


A I C O N P I T U R E S

F I L M S

W&P Coffee House
1000 1st Ave
New York, NY 10010

I'm here 

	Small	Large
COFFEE		
ESPRESSO	\$2.00	\$2.50
Americano	\$2.00	\$2.50
Hot Latte	\$2.50	\$3.00
Hot Mocha	2.50	3.00
Hot Cappuccino	2.50	3.00
Hot Flat White	2.50	3.00
TEA		
Assorted Brewed Tea	\$2.00	\$2.50
English Breakfast Tea	\$2.00	\$2.50
Traditional Earl Grey Tea	2.00	2.50
SAVORY		
Assorted Pastries	\$1.50	\$2.00
Croissants	1.50	2.00
Muffins & Buns	1.50	2.00
Handmade Breads	1.50	2.00
Hot Soup, Salad, Sandwiches	4.00	5.00
Specialty Beer		
Beer	\$3.00	\$4.00
Hard Cider	3.00	4.00
Wine	3.00	4.00
Large		
Large Coffee	3.00	3.50
Large Tea	2.50	3.00
Large Savory	2.00	2.50



머신러닝을 활용한 영화 흥행 예측

박경 * | 김병 * | 최혜 * | 김유 *



A table of Contents

1

기획

자료 조사 및 계획 수립, 역할 분담

2

데이터 수집

데이터 수집 과정 및 필터링

3

데이터 전처리

데이터에 있는 문자들을 수치화

4

ML(머신러닝)

머신러닝 과정 및 정확도

5

결론

영화에서 가장 중요한 변수 도출

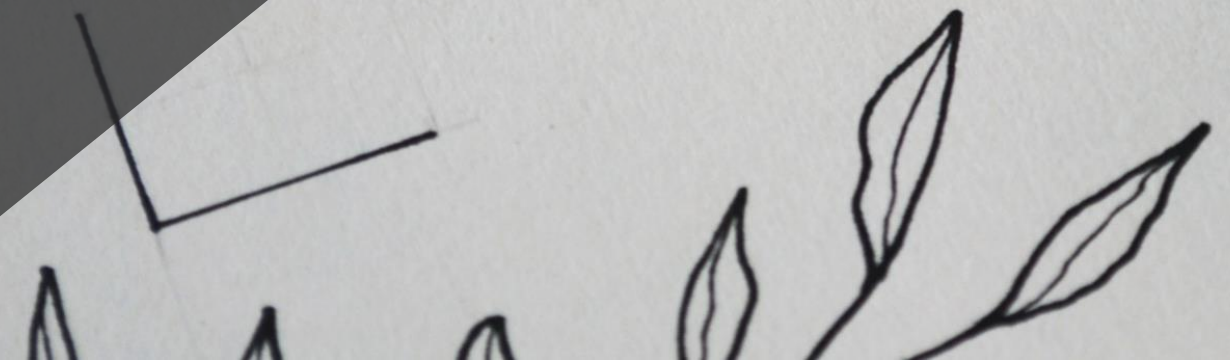
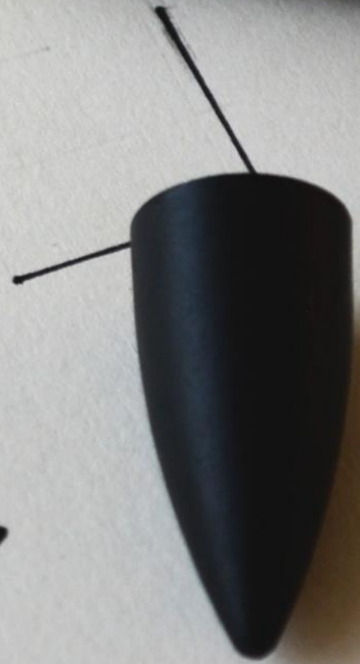
Part 1

기획

D W D V Z Z
4 5 6 7 8 9 3
11 12 13 14 15 16 1
18 19 20 21 22 23 24
25 26 27 28 29 30

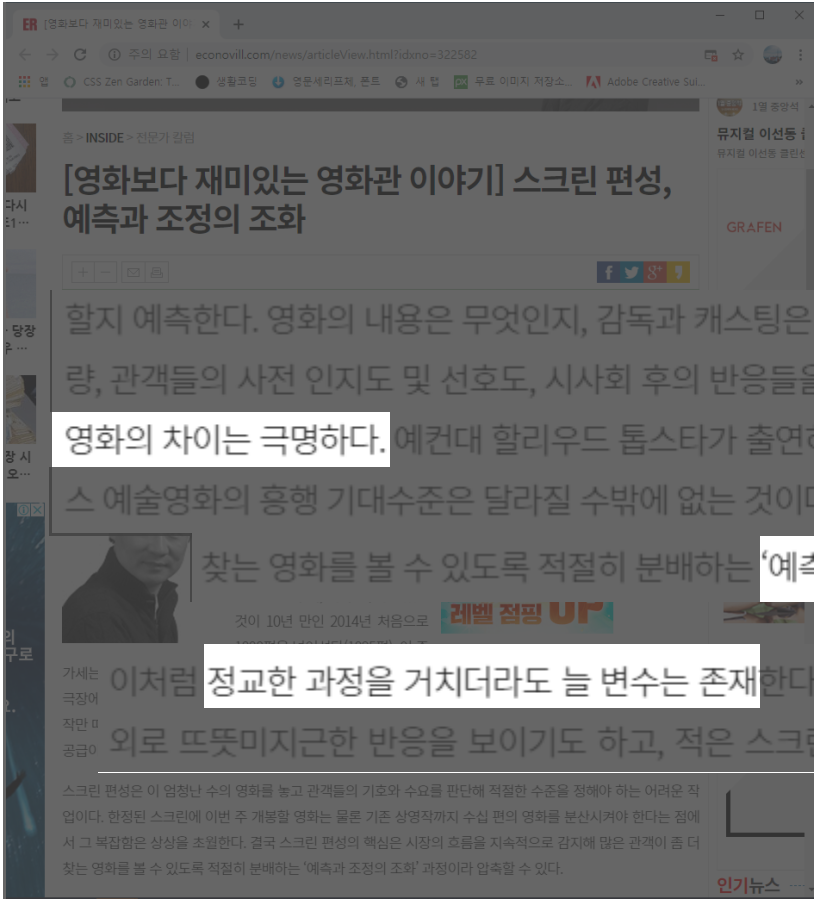
back to school
10-14

12	D
13	W
14	D
15	V
16	
17	
18	





영화산업과 AI



“영화 산업의 핵심인 스크린 편성”



“위너 브라더스, 영화 흥행 예측하는 AI 도입”

“ 머신 러닝을 활용한 ”
영화 흥행 예측



Part 2

데이터 수집







독립 변수 1

»» 제작국가

한국영화진흥위원회 API 서비스 이용, crawling

»» 관람 등급

한국영화진흥위원회 API 서비스 이용, crawling

»» 배급사

한국영화진흥위원회 API 서비스 이용, crawling





독립 변수 2

장르

한국영화진흥위원회 API 서비스 이용, crawling

감독 스타성

연출 영화 수집 후 각 영화의 관람객 수집 후 평균 냄

배우 스타성

출연 영화 수집 후 각 영화의 관람객 수집 후 평균 냄





종속 변수

» 관람객 수

API_KOBIS (영화권통합전산망)사이트 이용



KOBIS
영화관입장권통합전산망

영화정보 ▼ | 검색어를 입력하세요. 🔍

영화정보검색 박스오피스 테마동계 공식동계 온라인상영권 박스오피스 고객센터

독립·예술영화 박스오피스 +

3 시네마 천국
Cinema Paradiso

개봉일 1990-07-07
제작상태 개편
영화구분 예술영화

관람등급 전체관람가
상영시간 124분 0초
제작국가 프랑스

감독 중세페 토르나토레
장르 드라마, 멜로/로맨스
배급사 주식회사 왓차(주)립엔터...

