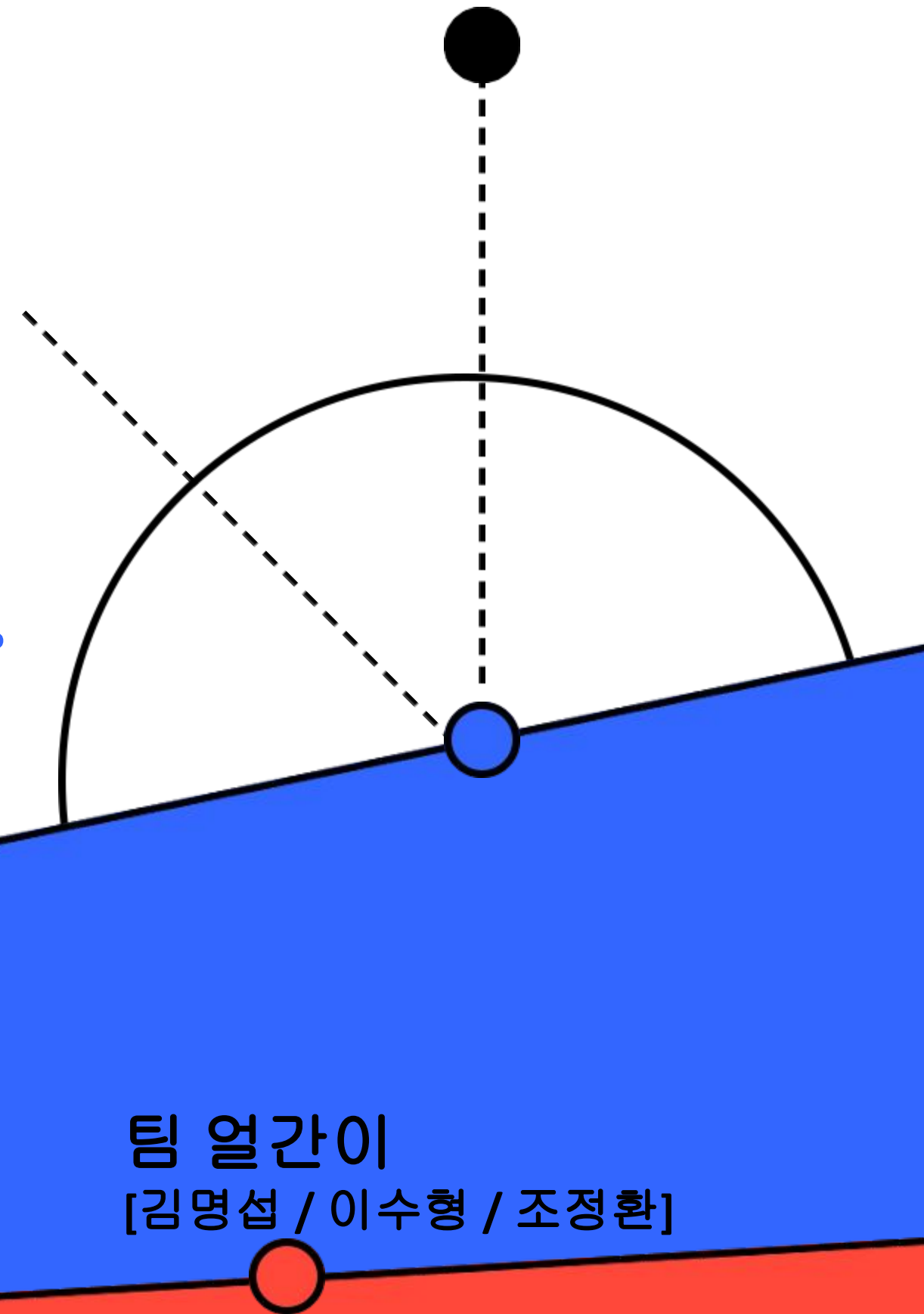
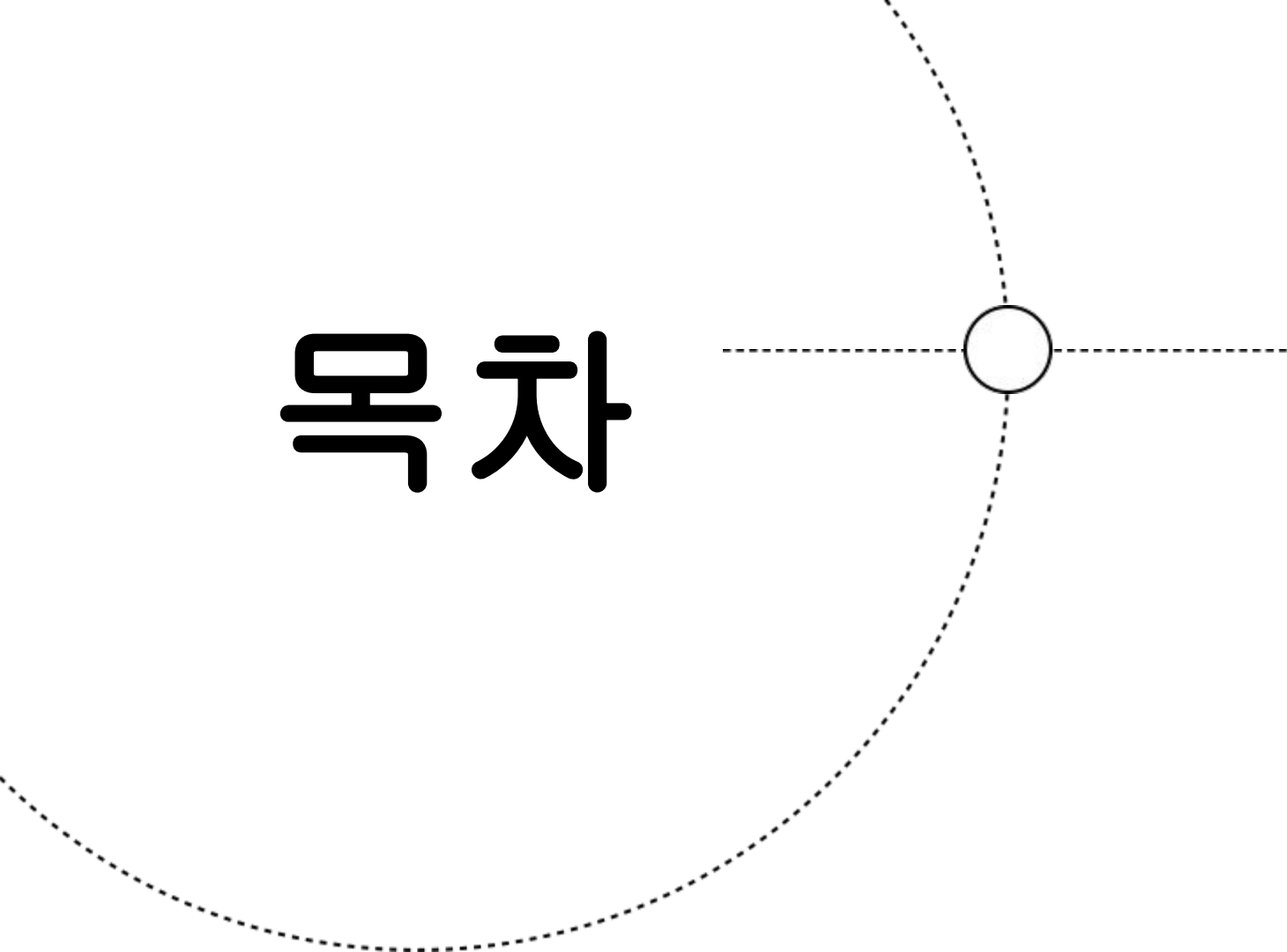


