



# 헌법소원판결 결정문 학습을 통한 헌법소원 판결 결과 예측

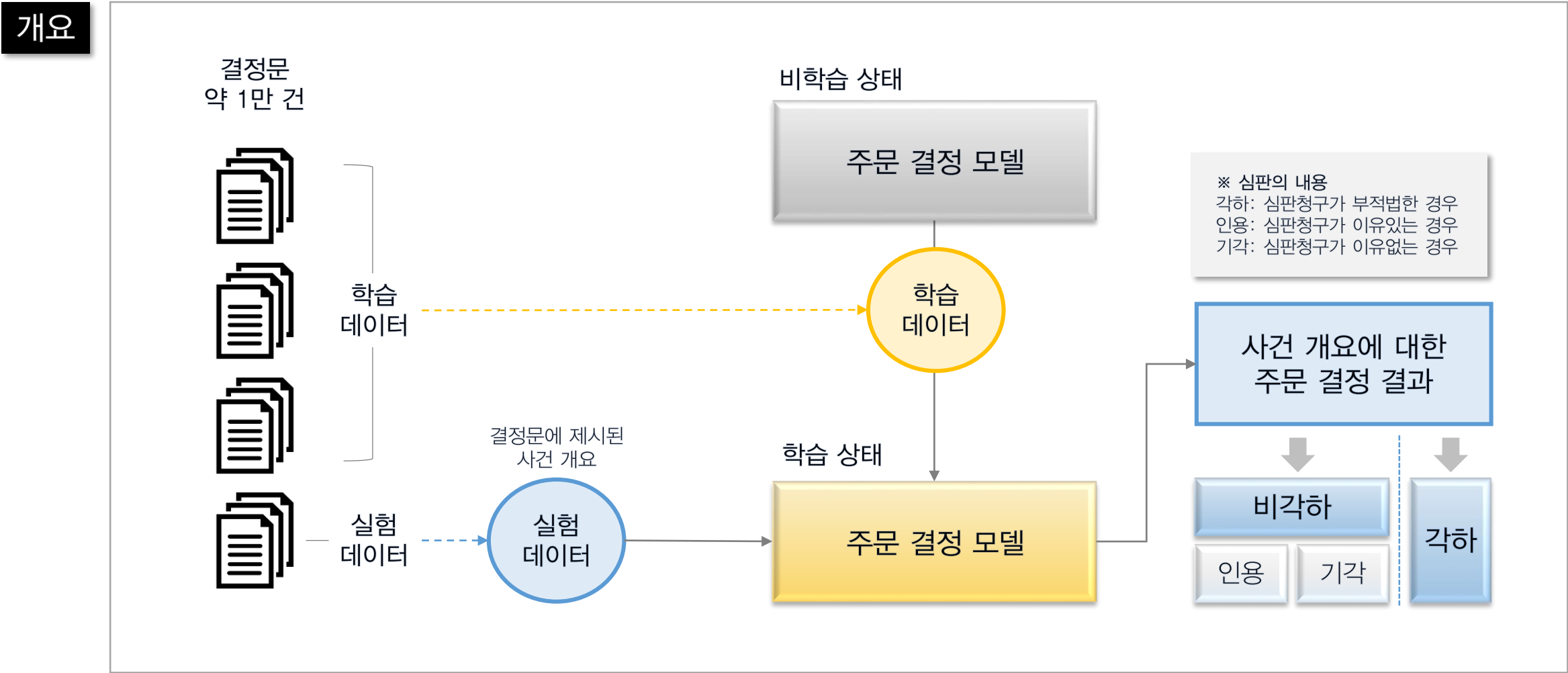
산업경영공학과 김선웅 정지혜

1. 프로젝트 개요
2. 프로젝트 진행 단계
3. 지난 발표에서 제공받은 피드백
4. Document Representation: Doc2Vec
5. Document Classification: k-NN/SVM
6. 예측 결과 및 분석
7. 결론 및 제언