영화명	매출액	관객수	증감률
1 어누-아시키 : 히어로...	6,977,840 원	966 명	-
2 서지 아웃	7,172,240 원	939 명	-
3 시네마 천국	4,385,000 원	868 명	-
4 건조 여름보	6,093,500 원	748 명	-
5 언어의 정원	2,540,000 원	502 명	-
6 유령선	3,488,840 원	434 명	+ 2
7 엠문4: 더 파이널	3,008,000 원	376 명	+ 1
8 라리퀸	2,934,040 원	367 명	+ 1
9 바이올렛 에버가든 ...	2,824,560 원	357 명	+ 6
10 범왕 벤베어	2,592,000 원	324 명	+ 8

2020년 04월 23일(목) 기준

박스오피스

주요지표

Chart

독립·예술영화

공지사항

박스오피스 | 독립·예술영화 박스오피스 | 실시간 예매율 | 독립·예술영화 예매율 | 좌석 점유율 | 온라인 이용순위 | 해외 박스오피스





이용 플랫폼

KOBIS
영화관입장권통합전산망

영화정보 ▼ | 검색어를 입력하세요.

 **NAVER**

Daum

 **영화관입장권통합전산망** **오픈API**



기본데이터 수집

KOFIC 영화관입장권통합전산망 오픈API

이용안내 | **OPEN API** | 개발급/관리 | 이용약관

제공 서비스 > 다운로드 > 튜토리얼

OPEN API > 제공서비스

제공 서비스
영화관입장권통합전산망이 제공하는 오픈API 서비스 모음입니다. 사용 가능한 서비스를 확인하고 서비스별 인터페이스 정보를 조회합니다.

1 박스오피스	• 일별 박스오피스 • 주간/주말 박스오피스
2 공동코드조회	• 공동코드 조회
3 영화정보	• 영화목록 • 영화 상세정보
4 영화사정보	• 영화사목록 • 영화사 상세정보
5 영화인정보	• 영화인목록 • 영화인 상세정보

일별 박스오피스 API 서비스

특정 일자 상영작들의 박스오피스 정보를 영화구분(다양성영화, 상업영화), 한국/외국 구분, 상영지역 등의 조건을 통해 조회합니다. REST/SOAP 방식 중 선택적으로 호출가능하며 REST 방식의 응답형식은 XML과 JSON를 지원합니다.(URI의 extension으로 구분)

1. REST 방식

- 기본 요청 URL : <http://www.kobis.or.kr/kobisopenapi/webservice/rest/boxoffice/searchDailyBoxOfficeList.xml> (또는 .json)
- 요청 parameter : 몇번의 요청 인터페이스 정보를 참조하여 GET 방식으로 호출

2. SOAP 방식

- 요청 URL : <http://www.kobis.or.kr/kobisopenapi/webservice/soap/boxoffice>
- WSDL URL : <http://www.kobis.or.kr/kobisopenapi/webservice/soap/boxoffice?wsdl>
- Operation : searchDailyBoxOfficeList

3. 인터페이스

- 요청 인터페이스

요청 변수	값	설명
key	문자열 (필수)	발급받은키 값을 입력합니다.

— 2016 ~ 2018 영화 데이터 수집

- 각 영화의 기본 데이터 확보
 - 장르
 - 제작국가
 - 배급사
 - 관람 등급

— 영화의 감독, 배우 데이터 리스트 확보

※ 자세한 내용은
<http://www.kobis.or.kr/kobisopenapi/homepg/main/main.do>
 (구글:영화진흥위원회 API 검색) 사이트에서 확인할 수 있습니다.



배우 및 감독 스타성 데이터 수집



로버트 다우니 주니어 영화



영화 인피니티 워 - Daum 검색

search.daum.net/search?mil_suggest=btn&w=tor&DA=SBC&q=영화+인피니티+워

영화 인피니티 워

통합검색 블로그 사이트 동영상 웹문서 백과사전 브랜치 카페 더보기 기간전체

관련 영화 어벤져스 인피니티 워 2020 개봉 영화 영화 어벤져스 엔드게임 어벤져스 영화 순서 영화 협탄 마블

다음 영화 어벤져스: 인피니티 워 Avengers: Infinity War. 2018 제작

요약 미국 액션 외 2018.04.25 개봉 12세이상관람가 149분

감독 안소니 루소 조 루소

출연 로버트 다우니 주니어, 조슈 브롤린, 마크 러팔로, 톰 히들스턴 [더보기](#)

줄거리 새로운 조항을 이룬 어벤져스, 역대 최강 빌런 타노스에 맞서 세계의 운명... [더보기](#)

누적관객수 11,212,710 명 (2020.04.23, 역대 20위, 영화진흥위원회 영화관입장권통합전산망) [연예 영화 순위](#)

배기건 [\[해외\] 빌런](#) 아마존과 페이스북이 '주목할 할프 2'에 등장한 이유

[영화보기](#) 3번 시사회

동영상 8건 포토 55건

원정 시리즈 [더보기](#)

평점&15자평 [평점추가](#)

★★★★★ 2020.04.22

★★★★★ 2020.04.21

7.6 2,829명 >

볼 때마다 눈물남 2020.04.09

블로그

[이웃 어드] 심베르스: 인피니티 워 <인피니티 스톤을 향한 무한 대결> 2020.04.03

영화 어벤져스: 인피니티 워 출연진 1/7 <>

로버트 다... 조슈 브롤린 마크 러팔로 톰 히들스턴 크리스 메...
아이언맨 토니 스타크 역 타노스 역 앙크루투스 로카 역 영단 이블리 키스틴 역

영화 어벤져스: 인피니티 워 메가전 1/20 <>

[\[해외\] 칼럼](#) 아마존과 페이스북이 '...
+ 빌런(Vulture)와 리프린트 계약을 맺고 번역한 콘텐츠들 편집한 글입니다. [주목할 할프 2] 인터넷 속으로(이 하 주목할 할프 2) 첫 장면에서...

