

4U

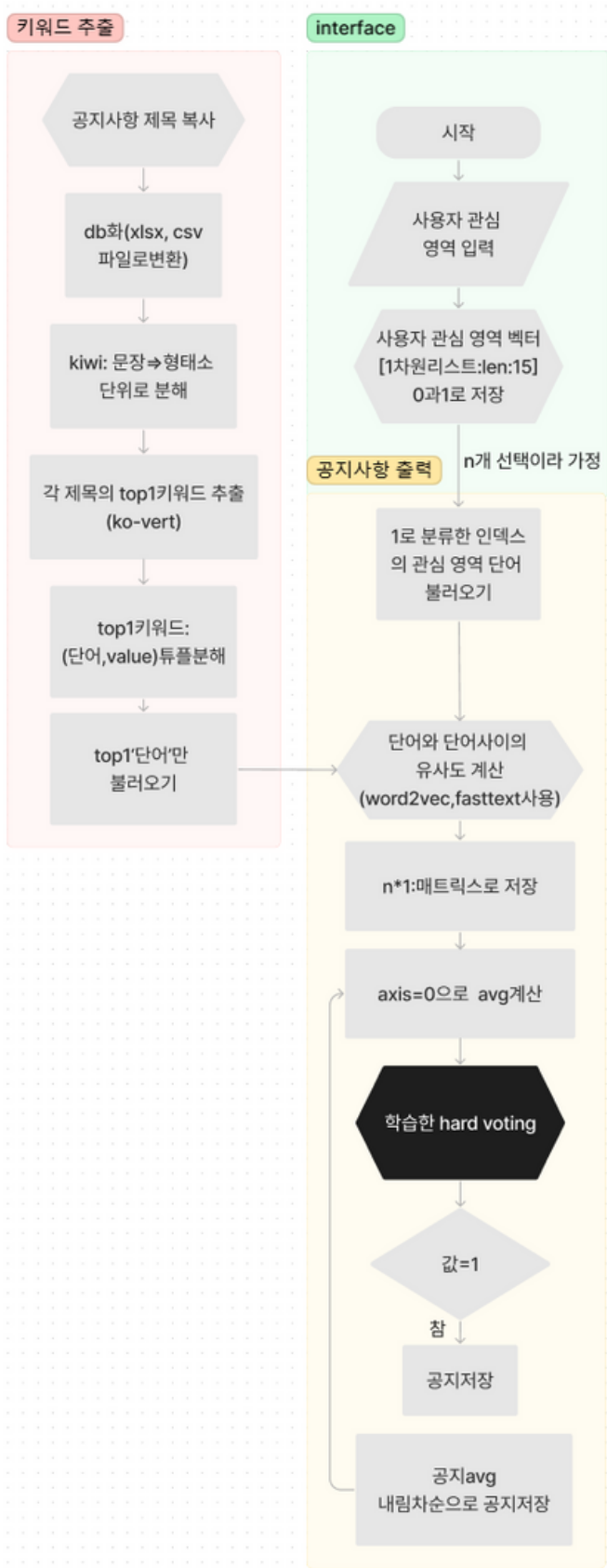
사용자 맞춤형 학교 공지사항 알리미

목차

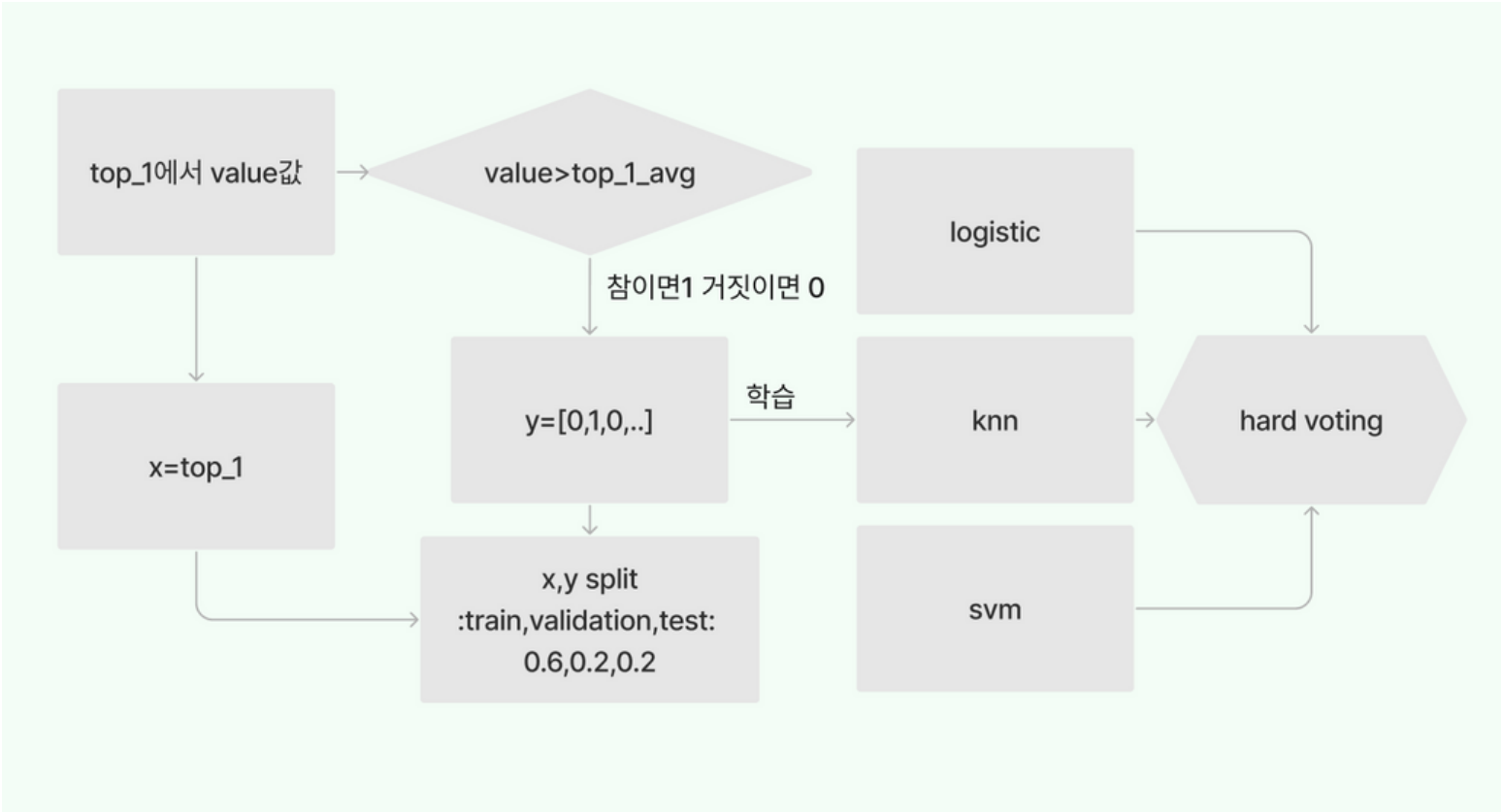
1. 알고리즘 및 시스템 구성도
2. db 입력 후 kiwi, bert 사용 결과
3. classification 성능 비교
4. 단어 유사도 출력 모델 성능 비교
(word2vec, fasttext)
5. 실행 결과

알고리즘 및 시스템 구성도

시스템 구성도



알고리즘 구성도



db입력후 kiwi,bert사용결과

kiwi사용결과: 형태소 분석

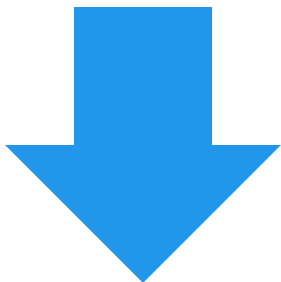
['인재', '개발원', '채용', '상담회', '한국', '경제인', '협회', '전국경제인연합회', '채용', '상담회']
['채용', '산학', '협력단', '직원', '채용', '광고']
['글로벌', '소통', '교육실', '겨울', '계절', '학기', '영어', '말하기', '프로그램', '등록', '안내']
['언어', '교육원', '학년도', '싱가포르', '공대', '위탁', '과정', '언어', '교육원', '한국어', '도우미', '모집']
['교육', '혁신', '센터', '융합', '수업', '운영', '사례', '공유', '융합', '수업', '운영', '사례', '발표', '전시', '안내']
['인재', '개발원', '학년도', '겨울', '방학', '현장', '실습', '모집', '안내']
['학년도', '전기', '졸업', '예정', '여부', '조회', '졸업', '의사', '신청', '관련', '안내']
['국제', '학생증', '발급', '지원', '행사', '안내']
['학년도', '전공', '변경', '시행', '안내']
['학년도', '시행', '안내']

bert사용결과:top1중요도를 출력

index 0: [('협회', 0.7234)]
index 1: [('광고', 0.7352)]
index 2: [('영어', 0.5199)]
index 3: [('싱가포르', 0.6966)]
index 4: [('수업', 0.5876)]
index 5: [('학년도', 0.8406)]
index 6: [('학년도', 0.8442)]
index 7: [('국제', 0.8671)]
index 8: [('학년도', 0.9189)]
index 9: [('학년도', 0.9569)]
index 10: [('싱가포르', 0.6962)]
index 11: [('안내', 0.7502)]
index 12: [('연구윤리', 0.8173)]

