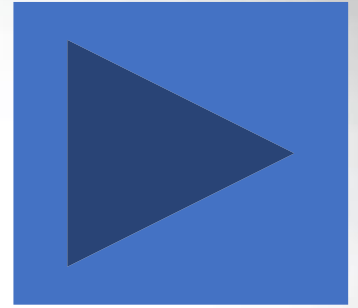


TensorFlow를 활용한 영화 포스터 디자인과 흥행의 상관관계 분석

유니월 빅데이터반 4기
안장현 이웅기 황윤상 김태효 김명중



연사 소개



1. 초딩도 이해하는 **분석 과정** 이 웅 기
2. 고딩도 이해하는 **기술 설명** 김 명 중
3. 유니월 학생을 위한 **기술 시연** 안장현 & 김태효

분석 동기

영화에게 포스터란?

그렇다면,
성공한 영화는
디자인 패턴이 있을까?

관객과의 최초 컨택 포인트

관객에게 전달되는 영화의 첫인상

영화 흥행을 위한 첫 단추

"포스터가 매력적이어야 관객이 온다"

흥행 여부와 직간접적으로 연결되기 때문에

포스터는 영화감독보다

제작자의 영향력이 강하게 작용한다

폰트 크기

인물 수

배경 명암

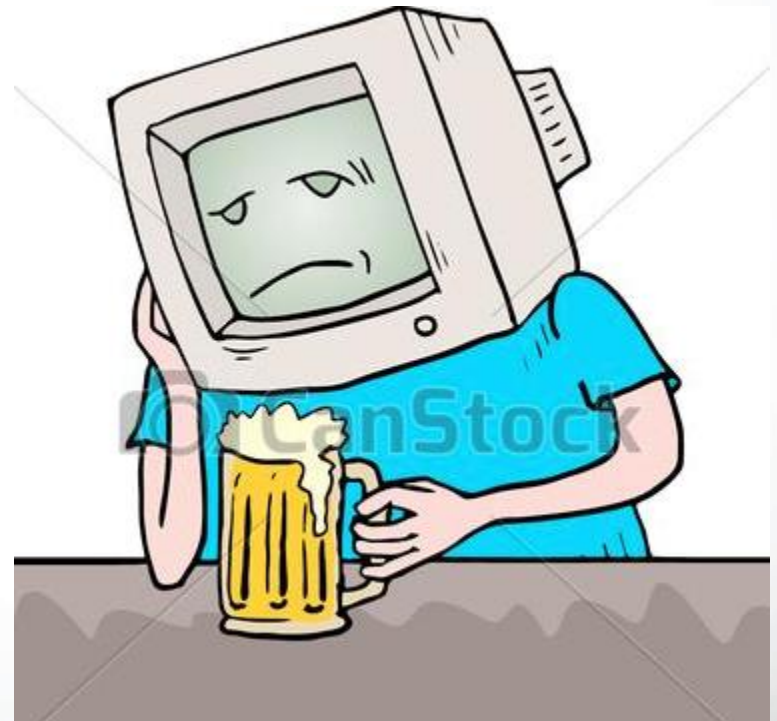
소품

ETC..

야 컴퓨터야!
우린 이미지를 줄게
니가 이미지 안에 뭐가 있는지
인식해봐!!

에이씨 ——
알았엉 해볼게
ㄸ.ㄸ

못하면 태효가 널
부셔 버린댔어!



1-1. 영화 제목 • 포스터 크롤링 : NAVER

인셉션

Inception, 2010

관람객 ? ★★★★★

기자·평론가 ★★★★★ 7.93

네티즌 ? ★★★★★ 9.21 내 평점 ★★★★★ 등록 >

개요 액션, 모험, SF, 스릴러 | 미국, 영국 | 147분 | 2010.07.21 개봉

감독 크리스토퍼 놀란

출연 레오나르도 디카프리오(코브), 와타나베 켄(사이토), 조셉 고든 ... 더보기 >

등급 [국내] 12세 관람가 [해외] PG-13 ?



