-border > div#whole-body > div#wrapper > div#twocols > div#content-area > div#postListBody > div#post_1 > div.post-back > table.post-body > tbody > tr > td.bcc > div.wrap_rabbit > div.se-viewer > div.se-main-container > div.se-

component.se-text').text

jaso = []

for i in driver.find_elements(By.CSS_SELECTOR,'div.se-main-container > div.se-component.se-text'):

for j in i.find_elements(By.CSS_SELECTOR,'div > div > div > p'):

jaso.append(j.text.strip())

프레임 빠져나오기(처음 상태로 이동)

driver.switch to.default content()









Introduction

Naïve Bayes

KNN

Travel

CNN

데이터 정제

<MBTI>

```
# 특수문자 제거 및 공백 제거
mbti.test = mbti.test.replace(r'[一-龥]|[A-Za-z0-9]|\\[.+?\\]|[\\(\\)]|[^\\w]',r'', regex=True)
mbti.test = mbti.test.replace(r'\s{2,}',r'', regex=True).str.strip()
```

<자소서>



데이터 정제







Introduction

Naïve Bayes

KNN

Travel

CNN

<형태소 분석>





MBTI 분류 나이브베이즈







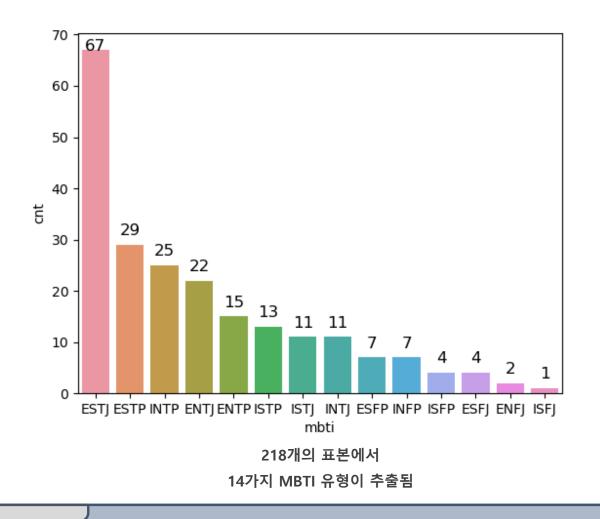
Introduction

Naive Bayes

KNN

Travel

```
from sklearn.feature extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import TfidfTransformer
# 학습데이터
cv = CountVectorizer(tokenizer=okt_pos) # okt활용하여 명사/형용사만 사전으로 만들기
x_train = cv.fit_transform(mbti_g['특성']) # 학습데이터
cv.get_feature_names() # 정보 확인
x_train.toarray()
y_train = mbti_g['MBTI']
# 테스트데이터
tfidf = TfidfTransformer()
x_train = tfidf.fit_transform(x_train) # 위에서 fit작업이 완료되어 사전이 만들어져 있고
transform 작업만하면 됨
# 학습 모델
nb = MultinomialNB()
nb.fit(x_train,y_train)
# 예측
a = []
for i in range(0,len(jasosu['text'])): #변수(자소전_b['text'])
  x_test = cv.transform(pd.Series(jasosu['text'][i]))
  x_test_tfidf = tfidf.transform(x_test)
  a.append(nb.fit(x_train_tfidf, y_train).predict(x_test_tfidf))
```