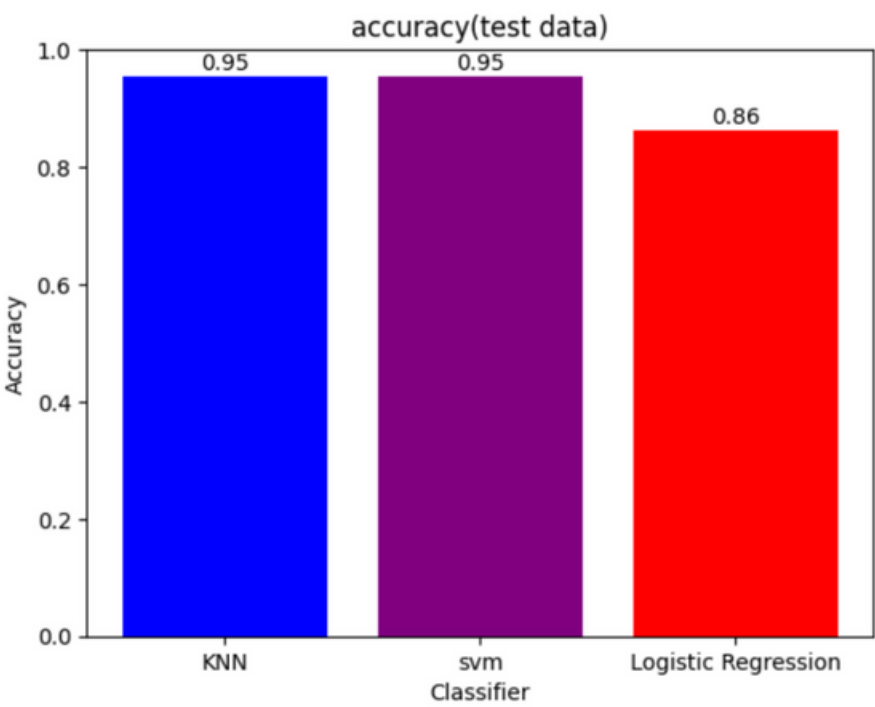
classification성능 비교

6개:로지스틱,knn,svm,나이브 베이즈,결정트리,gradient boosting



3개:knn,svm,로지스틱

정확도



mse&분류 성능

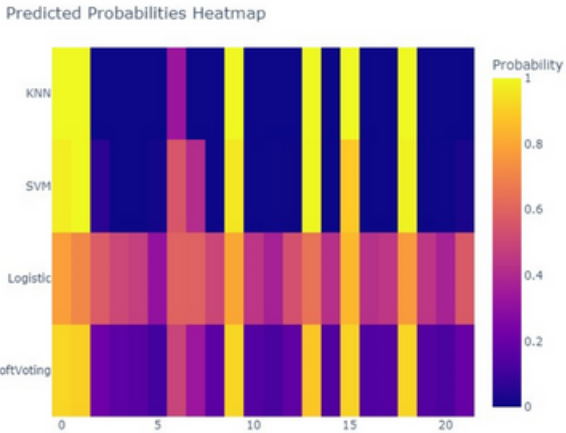
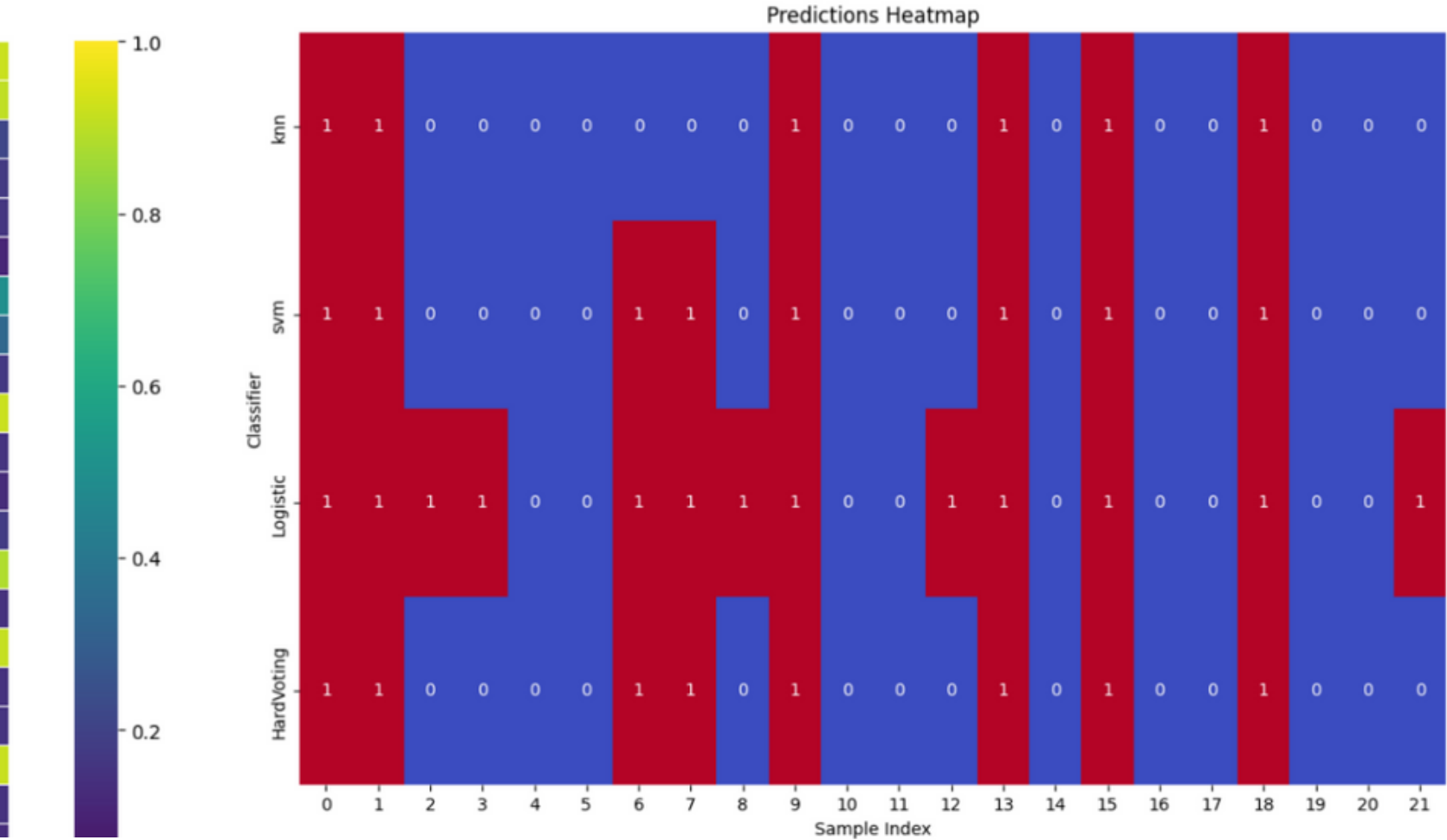
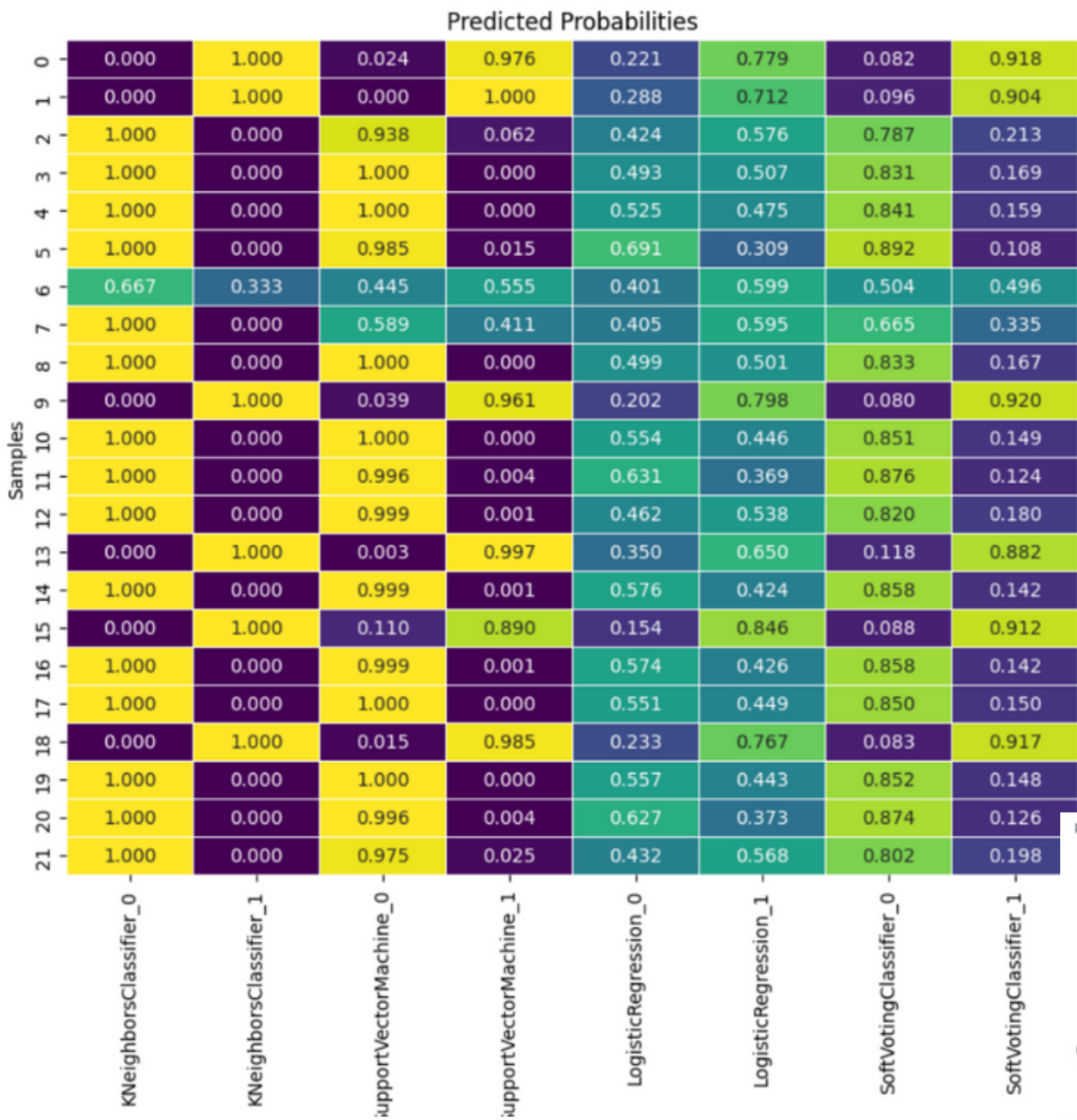
classifier	MSE	precision	recall	f1 score
knn	0.045455	1.00000	0.857143	0.923077
svm(support vector machine)	0.045455	0.875000	1.00000	0.933333
logistic regression	0.272727	0.538462	1.00000	0.700000

