

2023-2

텍스트마이닝

팀 프로젝트 최종보고서

---

# 취업플랫폼 채용 정보 및 리뷰 분석 모델 제안

박상학(2023522122)  
박세연(2023522114)  
황성아(2023522119)

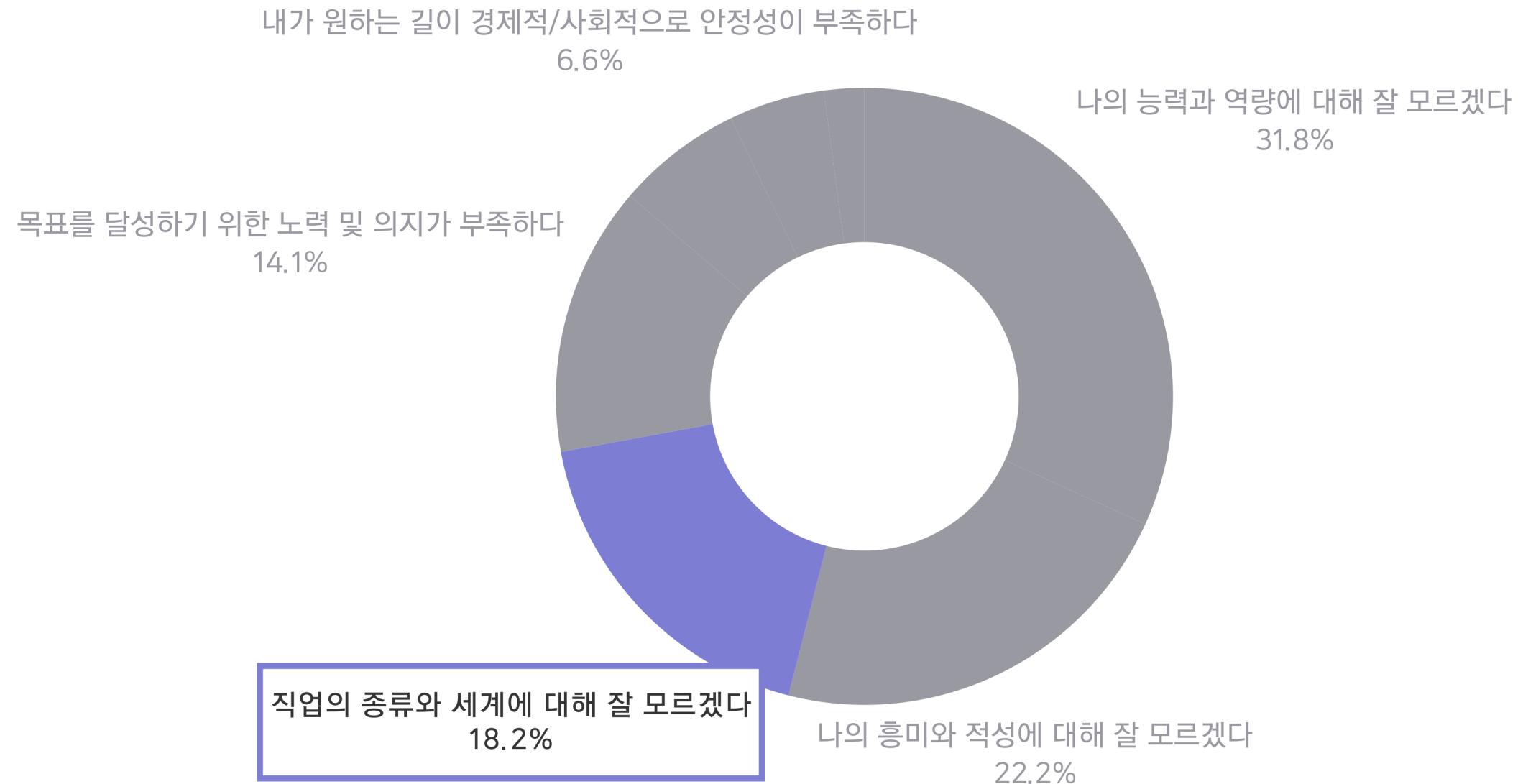
# 목 차

- I      연구배경 및 목적
- II     분석 주제
- III    데이터 수집 및 전처리
- IV    기업 리뷰 데이터 분석
- V     채용공고 데이터 분석
- VI    활용 방안
- VII   한계점
- VIII   향후 연구

## | 연구배경 및 목적

진로 및 직업 탐색을 할 때, 직종별 어떤 역량을 준비해야 할지 막막한 상황이 많음

진로 결정/선택에 어려움을 경험하고 있다면, 그 이유가 무엇이라고 생각합니까?



- 진로 및 직업 탐색자는 특정 기업의 공고나 리뷰를 보는 것보다는, 해당 직종의 공통적인 정보를 더욱 필요로 함
- 관심 직종의 전반적인 만족도는 어떠한지, 공통적으로 요구하는 역량에는 무엇이 있는지 등을 파악할 필요가 있음

## | 연구배경 및 목적

### 1) 흔히 접하는 정보와 실제 공고와의 불일치성

‘데이터 분석가 역량’에 대해 구글링한 결과, 엑셀 보고서를 잘 만드는 능력이 필요 요건 중 하나로 나옴

이러한 정보는 실제 요구 역량과 얼마나 일치할까에 대한 의문점

#### 3. 데이터 분석가에게 필요한 것은?

지금까지의 설명을 통해 데이터 분석가의 업무 내용을 대략 이해하고 있을 거라고 생각합니다. 그러면 데이터 공부는 어디서부터 시작해야 할까요? 다음 항목에서는 데이터 애널리스트에게 필요한 기술을 살펴보았습니다.