# 1. 프로젝트 개요

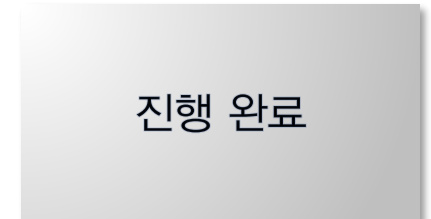
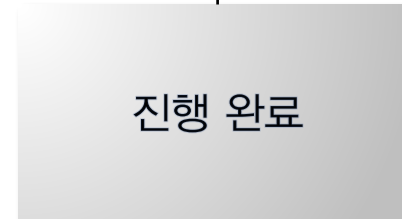
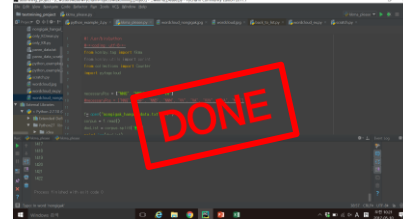
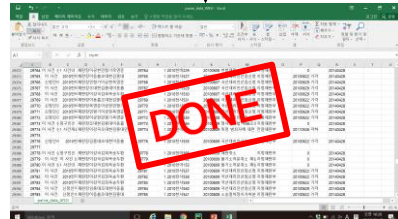
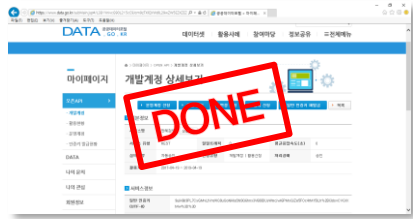
목표

헌법재판소의 헌법소원판결 결정문 학습을 통해 사건 개요를 입력할 시 각하/인용/기각 여부를 결정하는 모델 개발



## 2. 프로젝트 진행 단계

피드백: 판결과 관련된 텍스트 수작업으로 제거



### ① 데이터 수집

공공데이터 포털  
API 활용

XML 형식으로 정  
리된 결정문 수집

계정 당 1일 처리  
가능 건수 500건

### ② 데이터 전처리

메타정보를 포함  
필드 별로 정리

DB화 프로토콜을  
지키지 않은 오래  
된 데이터 제외

사건 개요가 작성  
되지 않은 데이터  
제외

### ③ 데이터 시각화



Python 활용해  
wordcloud 생성

Konlpy(Kkma)

Pytagcloud

### ④ 데이터 학습

Support Vector  
Machine

K-Nearest  
Neighbor

Classification  
Tree

### ⑤ 실험 및 분석

구현한 classifier  
에 대한 실증 데  
이터 적용

### 3. 지난 발표에서 제공받은 피드백

---

#### 피드백 #1

##### 1. 분류 방식의 문제점

- 판결의 유형중 기각된 판결과 그렇지 않은 판결로 나누는 것은 ‘각하’와 ‘인용’의 데이터 성질이 다르기 때문에 부적절해 보임
- 2-phase classification을 통해 실제 판결 방식과 마찬가지로 ‘각하’여부를 먼저 결정지은 후, 그렇지 않은 데이터에 대해 ‘인용’과 ‘기각’을 분류

#### 피드백 #2

##### 2. 사용 데이터의 문제점

- 판결문이기 때문에 결과를 예측하는 모형이 아닌 결과를 미리 학습하는 모형이 될 것으로 보임
- XML format 내부의 판결문의 형식이 규칙적이지 않아 코딩을 통해 데이터 처리 불가능
- 22000여개의 데이터 중 수작업을 통해 청구인 및 피청구인의 주장, 판단, 추가의견 등을 모두 삭제한 11000여개 데이터 이용