classification 성능 비교

3개:knn,svm,로지스틱

soft voting

hard voting



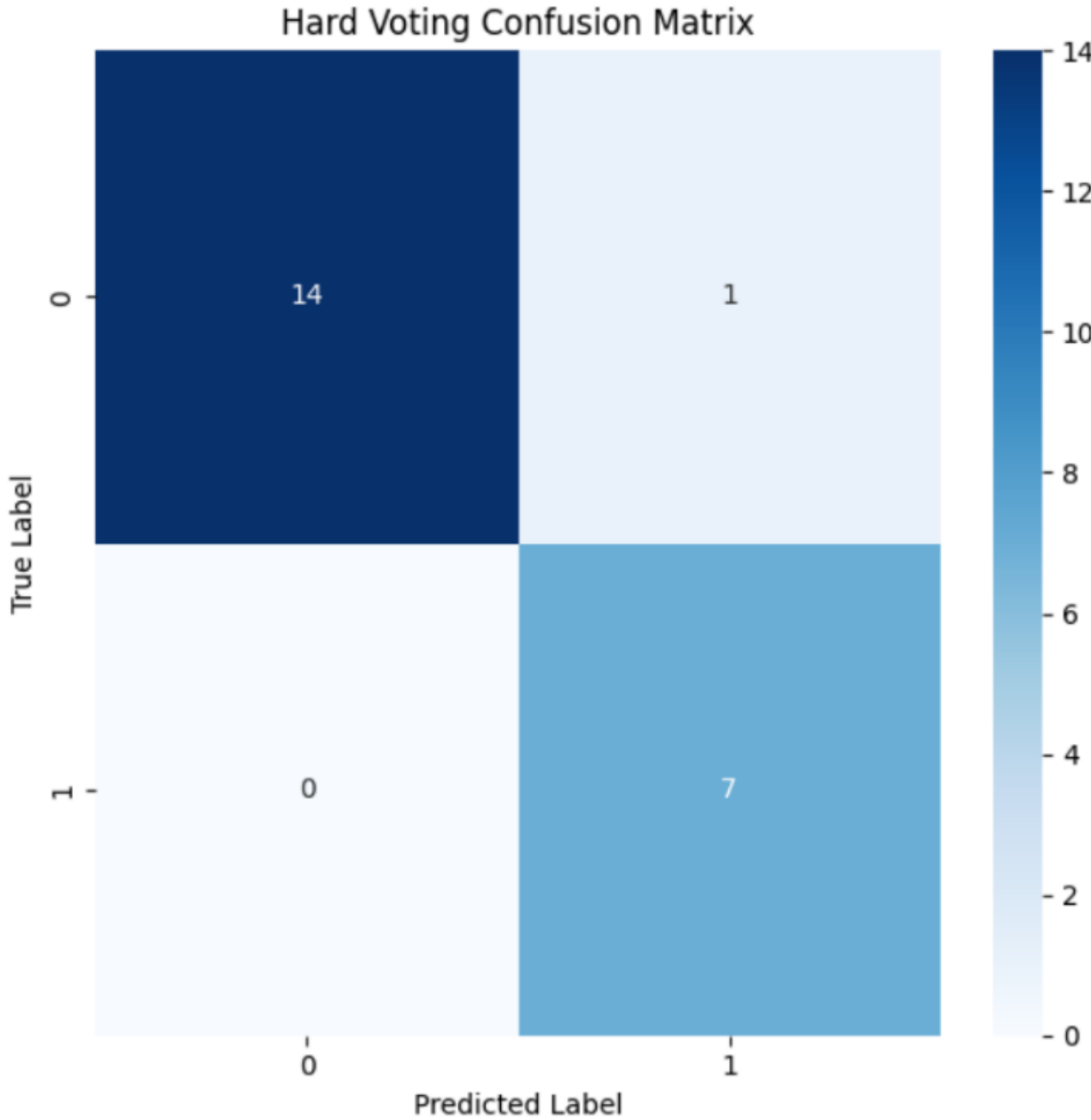
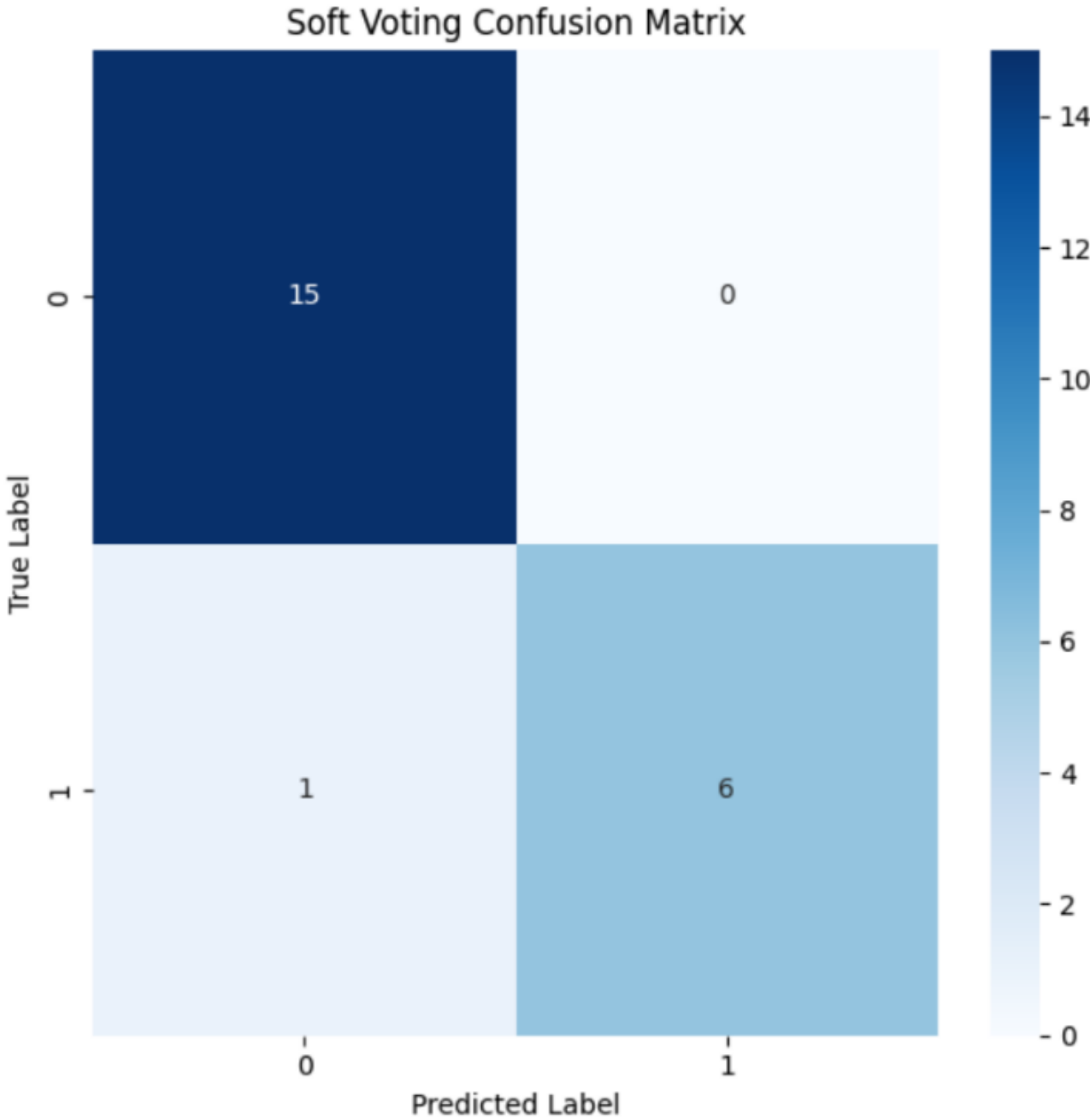
히트맵 비교

classification 성능 비교

3개:knn,svm,로지스틱

soft voting

hard voting



ConfusionMatrix

classification성능 비교



사용한 앙상블	softvoting	hardvoting
정확도	0.9545	0.9545
정밀도	1.0000	0.8750
재현율	0.8571	1.0000
f1	0.9231	0.9333 ✓