분류한 MBTI별 직무 분포





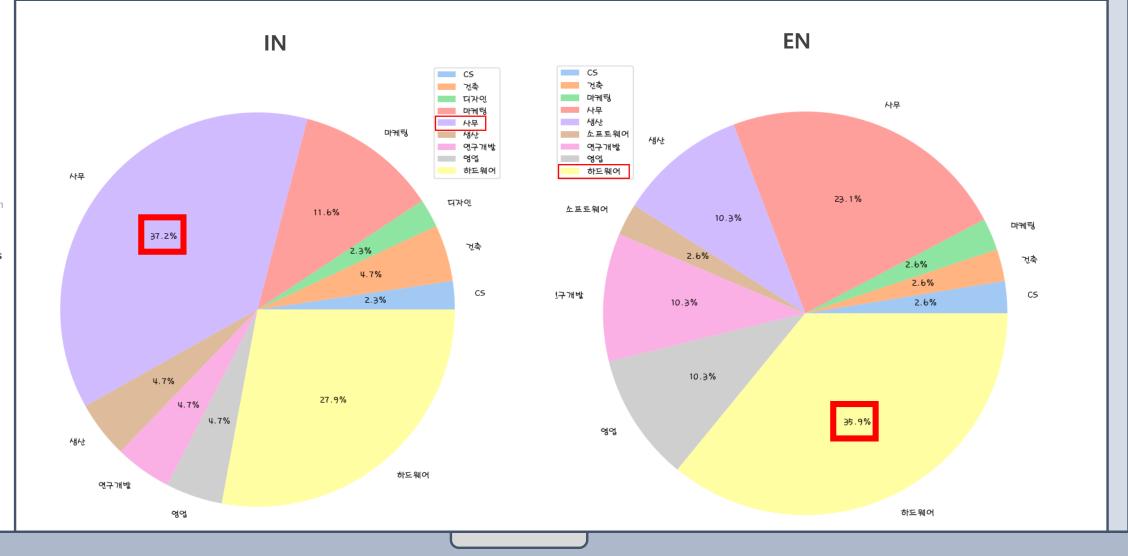


Introduction

Naive Bayes

KNN

Travel





분류한 MBTI별 직무 분포





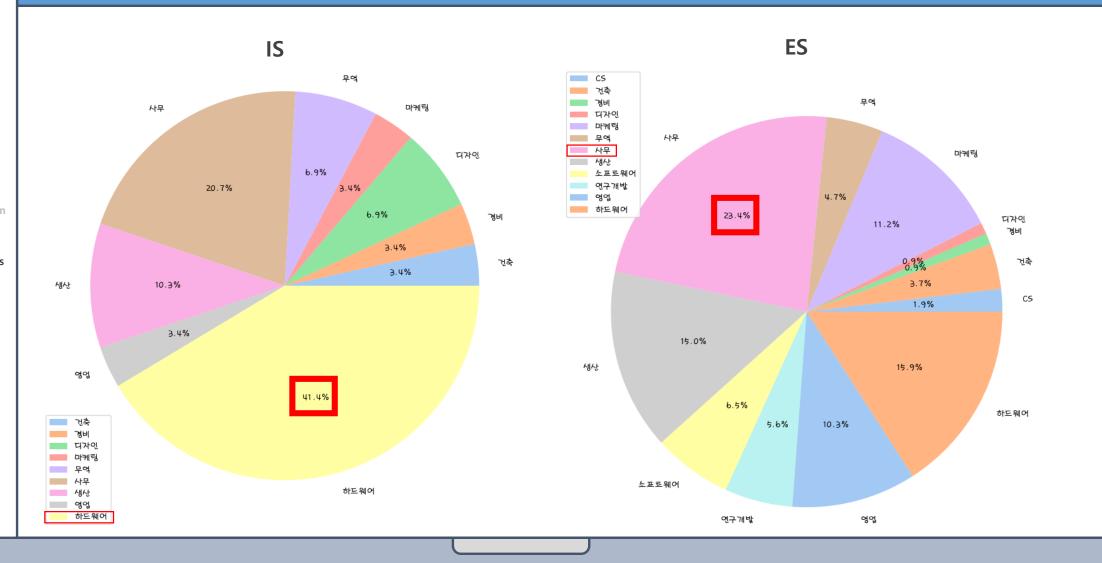


Introduction

Naive Bayes

KNN

Travel





나이브베이즈 분류 키워드 TOP10







Introduction

Naive Bayes

KNN

Travel

CNN