### 3. 지난 발표에서 제공받은 피드백

panre\_data\_my\_homework - Excel

Sunung Kim

파일 홈 삽입 그리기 페이지 레이아웃 수식 데이터 검토 보기 어떤 작업을 원하시나요? 공유

잘라내기 붙여넣기 복사 서식 복사 클립보드 글꼴 맞춤 텍스트 줄 바꿈 병합하고 가운데 맞춤 일반 표시 형식 스타일 셀 자동 합계 채우기 지우기 정렬 및 필터 찾기 및 선택 편집

C8

1 사건의 개요 청구인은 6 25 전쟁 참전용사인바 1986 경 국가유공자 등록신청을 하여 2005 5 20 재분류신체검사를 신청하기 전까지 상이등급 6급 1항 506 호에 해당하였다 그러나 위 재분류신체검사에서 6급 2항으로 상이등급이 하락하자 수원지방법원에 상이등급변경결정취소의 소를 제기 2005구단5073 하여 2007 11 13 기각되었고 이는 서울고등법원 2007누34097 을 거쳐 대법원 2008두9157 에서 확정되었다 이에 청구인은 법원이 청구인에게 불리하게 재판한 것은 국가유공자 등 예우 및 지원에 관한 법률 시행규칙 제8조의4 관련 별표 4 신체상이가 3 이상인 자에 대한 상이처 종합판정기준에서 7급 401호 및 7급 807 호 등을 포함시키지 않았기 때문이라고 주장하면서 2008 8 29 이 사건 헌법소원심판을 청구하였다 2 판단 국가유공자 등 예우 및 지원에 관한 법률 시행규칙 제8조의4 관련 별표 4 신체상이가 3 이상인 자에 대한 상이처 종합판정기준에서 7급 401호 및 7급 807호 등을 포함시켜야 하는 헌법에서 유래하는 행정입법의 작위의무가 존재하는지를 살펴보면 헌법 제34조 제5항 등 헌법규정이나 헌법해석상 이러한 구체적 입법의무가 존재한다고 볼 수 없으며 달리 그러한 의무를 구체적으로 발생시킨다고 볼 만한 법률상의 근거도 존재하지 않으므로 이 사건 헌법소원심판청구는 헌법소원의 대상이 될 수 없는 입법부작위를 그 심판대상으로 삼고 있어 부적법하다 가사 청구인의 주장을 부진정입법부작위로 보아 국가유공자 등 예우 및 지원에 관한 법률 시행규칙 제8조의4 관련 별표 4 자체를 대상으로 평등권 침해 등을 이유로 한 적극적인 헌법소원심판청구로 선해하여 살펴보더라도 청구인에 대한 재분류신체검사 결과통보가 있었던 2005 7 7 경 늦어도 청구인에 대한 수원지방법원의 기각판결이 있었던 2007 11 13 경에는 국가유공자 등 예우 및 지원에 관한 법률 시행규칙 제8조의4 관련 별표 4 에 의한 청구인의 기본권 침해사유가 발생하였음을 알았다고 할 것이므로 그로부터 90일의 청구기간이 경과한 후인 2008 8 29 에 제기된 이 사건 헌법소원심판청구는 청구기간을 도과하여 부적법하다 3 결론 그렇다면 이 사건 심판청구는 헌법재판소법 제72조 제3항 제2호 제4호에 따라 이를 각하하기로 하여 관여 재판관 전원의 일치된 의견으로 주문과 같이 결정한다