분류 성능지표

Fasttext

희귀 단어도 다른 단어의 n-gram과 겹치는 경우라면, Word2Vec과 비교하여 비교적 높은 임베딩 벡터값 얻음.
Fasttext를 학습할 때 단어 사전을 wiki_20190620_small.txt 을 소스로 하여 단어를 학습

```
Enter a word: 조교
Enter the number of similar words you want to find: 5
Words similar to '조교': [('프로그램', 0.1166064664721489), ('학생', 0.10648241639137268), ('인재', 0.09798437356948853), ('트랙', 0.08016421645879745), ('채용', 0.066685974597)]
Enter the first word: 채용
Enter the second word: 설명회
```



```
Enter a word: 조교
Enter the number of similar words you want to find: 5
Words similar to '조교': [('프로그램', 0.1166064664721489), ('학생', 0.10648241639137268), ('인재', 0.09798437356948853), ('트랙', 0.08016421645879745), ('채용', 0.066685974597)]
Enter the first word: 채용
Enter the second word: 설명회
Similarity between '채용' and '설명회': -0.04564239829778671
```

공지에 포함되지 않은 단어를 입력해도 값이 출력됨

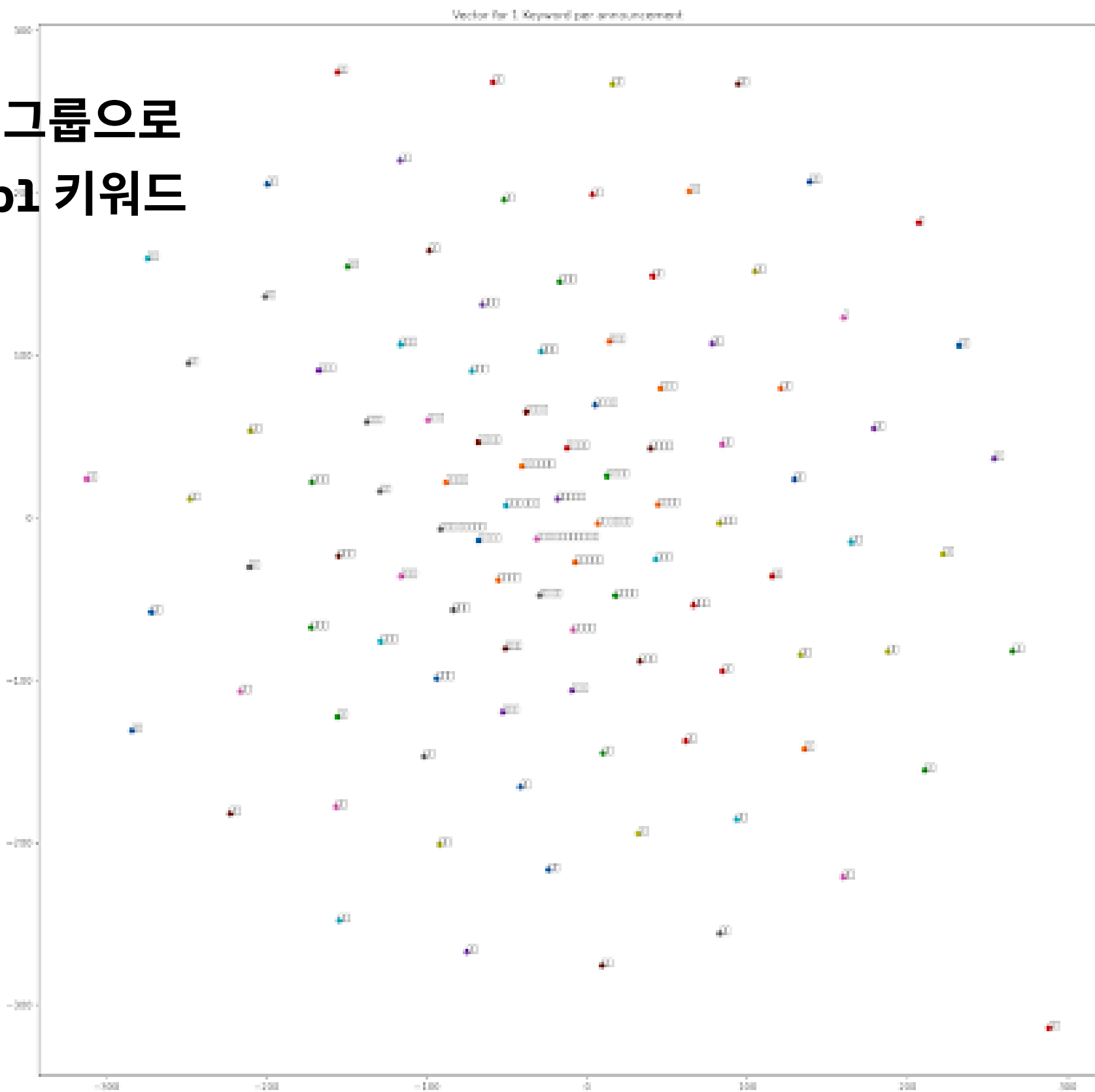
```
similar_words = saved_model.wv.most_similar("사용자관심태그 변수", topn=5) #사용자관심태그 변수로
print(similar_words)
```

```
[('학생', 0.09780142456293106), ('겨울', 0.09631039202213287), ('인재', 0.06986626982688904), ('계약', 0.06808039546012878), ('TEL0S', 0.06452154368162155)]
```

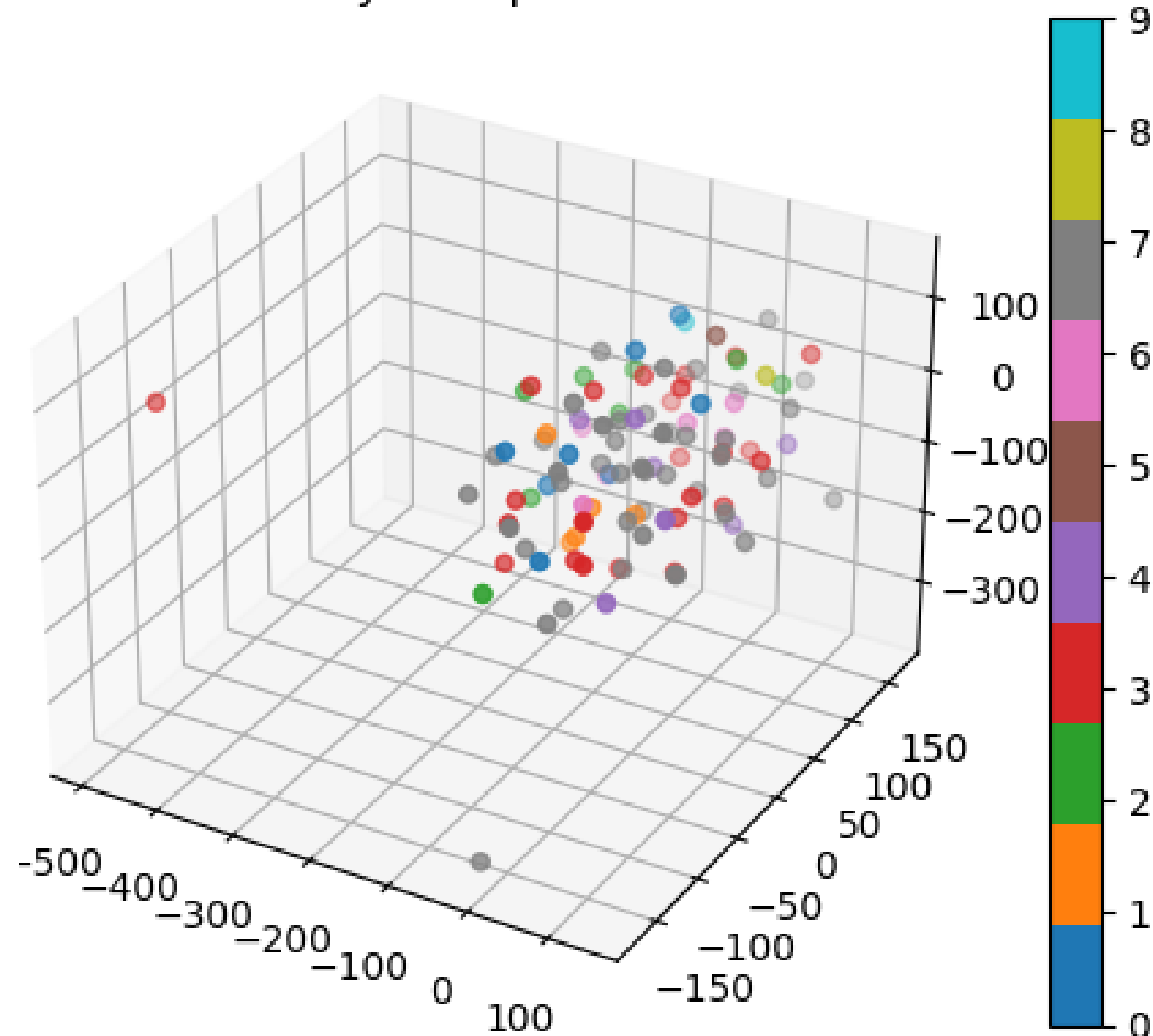
Fasttext&K-means clustering

혁신 : 3
영상 : 4
도서관 : 3
진로 : 4
채용 : 1
수강 : 2
두산연강재단 : 7
사회 : 6
온라인 : 3
관리자 : 7
인재 : 6
리더십 : 7
디자인 : 0
삼성 : 4
개발 : 0
계약 : 7
학기 : 7
기념품점 : 3
교육 : 4
행정실 : 7
원 : 6
식품부 : 3
문화예술 : 3
변리사 : 4
폭력 : 4
교과 : 7