# 기업 리뷰 데이터를 활용한 이직률 예측 모델 제안 및 기업 분석

팀 얼간이  
[김명섭 / 이수형 / 조정환]



# 목차



- 01 사업 개요
- 02 기술 개발 및 결과
- 03 데이터 분석
- 04 기술 활용방안
- 05 후행 연구 제안

# 01 사업 개요

02 기술 개발 및 결과

03 데이터 분석

04 기술 활용방안

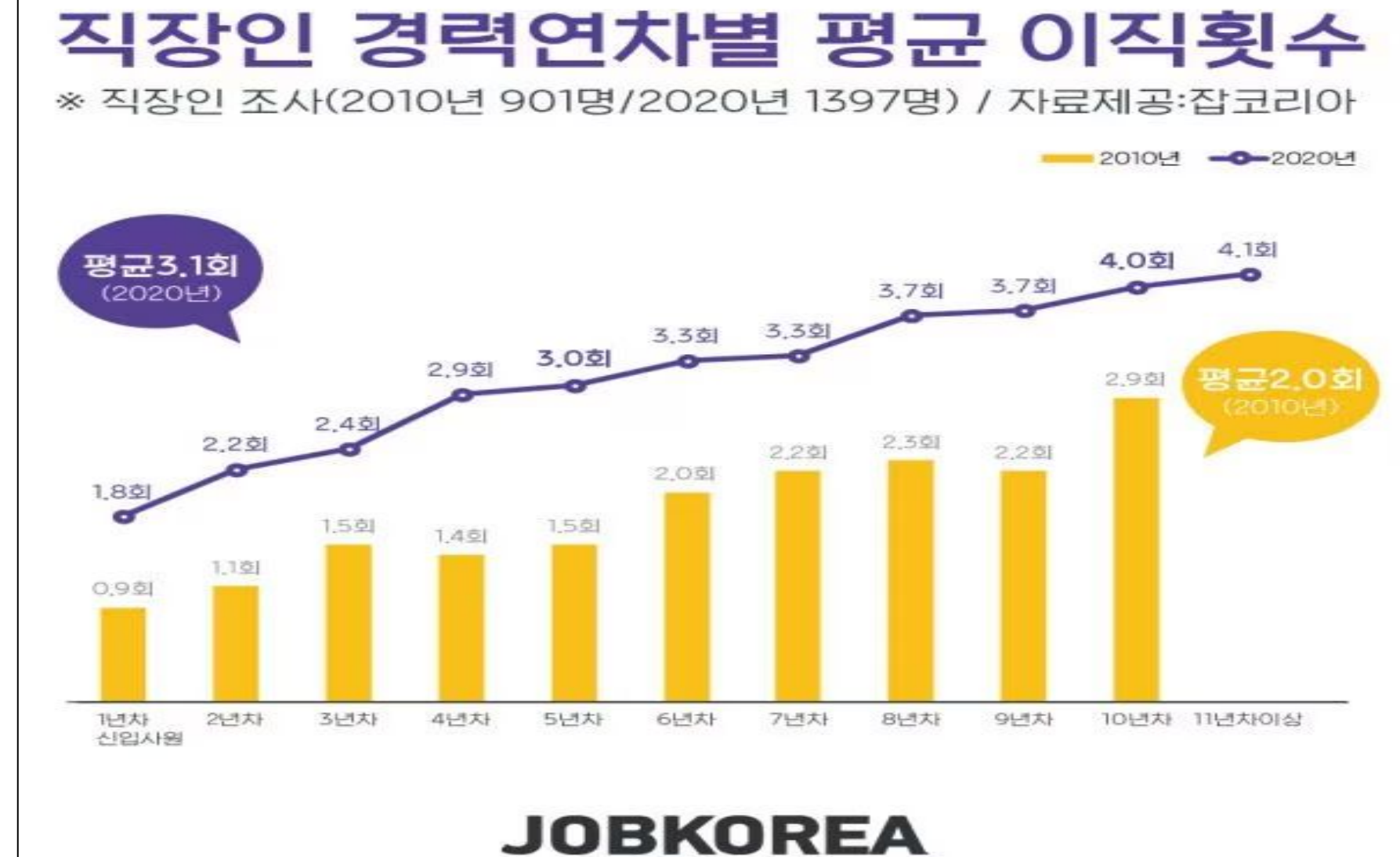
05 후행 연구 제안

# 연구 배경 및 필요성

## 조직문화의 변화



## 이직 횟수 증가 및 기업 대응 필요성



# 선행 연구

## 선행 연구 1

Journal of the Korean Data &  
Information Science Society  
2021, 32(5), 1035-1047

<http://dx.doi.org/10.7465/jkdi.2021.32.5.1035>  
한국데이터정보과학회지

### IT 기업 직원의 만족 및 불만족 요인에 따른 이직률 예측: 토픽모델링과 머신러닝을 활용하여<sup>†</sup>

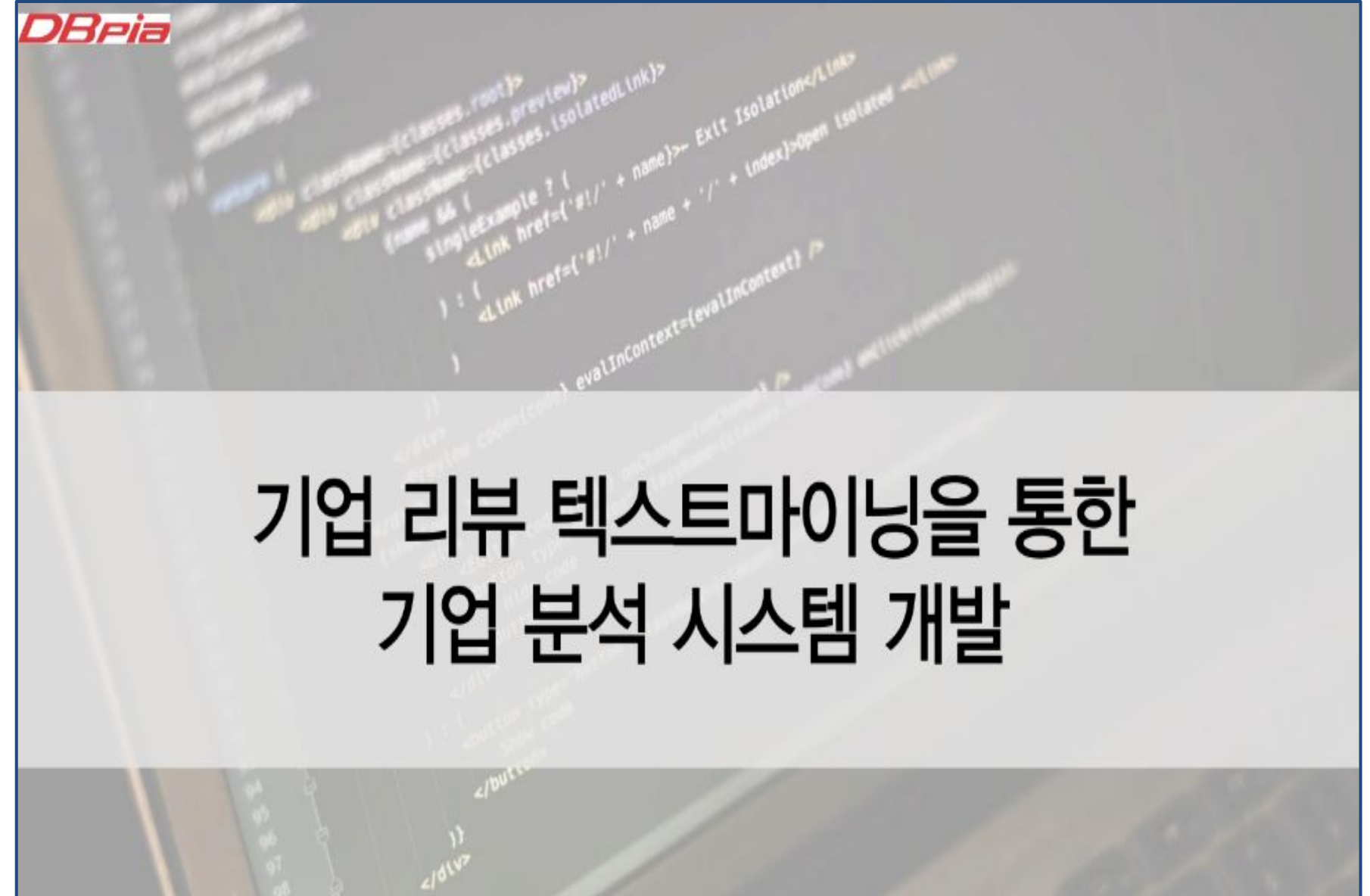
최진욱<sup>1</sup> · 신동원<sup>2</sup> · 이한준<sup>3</sup>