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
2	25273	이 사건 심	1 사건의	재판장 송두환 이공현 김종대			25273		1 2008헌바94		20080828	헌법재판소법 제41조	지정재판부		0			20140428
3	25274	신청인의	2008헌비	재판장 이강국 민형기 이동흙			25274		1 2008헌사388		20080828	국선대리인선임신청	지정재판부		20080909	기각		20140428
4	25275	신청인의	2008헌	재판장 김종대 이공현 송두환			25275		1 2008헌사389		20080828	국선대리인선임신청	지정재판부		20080909	기각		20140428
5	25276	신청인의	2008헌	재판장 송두환 이공현 김종대			25276		1 2008헌사390		20080828	국선대리인선임신청	지정재판부		20080909	기각		20140428
6	25277	청구인의	1 사건의	재판장 김희욱 조대현 목영준			25277		1 2008헌아102		20080828	불기소처분취소 재심	지정재판부		0			20140428
7	25278	신청인이	신청인의	재판장 목영준 조대현 김희욱			25278		1 2008헌사391		20080829	국선대리인선임신청	지정재판부		20080909	선정		20140428
8	25280	이 사건 심	재하지 않	재판장 이공현 김종대 송두환			25280		1 2008헌마550		20080829	입법부작위 위헌확인	지정재판부		0			20140428
9	25281	청구인의	1 사건의	재판장 김희욱 조대현 목영준			25281		1 2008헌마551		20080829	변지경려자로 미사제 지정재판부			0			20140428

panre\_data\_my\_homework

편집

## 4. Document Representation: Doc2Vec

```
python_example_2.py × back_to_list.py × wordcloud_ex.py × kkma_please.py ×  
1      # -*- coding: utf-8 -*-  
2  
3      import nltk  
4      import pandas as pd  
5      import numpy as np  
6      from gensim.models import Doc2Vec  
7      from collections import namedtuple  
8      from konlpy.tag import Twitter  
9      from konlpy.utils import pprint  
10  
11     # Import postagger  
12     pos_tagger = Twitter()  
13  
14     # Read data(csv) and convert into list form  
15     panres = pd.read_csv('panre_jihye.csv', encoding='cp949')  
16     panres.head()  
17     reasons_list = list(panres['reason'])  
18     reasons = reasons_list[:18]  
19     tags = [str(i+1) for i in range(len(reasons))]  
20     print tags  
21
```

### (◀ 코드 일부 캡처)

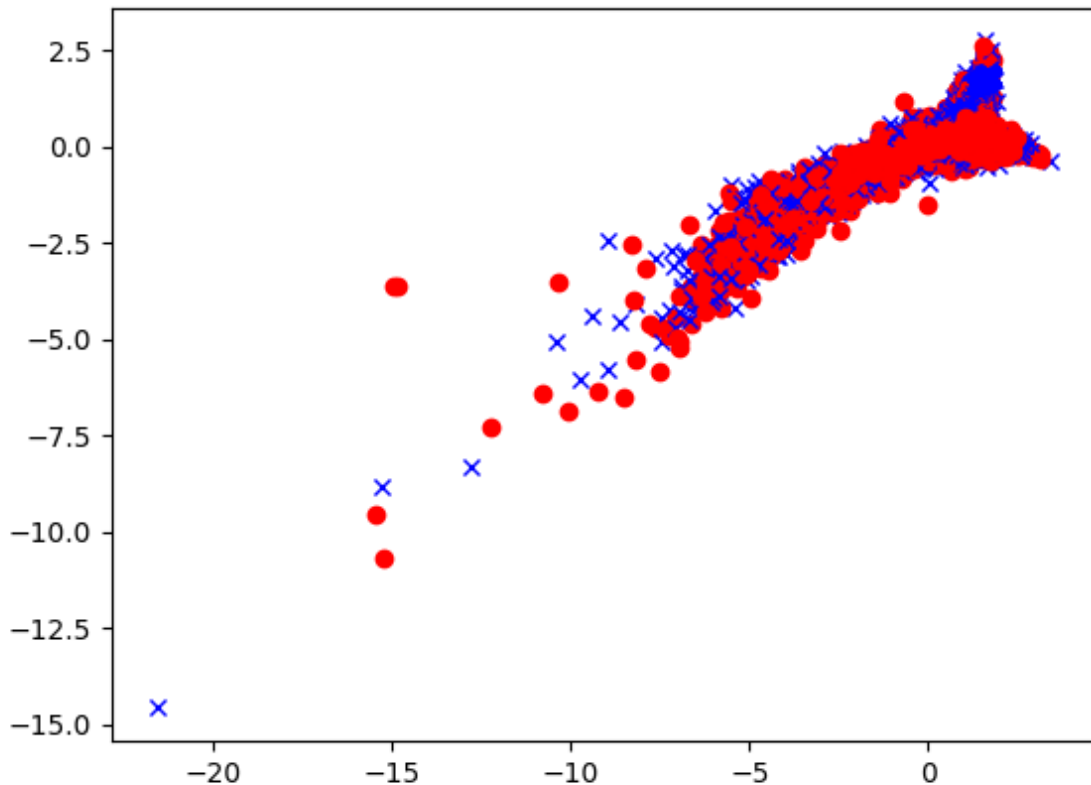
- 총 10956개 document에 대해 파이썬 이용해 Doc2Vec 구현
- Gensim 라이브러리의 doc2vec 모듈 사용
- 전처리가 끝난 문서에 대해 konlpy.tag의 Twitter 형태소 분석기를 이용해 tokenize 및 stemming
- Tokenize된 document를 negative sampling을 적용한 doc2vec을 통해 distributed representation으로 변환
- Negative sampling의 옵션 값은 genism 공식 튜토리얼에서 제공하는 추천 값 사용
- 변환된 벡터를 다음 단계의 classification에 사용