다운로드

공식사이트 ▾

❤ 4,210



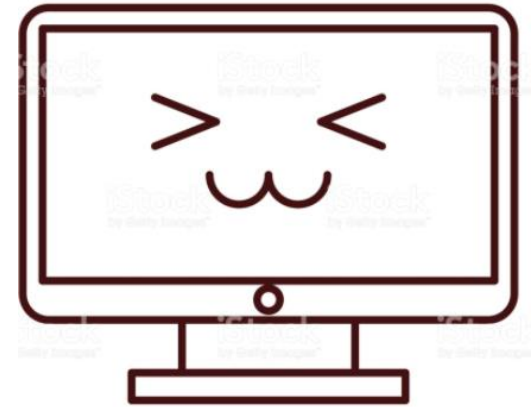
1-2. 영화 내용 크롤링 : 영화진흥위원회

순위	영화명	개봉일	매출액 ₩	매출액 점유율 %	관객수 명	스크린수	상영횟수
1	아바타	2009-12-17	81,455,728,000	7.0%	8,151,952	714	111,860
2	아저씨	2010-08-04	47,101,332,000	4.1%	6,178,248	501	109,853
3	인셉션	2010-07-21	43,394,990,000	3.7%	5,827,444	531	81,891
4	크리미널 마인드	2010-02-04	40,135,723,000	3.5%	5,718,812	583	118,243
5	아이언맨 2	2010-04-29	32,635,195,500	2.8%	4,425,003	921	87,665
6	전우치	2009-12-23	26,345,113,000	2.3%	3,611,472	556	73,175
7	이끼	2010-07-14	25,437,300,500	2.2%	3,350,303	740	58,198
8	포화속으로	2010-06-16	23,831,958,600	2.1%	3,330,324	742	80,091
9	하모니	2010-01-28	21,641,750,000	1.9%	3,018,131	384	77,723
10	범죄자	2010-06-02	22,490,024,000	1.9%	2,985,483	495	67,562

순위	영화명	개봉일	매출액	매출 액 점 유 율	관객수	스크린수	상영횟수
3	인셉션	2010-07-21	43,394,990,000	3.7%	5,877,444	531	81,891

2. 뒤에 [아주 좋은] 컴퓨터를 굴려보자!

칙칙



푹푹

지잉!