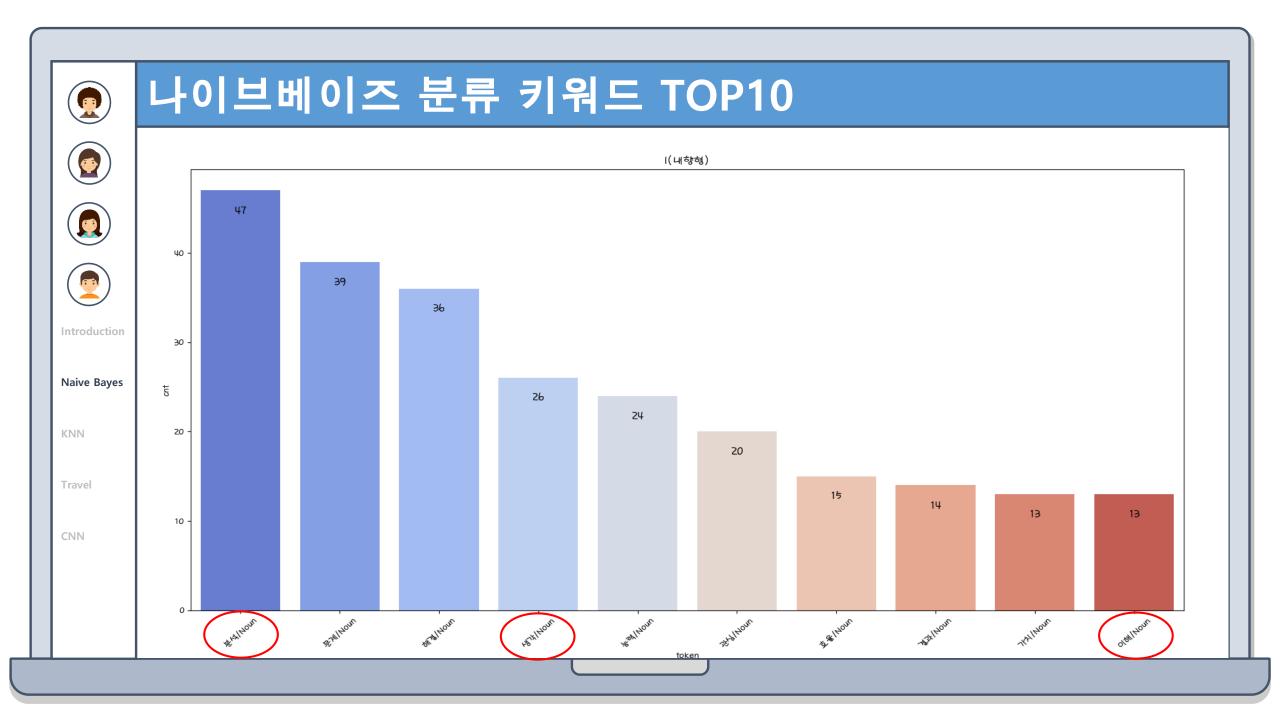
자소서의 단어를 mbti의 특성으로 분류할 때 E특성이랑 I특성에 차이를 알아보기 위해 어떤 단어들을 이용해서 분류했고, 어떤 단어들이 가장 많이 나왔는지 알기 위한 코드 작업 진행

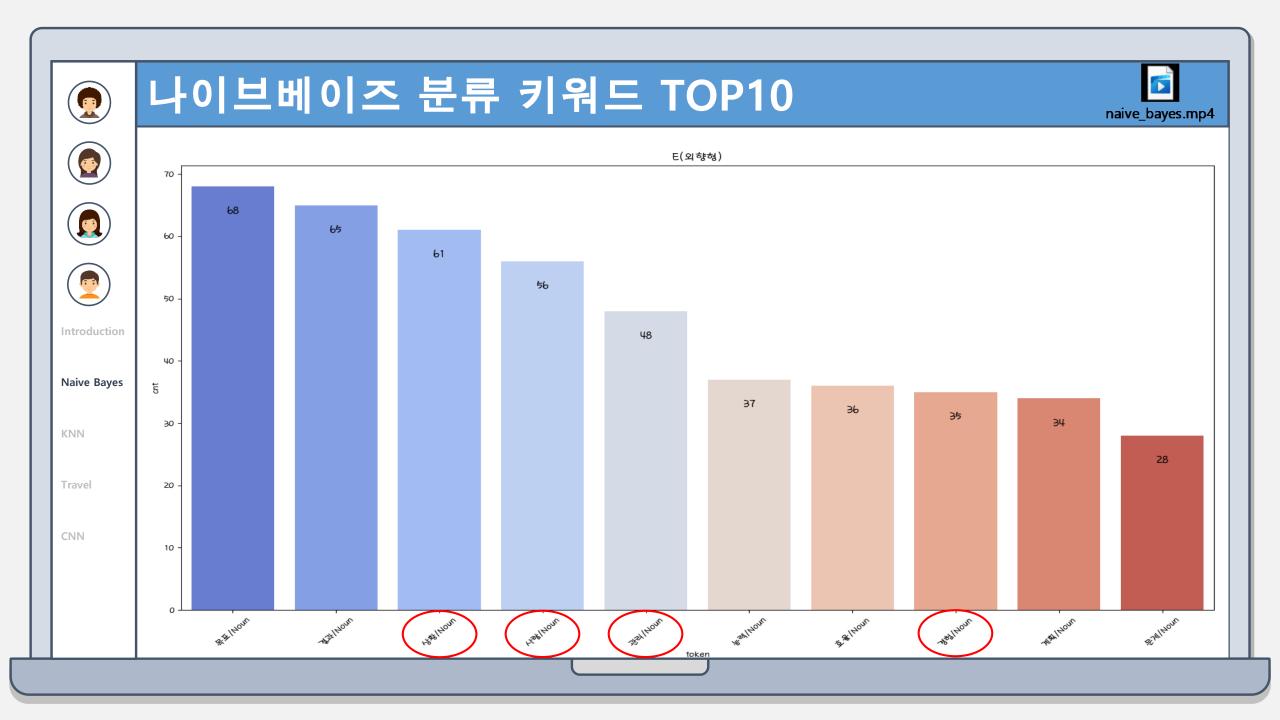
나이브베이즈를 통해 자소서와 훈련데이터인 MBTI 특성을 I(내향형)와 E(외향형)로 나눈 다음, okt_pos를 이용해 각각의 자소서 내용과 특성을 토큰 및 데이터프 레임화