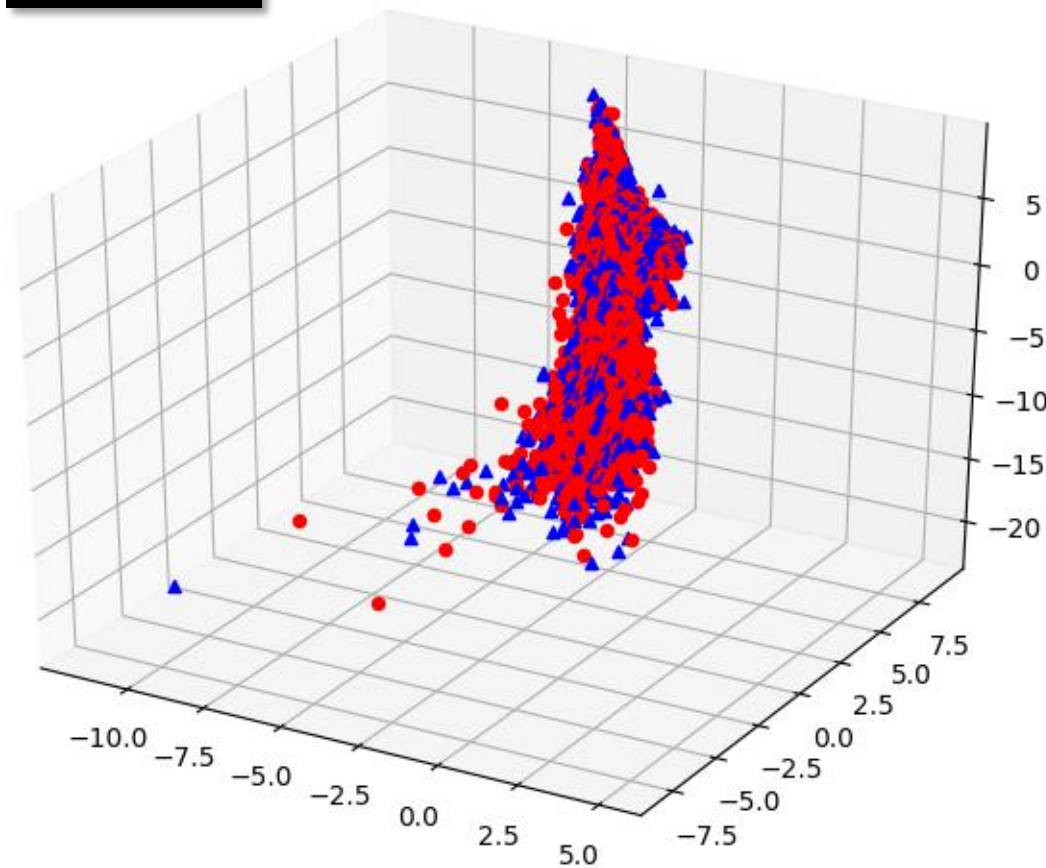
## 4. Document Representation: Doc2Vec

- 각 2차원과 3차원에 document를 embedding한 결과
- 차원을 낮추면 정보량의 손실이 클 것으로 예상해 vector의 차원을 순차적으로 최대 1000차원까지 늘려 document vector를 생성 후 각 결과를 이용해 classification에 이용

2D plot



3D plot





## 5. Document Classification: k-NN/SVM

```
1  # -*- coding: utf-8 -*-
2
3  import ...
4
5
6
7
8
9
10
11
12  data = np.loadtxt('DocumentVectors_1000.txt')
13  print np.shape(data)
14  panre = pd.read_csv('on_test.csv', encoding='cp949')
15  panre.head()
16  rsta = list(panre['rstaRsta'])
17  reasons = list(panre['reason'])
18
19  for i in range(len(reasons)):