##### Excel- 엑셀 보고서

Excel은 표 계산 소프트웨어로서 널리 이용되면서 데이터 분석에도 중요한 역할을 하고 있습니다. Excel은 데이터 처리와 계산, 간단한 그래프 작성뿐 아니라 피벗테이블 만들기이나 Vlookup 등의 복잡한 함수를 포함한 고도의 기능도 구현할 수 있습니다. 데이터 애널리스트는 다른 데이터 분석이나 시각화 툴을 배우기 전에 Excel을 잘 사용해야 합니다.

출처: <https://m.site.naver.com/1gR4Q>

##### 자격 요건

- 이런 분과 함께하고 싶어요
  - 데이터 분석 관련 실무 경력 2년 이상이신 분
  - 데이터 분석, 머신러닝 또는 데이터 기반 마케팅에 대한 전반적인 지식을 보유하신 분
  - 데이터 추출, 전처리, 분석, 인사이트 도출까지 데이터 분석 전과정을 경험해보신 분
  - SQL을 자유자재로 활용하여 Raw Data 처리 및 분석이 가능하신 분
  - 분석 결과를 통한 결과 도출 및 논리적인 개선사항을 제시할 수 있는 분
  - 데이터 시각화(Tableau, Amplitude 등)에 대한 경험을 보유하신 분
  - 기획, 마케팅, 개발 등 다양한 직군과 원활한 커뮤니케이션 및 협업이 가능하신 분

##### 우대사항

- 이런 분이면 더 좋아요
  - 데이터 기반의 인사이트를 통해 서비스 지표 상승이나 문제 해결의 경험이 있으신 분
  - 평소에 꼼꼼하게 계산하는 것을 좋아하시는 분
  - 데이터 기반 인사이트를 누구나 이해할 수 있도록 쉽게 커뮤니케이션 할 수 있으신 분
  - 데이터 관련 전공 (통계학/컴퓨터공학/산업공학 등) 혹은 데이터 관련 교육을 이수하신 분
  - 데이터분석준전문가(ADsP), 빅데이터 분석기사 등 분석 관련 자격증 보유자
  - 마케팅 지표(first open, retention, conversion rate)에 대한 이해도가 높으신 분

출처: [https://www.jobplanet.co.kr/job/search?posting\\_ids%5B%5D=1258612](https://www.jobplanet.co.kr/job/search?posting_ids%5B%5D=1258612)

## | 연구배경 및 목적

### 2) 직종별 전반적인 공통 정보(요구 역량, 맡는 업무, 만족도 등)의 부재

취업 플랫폼에서는 현재 각 기업별 채용 공고와 리뷰는 확인할 수 있지만

직종별로 공통적으로 요구하는 역량이나, 직종별 만족도 등의 정보는 확인할 수 없음

The screenshot shows a grid of 12 company profiles from the '인력UP은 포지션 MATCH' platform. Each profile includes the company logo, name, brief description, rating, review count, and a 'View Details' button.