자소서 및 특성 토큰이 각각의 MBTI랑 일치할 수 있게 MBTI를 기준으로 merge를 이용해 합치고 자소서와 특성 사이에 일치하는 토큰을 추출 및 데이터프레임화

위 데이터프레임에서 나온 토큰들을 Counter().most_common()을 이용해 토큰 일치 빈도 상위 20개까지 추출

상위 20개 중 유의미하지 않은 단어를 제거해 I랑 E에서 각각 10개씩 추출







KNN/Kmeans 시각화

KNN 알고리즘 계산식







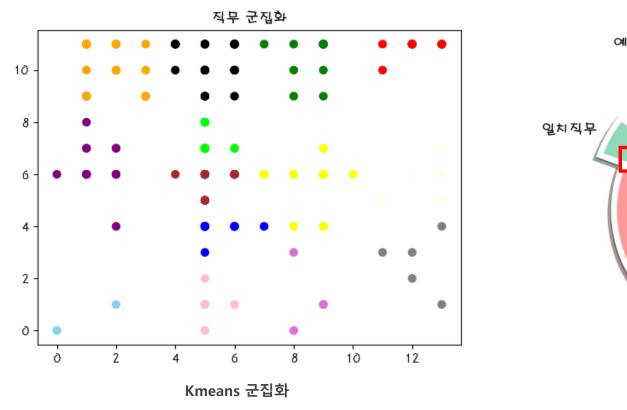
Introduction

Naïve Bayes

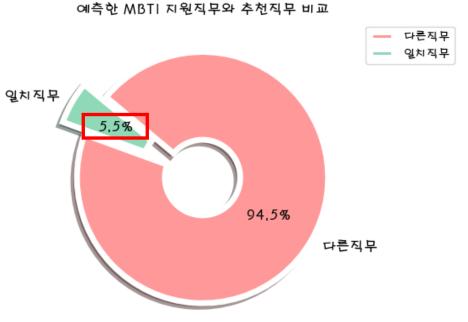
KNN

Travel

CNN



work_mbti_a['result'] = np.sqrt((work_mbti_a['work'] -0)**2 + (work_mbti_a['recommend'] - 0)**2).astype(int)