20      if pd.isnull(reasons[i]) is True:
21          rsta.remove(rsta[i])
22
23      is_gigak = np.zeros(len(rsta))
24      is_inyong = np.zeros(len(rsta))
25
26      # for row in rsta:
27      #     if row != '기각' and row != '합헌' and row != '선정' and pd.isnull(row) is False:
28      #         print row
29
30      for i in range(len(rsta)):
31          if pd.isnull(rsta[i]) is True or rsta[i] == '각하':
32              is_gigak[i] = 1
33
34          if rsta[i] == '선정' or rsta[i] == '위헌' or rsta[i] == '위헌확인' or rsta[i] == '인
35
36  print sum(is_inyong[i] for i in range(len(is_inyong)))
```

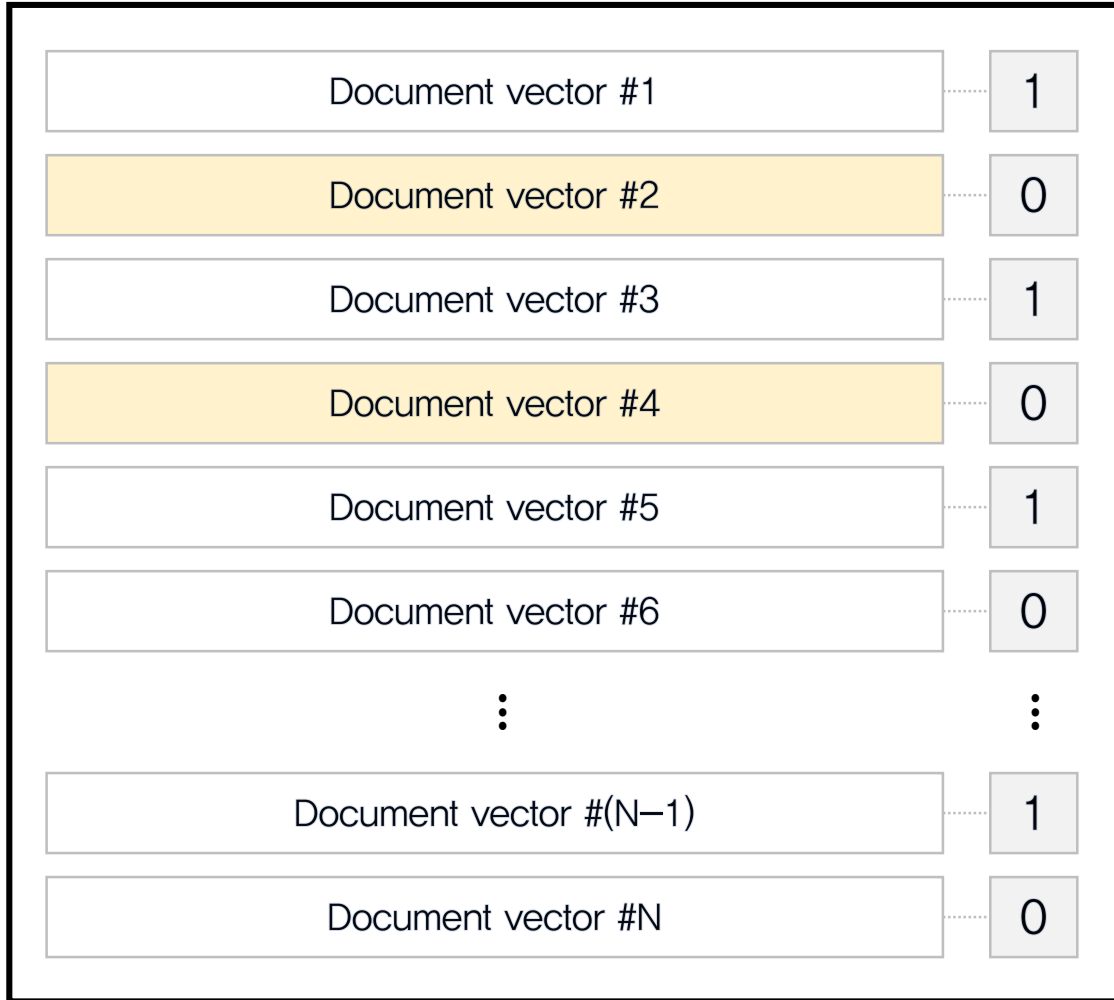
### (◀ 코드 일부 캡처)