<-10개의 그룹으로
분류된 top1 키워드



Vector for 1 Keyword per announcement



Fasttext로 돌린 K-means clustering 2차원, 3차원 그래프

비지도 학습으로 10개 영역 분류&예측

X1, X2, X3: 상위 3개의 키워드의 중요도 값
Y: 각 공지사항의 분류번호

Bert의 중요도 값을 사용하여
두 가지 방식으로 (K-Means, 계층적 군집화)

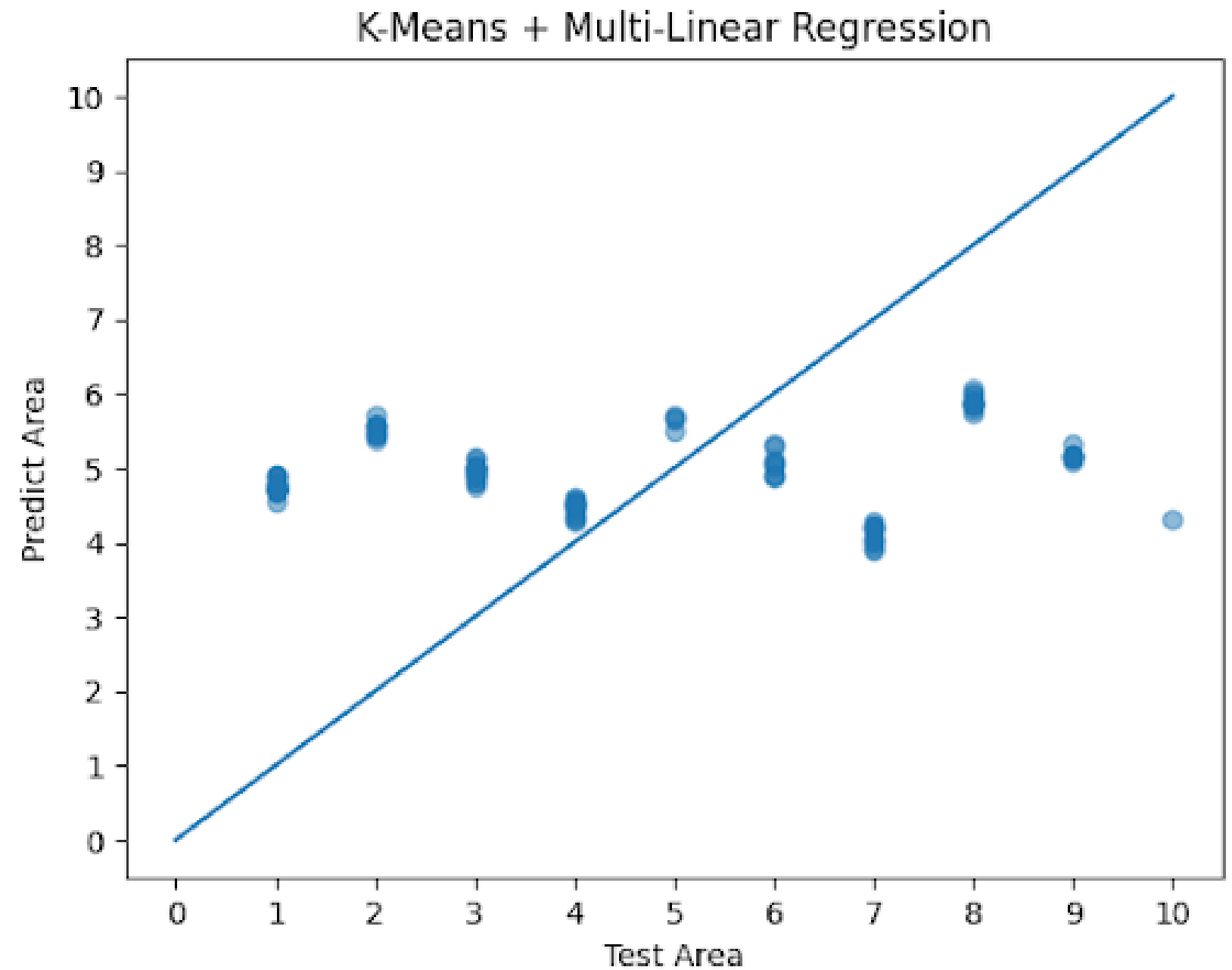
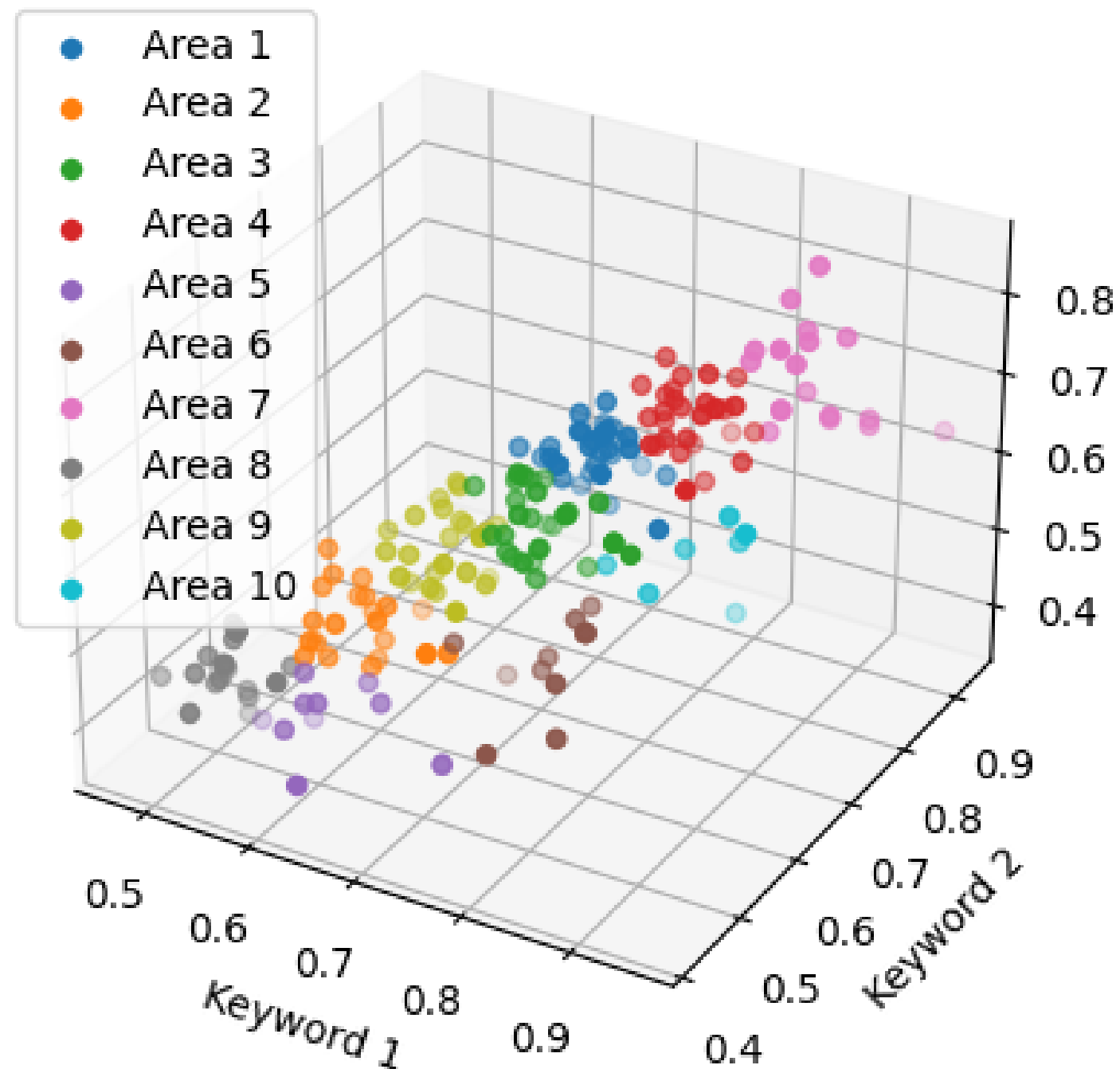
$Y=aX1+bX2+cX3$ 의 다중선형회귀를 진행