KNN 시각화





MBTI별 여행지 추천







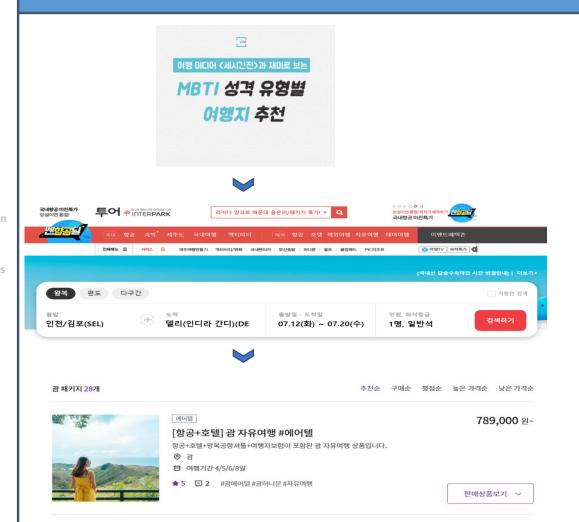
Introduction

Naïve Bayes

KNN

Travel

CNN



MBTI별 여행지 추천을 검색해 각 MBTI별로 어울리는 여행지를 확인함

인터파크에서 평일(7/12~7/20), 주말(7/16~7/24) 나눠서 각 추천 여행 지별 항공권들 최저 가격으로 3개씩 크롤링

패키지 여행을 가고 싶은 분들을 위해서 하나투어에서 추천 여행지별 7월달 패키지 상품을 추천순으로 3개씩 크롤링



MBTI별 여행지 추천







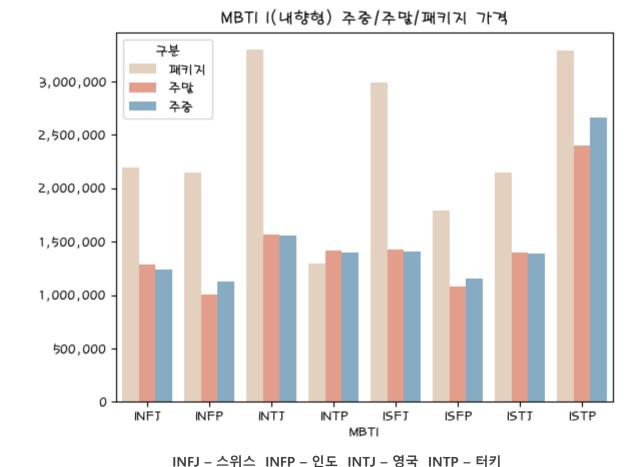
Introduction

Naïve Bayes

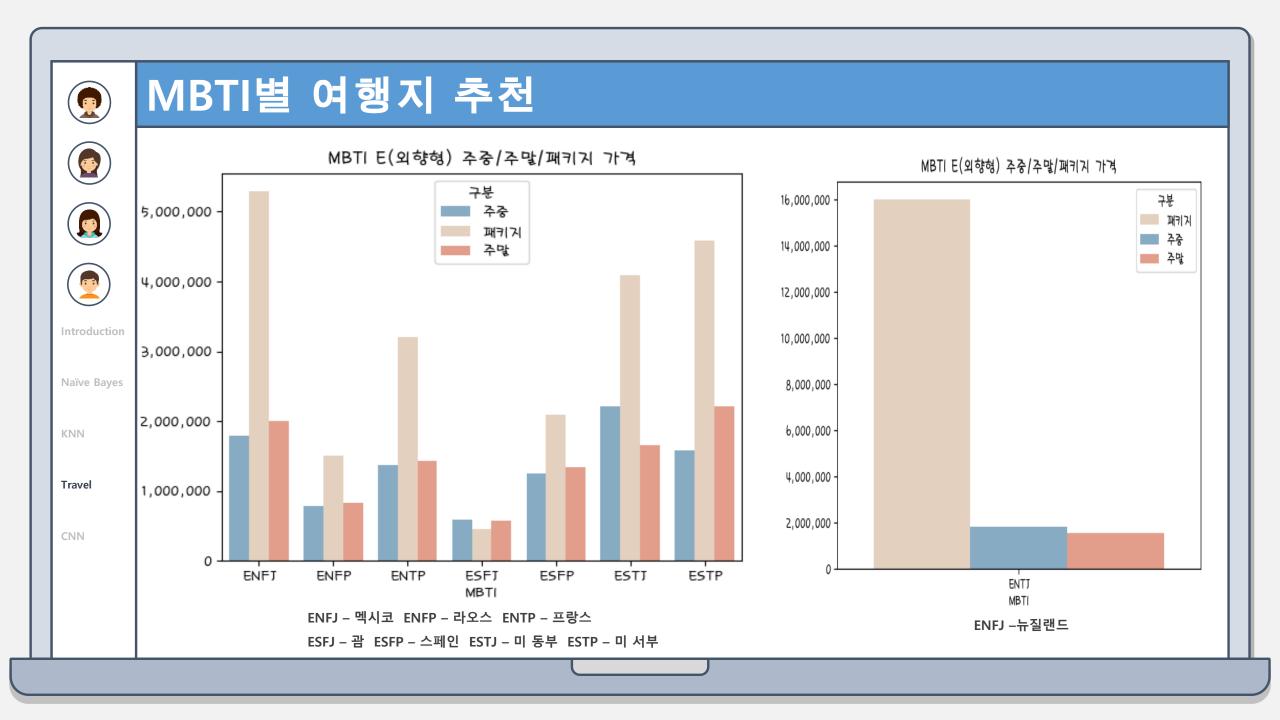
KNN

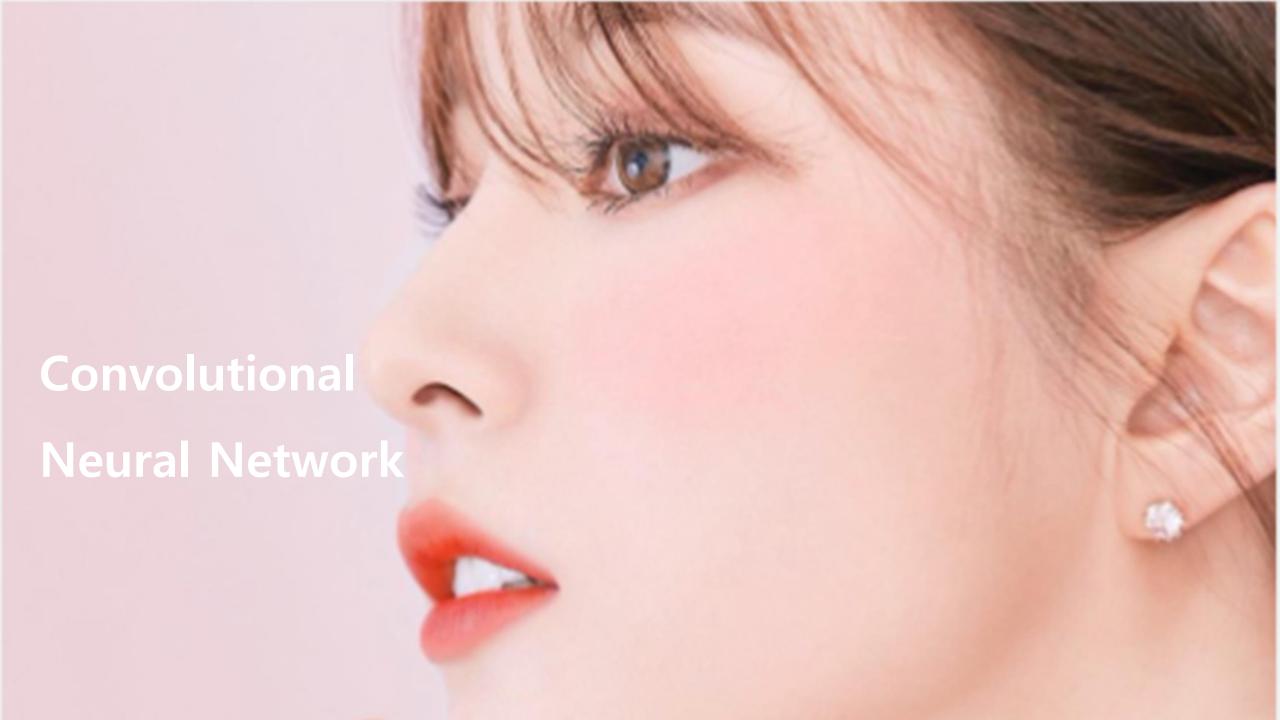
Travel

CNN



ISFJ – 체코 ISFP – 발리 ISTJ – 이탈리아 ISTP - 아이슬란드







CNN 이미지 자료 구성도







Introduction

Naïve Bayes

KNN

Travel



7450.jpg



7451.jpg



7452.jpg



7477.jpg



7479.jpg



7484.jpg



7509.jpg



7511.jpg



7512.jpg



1209.jpg



1212.jpg



1215.jpg



1229.jpg



1234.jpg



1255.jpg



1298.jpg



1301.jpg



1312.jpg









Introduction

Naïve Bayes

KNN

Travel

CNN

```
CNN 코드
```

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Input, Dense, Flatten, Activation

from tensorflow.keras.optimizers import Adam

from tensorflow.keras.applications import VGG16

from tensorflow.keras.applications.vgg16 import preprocess_input

from tensorflow.keras.preprocessing.image import ImageDataGenerator, img_to_array, load_img

import numpy as np

import tensorflow as tf

from tensorflow import keras

from tensorflow.keras.callbacks import ReduceLROnPlateau

model = Sequential([VGG16 MODEL, Flatten(), Dense(500, activation = 'relu'), Dense(16, activation='softmax')])

rlp_cb = ReduceLROnPlateau(monitor='val_loss', mode='min', patience=1, factor=0.3, verbose=1)