영화순위 1/5 <>

저 산 너머	어벤져스: 엔드...	어벤져스: 인피...	미어 스테이
★ 7.9 12.5%	★ 7.8 10.4%	★ 7.6 8%	★ 6.6 5.3%
20.04.30 개봉	19.04.24 개봉	18.04.25 개봉	20.04.22 개봉
억1989만명 추계	억1121만명 추계		

2020.04.24, 10:30 기준, 영화진흥위원회 제공

관련 현재상영작 개봉예정작 역대관객순위 영화관

분야별 검색어 04.24, 11:18

— 2016 ~ 2018 영화 리스트 사용

— Crawling Selenium
기본 URL 검색 page



배우 및 감독 스타성 데이터 수집

— 주연 배우의 주연작품 수집
감독 연출작품 수집

DoM 로버트 다우니 주니어 영화

통합검색 블로그 이미지 사이트 동영상 카페 브런치 뉴스 더보기

관련 로버트 다우니 주니어 출연작 로버트 다우니 주니어 리즈 로버트 다우니 주니어 부인 로버트 다우니 주니어 가족 로버트 다우니 주니어 키

인물 마이프로필 관리

로버트 다우니 주니어 (Robert Downey Jr.) 영화배우

출생 1965년 4월 4일 (만 55세), 미국 뉴욕
신체 174cm
가족 아버지 로버트 다우니 시니어, 배우자 수잔 다우니
학력 산타모니카 고등학교 중퇴
데뷔 1970년 영화 '파운드'
수상 2019 톰 초이스 어워드 영화 액션 남자배우 부문 외 13건
사이트 인스타그램, 트위터, 페이스북, 팬카페
사진제공 : Gettyimageskorea

영화 ^ 방송 v 앨범 v

설록 홈즈 3 설록 홈즈역 2021

닥터 두리틀 2020.01.08

어벤저스: 엔드게임 토니 스타크/ 아이언맨 역 2019.04.24

영화 75건 전체보기 >

주니어 | 다중영화 x +

personid=8357

영화 연예

홈 현재상영/개봉예정 박스오피스 빠른예매 뉴스 내방점 예매내역

메인 영화참여작

로버트 다우니 주니어 영화참여작 (75) 역렬순

주연 (56)

설록 홈즈 3, 2021 설록 홈즈 역 (주연)

닥터 두리틀, 2020 존 두리틀 박사 역 (주연) (기획)

어벤저스: 엔드게임, 2019 토니 스타크/ 아이언맨 역 (주연) [수상] 2019 MTV 영화 & TV 상 최고의 영웅상

어벤저스: 인피니티 워, 2018 아이언맨/토니 스타크 역 (주연)

코로나19 극복을 위한 일자리 안정자금 추가지원
10인 미만 사업장 월 최대 7만원, 10인 이상 사업장 월 최대 4만원

검색하려면 여기에 입력하십시오.

오전 11:17 2020-04-24



관람객수 데이터 수집

The screenshot shows the KOBIS (Korea Copyright Commission) website interface for movie data collection. The main focus is on the movie '어벤저스: 인피니티 워 (Avengers: Infinity War)'. The interface includes a search bar, navigation tabs, and a detailed data table for the movie's performance.

어벤저스: 인피니티 워 (Avengers: Infinity War)

기본정보 | **통계정보** | 상영현황정보

박스오피스

일지별 스크린수, 상영횟수, 관객수, 매출액 등의 통계정보입니다. 엑셀을 통해서 세부적인 통계를 확인할 수 있습니다.

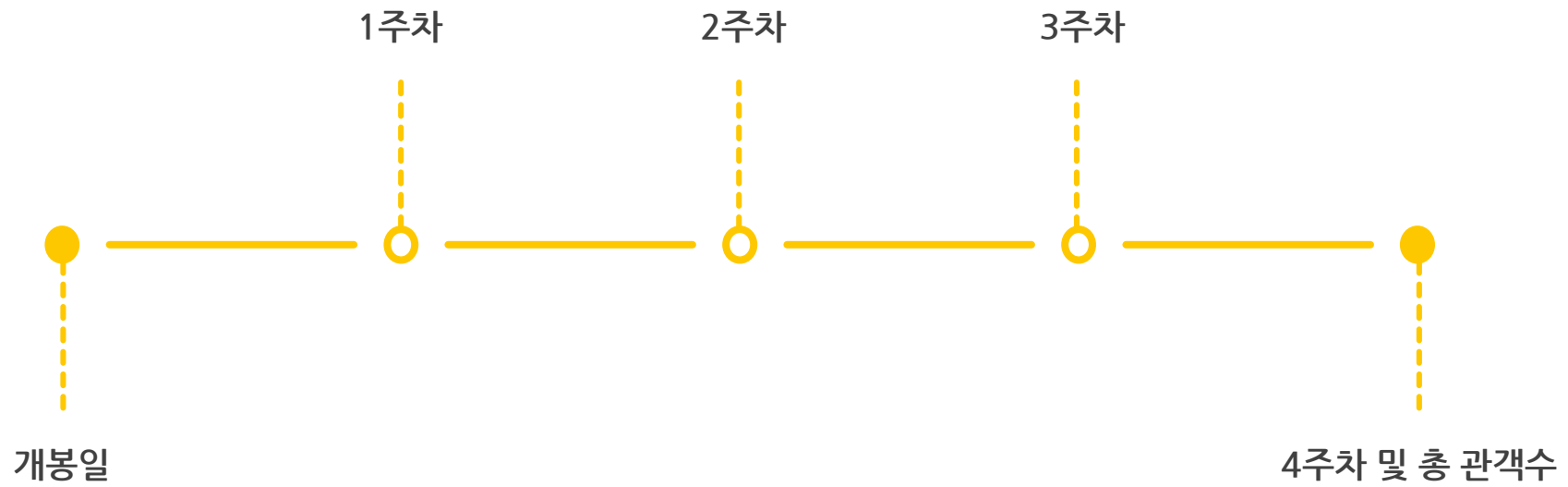
개봉 10일 | 최신 10일 *출처: KOBIS(발권)통계 *통계기준일: 2020.04.23 **엑셀**

날짜	스크린수	상영횟수	매출액	관객수	누적매출액	누적관객수	순위
개봉이전	1	1	9,360,000	624	9,360,000	624	N/A
개봉1일(04/25)	2,460	11,429	6,929,271,100	980,042	6,938,631,100	980,666	1
개봉2일(04/26)	2,394	11,628	5,195,838,900	595,668	12,134,470,000	1,576,334	1
개봉3일(04/27)	2,460	12,188	7,026,026,870	739,908	19,160,496,870	2,316,242	1
개봉4일(04/28)	2,553	13,184	12,300,421,261	1,333,307	31,460,918,131	3,649,549	1
개봉5일(04/29)	2,548	12,945	10,307,417,400	1,114,946	41,768,335,531	4,764,495	1
개봉6일(04/30)	2,462	12,255	4,360,209,000	495,787	46,128,544,531	5,260,282	1
개봉7일(05/01)	2,141	10,730	6,013,244,200	717,324	52,141,788,731	5,977,606	1
개봉8일(05/02)	2,114	10,387	2,454,649,900	276,967	54,596,438,631	6,254,573	1
개봉9일(05/03)	2,138	10,189	2,340,404,400	265,430	56,936,843,031	6,520,003	1
개봉10일(05/04)	2,156	10,648	3,916,977,653	419,443	60,853,820,684	6,939,446	1