Building

Person

Person

Text

Tree

Person

Car

Person

Text





Building

다크나이트의 스케일과
매트릭스의 미래가 만났다
생각을 훔치는 거대한 전쟁

Text

Tree

Person

Person

Car

Person

Person

Text

인셉션
INCEPTION

2. 뽑아낸 데이터를 데이터프레임에 저장

Text

Building

Tree

갯수	5	1	1	1	어찌구저찌구

Person

Person

Car

Person

컬럼은 사람이 아닌
컴퓨터가 정의한다

Person

Text

영화제목	관객 수	사람	건물	나무	자동차	비고
인셉션	5,877,444	5	1	1	1	

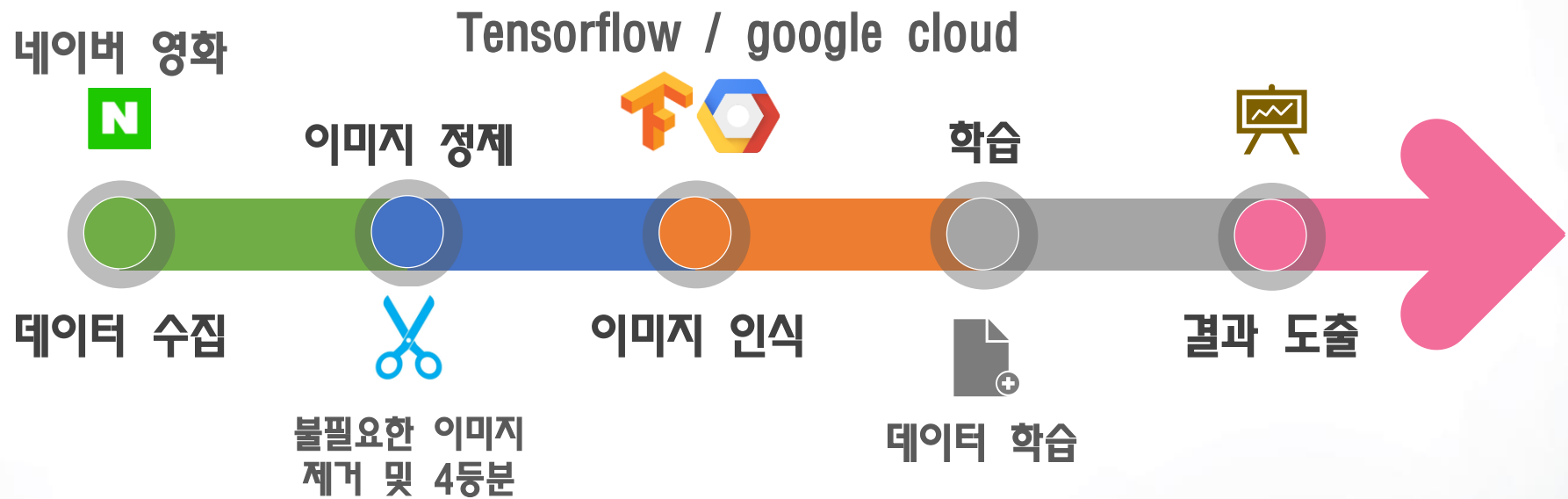
각기 다른 컬럼의
29,000개 포스터 데이터 수집

위의 Train Data를 활용,
미래에 개봉할 영화 관람객 수를 예측!

2. 고딩도 이해하는 기술 설명

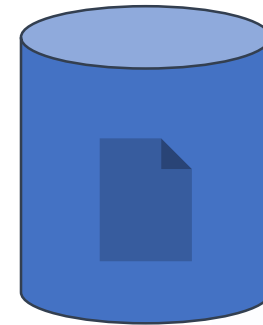
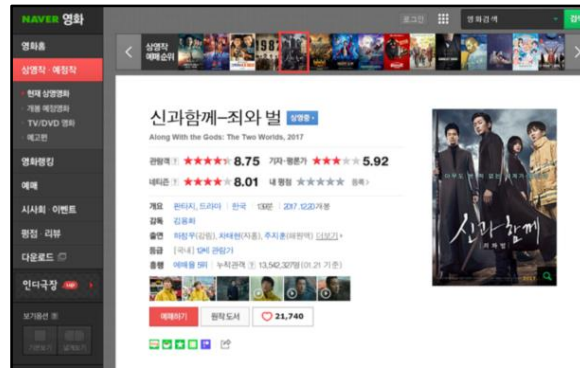
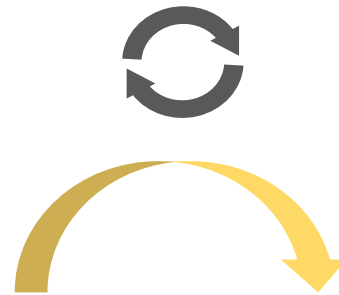
김 명 중

진행 순서



Selenium

포스터 수집



65,000개

문제점

```
***STOP: 0x000000D1 (0x00000000, 0xF73120AE, 0xC0000008, 0xC0000000)

A problem has been detected and Windows has been shut down to prevent damage
to your computer.

DRIVER_IRQL_NOT_LESS_OR_EQUAL

If this is the first time you've seen this Stop error screen, restart your
computer. If this screen appears again, follow these steps:

Check to make sure any new hardware or software is properly installed. If this is a
new installation, ask your hardware or software manufacturer for any Windows updates
you might need.

If problems continue, disable or remove any newly installed hardware or software.
Disable BIOS memory options such as caching or shadowing. If you need to use Safe
Mode to remove or disable components, restart your computer, press F8 to select
Advanced Startup Options, and then select Safe Mode.

*** WXYZ.SYS - Address F73120AE base at C00000000, DateStamp 36b072a3

Kernel Debugger Using: COM2 (Port 0x2f8, Baud Rate 19200)
Beginning dump of physical memory
Physical memory dump complete. Contact your system administrator or
technical support group.
```