<sup>1</sup>고려대학교 스마트미디어 서비스 연구센터 · <sup>2</sup>명지대학교 산학협력단 · <sup>3</sup>명지대학교 경영정보학과  
접수 2021년 8월 27일, 수정 2021년 9월 7일, 게재확정 2021년 9월 7일

#### 요약

IT 산업에서의 이직은 기술 축적과 개발 및 관리에 있어 기업들에게 큰 어려움이 되고 있어 이에 대한 연구가 긴요한 실정이다. 하지만 아직까지 IT 분야의 이직 및 인력 관리와 관련된 연구는 많지 않다. 이에 본 연구에서는 이직률에 영향을 미치는 직원들의 만족 요인과 불만족 요인들을 살펴보고 이를 활용한 IT 기업의 이직률 예측 모델을 제안하고자 한다. 이를 위해 온라인 기업 리뷰 사이트인 잡플래닛에서 국내 유가증권시장에 상장된 129개 IT 기업에 종사하는 직원들의 리뷰 21,589건을 수집하고 토픽모델링을 실시하였다. 추출된 토픽들을 활용하여 이직률에 대한 머신러닝 기반의 예측 모델을 제안하였다. 또한 변수 중요도 평가를 통하여 이직률에 영향을 미치는 직원들의 만족 및 불만족 요인별 영향 정도를 분석하여 제시하였다.

## 선행 연구 2





# 사업 목표

## 개발 목표

포스트 코로나 시대와 MZ  
세대들의 취업시장 진입으로 인한  
변화 관측

실근무자의 평가가 반영된  
리뷰데이터를 활용하여 이직률 예측  
모델 개발

## 활용 방안

이직률 예측을 통해  
기업 운영 평가와 방향 수정

각각의 토픽에서  
기업의 객관적인 위치 확인 가능

01 사업 개요

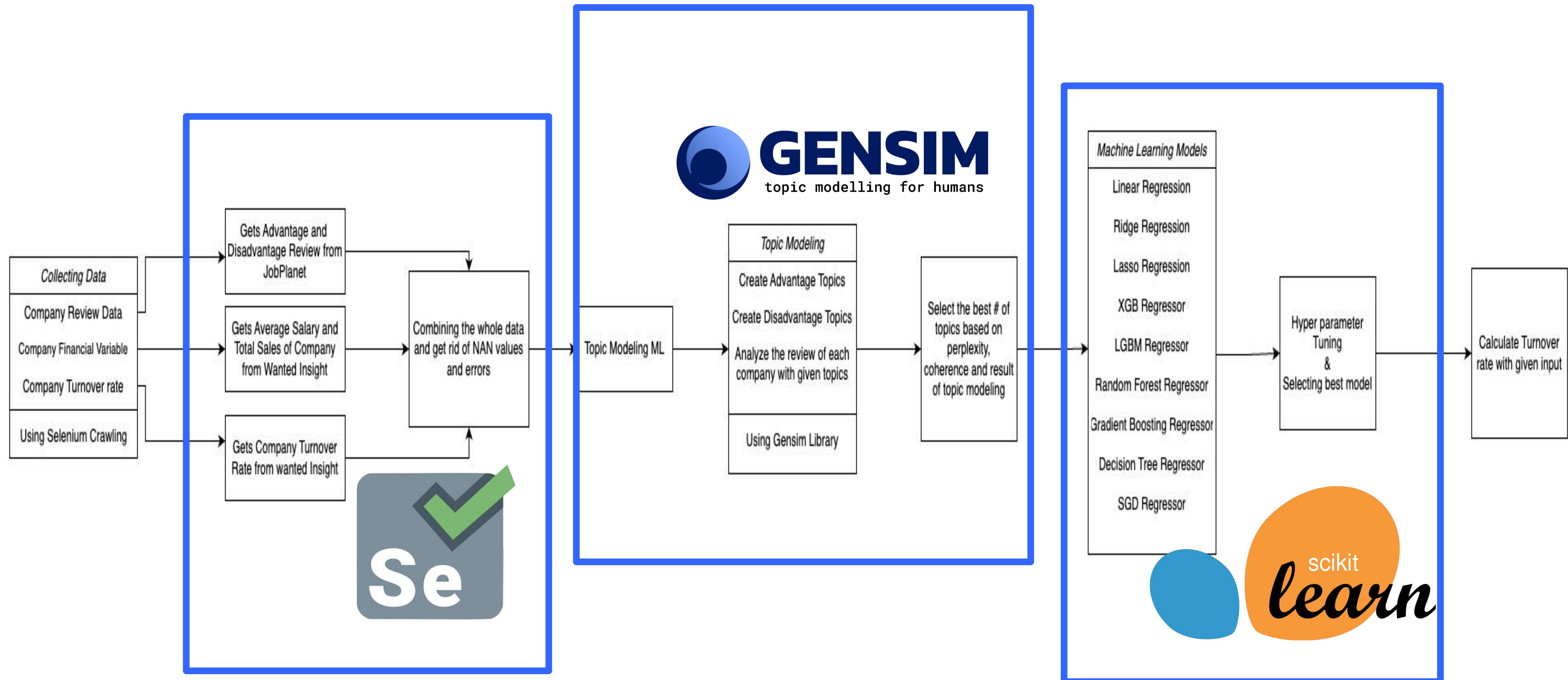
## 02 기술 개발 및 결과

03 데이터 분석

04 기술 활용방안

05 후행 연구 제안


# 기술 개발





# 데이터 수집

## 기업 리뷰 데이터

 서울 | 2023. 09

★★★★★

"안정적인 통신산업군의 대기업으로 탄탄한 사업구조"

승진 기회 및 가능성

■■■■■

복지 및 급여

■■■■■

업무와 삶의 균형

■■■■■

사내문화

■■■■■

경영진

■■■■■

장점

재택(팀바탕), 기타 자잘한 복지, 통신비 지원, 복포, 의료비 지원 등 전반적으로 업무 강도가 그렇게 강하진 않음

단점

연봉이 너무 째  
불필요한 보고, 광팔이들이 난무함, 일 처리가 답답함

경영진에 바라는 점

타사 대비 연봉 경쟁력이 약합니다.  
너무 광팔이식 신사업에 휘둘리지 말고 기업을 탄탄하게 유지했으면 좋겠습니다.