— 주 차별 관람객 추이 확인을 위한
박스 오피스 데이터 수집



영화 관객수 데이터 수집



Part 3

데이터 전처리





독립변수

movieNm	genres	nations	companys	audits	actors	directors
셜록: 유령신부	드라마범죄	영국	메가박스중앙(주)플러스엠	12세이상관람가	베네딕트 컴버배치 마틴 프리먼	더글라스 맥키넨
굿 다이노	코미디 애니메이션 어드벤처	미국	월트디즈니컴퍼니코리아(주)	전체관람가	레이몬드 오초아 잭 브라이트	피터 손
나를 잊지 말아요	멜로/로맨스	한국	씨제이이엔엠(주)	15세이상관람가	정우성 김하늘	이윤정
유스	드라마	영국 이탈리아 프랑스 스위스	그린나래미디어(주)	15세이상관람가	마이클 케인 하비 케이틀 레이철 와이즈 폴 다노 제인 폰다	파올로 소렌티노
헤이트풀8	서부극(웨스턴) 액션 스릴러	미국	(주)누리픽쳐스	청소년관람불가	사무엘 L. 잭슨 커트 러셀 제니퍼 제이슨 리 윌튼 고킨스 데미안 비쉬어 팀 로스 ...	쿠엔틴 타란티노



독립변수

movieNm	genres	nations	companys	audits	actors	directors
셜록: 유령신부	드라마범죄	영국	메가박스중앙(주)플러스스엠	12세이상관람가	베네딕트 컴버배치 마틴 프리먼	더글라스 맥키닌
굿 다이노	코미디 애니메이션 어드벤처	미국	월트디즈니컴퍼니코리아(주)	전체관람가	레이몬드 오조아잭 브라이트	피터 손
나를 잊지 말아요	멜로/로맨스	한국	씨제이이엔엠(주)	15세이상관람가	정우성 김하늘	이윤정
유스	드라마	영국 이탈리아 프랑스 스위스	그린나래미디어(주)	15세이상관람가	마이클 케인 하비 케일러 레이철 와이즈 폴 다노 제인 폰다	파올로 소렌티노
헤이트풀8	서부극(웨스턴) 액션 스릴러	미국	(주)누리픽처스	청소년관람불가	사무엘 L. 잭슨 쿠트 러셀 제니퍼 제이슨 리 윌튼 고킨스 데미안 비쉬어 팀 로스 ...	쿠엔틴 타란티노

movieNm	nations	companys	audits	actors	directors
셜록: 유령신부	1	3	2	592948	1278885
굿 다이노	2	1	1	997635	1330181
나를 잊지 말아요	3	3	3	1181598	427107
유스	1	1	3	17616	35836
헤이트풀8	2	1	4	48565	137580

종속변수

movieNm	w1	w2	w3	w4
셜록: 유령신부	2016-01-09	2016-01-16	2016-01-23	2016-01-30
굿 다이노	2016-01-14	2016-01-21	2016-01-28	2016-02-04
나를 잊지 말아요	2016-01-14	2016-01-21	2016-01-28	2016-02-04
유스	2016-01-14	2016-01-21	2016-01-28	2016-02-04
헤이트풀8	2016-01-14	2016-01-21	2016-01-28	2016-02-04

movieNm	w1	w2	w3	w4	tot
셜록: 유령신부	1086912	1251033	1275444	1278172	1278894
굿 다이노	736477	1103163	1294868	1325983	1330183
나를 잊지 말아요	391766	423624	425020	427057	427197
유스	32158	52365	65062	74440	87204
헤이트풀8	92834	111986	116484	117396	122266

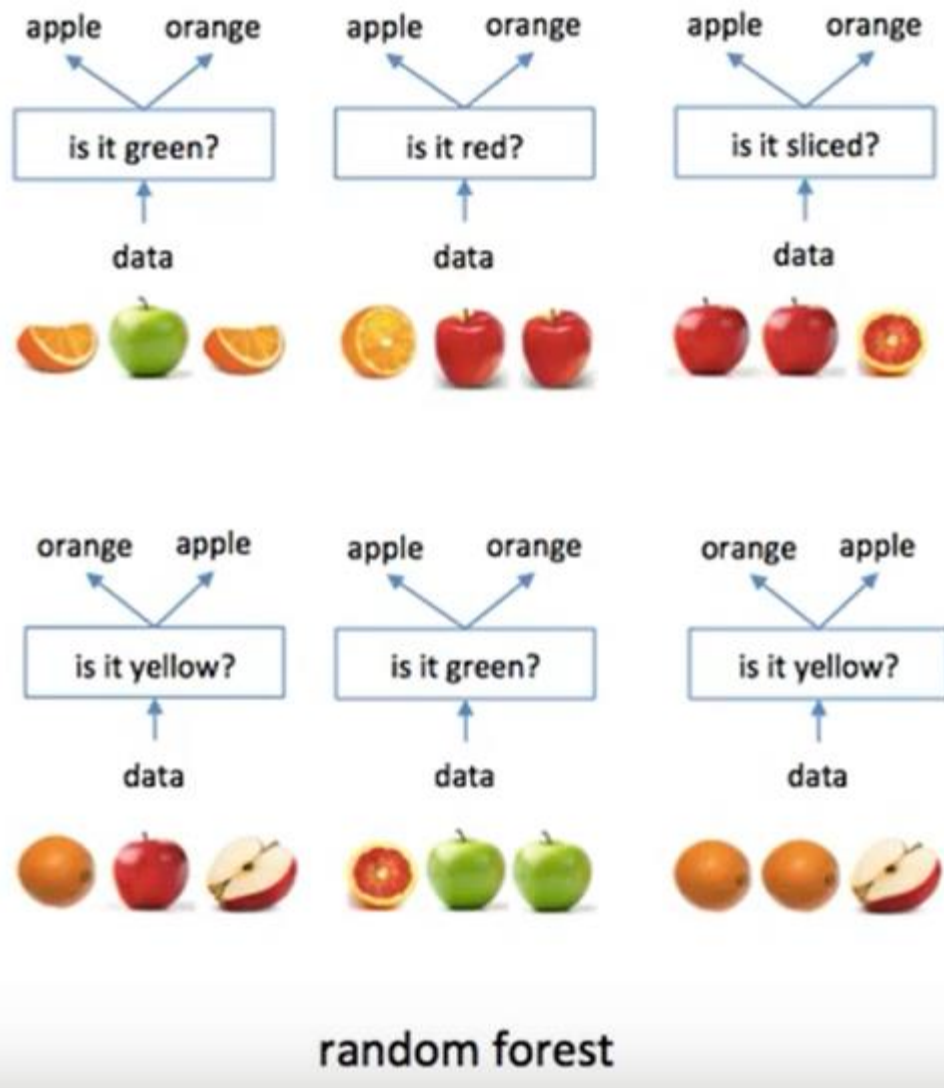
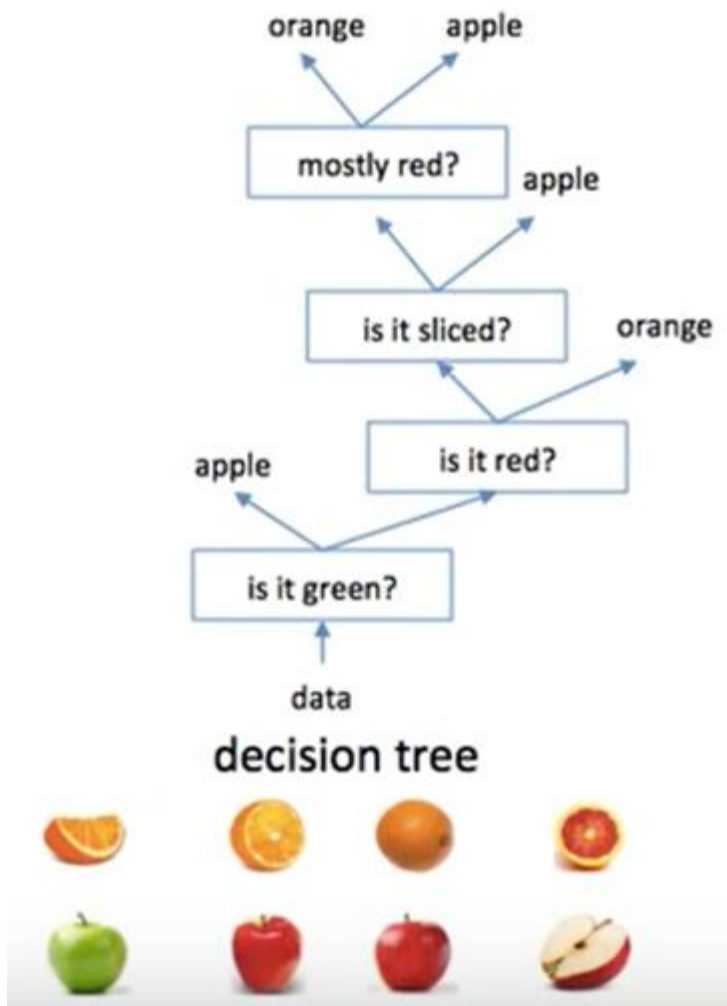
Part 4