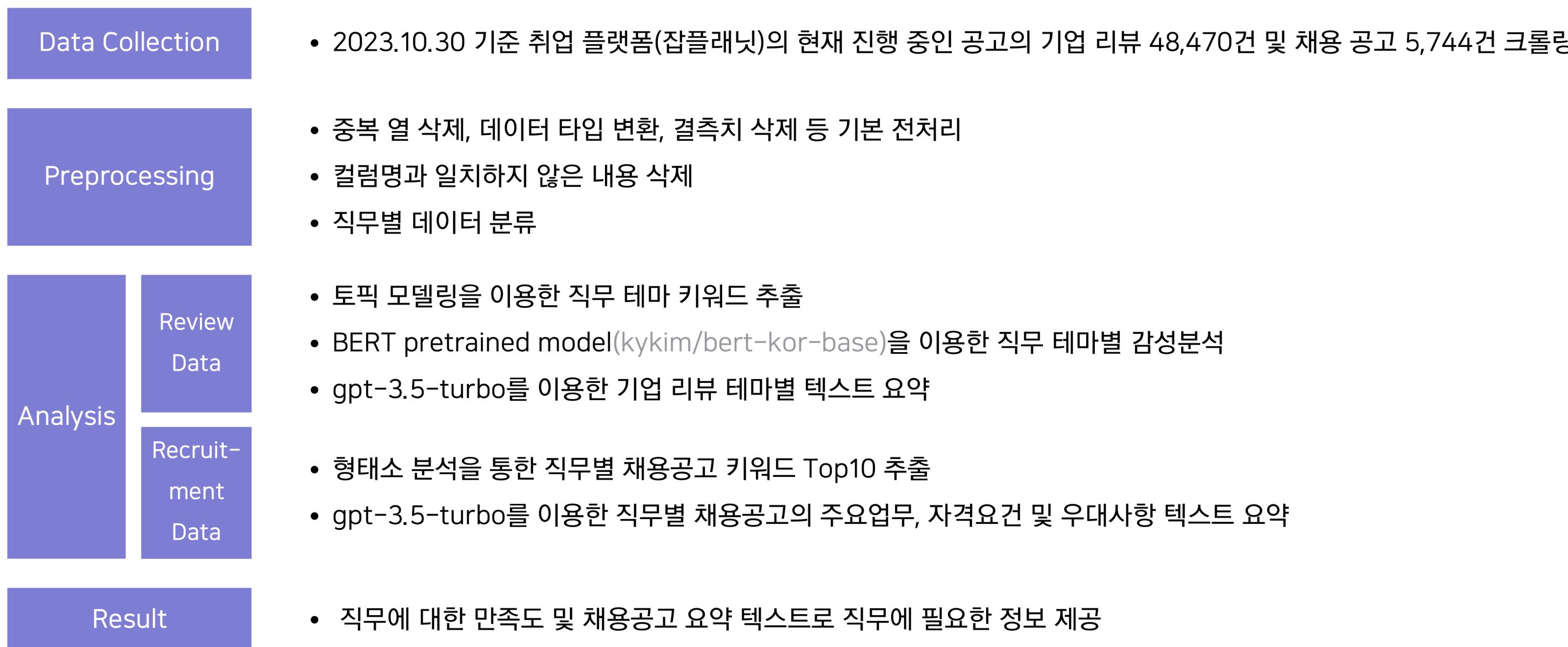
- NAVER**: 네이버 검색광고 공식대행사 채용관. Rating: ★☆☆☆☆, Reviews: D-3.
- 빙그레**: (주)빙그레 하반기 신입 수시채용 (회계 직무). Rating: ★☆☆☆☆, Reviews: D-3.
- 밀당PT**: 아이헤이트플라잉버그스(주) 부문 대규모 채용(신입 우대). Rating: ★☆☆☆☆, Reviews: D-9.
- koreaBio**: 한국바이오헬스 바이오헬스 온라인 채용관. Rating: ★☆☆☆☆.
- GC녹십자**: (주)녹십자 GC녹십자 계약직 채용. Rating: ★☆☆☆☆, Reviews: D-5.
- 점핏!**: 점핏 IT 개발자만을 위한 전문 채용관. Rating: ★☆☆☆☆, Reviews: D-12.
- UNISEM**: 유니셈(주) 반도체 엔지니어 외 각부문 채용. Rating: ★☆☆☆☆, Reviews: D-12.
- BORYUNG**: (주)보령 Onco부문 항암제 영업직. Rating: ★☆☆☆☆, Reviews: D-6.
- WOOJIN INC.**: (주)우진 2024년 신입/경력 공개채용. Rating: ★☆☆☆☆, Reviews: D-0.
- 우미건설**: 우미건설(주) 2024년 상반기 대졸 신입사원 채용. Rating: ★☆☆☆☆, Reviews: D-0.
- ITP 인천테크노파크**: 인천테크노파크 인천 우수기업 온라인 채용관. Rating: ★☆☆☆☆, Reviews: D-0.
- Synapse imaging**: 시냅스이미징(주) [시냅스 이미징] 연구소장 / 개발 / ... Rating: ★☆☆☆☆, Reviews: D-0.

출처: 사람인

## II 분석 주제

진로(직업) 탐색자를 위한 직무 비교, 직종별 주요 업무 및 자격 요건 요약, 만족도 정보 제공

### Research of Procedure



## III

## 데이터 수집 및 전처리

### 1) 기업 채용공고 데이터 크롤링

**Jobplanet** 기업 랭킹 채용 뉴스 기업 비교 멤버십

직종, 업무, 스킬, 회사 등

개발 경력 근무 지역 고용 형태 학력 초기화

총 999+건 정확도순

- QA 담당자 [ZimGo]**  
(주)트레드링스 | 서울  
복지 및 급여 : 4.0
- SW 품질 보증 신입/경력**  
(주)아이나비시스템즈  
업무와 삶의 균형 : 3.4
- Python Django 백엔드 개발자**  
(주)똑똑한개발자  
사내문화 : 4.3
- [신입] ASP.NET개발자**  
오엠티아이코리아(주)  
사내문화 : 5.0

**QA 담당자 [ZimGo]**

(주)트레드링스 | 서울

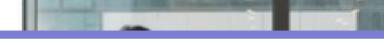
**상세** 리뷰 면접

**요약**

- 마감일: 2023.12.30 D-23
- 직무: QA(Quality Assurance)
- 경력: 3년 이상
- 고용형태: 정규직
- 근무지역: 서울
- 스킬: QA, istqb, 테스트, 검증

**기업 이미지**






**이름** 데이터 사이언티스트

이력서 10% 완성

잡플래닛에서 지원하면?  
당신의 합격을 위해 준비했어요.

- 1차 합격 시 면접 가이드
- 최종 합격 시 연봉 협상 가이드
- 취업 축하금 200만원

어디에서 신청하면 되나요?

**공고 저장**

**지원하기**

?

7

### III 데이터 수집 및 전처리

#### 2) 기업 리뷰 데이터 크롤링

The screenshot shows a company review card for 'QA 담당자 [ZimGo]' from '(주)트레드링스'. The review has a rating of 4.0. The card includes sections for '승진 기회 및 가능성', '복지 및 급여', '업무와 삶의 균형', '사내문화', and '경영진'. A purple box highlights the review content.