메모리 때문에 멈춰버림

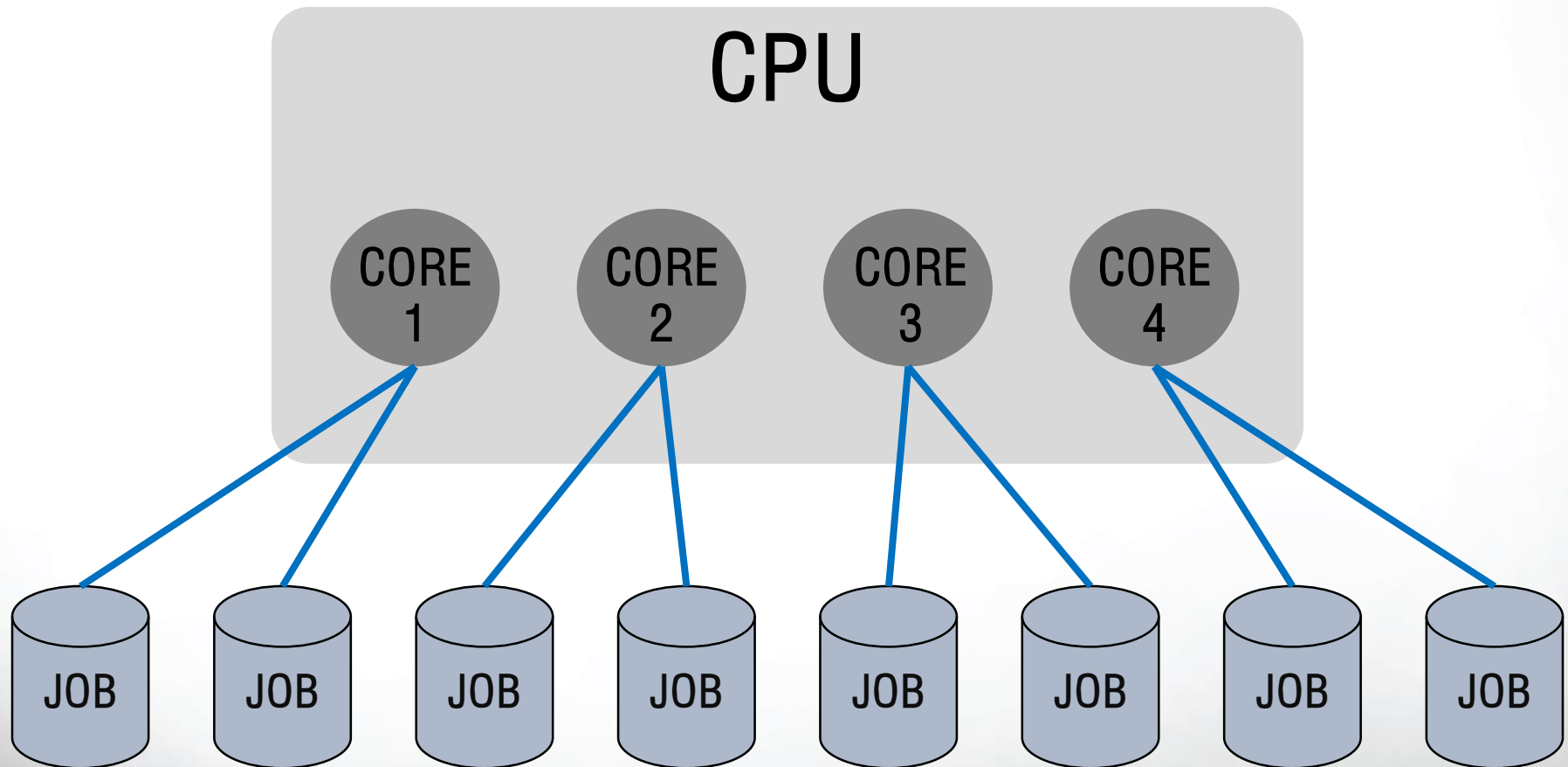


너무 오래걸림

$(75791\text{개} * 2\text{초}) / 60(\text{초}) / 60(\text{분}) = \text{대략 } 42\text{시간}$

멀티프로세싱(multi-processing)