이 기업은 1년 후 성장하고 있을 것이다.

이 기업을 추천 합니다!

잡플래닛

## 재무제표 데이터



예상평균연봉

12,848만원 상위 1%

월 세전 1,071만원

출처 : 금감원 2023.03



총 매출액

199.7조원 상위 1%

1인당 매출액 17.9억원  
매출 대비 임금 3%

출처 : 금감원 2021.12

원티드 인사이트

## 이직률 데이터



총 인원

35,448명 상위 1%

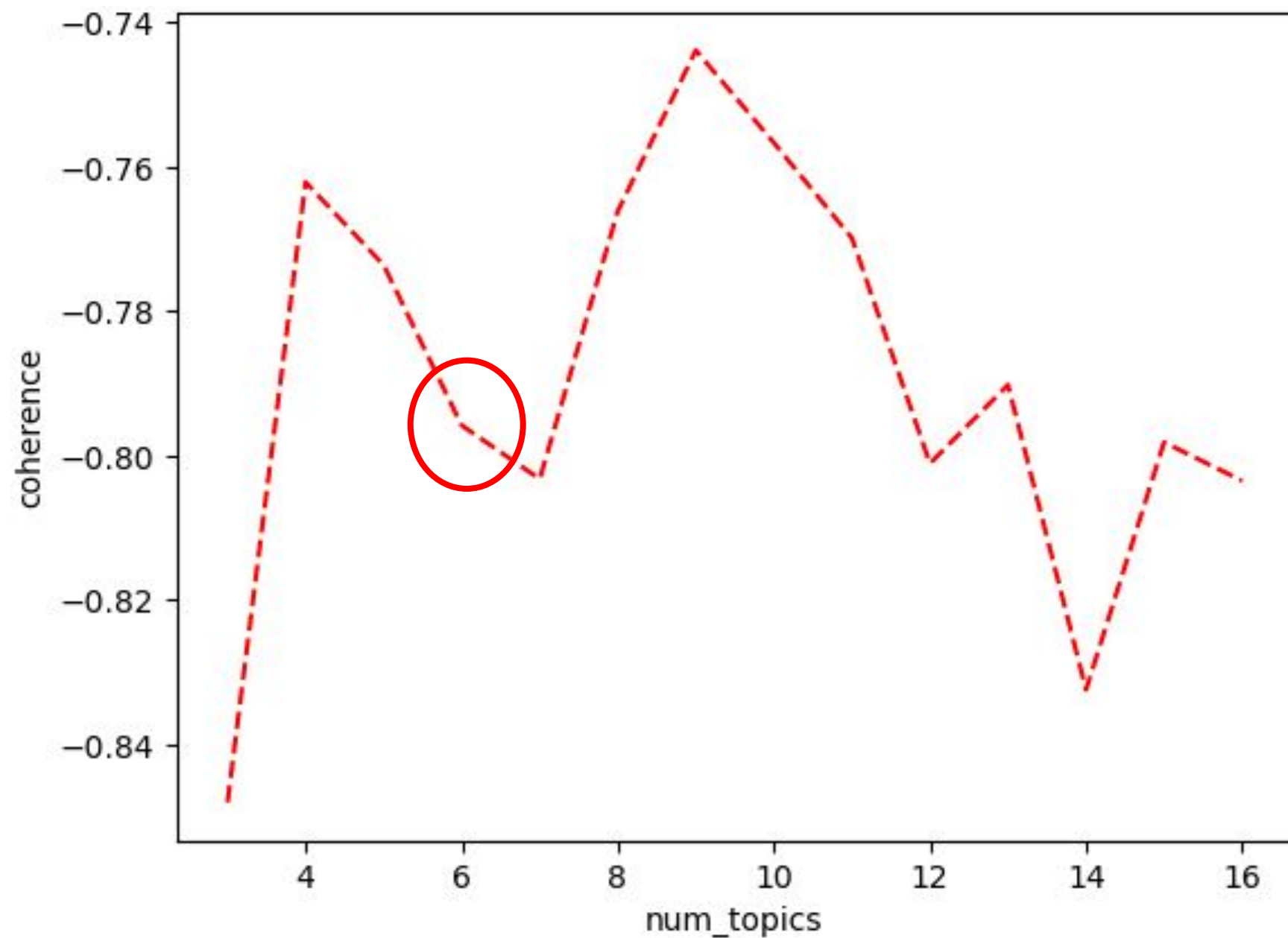
퇴사 3,864명 (11%)  
입사 4,330명 (12%)