테스트 데이터와 예측 데이터를 비교하여 같은지 확인

	X1(keyword1)	X2(keyword2)	X3(keyword3)
1번째 공지사항	0.7976	0.7846	0.6537
2번째 공지사항	0.7234	0.6864	0.6608
3번째 공지사항	0.7352	0.6272	0.5600
4번째 공지사항	0.5199	0.4851	0.4582
5번째 공지사항	0.6966	0.6114	0.5476
...
216번째 공지사항	0.7370	0.7176	0.6357
217번째 공지사항	0.7515	0.6480	0.6328
218번째 공지사항	0.7438	0.7084	0.6847
219번째 공지사항	0.8285	0.7389	0.7291
220번째 공지사항	0.5133	0.4983	0.4409

220 rows × 3 columns

비지도 학습으로 10개 영역 분류-K-Means



유효한 예측 수: 21
전체 훈련 횟수: 110
예측 적중 확률: 19.0909%

K-Means로 분류한 10개 영역별 WordCloud 결과

Word Cloud Area 1



Word Cloud Area 2



Word Cloud Area 3



Word Cloud Area 4



Word Cloud Area 5



Word Cloud Area 6



Word Cloud Area 7



Word Cloud Area 8



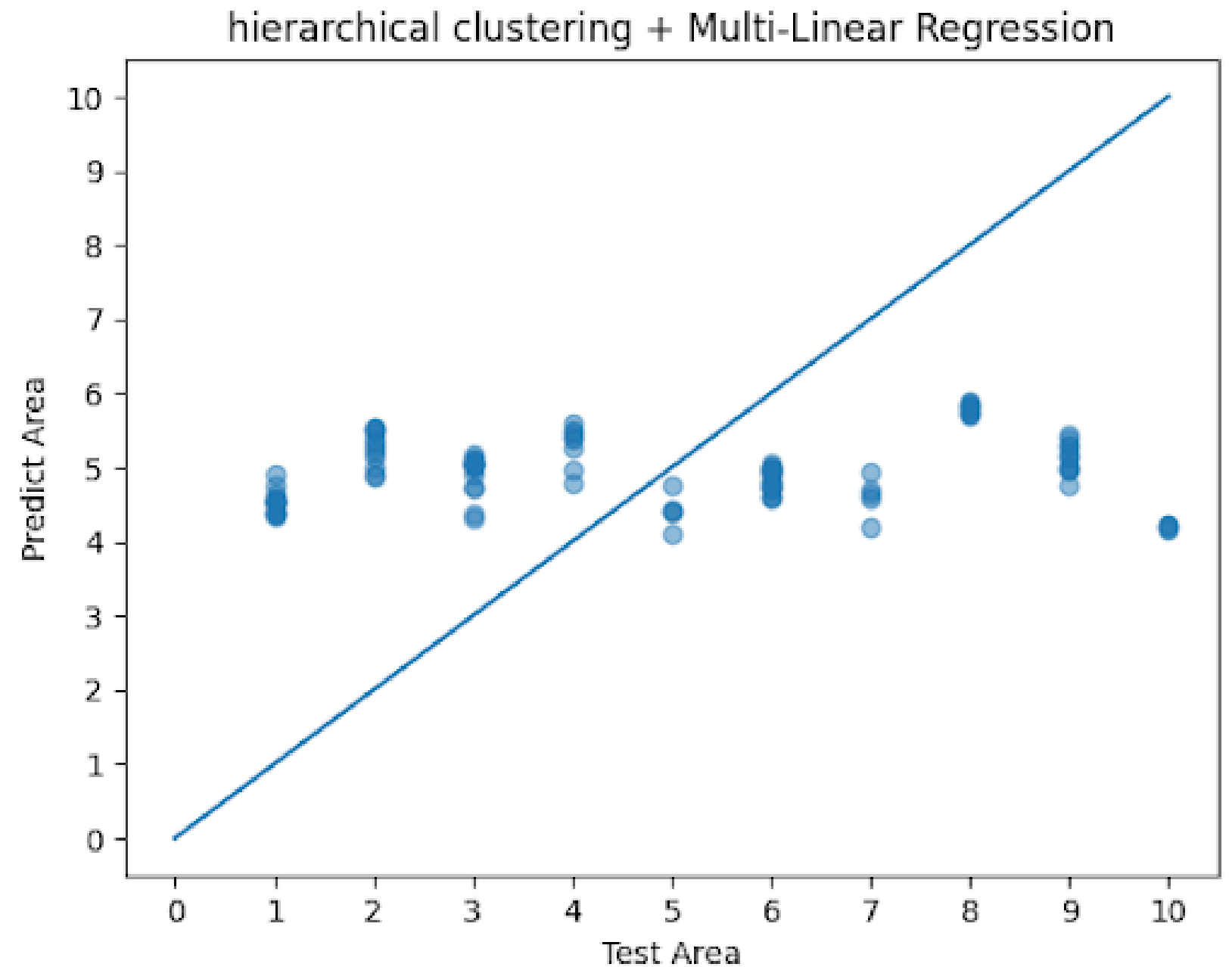
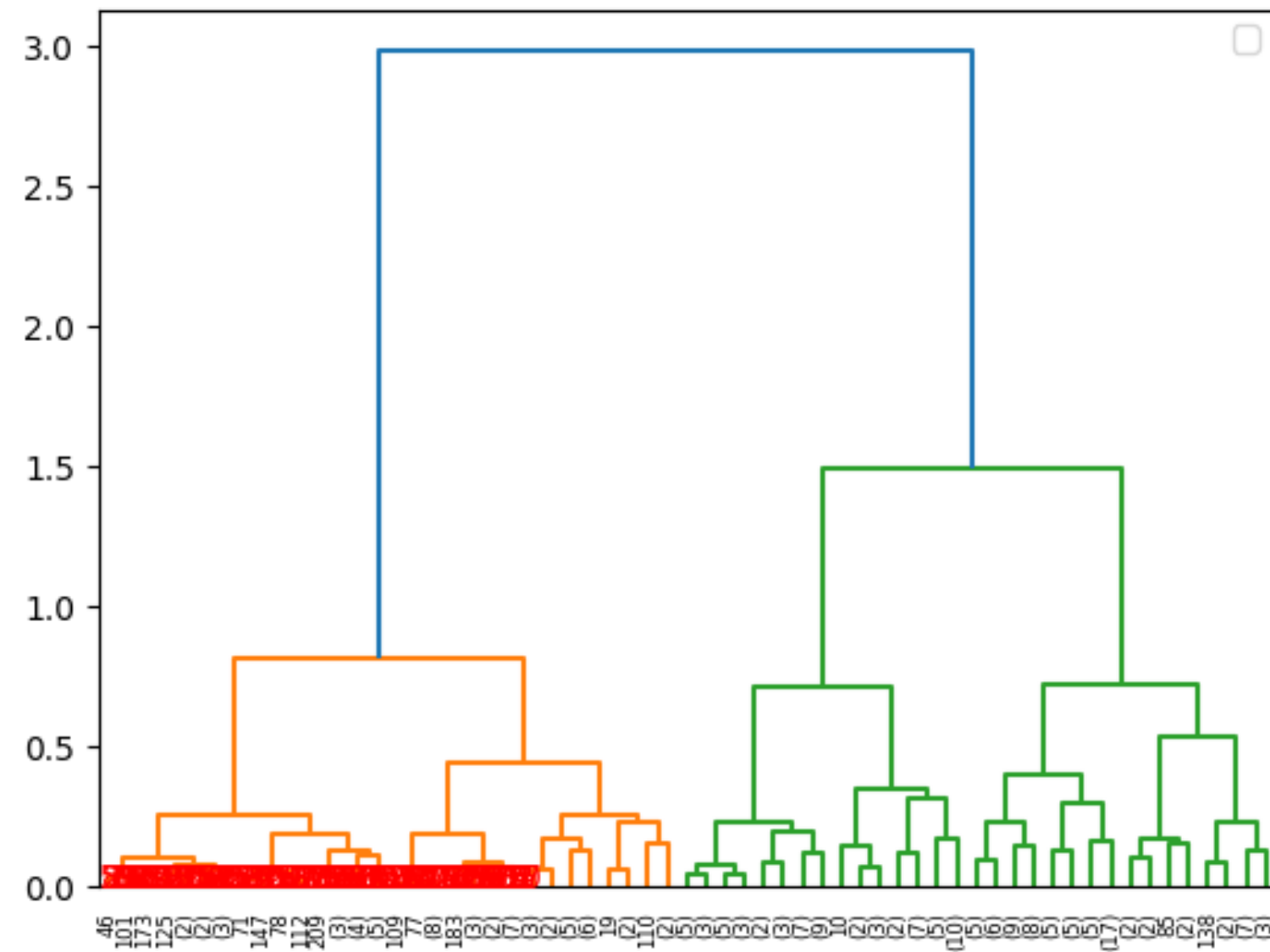
Word Cloud Area 9