Jobplanet 웹사이트 스크린샷입니다. 중앙에 표시된 회사는 '(주)트레드링스'로, 직무는 'QA 담당자 [ZimGo]'입니다. 평점은 4.0입니다. 리뷰 내용은 다음과 같습니다.

기획/경영 | 현직원 | 경기 | 2023. 10

★★★★★ ★

"조용한 도서관 같은 곳, 딱히 시끄럽지 않고 조용히 지낼 수 있음"

승진 기회 및 가능성

복지 및 급여

업무와 삶의 균형

사내문화

경영진

장점

단점

이름: 데이터 사이언티스트

이력서 10% 완성

잡플래닛에서 지원하면?

1차 합격 시 면접 가이드

최종 합격 시 연봉 협상 가이드

취업 축하금 200만원

어디에서 신청하면 되나요?

공고 저장

지원하기

## III

## 데이터 수집 및 전처리

Descriptive Statistics of Company Review Data

Column Name	Data Type	Count	Total Token	Noun Token
직무	string	48,132	-	-
고용 환경	string	48,132	-	-
근무지	string	48,132	-	-
날짜	datetime	47,266	-	-
별점	int	48,470	-	-
요약	string	48,470	1,006,746	411,109
장점	string	48,470	1,453,803	613,439
단점	string	48,470	1,905,969	779,140
경영진에게 바라는 점	string	48,470	1,202,806	456,641
Total	-	-	5,569,324	2,260,329

Descriptive Statistics of Recruitment Post Data

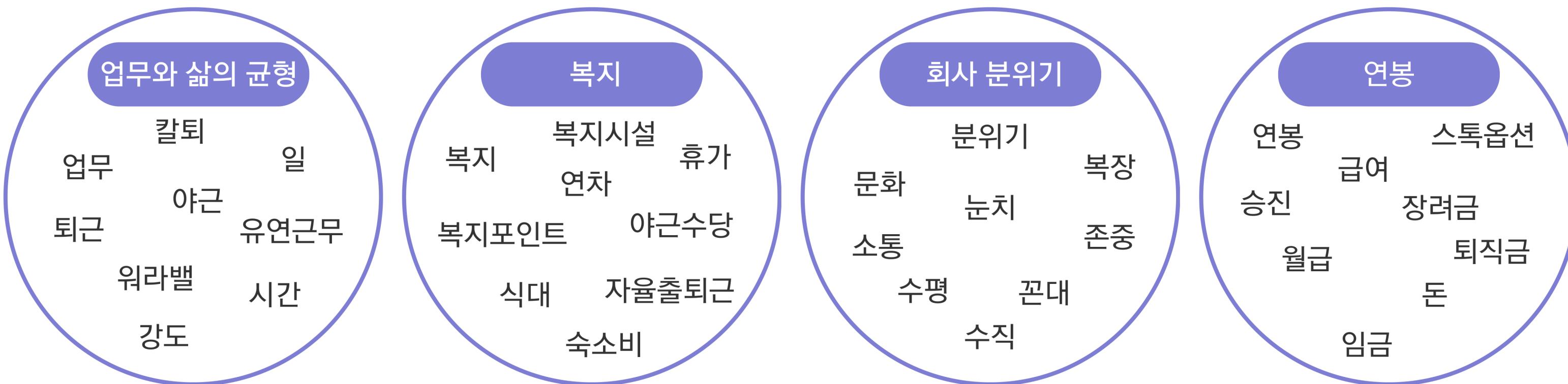
Column Name	Data Type	Count	Total Token	Noun Token
회사명	string	5,744	-	-
고용 형태	string	5,744	-	-
스킬	string	5,744	-	-
기업 소개	string	5,207	747,322	325,003
주요 업무	string	5,744	277,018	140,643
자격 요건	string	5,744	199,072	90,880
우대 사항	string	5,331	162,680	77,661
채용 절차	string	3,814	177,577	89,778
복리후생	string	5,289	610,718	301,743
회사위치	string	5,708	-	-
직무(복수선택)	string	5,744	-	-
직무_1v1	string	5,744	-	-
직무_1v2	string	5,744	-	-
Total	-	-	2,174,387	1,025,708

## 3) 데이터 전처리

- 중복 열 삭제, 데이터 타입 변환, 주요 컬럼 제외 결측값 존재 행 삭제 등 기본적인 전처리 완료
- 컬럼과 일치하지 않은 내용이 다수 존재하므로, 해당 컬럼 관련 주제가 아닌 다른 내용은 삭제
- 분석에 사용할 컬럼에 대해 문자열 정제

## IV 기업 리뷰 데이터 분석

1) 테마별 감성분석을 위해 토픽 모델링(num\_topics=4)을 통해 테마 키워드 생성



개발 | 전직원 | 서울 | 2023. 11

★★★★★

"근무시간 잘 지켜지고 복지가 좋습니다"

승진 기회 및 가능성



복지 및 급여



업무와 삶의 균형



사내문화



경영진



장점

근무시간이 잘 지켜진게 최대 장점이라고 생각합니다

단점

업무 강도가 높지만 전반적으로 단점은 별로 없는 곳이라고 생각합니다

경영진에 바라는 점

재택 등 직원 복지 확충에 노력해주세요

이 기업은 1년 후 비슷할 것이다.