출처 : 국민연금 2023.09 ⓘ

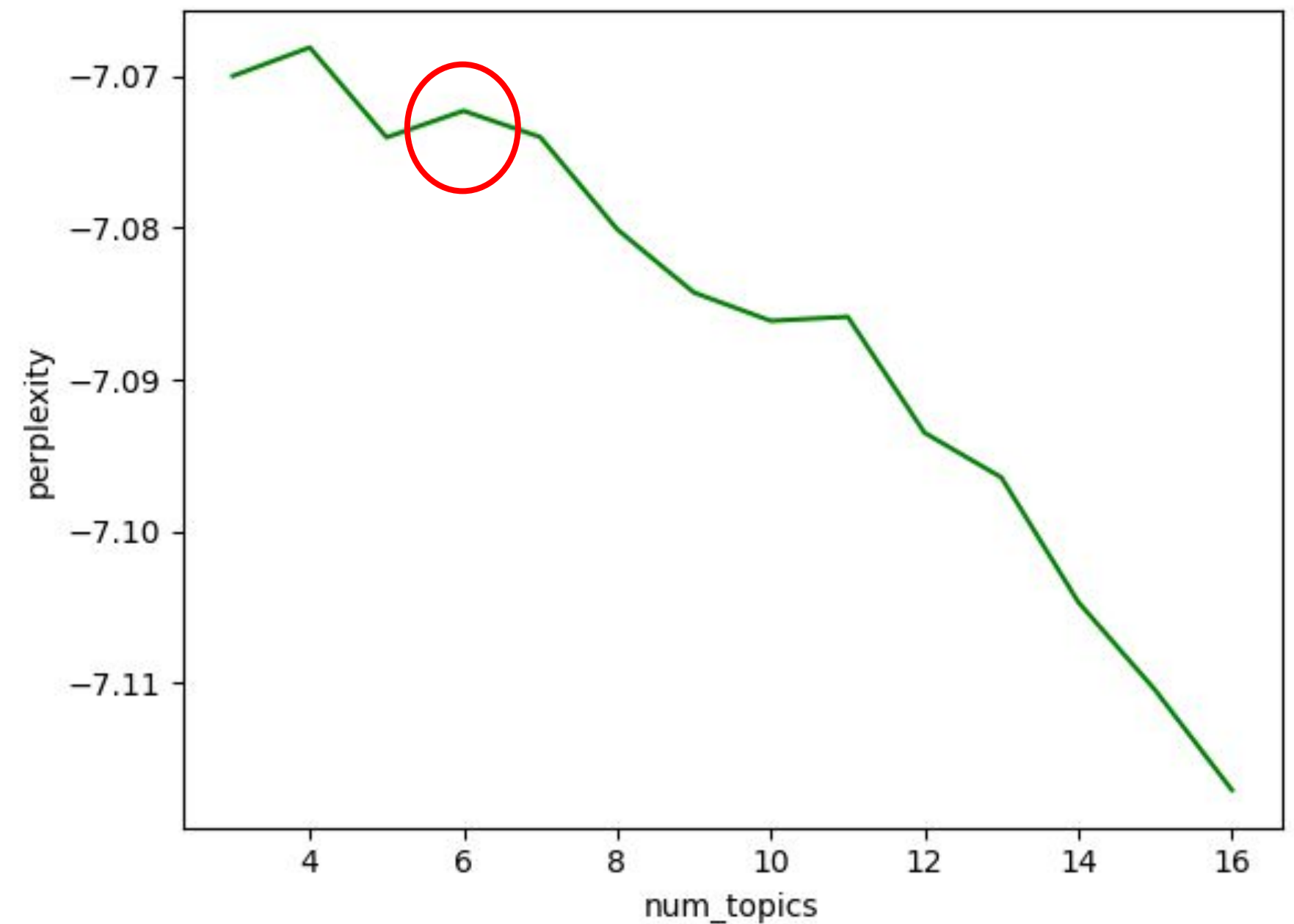
원티드 인사이트

# 토픽 모델링 결과

Advantage Coherence

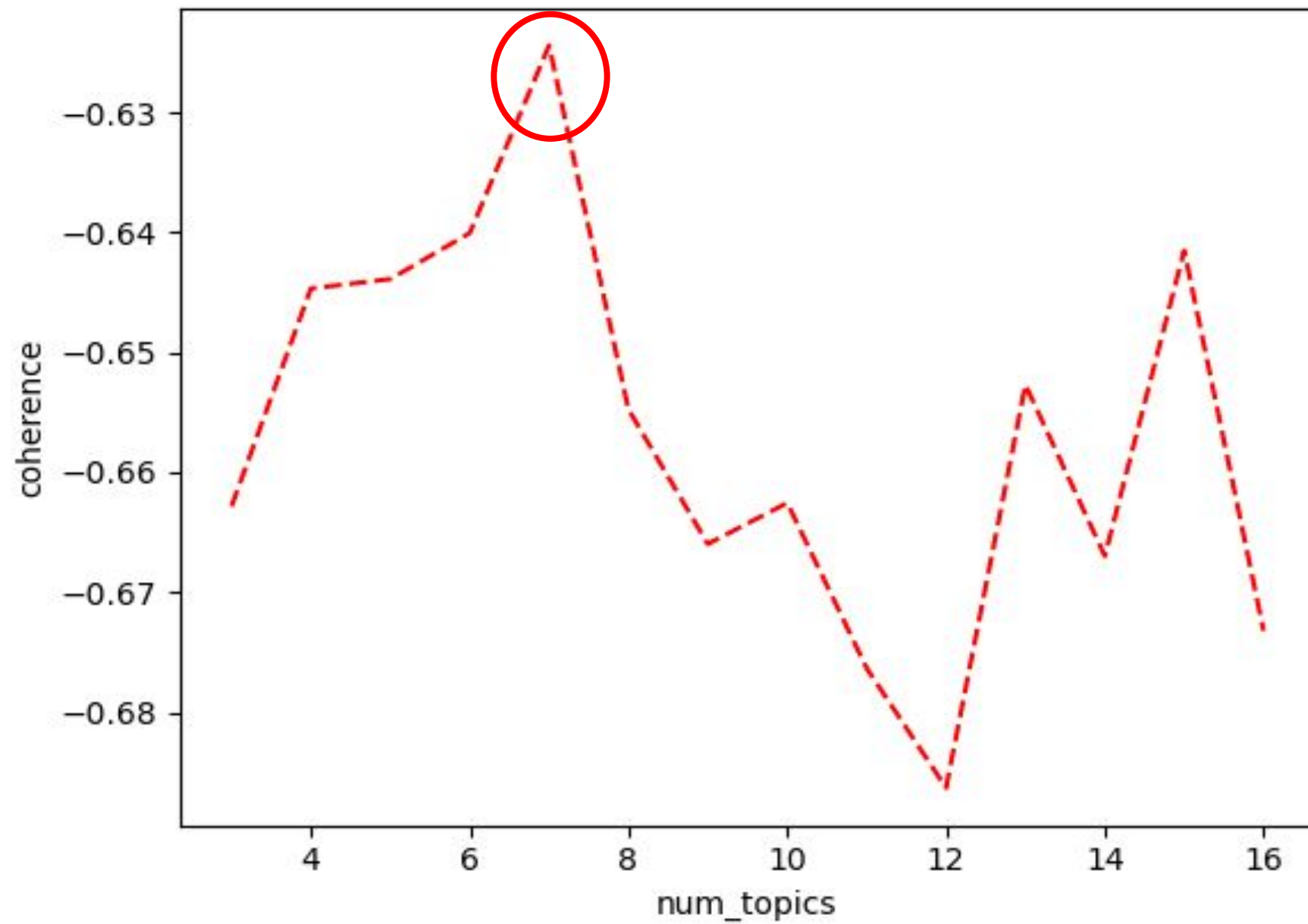


Advantage Perplexity

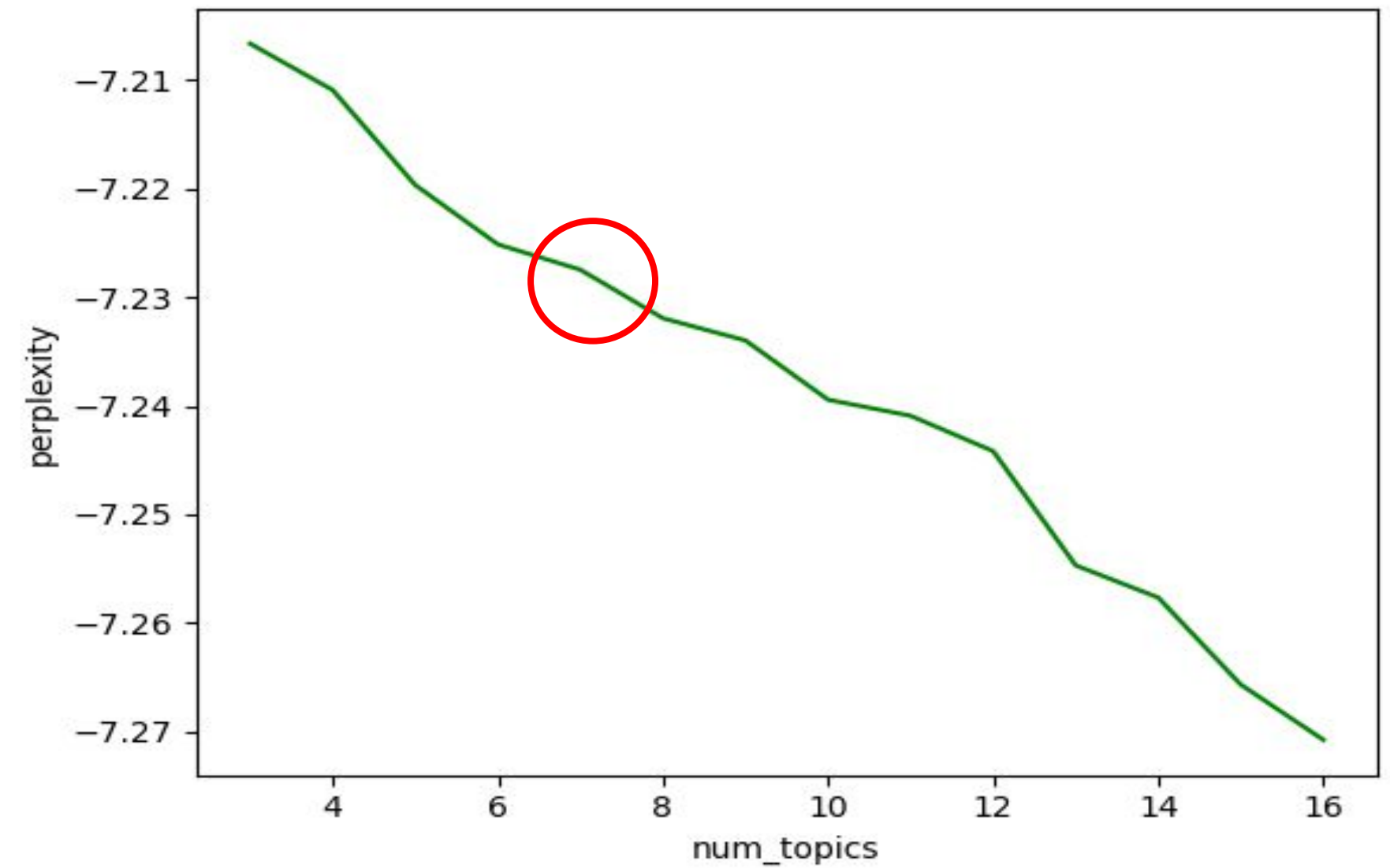


# 토픽 모델링 결과

Disadvantage Coherence



Disadvantage Perplexity



# 장점 토픽 모델링 결과

장점 토픽	토픽모델링 결과물
잘 설립된 복지 시스템	0.011*"안정" + 0.009*"공공기관" + 0.009*"서울" + 0.008*"부바" + 0.007*"유연근무제" + 0.007*"육아휴직" + 0.006*"정년" + 0.006*"강도" + 0.006*"높은" + 0.005*"공기업"
회사 시설 및 제도	0.013*"점심" + 0.009*"저녁" + 0.009*"버스" + 0.009*"기숙사" + 0.008*"통근" + 0.008*"높은" + 0.008*"식당" + 0.007*"맛있음" + 0.006*"수당" + 0.006*"아침"
여가 지원	0.013*"데이" + 0.013*"금요일" + 0.012*"높은" + 0.008*"포인트" + 0.008*"제도" + 0.008*"리프" + 0.007*"패밀리" + 0.007*"업계" + 0.006*"대비" + 0.006*"여름"
자율적 근무환경	0.013*"재택근무" + 0.008*"재택" + 0.008*"대기업" + 0.007*"동료" + 0.007*"포인트" + 0.007*"기회" + 0.006*"업계" + 0.006*"사내" + 0.005*"개인" + 0.005*"교육"
만족스러운 급여	0.008*"수당" + 0.006*"강도" + 0.006*"대기업" + 0.005*"따라" + 0.005*"명절" + 0.005*"때문" + 0.005*"나눔" + 0.005*"지급" + 0.005*"사업" + 0.005*"없고"
직원 친화적 분위기	0.012*"카페" + 0.012*"커피" + 0.011*"건물" + 0.010*"점심" + 0.009*"간식" + 0.009*"할인" + 0.009*"사내" + 0.009*"식대" + 0.008*"사옥" + 0.007*"사무실"