Word Cloud Area 10



비지도 학습으로 10개 영역 분류-계층적 군집화

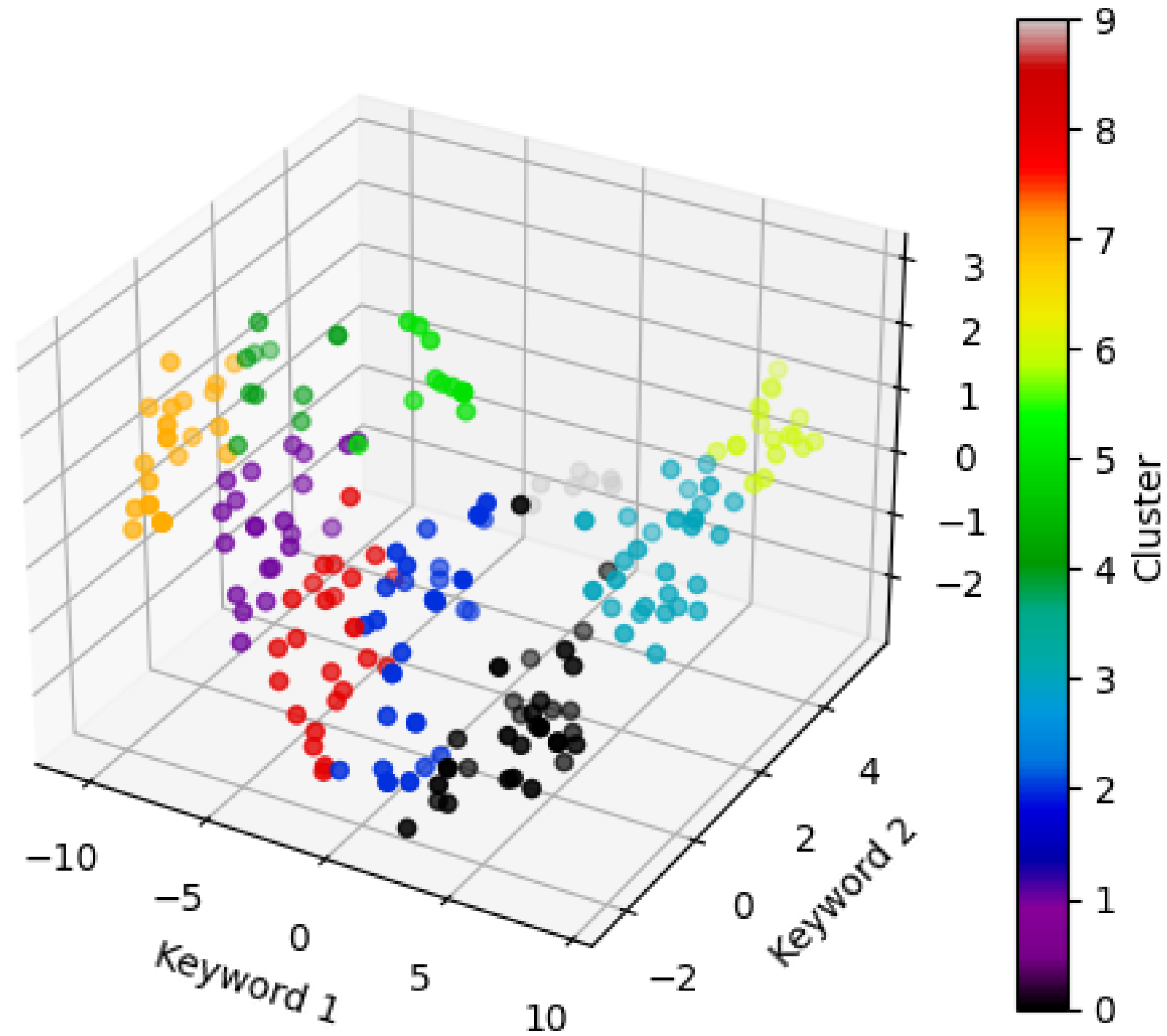


유효한 예측 수: 1
전체 훈련 횟수: 110
예측 적중 확률: 0.9091%

비지도 학습으로 10개 영역 분류-(추가) T-SNE

K-Means군집화를 사용하여 데이터를 10개의 클러스터로 나누고 3차원으로 시각화

t-SNE Visualization with 10 Clusters



한국어 Word2Vec 만들기

1.형태소 분석기를 사용하여 명사를 추출(kkma, komoran)

수정된 제목(kkma)	수정된 제목(komoran)	일치여부	불일치 키워드
채용 본교 인재 개발원 연구원 채용 공고	채용 본교 인재 개발원 연구원 채용 공고	O	
인재 개발원 채용 상담회 한국 경제인 협회 구 전국 경제인 연합회 채용 상담회 수	인재 개발원 채용 상담회 한국 경제인 협회 구 전국 경제인 연합회 채용 상담회 수	O	
채용 산학 협력단 직원 채용 공고	채용 산학 협력단 직원 채용 공고	O	
글로벌 소통 교육실 겨울 계절 학기 영어 말하기 프로그램 등록 안내	글로벌 소통 교육실 겨울 계절 학기 영어 말하기 프로그램 등록 안내	O	
언어 교육원 학년도 싱가포르 난양 공대 위탁 과정 언어 교육원 한국어도 우미 모집	언어 교육원 학년도 싱가포르 난양 공대 위탁 과정 언어 교육원 한국어도 우미 모집	O	
교육 혁신 센터 융합 수업 운영 사례 공 유희 융합 수업 운영 사례 발표 및 전시 안내	교육 혁신 센터 융합 수업 운영 사례 공 유희 융합 수업 운영 사례 발표 및 전시 안내	O	
인재 개발원 학년도 겨울 방학 현장 실습 모집 안내	인재 개발원 학년도 겨울 방학 현장 실습 모집 안내	O	
학년도 전기 년 월 졸업 예정 여부 조회 및 졸업 의사 신청 관련 안내	학년도 전기 년 월 졸업 예정 여부 조회 및 졸업 의사 신청 관련 안내	O	
국제 학생증 발급비 지원 행사 안내	국제학생증 발급비 지원 행사 안내	X	(‘학생증 지원 행사’, ‘발급비 지원 행사’)
학년도 전공 변경 시행 안내	학년도 전공 변경 시행 안내	O	
학년도 전과 시행 안내	학년도 전과 시행 안내	O	
국제 교류팀 싱가포르 교환 프로그램 설명회	국제 교류팀 싱가포르 교환 프로그램 설명회	O	
대학 건강 센터 월 학생 건강 검사 시행 안내	대학 건강 센터 월 학생 건강 검사 시행 안내	O	
연구 윤리 센터 월 연구 윤리 특강 신청 안내	연구 윤리 센터 월 연구 윤리 특강 신청 안내	O	
학부 학년도 학기 이화 플러스 장학금 신청 안내	학부 학년도 학기 이화 플러스 장학금 신청 안내	O	
다 문화 연구소 상호 문화 학술 대회 외국에서 바라보는 한국 상호 문화 교육적 관점에서	다 문화 연구소 상호 문화 학술 대회 외국에서 바라보는 한국 상호 문화 교육적 관점에서	O	
학부 년도 방일영 문화 재단 장학생 선발 안내 북한 이탈 주민 대학생 지원	학부 년도 방일영 문화 재단 장 학생 선발 안내 북한이 탈주민 대 학생 지원	X	(‘년도 주민 대학생’, ‘학생 년도 방일’)
글로벌 소통 교육실 겨울 계절 학기 영어 말하기 프로그램 등록 안내	글로벌 소통 교육실 겨울 계절 학기 영어 말하기 프로그램 등록 안내	O	
글로벌 소통 교육실 겨울 계절 학기 고급 영어 글쓰기 등록 안내	글로벌 소통 교육실 겨울 계절 학기 고급 영어 글쓰기 등록 안내	O	
인권 센터 인권 센터 젠더 교육 플랫폼효재 폭력 예방 특강 신청 안내	인권 센터 인권 센터 젠더 교육 플랫폼효재 폭력 예방 특강 신청 안내	O	
인재 개발원 진로 디자인 일 세계에서 나의 진로 찾기 특강 금	인재 개발원 진로 디자인 일 세계에서 나의 진로 찾기 특강 금	O	
학부 년도 전기 공사 공제 조합장 학회 장학생 선발 안내	학부 년도 전기 공사 공제 조합장 학회 장 학생 선발 안내	X	(‘년도 장학생 선발’, ‘선발 년도 전기’)
학생 서비스 센터 잃어버린 물건을 확인하십시오	학생 서비스 센터 잃어버린 물건을 확인하십시오	O	
미래 혁신 센터 대학 혁신 지원 사업 재학생 인식 조사 시행 안내	미래 혁신 센터 대학 혁신 지원 사업 재 학생 인식 조사 시행 안내	O	
채용 사회 과학 대학 사회 학과 사무실 계약 직원 채용 공고	채용 사회과학 대학 사회학과 사무실 계약 직원 채용 공고	X	(‘채용 사회 계약’, ‘채용 사회학과 사무실’)
학년도 전공 학과 결정 시행 안내	학년도 전공 학과 결정 시행 안내	O	
출판 문화원 이화 플래너 이화 달력 출시 기념 팝업 스토어	출판 문화원 이화플래너 이화 달력 출시 기념 팝업 스토어	X	(‘문화원 이화 팝업’, ‘팝업 문화원 이화플래너’)
조교 모집 학년도 학기 체육 과학부 행정실 급 조교 모집	조교 모집 학년도 학기 체육 과학부 행정실 급 조교 모집	O	
학년도 비 사범대 교직 과정 이수 신청 안내	학년도 비 사범대 교직과정 이수 신청 안내	O	
모집 년도 창업 교육 혁신 선도 대학 대학생 스타트업 애널리스트 기 모집	모집 년도 창업 교육 혁신 선도 대학 대학생 스타트업 애널 리스트 기 모집	X	(‘모집 년도 기’, ‘선도 대학 대학생’)
공연 문화 연구 센터 평화 음악회 대강당 공연 안내 월 일 토	공연 문화 연구 센터 평화 음악회 대 강당 공연 안내 월 일 토	X	(‘공연 월 일’, ‘공연 센터 평화’)