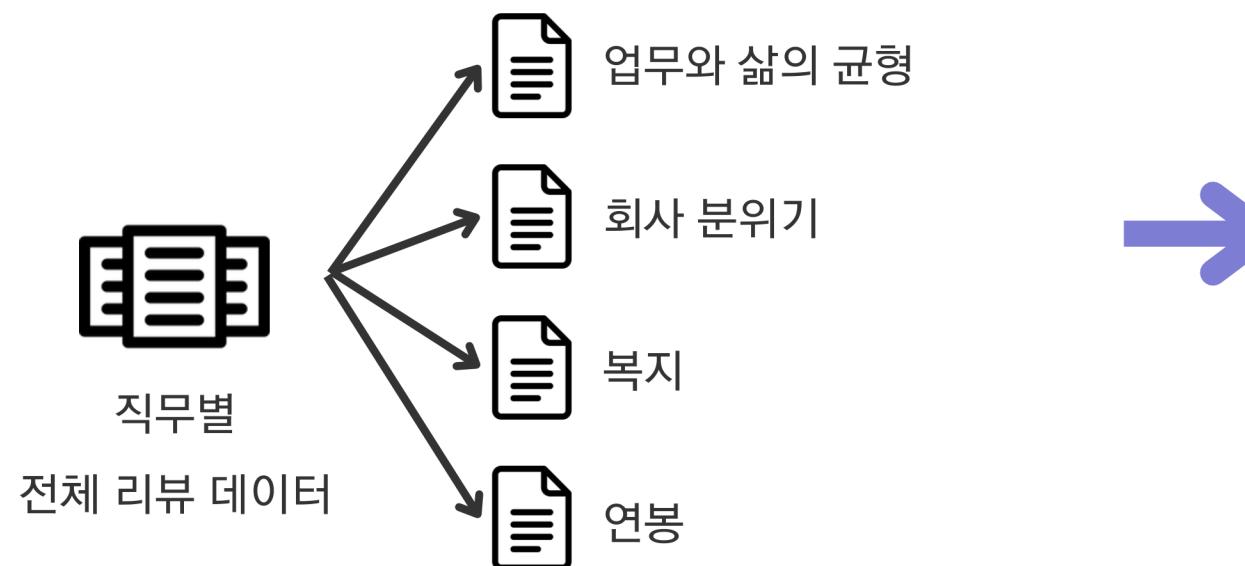
이 기업을 추천 합니다!

출처: 잡플래닛 기업리뷰, 네이버

개선점) 기존 잡플래닛의 테마는 사람이 직접 별점을 매기는 방식으로 작성한 글과의 불일치 가능성 존재  
토픽모델링으로 실제 문장과 일치하는 테마별 감성분석, 및 요약 정보 제공

## 2) 직무(큰 카테고리)의 테마별 감성 스코어 계산

## IV 기업 리뷰 데이터 분석



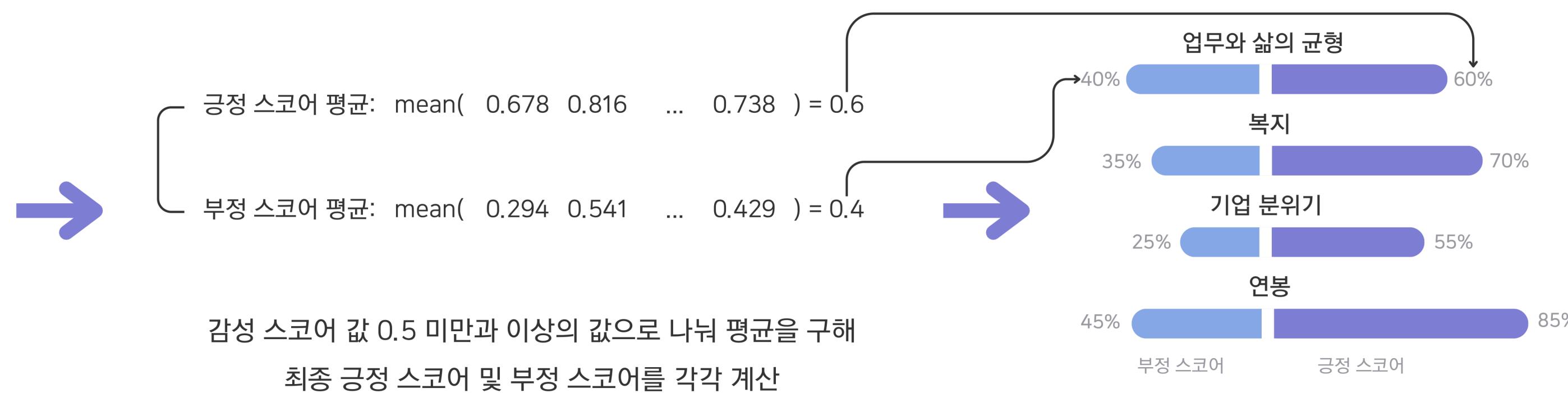
## 직무의 전체 리뷰 문장을 4가지 테마에 맞게 분류

테마1	BERT pretrained model
	<u>감성 스코어</u>
• 리뷰 데이터1	0.678
• 리뷰 데이터2	0.816
• 리뷰 데이터3	0.294
...	...
• 리뷰 데이터K	0.429