- sklearn 라이브러리의 svm, neighbors 모듈 사용
- 각하여부 - 인용여부에 걸친 2-phase 분석 시도
- SVM을 이용하여 doc2vec의 vector size를 20~1000까지 조절하여 적절한 matrix size 탐색 시도
- k-NN을 이용하여 matrix size를 100으로 고정해 두고, n-neighbors를 조절하여 적절한 정확도를 주는 n 탐색 시도
- Training data:Test data 각 8:2의 비율로 10 fold cross validation을 통해 실험 계획

## 5. Document Classification: 2-Phase Classification

### 1<sup>st</sup> Phase Classification

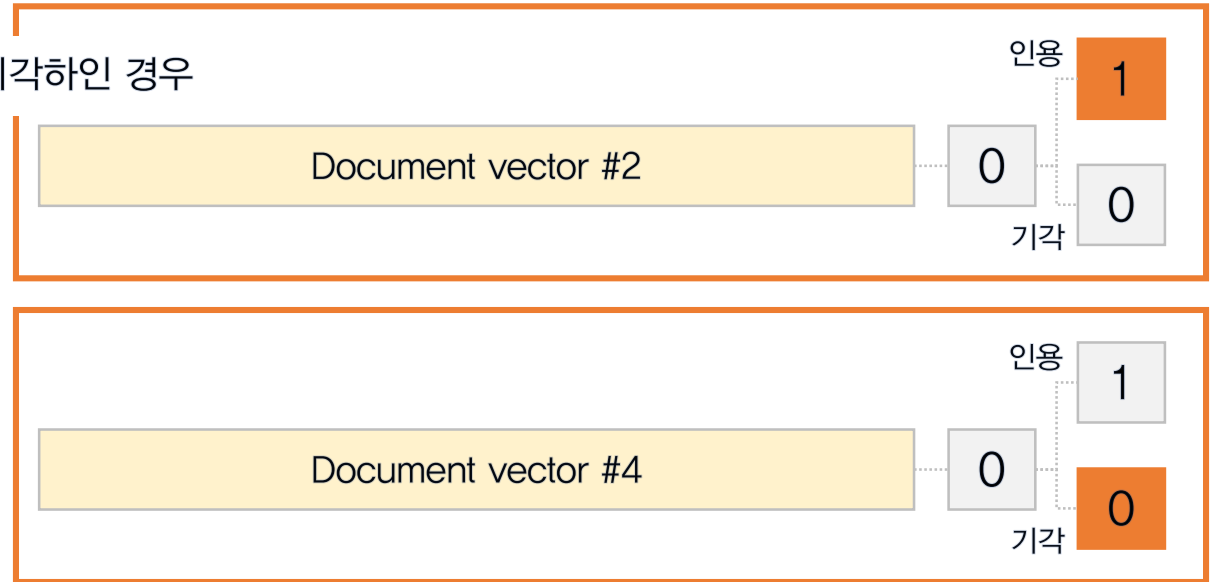
각하/비각하의 여부 판단  
: SVM/k-NN



### 2<sup>nd</sup> Phase Classification

인용/기각의 여부 판단  
: SVM

비각하인 경우



## 6. 예측 결과 및 분석: Phase 1/SVM

- 모든 실험에서 큰 차이 없이 50%를 웃도는 정확도를 보였으나, 60% 이상의 성능을 낸 경우는 없었음
- Vector size에 따라 연산 시간이 증가하는 것을 관찰할 수 있음

Vec = 20	실제		
예측		각하	비각하
	각하	555.3	475.9
	비각하	519.4	641.4
정확도	54.59%	시간	4.43

Vec = 100	실제		
예측		각하	비각하
	각하	610.5	538
	비각하	464.2	579.3
정확도	54.28%	시간	15.54

Vec = 200	실제		
예측		각하	비각하
	각하	584.3	527
	비각하	490.4	590.3
정확도	53.59%	시간	19.34

Vec = 1000	실제		
예측		각하	비각하
	각하	411.7	364.8
	비각하	663	752.5
정확도	53.11%	시간	74.50