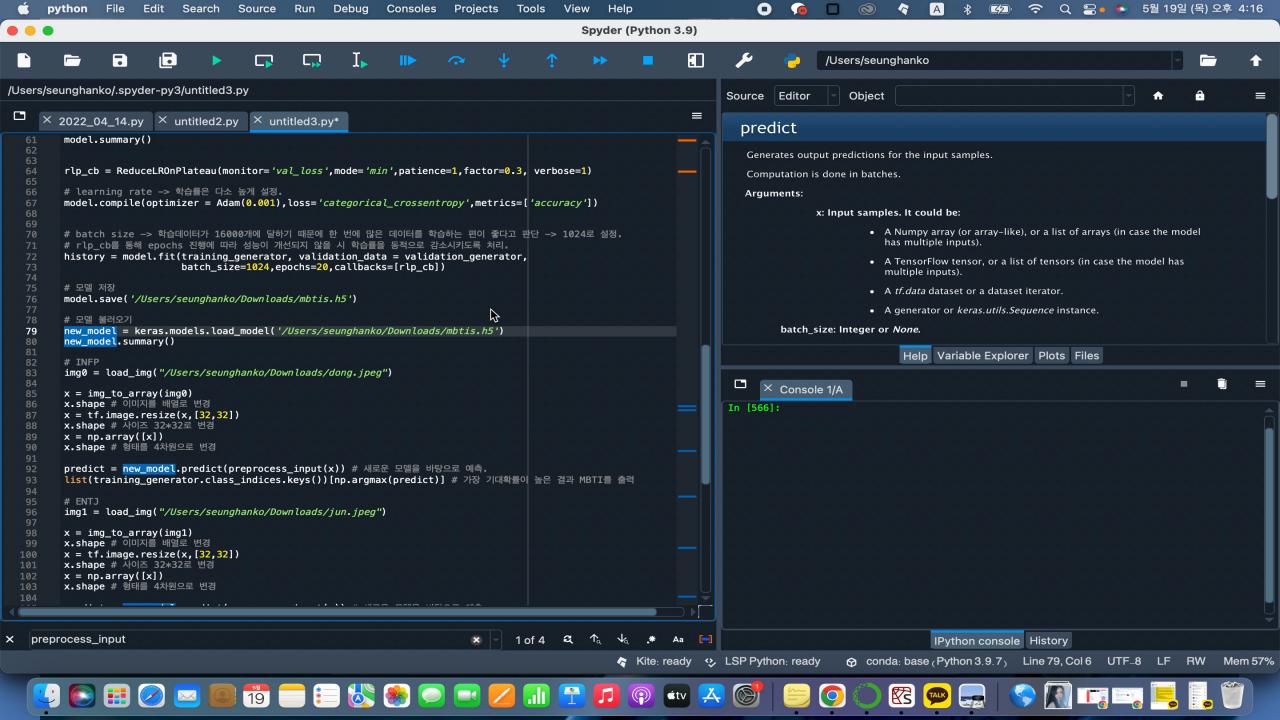
learning rate -> 학습률은 다소 높게 설정.

model.compile(optimizer = Adam(0.001), loss='categorical_crossentropy', metrics=['accuracy'])

batch size -> 학습데이터가 16000개에 달하기 때문에 한 번에 많은 데이터를 학습하는 편이 좋다고 판단 -> 1024로 설정.

rlp_cb를 통해 epochs 진행에 따라 성능이 개선되지 않을 시 학습률을 동적으로 감소시키도록 처리

history = model.fit(training_generator, validation_data = validation_generator, batch_size=1024, epochs=20, callbacks=[rlp_cb])





CNN 결과







Introduction

Naïve Bayes

KNN

Travel

CNN

<MBTI 예측 결과>









실제 : INFP 예측 : INFP 실제 : ENTJ 예측 : ENTJ 실제 : ISFJ 예측 : ISFP 실제 : INFP

예측 : INFP



Notion 링크

https://radical-paint-3ce.notion.site/MBTI-b17d12e5a7df42ec9405785be8934bc6