여러 개의 프로세서가 작업을 병렬처리 하는것

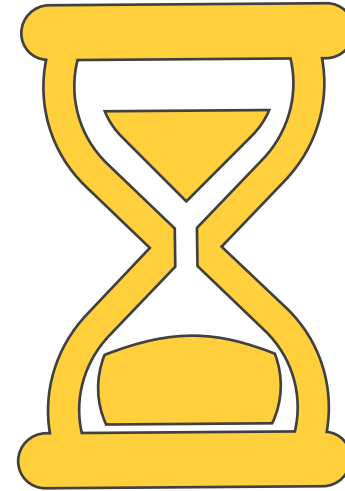


데이터 수집 소요 시간

계산

$(75791\text{개} * 2\text{초}) / 60(\text{초}) / 60(\text{분}) = 42\text{시간}$

8개 프로세스 사용 $42/8 = 5.25\text{시간}$



사용전 42시간



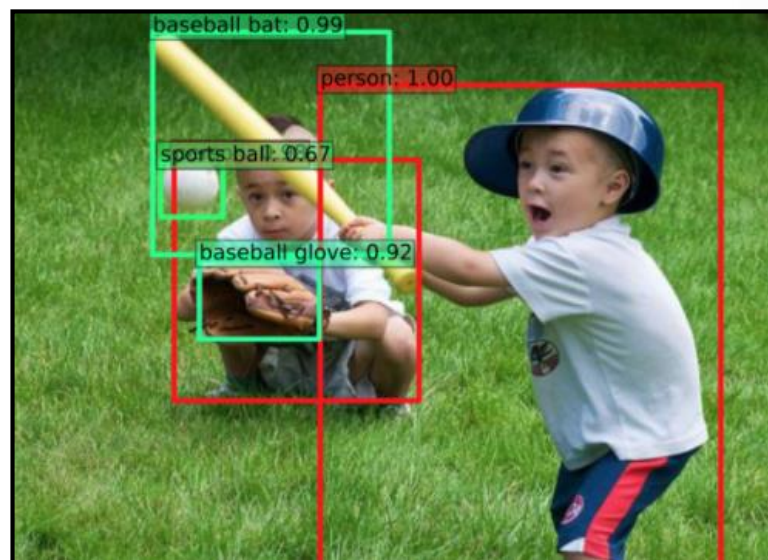
사용후 6시간

현실결과

$30\text{분} + 40\text{분} + 1\text{시간} + 1\text{시간} + 1\text{시간} + 1\text{시간} + 1\text{시간} + 40\text{분} = 6\text{시간}$



500size 이하

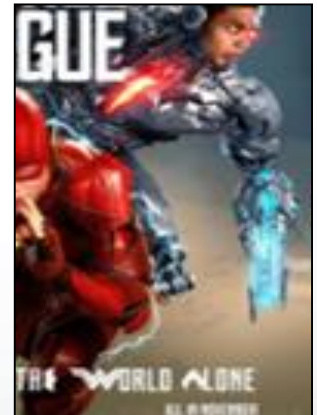
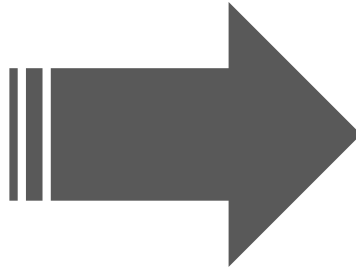


큰 사이즈



≠





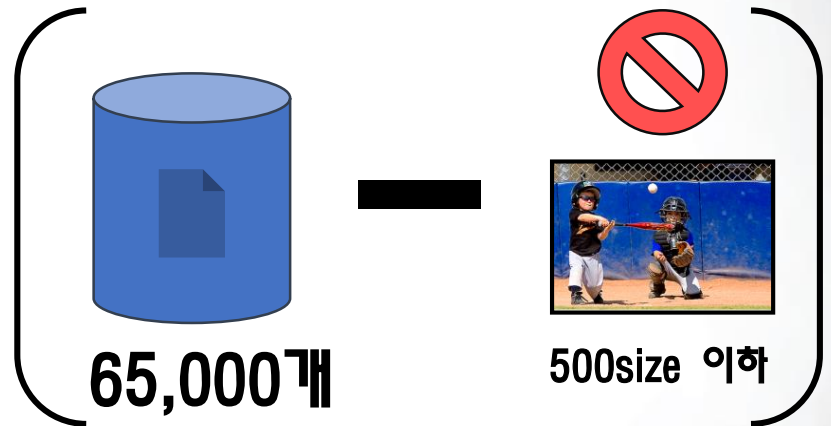


29,000개

=

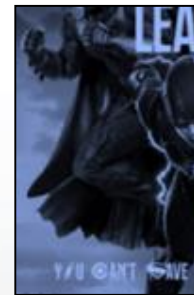
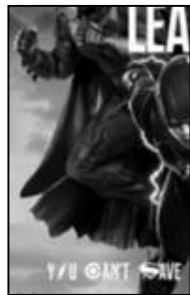
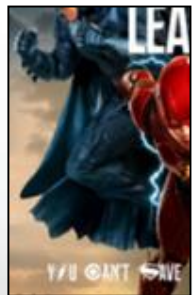