ML(머신러닝)





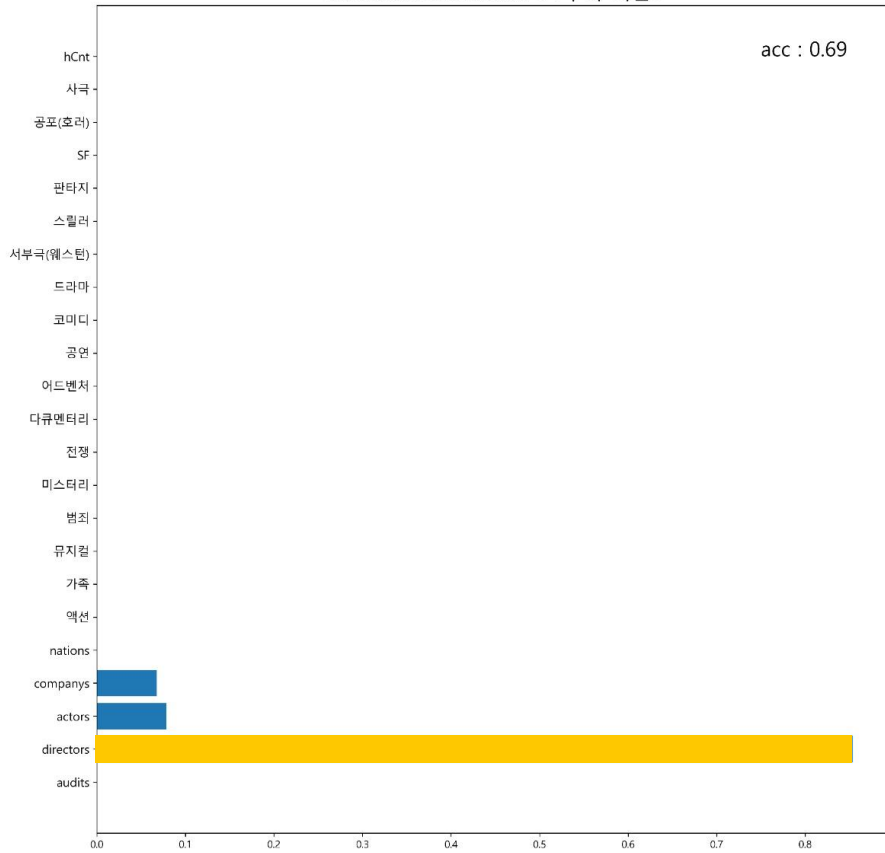
Decision Tree & Random Forest



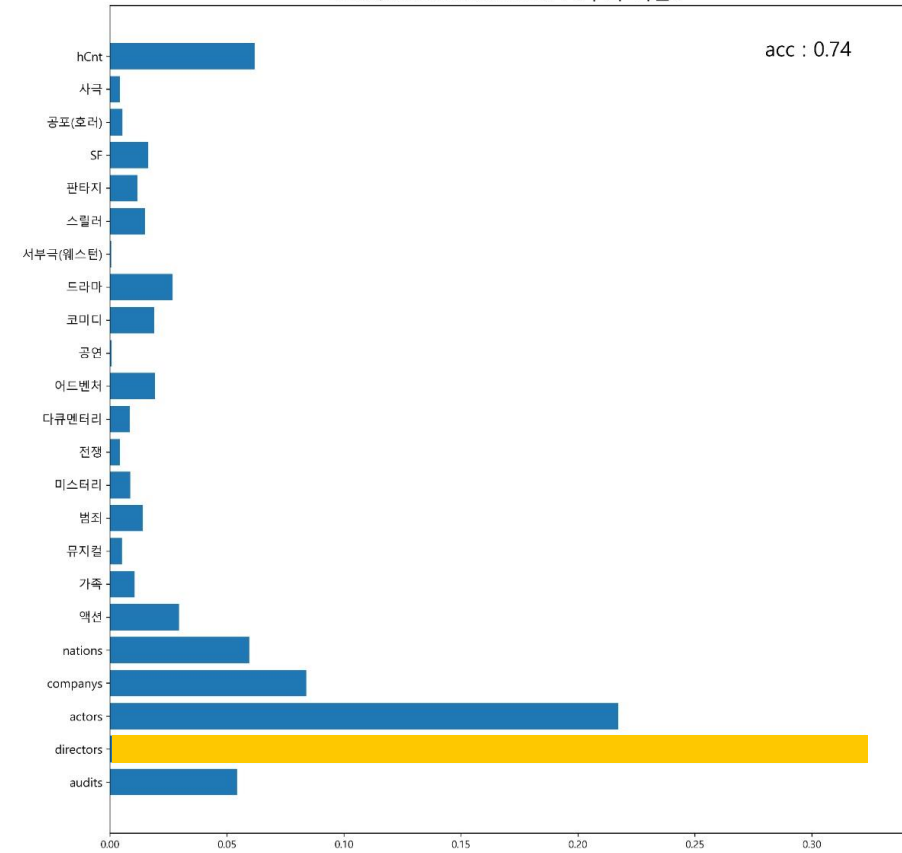


개봉 초기 ~ 1주차

<DecisionTreeClassifier : 1주차 기준>



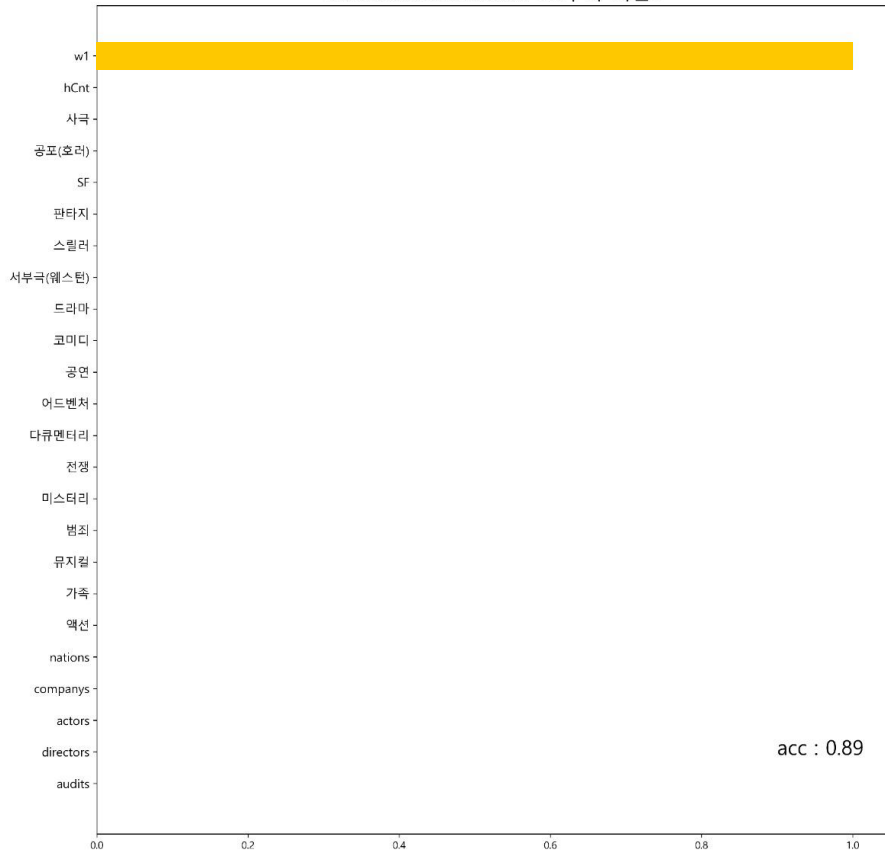
<RandomForestClassifier : 1주차 기준>



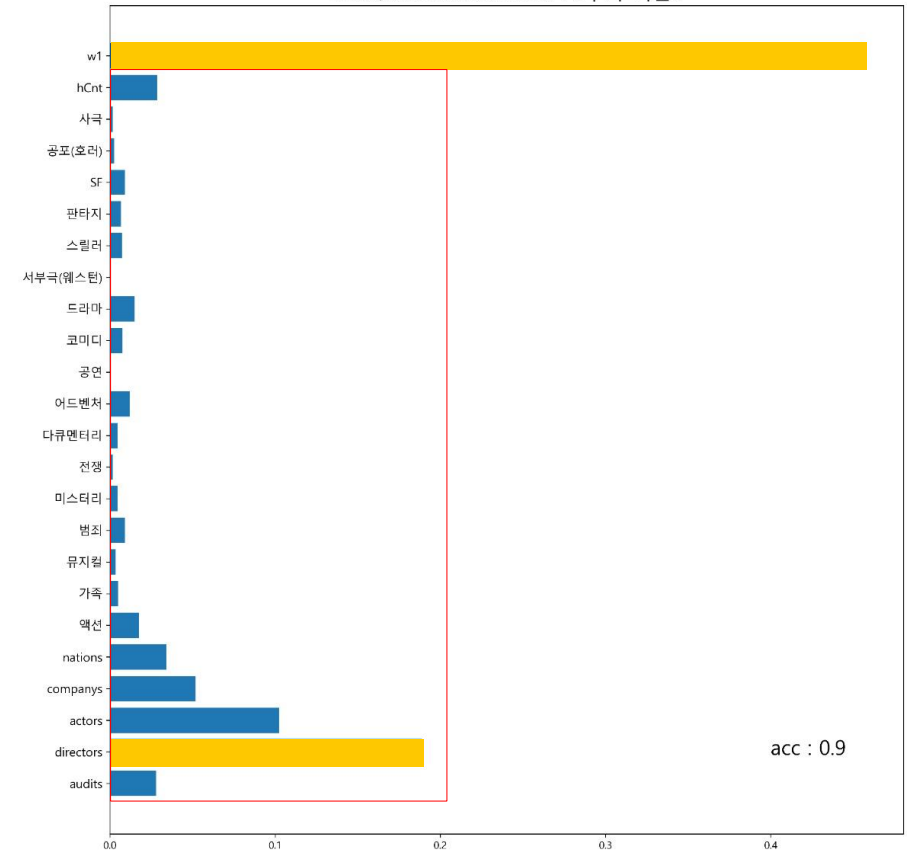


2주차