Model	Derived model (Pre-trained language model)	accuracy(val_acc) unit : %		
		NSMC	Steam	Movie
RNN	BILSTM	91.37	77.48	85.74
	Bi-GRU	91.43	77.51	85.28
BERT	M-BERT	88.87	77.46	85.76
	KLUE-BERT	91.20	82.12	90.25
	LMkor-BERT	93.86	83.41	90.30
	LMkor-ALBERT	93.15	82.53	88.97
	KLUE-RoBERTa	93.01	82.23	89.66
	XLM-R	93.04	80.92	87.21
GPT	KoGPT2	92.01	79.87	86.90
	GPT-3	94.04	78.03	90.35

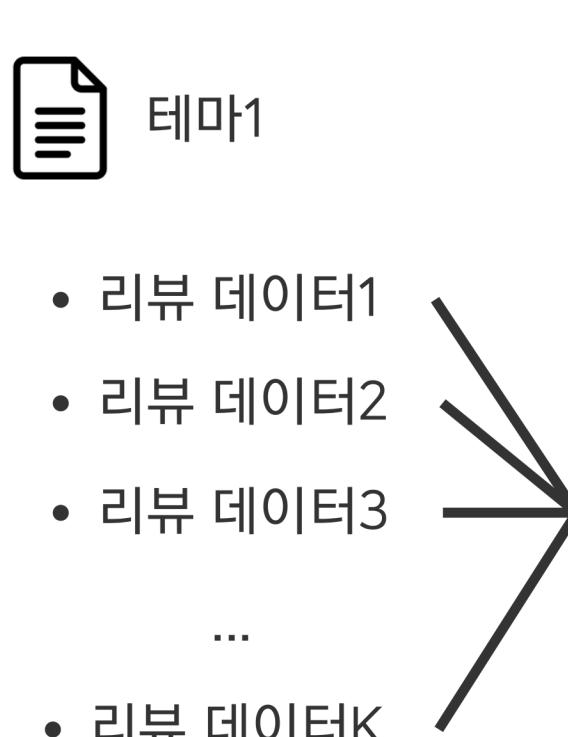
출처: <https://github.com/kiyoungkim1/LMkor>

이재홍. (2023). RNN과 트랜스포머 기반 모델들의 한국어 리뷰 감성분류 비교. *한국전자통신학회 논문지*, 18(04), 693-700.



## IV 기업 리뷰 데이터 분석

### 3) 직무(큰 카테고리)의 테마별 리뷰 데이터 요약



gpt-3.5-turbo

i.e. 개발 - Salary

급여 높은편, 복지포인트, 숙소비, 적당한 급여, 리플래쉬 포인트, 낮지않은 연봉

i.e. 개발 - Environment

회사 분위기는 부서에 따라 다르며, 워라밸을 중시하는 사람들에게는 힘들 수 있는 환경이며, 상위 직원간의 정치적인 문제와 복지, 보상 등이 부족하여 직원들의 이직률이 높음.