한국어 Word2Vec 만들기

2.한국어 Word2Vec 모델 생성(kkma)

불일치하는경우를 살펴보았을 때 kkma가 단어를 더 조밀하게 파악하는 경향이 보임
gensim으로부터 Word2Vec를 импорт 하여 한국어 Word2Vec 모델을 생성
교내공지사항 제목 220건을 데이터로 하고, 단어사전에 관심분야 15개의 단어를 추가

```
model1 = Word2Vec(sentences = totalnoun_list_kkma, vector_size = 100,  
window = 10, min_count = 1, workers = 2, sg = 1)
```

Fasttext로 벡터화

코사인 유사도 평균 계산

사용자가 관심분야 15개 중 공학, 공연, 역사, 여행을 선택했다고 가정
선택된 관심분야와 제목 키워드 한 개 간의 유사도 계산하여 평균 구하기

	0	1	2	3	4	5	6	7	8	9	...	210	211	212	213	214	215
여행	0.205768	0.102998	0.205768	0.077157	-0.133858	0.007201	0.005224	0.005224	0.122934	0.005224	...	0.114689	-0.053919	-0.167223	-0.104457	0.103384	-0.082038
공연	0.051580	0.125375	0.051580	-0.055467	-0.070215	-0.001064	0.018979	0.018979	-0.211778	0.018979	...	0.112831	-0.268537	-0.047154	-0.086210	0.009751	0.009619
공학	-0.020865	-0.124458	-0.020865	-0.016177	-0.008820	-0.110134	0.103689	0.103689	-0.132556	0.103689	...	-0.052305	0.101436	-0.038019	0.039893	0.060229	0.087719
역사	0.001058	0.002924	0.001058	-0.010737	0.156090	0.021628	-0.087177	-0.087177	0.227936	-0.087177	...	-0.134031	-0.006973	0.044244	-0.025853	0.057838	0.122468
평균	0.059386	0.026710	0.059386	-0.001306	-0.014201	-0.020592	0.010179	0.010179	0.001634	0.010179	...	0.010296	-0.056998	-0.052038	-0.044157	0.057801	0.034442

5 rows × 220 columns

Fasttext로 벡터화

HardVoting 가중치를 활용하여 추천 실행

코사인 유사도 평균값과 HardVoting 값을 곱하여 0인 경우 추천에서 제외
최종값이 큰 순서대로 내림차순하여 공지사항 리스트를 출력

	knn	svm	Logistic	HardVoting	cosine_similarity_avg	Hard*cosine
0	1	1	1	1	0.059386	0.059386
1	1	1	1	1	0.026710	0.026710
2	1	1	1	1	0.059386	0.059386
3	0	0	0	0	-0.001306	-0.000000
4	0	0	1	0	-0.014201	-0.000000
...
215	1	1	1	1	0.034442	0.034442
216	1	1	1	1	-0.020373	-0.020373
217	1	1	1	1	-0.052038	-0.052038
218	1	1	1	1	-0.052038	-0.052038
219	0	0	0	0	-0.003116	-0.000000

220 rows × 6 columns

	knn	svm	Logistic	HardVoting	cosine_similarity_avg	Hard*cosine
119	1	1	1	1	0.127496	0.127496
161	1	1	1	1	0.127496	0.127496
201	1	1	1	1	0.082412	0.082412
0	1	1	1	1	0.059386	0.059386
2	1	1	1	1	0.059386	0.059386
...
139	1	1	1	1	-0.088980	-0.088980
86	1	1	1	1	-0.088980	-0.088980
142	1	1	1	1	-0.088980	-0.088980
13	1	1	1	1	-0.096597	-0.096597
47	1	1	1	1	-0.122139	-0.122139

133 rows × 6 columns

최종 결과&설문조사 결과

Fasttext로 돌린 결과가 A, Word2vec로 돌린 결과가 B

다음 중 ‘공학’, ‘공연’, ‘여행’, ‘역사’와
더 관련이 높아보이는 것은?

A

B

1		1
2	119	[인재개발원] 23-2 이화멘토링데이 - 영업-마케팅-MD (10/7)
3	161	[인재개발원] 23-2 이화멘토링데이 - 영업-마케팅-MD(global.ver) (10/7)
4	201	2023 문화예술교육사 역량강화특강 <문화예술 치유프로그램 이해하기>안내
5	0	[채용] 본교 인재개발원 연구원 채용 공고
6	2	[채용] 산학협력단 직원 채용 공고
7	78	[채용] 이화여자대학교 기숙사 행정실 계약직원 채용 공고
8	92	[인사팀] 2023년 이화여자대학교 직원 채용 공고(2024.1.1.자)
9	176	[TELOSTRACK] 2023학년도 2학기 TELOSTRACK 신청 안내(~10/6)
10	195	[TELOSTRACK] 2023학년도 2학기 TELOSTRACK 신청 안내(~10/6)
11	141	[TELOSTRACK] 2023학년도 2학기 TELOSTRACK 신청 안내(~10/6)
12	164	[TELOSTRACK] 2023학년도 2학기 TELOSTRACK 신청 안내(~10/6)
13	214	[TELOSTRACK] 2023학년도 2학기 TELOSTRACK 신청 안내(~10/6)
14	115	[교육실]2023-2학기 졸업생을 위한 채플 소감문 모집(~10/30)
15	144	제11회 홍보아이디어 공모전 개최(모집: ~10.3(화)까지)
16	197	[모집] 이화 랩 스타트업 캠프(Ewha Lab Start-Up Camp) 참여팀-대학원 창업동아리 모집
17	111	[모집] 이화 스타트업 아카데미 'Lab to Scale Up' 수강생 모집
18	68	[학부] 2024년도 두산연강재단 대학 장학생 선발 안내(생활비)
19	203	[조교모집] 2024학년도 전기 일반대학원 전형관리조교 모집: ~ 9/13(수)
20	215	[채용]이화여자대학교 산학협력단 직원(사업관리 및 정산) 채용 공고
21	148	[이수매니지먼트 주식회사] 관리자 채용 공고
22	140	제2회 이화문예상 공모전 (시/시조/단편소설/수필)(~10.15.마감)

1		1
2	142	[대학건강센터] 10 월 피부과 (1차) 상담클리닉 안내
3	49	[이화인의 나눔가게] 랜앤루 '기부 새옷' 특판 안내
4	139	[학부] 2024년 STX장학재단 장학금 신청 안내
5	136	이화 저널리즘데이! 참가신청 안내 - 기자 선배와의 Q&A 등 (10.13.)
6	130	[학생지원팀] 학관 1층 연습실1호~3호 유레카 공간사용신청 관련 안내(10.5.목 12:00부터 유
7	44	[대학건강센터] 11월 피부과 1차 상담클리닉 안내 (11/3)
8	126	[안전팀] ECC 승강기 운행 중단 안내
9	48	[글로벌소통교육실] 인문계열 및 예체능계열 대상 계절학기 필수교양영어 미개설 사전 안내
10	123	[학사] 2023년 제4차 평생교육사 자격증 발급 신청 안내
11	145	[대학건강센터] 10월 학생건강검사 시행 안내
12	93	[학부] 2024년도 대산농촌재단 장학생 선발 안내
13	90	ALPS 동창회 기금 및 재학생 장학기금 마련 기증 비자회 개최 안내 (10/19, 목)
14	86	[인재개발원] 23-2 <취업 MC> 미리 경험해 보는 실무진 ONEDAY 모의 면접 안내
15	61	[강영대학행정실] 이화 CPA Day 안내 (11/14.화.17시)
16	81	[학부]「2023년도 충남 한뿌리 장학생」선발 안내
17	80	[학부] 2023년도 제4기 강원장학재단 장학생 선발 안내
18	67	[대학건강센터] 10월 2차 피부과 상담클리닉 안내 (10/20)
19	79	[인재개발원] 2023-겨울방학 현장실습 설명회 진행 안내: 10.27.(금) 15:00-16:30
20	149	* 교내 셔틀버스 미운행 기간 안내 - 9.28(목)~10.3(화) *
21	172	[학부] 2023년 2학기 고양시 대학생 본인부담 등록금 지원 안내
22	207	[출판문화원 신간안내] <시민생활과 법: 민사 생활법률 강의> (제4판)

총 13명

총 2명

Fasttext로 결정

감사합니다