<DecisionTreeClassifier : 2주차 기준>



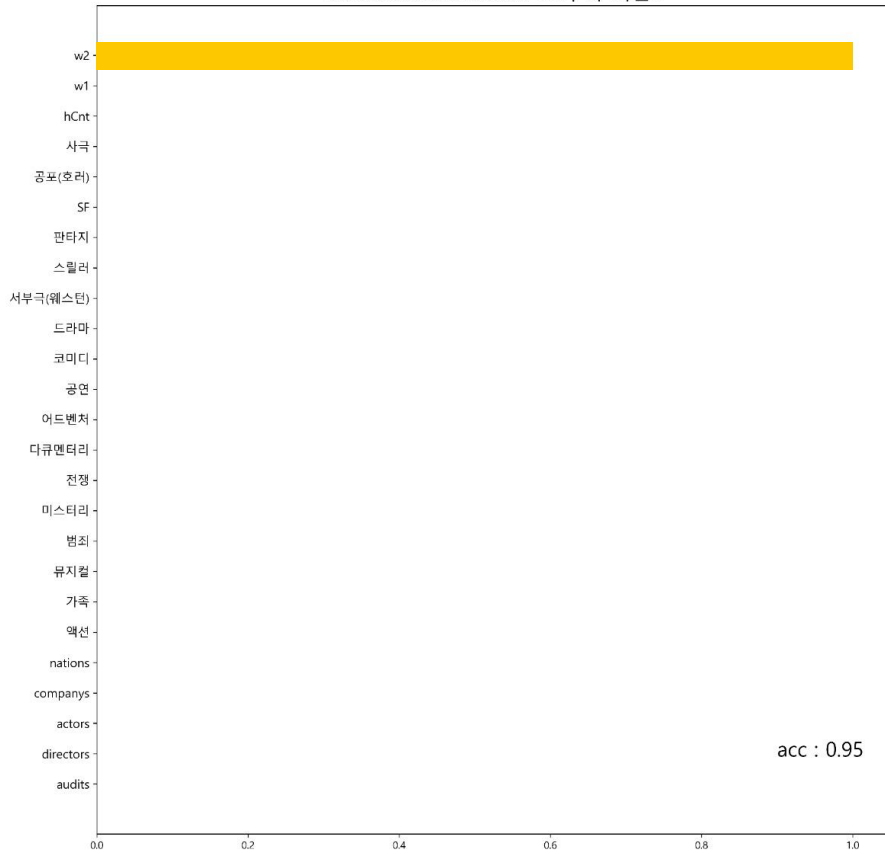
<RandomForestClassifier : 2주차 기준>



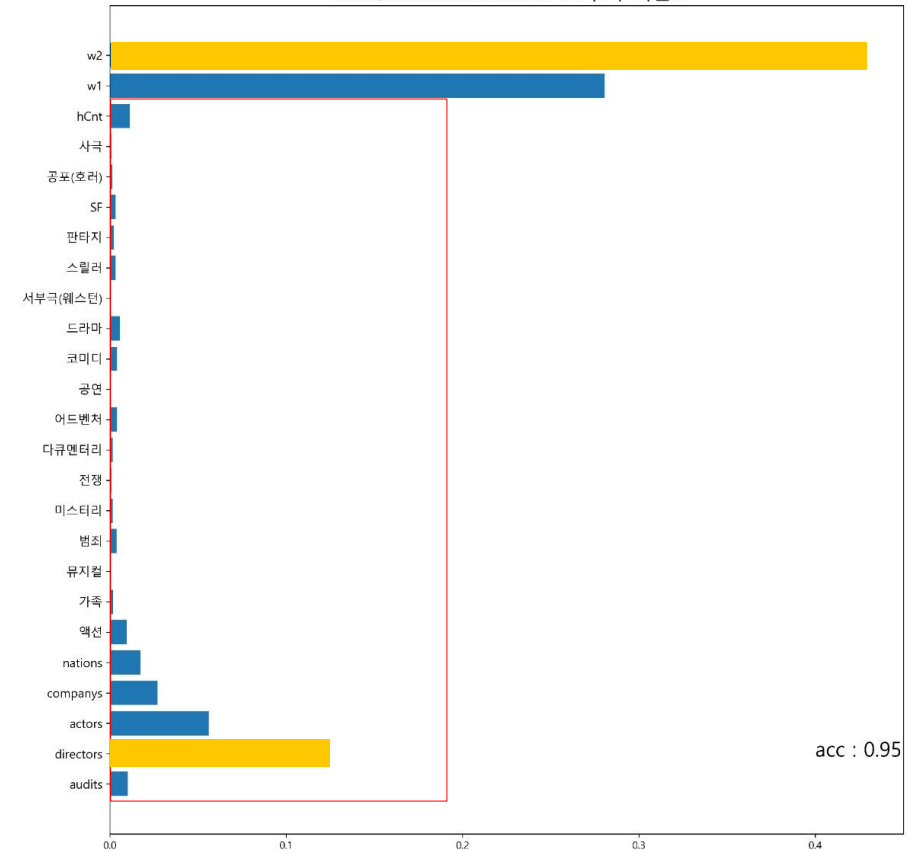


3주차

<DecisionTreeClassifier : 3주차 기준>



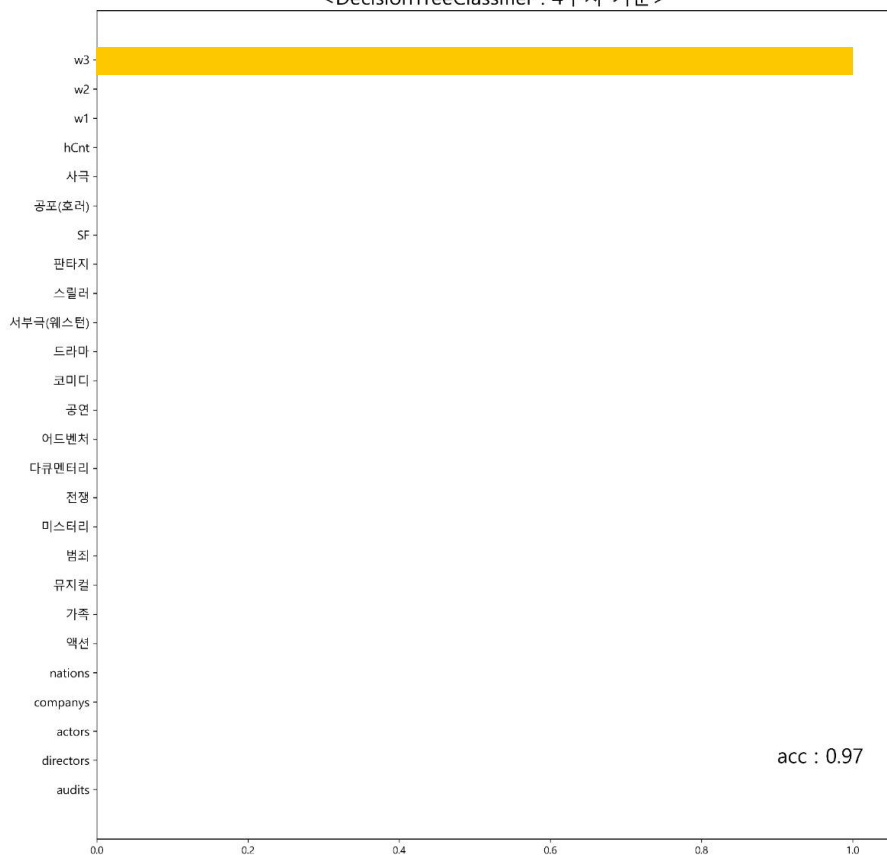
<RandomForestClassifier : 3주차 기준>



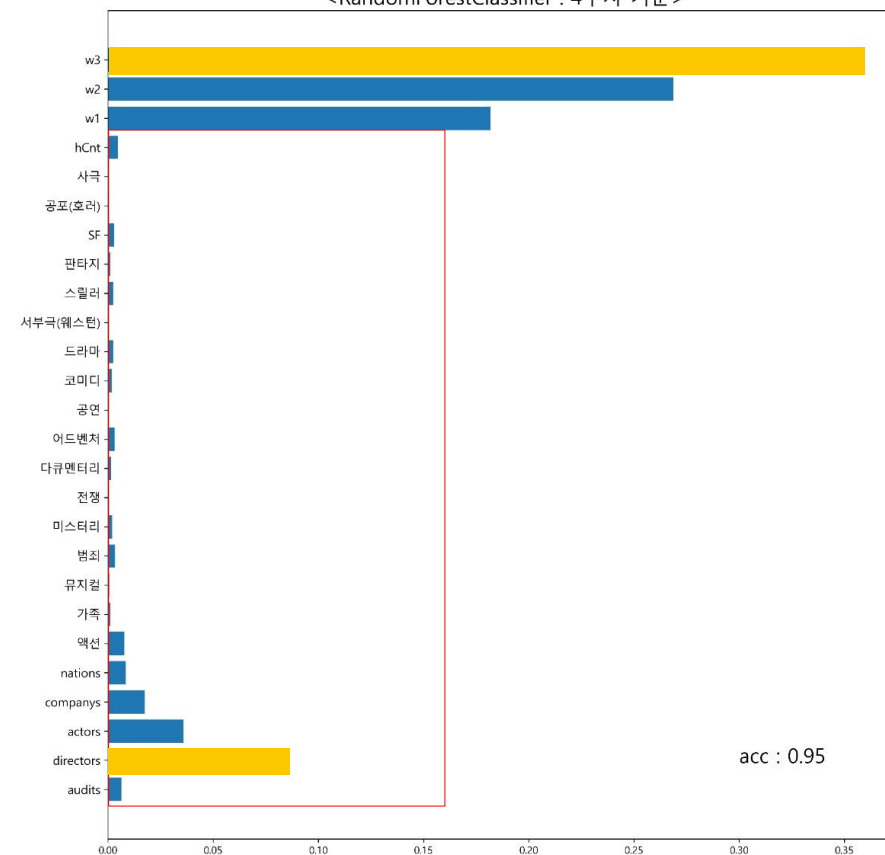