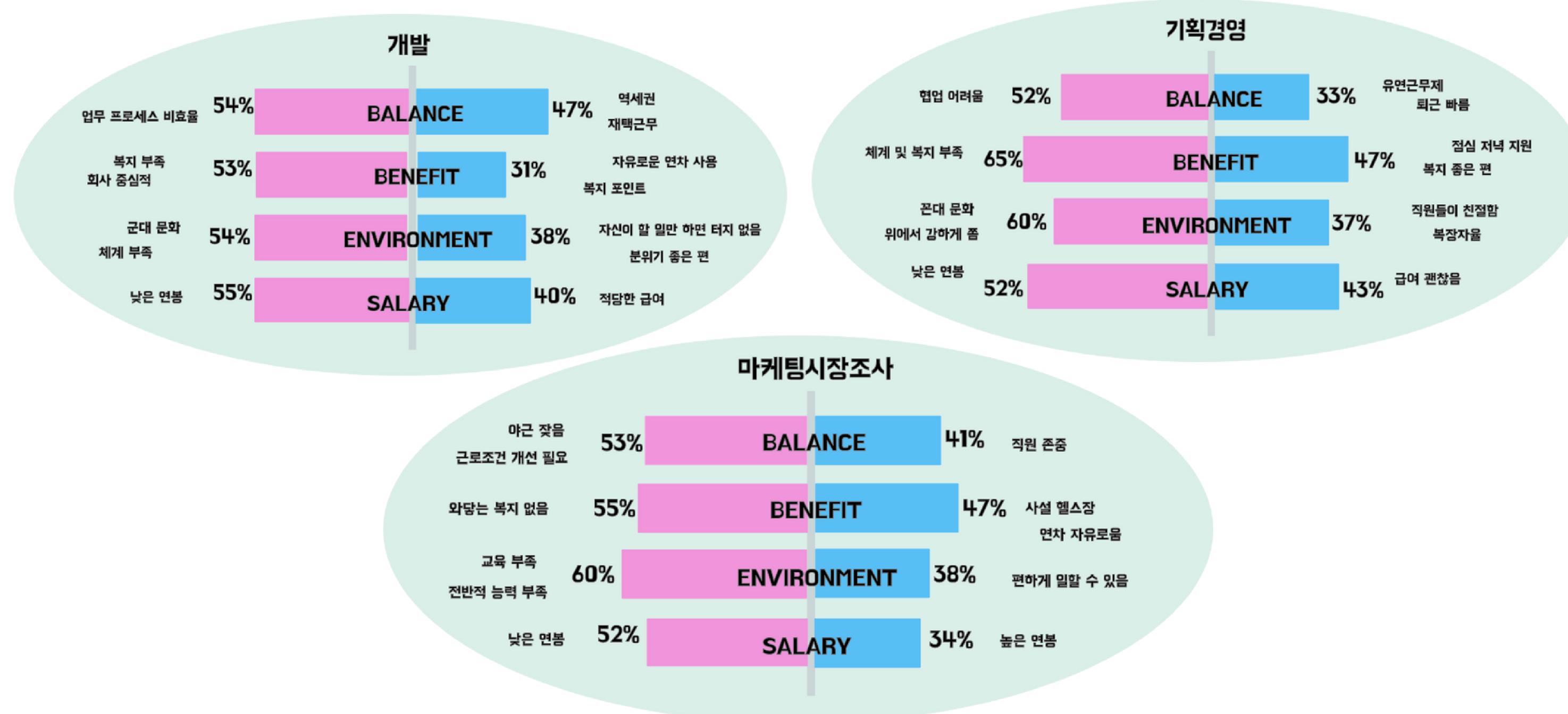
i.e. 개발 - Balance

회사의 복지 제도가 부족하며, 특별한 복지나 보상이 없어 직원들의 불만이 많다는 점이 도드라짐

ChatGPT를 이용해 지시문 방향성 제시 및 테마별 리뷰 요약글 생성

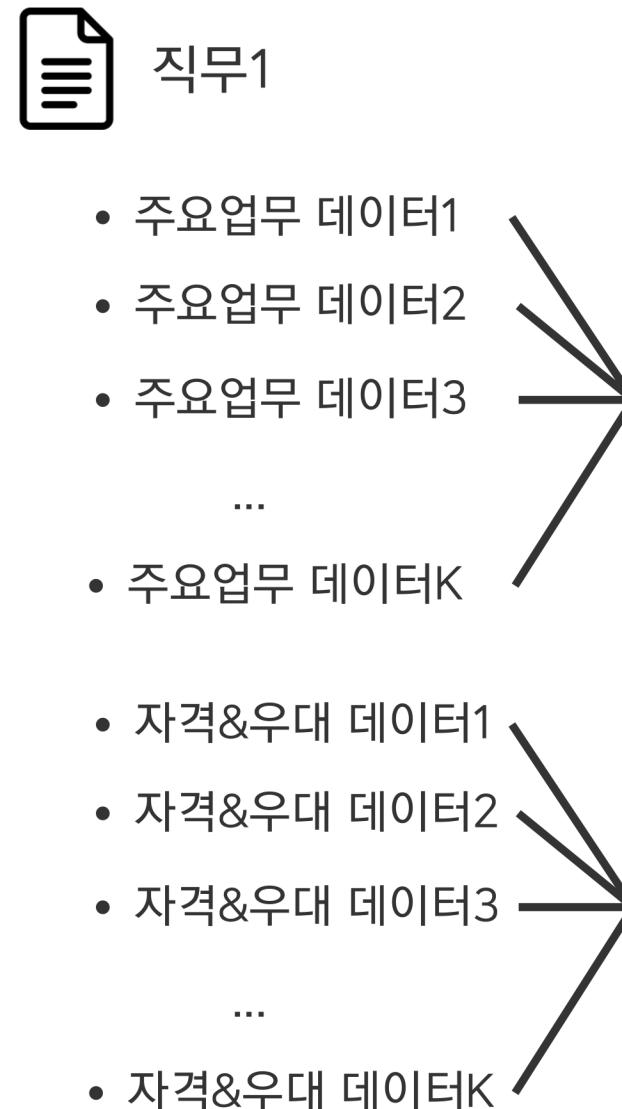
## IV 기업 리뷰 데이터 분석

### 직무별 재직자 만족도

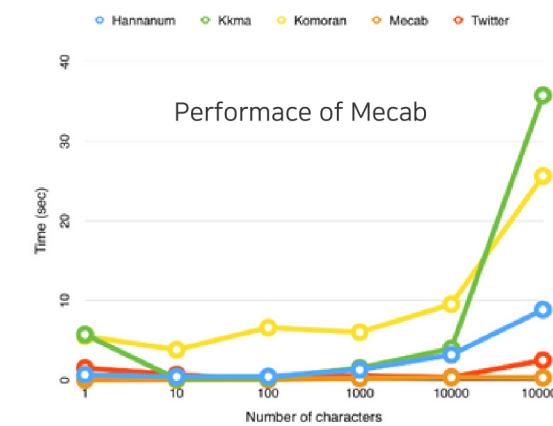


## V 채용공고 데이터 분석

### 1) 직무(하위 카테고리)의 채용공고(주요업무, 자격요건및우대사항)의 키워드(한글, 영어) Top10 추출



직무별 주요업무, 자격요건및우대사항 데이터 합치기



Mecab  
nltk

한글, 영어 명사 형태소  
종합 Top 10 추출



e.g. 개발 - ios Top10 키워드

관리 보수 운영 개선

어플리케이션 Native 유지

앱 플랫폼 모바일

주요업무 Top10

배포 프로젝트

설계 앱 MVVM

커뮤니케이션 관리

디자인 Swift iOS

자격요건 및 우대사항 Top10

## V 채용공고 데이터 분석



## V 채용공고 데이터 분석

### 2) 직무(하위 카테고리)의 채용공고(주요업무, 자격요건및우대사항) 데이터 요약



vi

활용 방안

- 실제 직종별 현업 종사자의 리뷰 및 실제 채용 공고에 대한 요약 및 정리를 제공함으로써 진로 탐색자에게 보다 신뢰도 높은 정보 제공
  - 직무별 정보를 얻기까지 정보 탐색 시간의 단축 효과
  - 기업별 정보에서 벗어나 직무별 공통적인 정보를 한눈에 파악 가능