## 6. 예측 결과 및 분석: Phase 1/k-NN

- 모든 실험에서 큰 차이 없이 50%를 웃도는 정확도를 보였으나, 60% 이상의 성능을 낸 경우는 없었음
- Vector size를 100으로 고정한 후 neighbor의 개수를 변경해가며 실험 진행

n = 5	실제		
예측		각하	비각하
	각하	526.4	507.4
	비각하	548.3	609.9
정확도	51.84%	시간	2.91

n = 30	실제		
예측		각하	비각하
	각하	510.7	468.3
	비각하	564	649
정확도	52.91%	시간	3.02

n = 10	실제		
예측		각하	비각하
	각하	530.1	497.6
	비각하	544.6	619.7
정확도	52.45%	시간	2.86

n = 50	실제		
예측		각하	비각하
	각하	483.4	454.1
	비각하	591.3	663.2
정확도	52.31%	시간	3.21

## 6. 예측 결과 및 분석: Phase 2/SVM

Vec = 20	실제				
예측			각하	비각하	
				인용	기각
	각하		501.8	427.5	
	비 각 하	인용	467.7	0	0
		기각		62.2	513.6
정확도	46.32%		시간	4.96	

Vec = 100	실제				
예측			각하	비각하	
				인용	기각
	각하		547.6	483.4	
	비 각 하	인용	421.9	0	0.8
		기각		56.3	462.8
정확도	46.09%		시간	11.52	

## 7. 결론 및 제언

---

- 각 classification 실험에서 대부분의 케이스가 정확도 50% 근방의 결과를 보임
- 데이터 수작업으로 인한 유의미한 데이터의 손실 가능성 존재
- 인용의 경우 해당 레이블이 부착된 데이터가 절대적으로 부족해 2<sup>nd</sup> phase의 정확도를 기대할 수 없음  
(Training data의 비율을 줄일 경우 인용 레이블이 부착된 데이터가 training set에 아예 포함되지 않는 경우도 존재)
- 공공 데이터로서 제공되는 판결문의 정형화된 프로토콜이 존재할 경우 더 정확도 높은 결과를 기대할 수 있음



경청해주셔서  
감사합니다