4주차

<DecisionTreeClassifier : 4주차 기준>

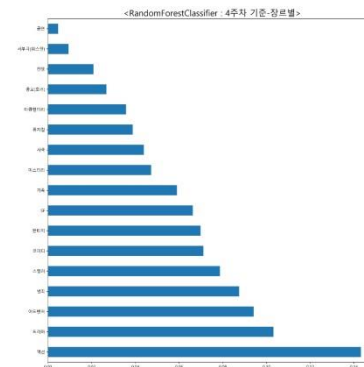
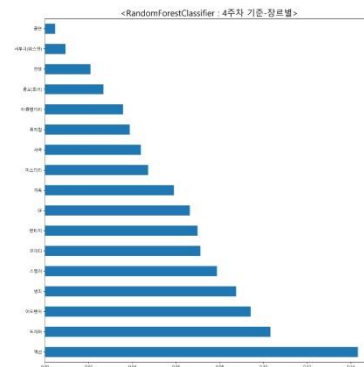
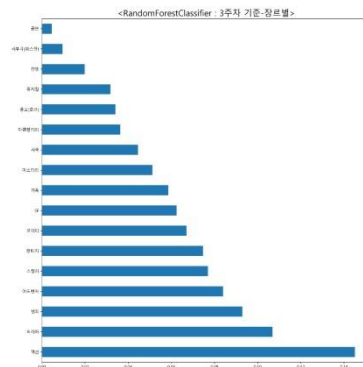
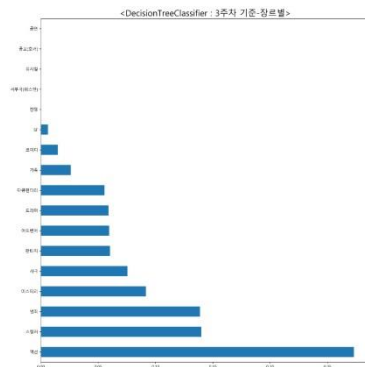
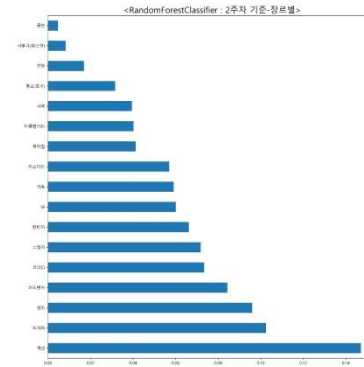
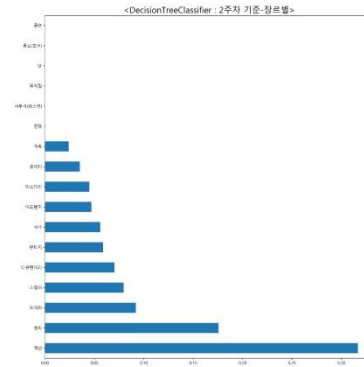
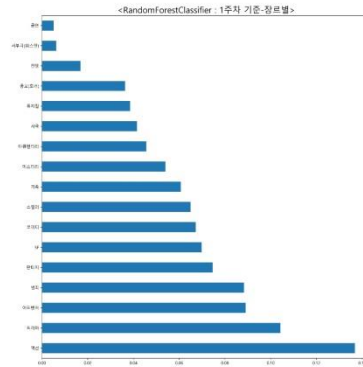