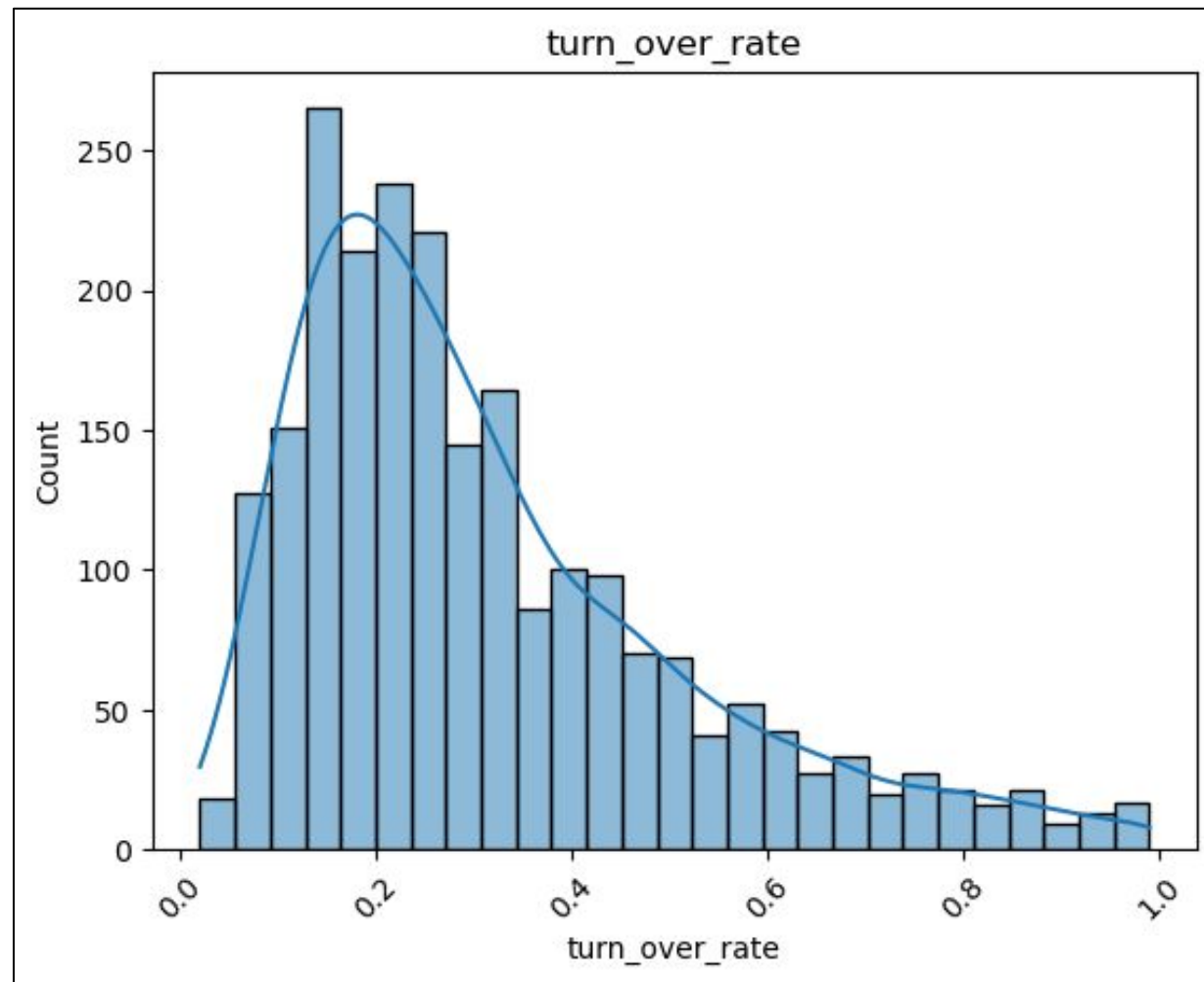
# 단점 토픽 모델링 결과

단점 토픽	토픽모델링 결과물
보수적인 문화 및 진급의 어려움	0.010* <b>"생산"</b> + 0.009* <b>"진급"</b> + 0.006* <b>"군대"</b> + 0.005* <b>"사원"</b> + 0.005* <b>"공장"</b> + 0.005* <b>"수직"</b> + 0.005* <b>"보수"</b> + 0.005* <b>"출근"</b> + 0.005* <b>"곤대"</b> + 0.005* <b>"승진"</b>
정규직 전환의 어려움	0.026* <b>"계약"</b> + 0.015* <b>"정규직"</b> + 0.008* <b>"현장"</b> + 0.007* <b>"강도"</b> + 0.006* <b>"사업"</b> + 0.005* <b>"전환"</b> + 0.005* <b>"수직"</b> + 0.005* <b>"교대"</b> + 0.005* <b>"군대"</b> + 0.005* <b>"차이"</b>
열악한 업무 프로세스와 조직구조	0.007* <b>"프로젝트"</b> + 0.006* <b>"무량"</b> + 0.006* <b>"평가"</b> + 0.005* <b>"조직"</b> + 0.004* <b>"소통"</b> + 0.004* <b>"리더"</b> + 0.004* <b>"팀바팀"</b> + 0.004* <b>"부분"</b> + 0.004* <b>"좋은"</b> + 0.004* <b>"성과"</b>
근무지 불만	0.012* <b>"위치"</b> + 0.006* <b>"행정"</b> + 0.006* <b>"연구원"</b> + 0.006* <b>"출퇴근"</b> + 0.005* <b>"수당"</b> + 0.005* <b>"식당"</b> + 0.005* <b>"인상"</b> + 0.005* <b>"교통"</b> + 0.005* <b>"연구"</b> + 0.005* <b>"주변"</b>
불확실한 성장 전망	0.009* <b>"임원"</b> + 0.007* <b>"사업"</b> + 0.007* <b>"영진"</b> + 0.006* <b>"영업"</b> + 0.005* <b>"경영"</b> + 0.005* <b>"미래"</b> + 0.005* <b>"매출"</b> + 0.005* <b>"조직"</b> + 0.004* <b>"성장"</b> + 0.004* <b>"정치"</b>
고객 응대 + 성과에 대한 압박	0.011* <b>"고객"</b> + 0.009* <b>"승진"</b> + 0.008* <b>"영업"</b> + 0.007* <b>"민원"</b> + 0.007* <b>"본사"</b> + 0.007* <b>"실적"</b> + 0.007* <b>"스트레스"</b> + 0.007* <b>"순환"</b> + 0.006* <b>"매장"</b> + 0.006* <b>"압박"</b>
수당 + 회사 시설 및 제도의 열악함	0.007* <b>"수당"</b> + 0.006* <b>"출근"</b> + 0.006* <b>"사용"</b> + 0.005* <b>"대표"</b> + 0.005* <b>"관리자"</b> + 0.004* <b>"교육"</b> + 0.004* <b>"관리"</b> + 0.004* <b>"진짜"</b> + 0.004* <b>"주말"</b> + 0.004* <b>"사원"</b>