+



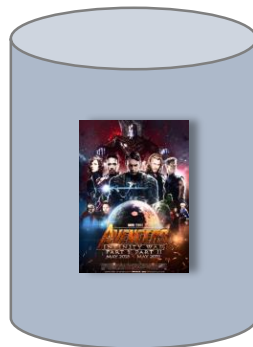


Object detection

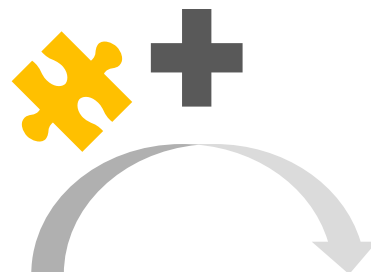




Object detection



이미지 데이터



dictionary

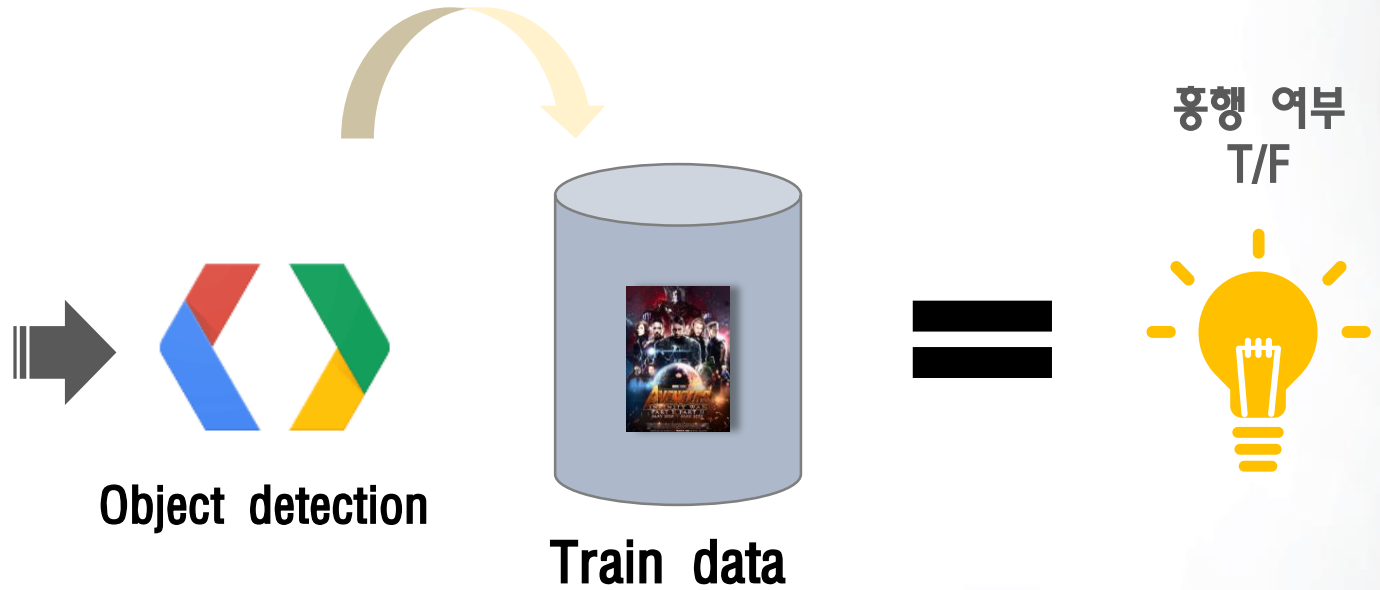
학습 train



제목	사람	사람 사이즈	제목 사이즈	동물	사물	관객수	흥행 T/F
스펙터	2명	1200x1600	120x220	null	null	200	T

개수나, 포스터의 차지하는
비율 등으로 컬럼 나열

개봉예정영화



3. 유니월 학생을 위한 기술 시연

안 장 현 & 김 태 효

소스코드(1)

```
# step.2 : tensorflow로 분석(Softmax Classifier)

X = tf.placeholder(tf.float32, [None, 82]) # 열 갯수 82에 맞춤
Y = tf.placeholder(tf.int32, [None, 1]) # 열 갯수 1에 맞춤

Y_one_hot = tf.one_hot(Y, 7) # 0 ~ 6 : 총 7개
Y_one_hot = tf.reshape(Y_one_hot, [-1, 7])

W = tf.Variable(tf.random_normal([82, 7]), name='weight')
b = tf.Variable(tf.random_normal([7]), name='bias')

logits = tf.matmul(X, W) + b
hypothesis = tf.nn.softmax(logits) # softmax 사용

cost_i = tf.nn.softmax_cross_entropy_with_logits(logits=logits, labels=Y_one_hot)
cost = tf.reduce_mean(cost_i)

train = tf.train.GradientDescentOptimizer(learning_rate=0.0001).minimize(cost) # 학습률 0.0001 설정

prediction = tf.argmax(hypothesis, 1) # 예측결과 확률값이 제일 크면 1 리턴
correct_prediction = tf.equal(prediction, tf.argmax(Y_one_hot, 1)) # 예측값과 실제값 일치정도
accuracy = tf.reduce_mean(tf.cast(correct_prediction, tf.float32)) # 정확도

###추측해볼 데이터
test = pd.read_csv("/Users/janghyeonan/it_project/src/data/youuplz.csv")
test1 = test.set_index('id')
test2 = test1.reindex(columns = res3.columns).loc[:, 'zebra'].fillna(0)
test2.values.shape
test2
```

소스코드(2)

```
#둘러보자
with tf.Session() as sess:
    sess.run(tf.global_variables_initializer())
    #print(sess.run(Y_one_hot, feed_dict = {Y:y_data}))
    for step in range(100001):
        sess.run(train, feed_dict={X: x_data, Y: y_data})
        if step % 1000 == 0:
            loss, acc = sess.run([cost, accuracy], feed_dict={X: x_data, Y: y_data})
            print("Step: {:5}\tLoss: {:.3f}\tAcc: {:.2%}".format(step, loss, acc))

# 예측 부분
pre = sess.run(hypothesis, feed_dict = {X: test2.values}) # test(개봉예정작)
n = [i for i in range(7) if pre[0][i] == max(pre[0])]
if n[0] == 0:
    print('\n예상관객수 : 100명 이하')
elif n[0] == 1:
    print('\n예상관객수 : 100명 초과, 100,000명 이하')
elif n[0] == 2:
    print('\n예상관객수 : 100,000명 초과, 500,000명 이하')
elif n[0] == 3:
    print('\n예상관객수 : 500,000명 초과, 1,000,000명 이하')
elif n[0] == 4:
    print('\n예상관객수 : 1,000,000명 초과, 5,000,000명 이하')
elif n[0] == 5:
    print('\n예상관객수 : 5,000,000명 초과, 10,000,000명 이하')
else:
    print('\n예상관객수 : 10,000,000명 초과')
```


예측 샘플



결과는?

Step: 87000	Loss: 1.892	Acc: 50.58%
Step: 88000	Loss: 1.892	Acc: 50.58%
Step: 89000	Loss: 1.892	Acc: 50.58%
Step: 90000	Loss: 1.891	Acc: 50.58%
Step: 91000	Loss: 1.891	Acc: 50.58%
Step: 92000	Loss: 1.891	Acc: 50.58%
Step: 93000	Loss: 1.891	Acc: 50.58%
Step: 94000	Loss: 1.891	Acc: 50.58%
Step: 95000	Loss: 1.890	Acc: 50.58%
Step: 96000	Loss: 1.890	Acc: 50.58%
Step: 97000	Loss: 1.890	Acc: 50.58%
Step: 98000	Loss: 1.890	Acc: 50.58%
Step: 99000	Loss: 1.890	Acc: 50.58%
Step: 100000	Loss: 1.889	Acc: 50.63%

예상관객수 : 100명 초과, 100,000명 이하

감사합니다