취업플랫폼 채용 정보 및 리뷰 분석

직무성향파악

Search

Search

 **직무 성향**

기업리뷰 데이터를 통한 테마별 감성분석 스코어 및 요약 정보 제공

**직무**

개발 교육 금융/재무

기획/경영 데이터 디자인

마케팅/시장조사 법률/법무

미디어/홍보 생산/제조

생산관리/품질관리

서비스/고객지원 엔지니어링

연구개발 영업/제휴

유통/무역 의약 인사/총무

전문직 특수계층/공공

**직무의 테마별 감성분석**

업무와 삶의 균형

40/100

복지

41/100

분위기

40/100

연봉

41/100

회사 위치가 좋고 사람들이 무난하고 서울 중심에 위치해 출퇴근이 편리하여 연차를 자유롭게 쓸 수 있고 멀스케어라는 복지가 있다.

연차를 눈치 보지 않고 쓸 수 있고 체력단련 휴가를 더 주며 매년 복지 포인트를 지급한다.

회사 분위기가 좋은 편이며 육아병 험할 수 있는 제도를 사용할 수 있고 자유로운 분위기이며 자신이 할 일만 하면 크게 퇴치당할 일이 없다.

칼퇴 가능한 부서가 많고 학급자가 다니기에 문화가 원활 내부 분위기 비교적 자유롭고 연봉도 나쁘지 않은 수준이다.

대기업의 단점과 스타트업 단점이 있는 곳 부서간 이해관계나 금융업계의 보수적인 업무처리 방식등이 단점으로 생각된다.

사내 폭지를 못 누린다고 생각하면 되고 정규직은 시장 평균 연봉에 못 미치는 경우가 대다수이다.

스타트업의 단점과 기존 금융사 단점이 격렬히 합쳐져서 시작한지 얼마 되지 않아 체계가 없고 복지가 안 좋으며 성과가 적고 직원들한테 돌아오는 게 적은 느낌이 든다.

금융권 치고 낮은 연봉 부서마다 분위기가 너무 다른 업무 프로세스도 비효율 본부 별 편입 및 의사소통 비원활 금융권 치고 낮은 연봉 부서마다 분위기가 너무 다른 업무 강도 를 높인 것 같음

**직무 분석**

직무별 채용공고 데이터를 통한 키워드 추출 및 요약 정보 제공

취업플랫폼 채용 정보 및 리뷰 분석

직무성향파악

직무분석

Search

직무 분석

직무별 채용공고 데이터를 통한 키워드 추출 및 요약 정보 제공

## 세부 직무

- > 개발
- > 교육
  - 교육기획/고재개발
  - 교직원
  - 대학교수
  - 외국어/여학원 강사
  - 유치원/보육 교사
  - 입시/보습/학원 강사
  - 전문강사
  - 초등고/특수학교 교사
  - 학습지/과외/방문 교사
- > 금융/재무
- > 기획/경영
- > 데이터
- > 디자인
- > 마케팅/시장조사
- > 미디어/홍보
- > 법률/법무
- > 생산/제조
- > 생산관리/품질관리

### 직무의 인기 키워드 Top 10

keyword1 키워드가 나온 문장 하나 정도 보여주기	keyword2 키워드가 나온 문장 하나 정도 보여주기	keyword3 키워드가 나온 문장 하나 정도 보여주기	keyword4 키워드가 나온 문장 하나 정도 보여주기	keyword5 키워드가 나온 문장 하나 정도 보여주기
keyword6 키워드가 나온 문장 하나 정도 보여주기	keyword7 키워드가 나온 문장 하나 정도 보여주기	keyword8 키워드가 나온 문장 하나 정도 보여주기	keyword9 키워드가 나온 문장 하나 정도 보여주기	keyword10 키워드가 나온 문장 하나 정도 보여주기

### 주요업무와 자격요건 및 우대사항

- 주요업무**  
백엔드 시스템 개발과 운영을 담당하는 Who We AreSaaS Platform 팀은 SSPSupplySide Platform의 백엔드 시스템 개발과 운영을 담당한다.
- 자격요건 및 우대사항**  
하나 이상의 서버 프로그래밍 언어에 능숙하며 테스트하기 좋은 코드를 작성하고 문제 해결 후 재발 방지를 위한 해법을 고민한다.

### 관련 채용공고 알아보기

buzzvil  
No.1 리워드 광고 플랫폼

humusOn  
유니온

번개장터

## VII 한계점

- 리뷰 데이터는 채용공고 데이터와 다르게 직무의 큰 카테고리 데이터가 존재해, 세부 직무별 감성 분석을 하진 못함
- Gpt3.5-turbo의 hallucination 문제로 인한 아쉬운 성능의 요약문 생성
- 토픽모델링으로 분류된 토픽의 모호함
- 직무별 감성분석 결과의 편차가 생각보다 적음

## VIII 향후 연구

- Aspect-based sentiment analysis 방법으로 감성분석
- 새로운 기준으로 직무 카테고리 생성
- ChatGPT의 hallucination 완화에 대한 연구

---

**감사합니다**