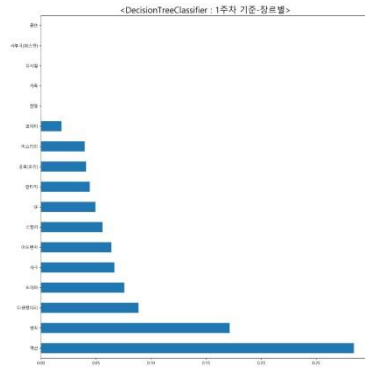


<RandomForestClassifier : 4주차 기준>





장르

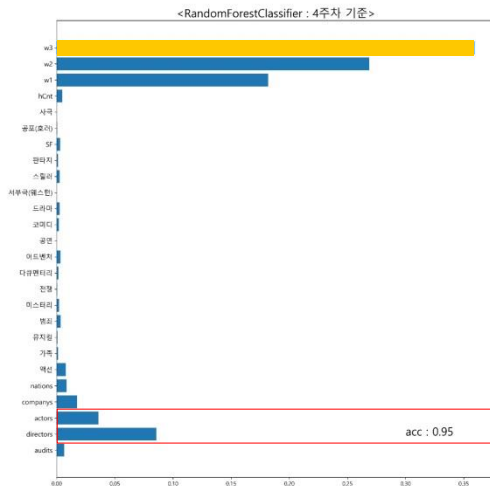
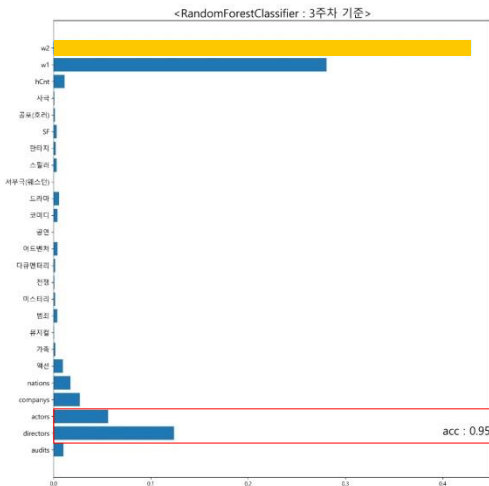
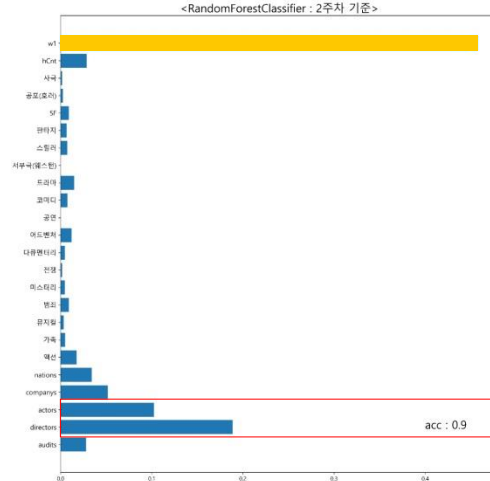
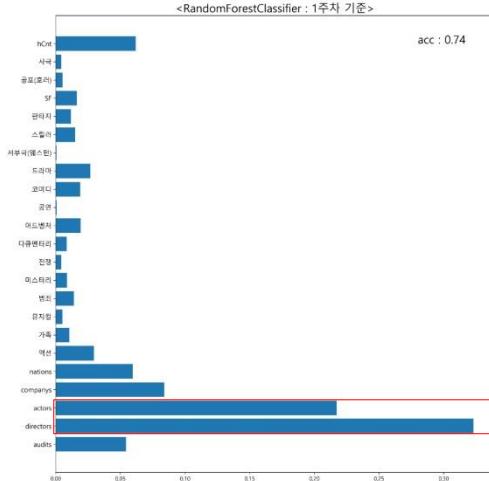


Part 5

결론



결론



“ 감독과 배우의 스타성 ”

Thank You