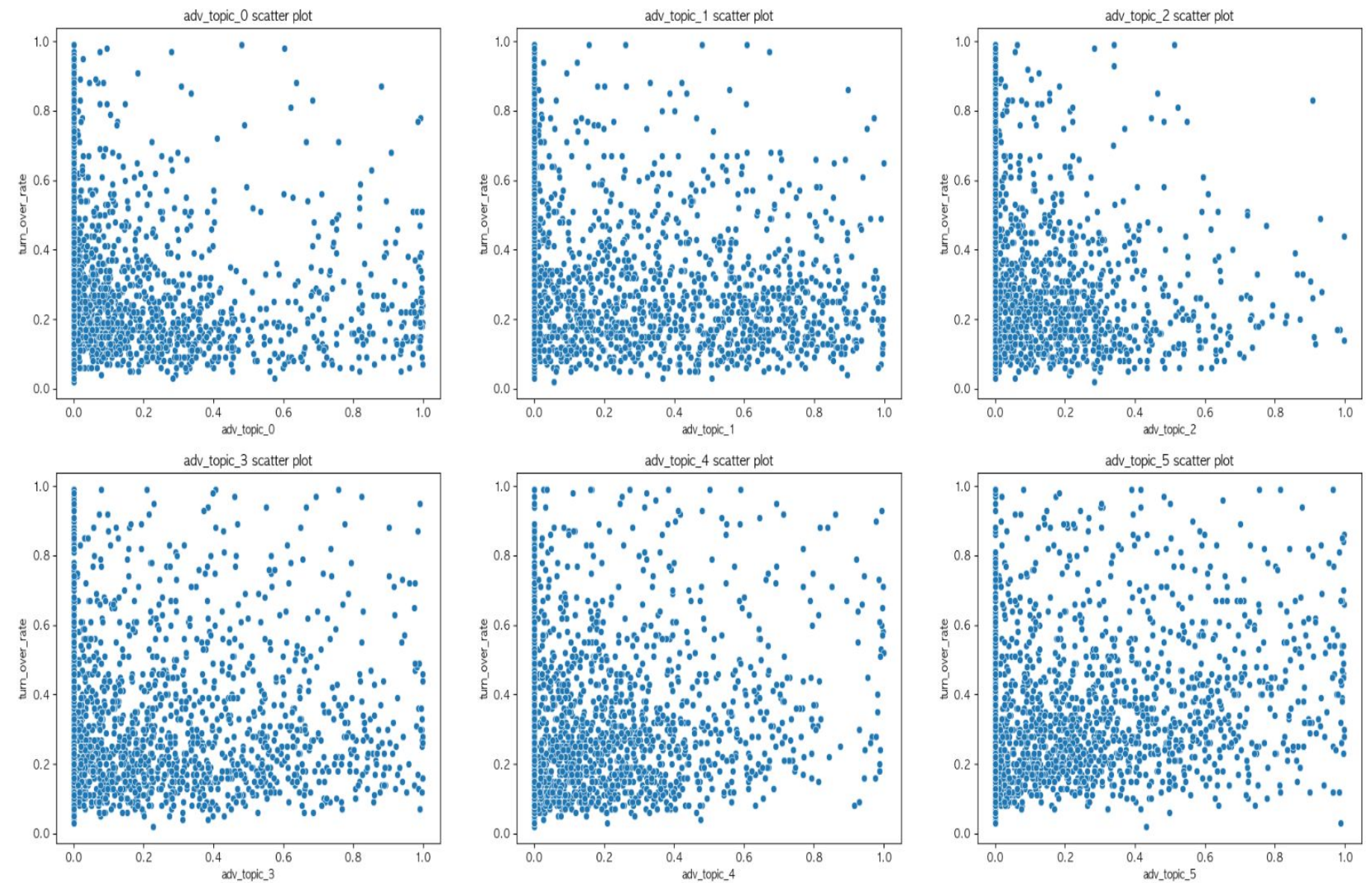


# 데이터 분포 형태

## 이직률 데이터 분포



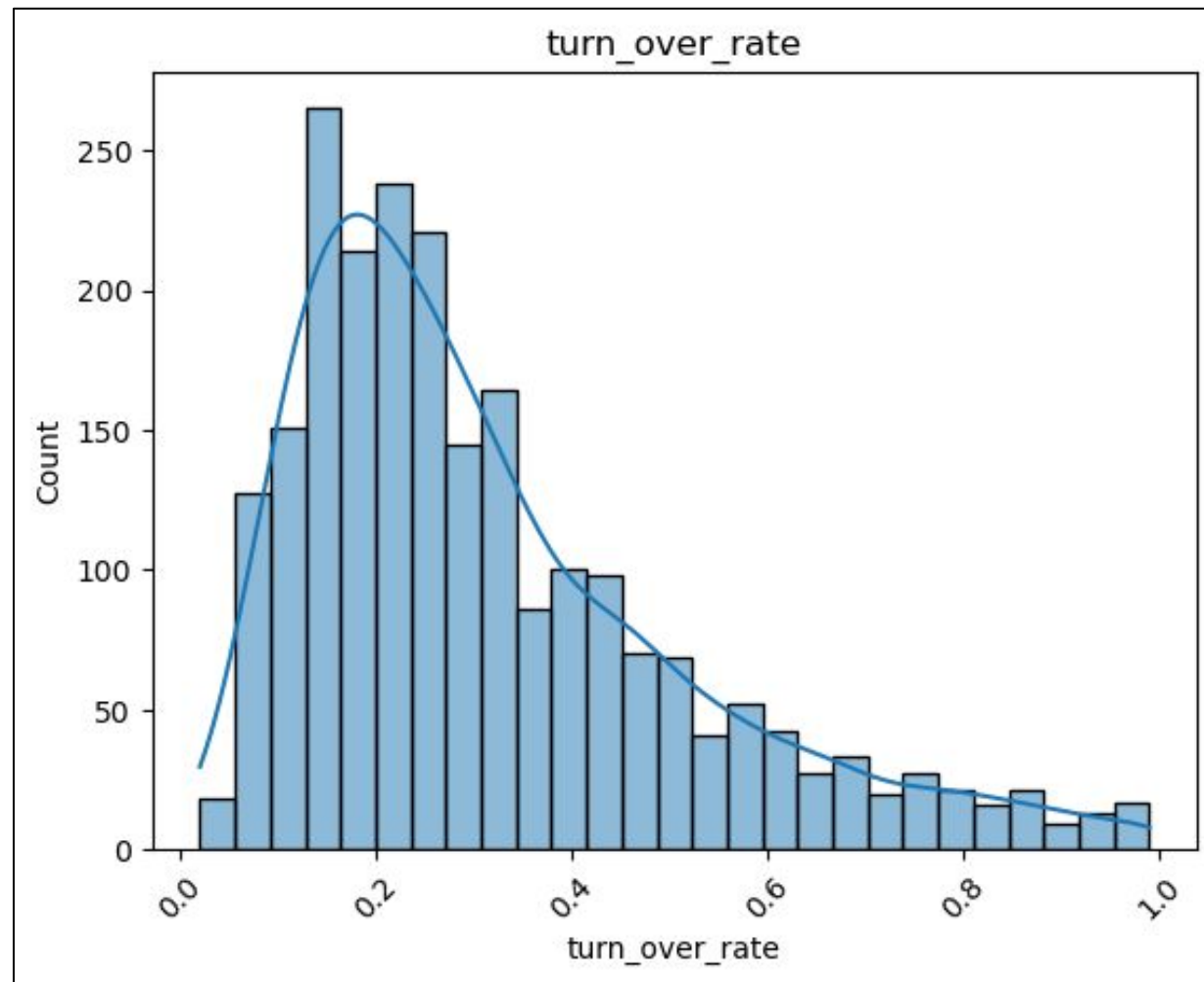
## 장점 리뷰 토픽모델링 분포



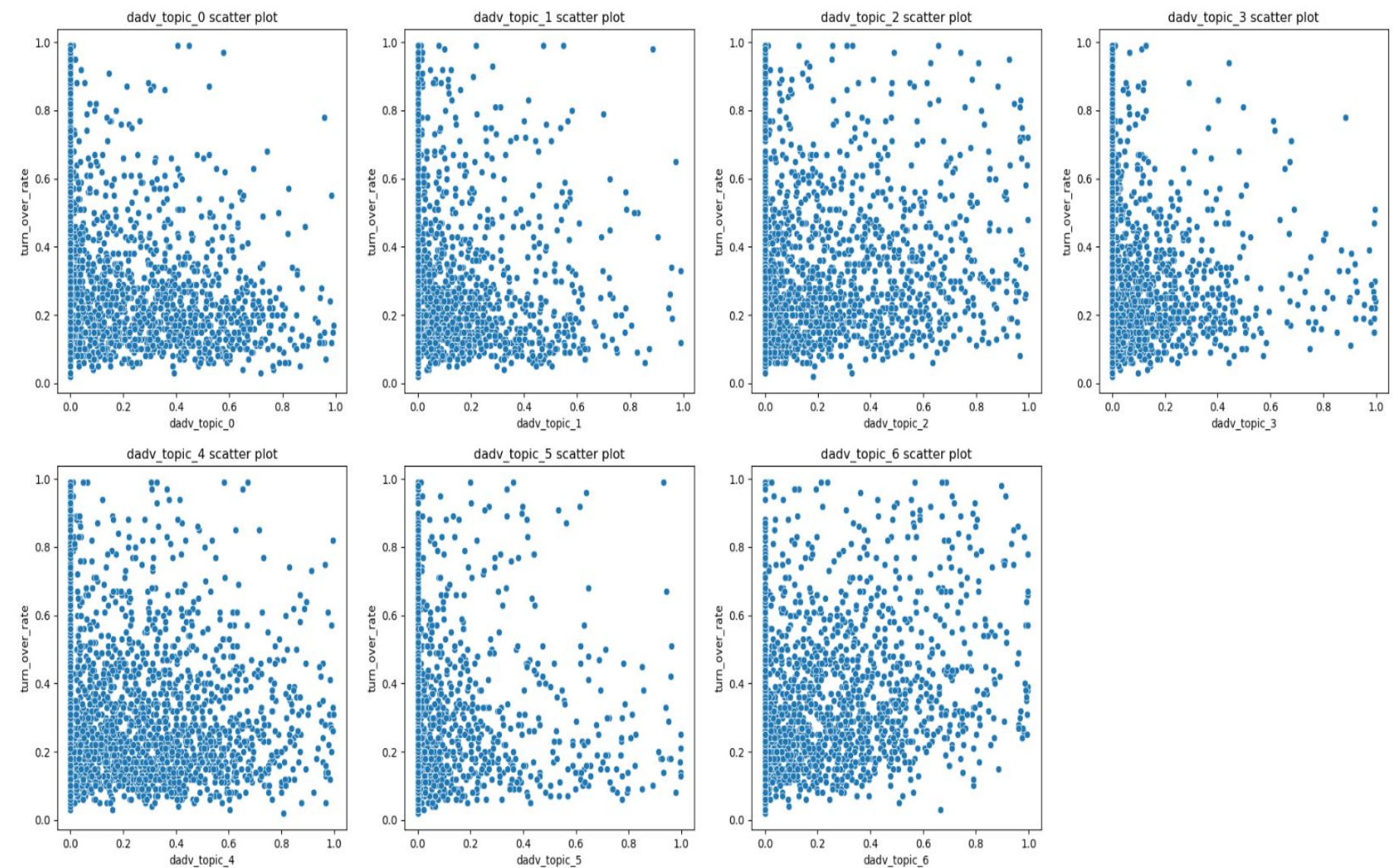


# 데이터 분포 형태

## 이직률 데이터 분포

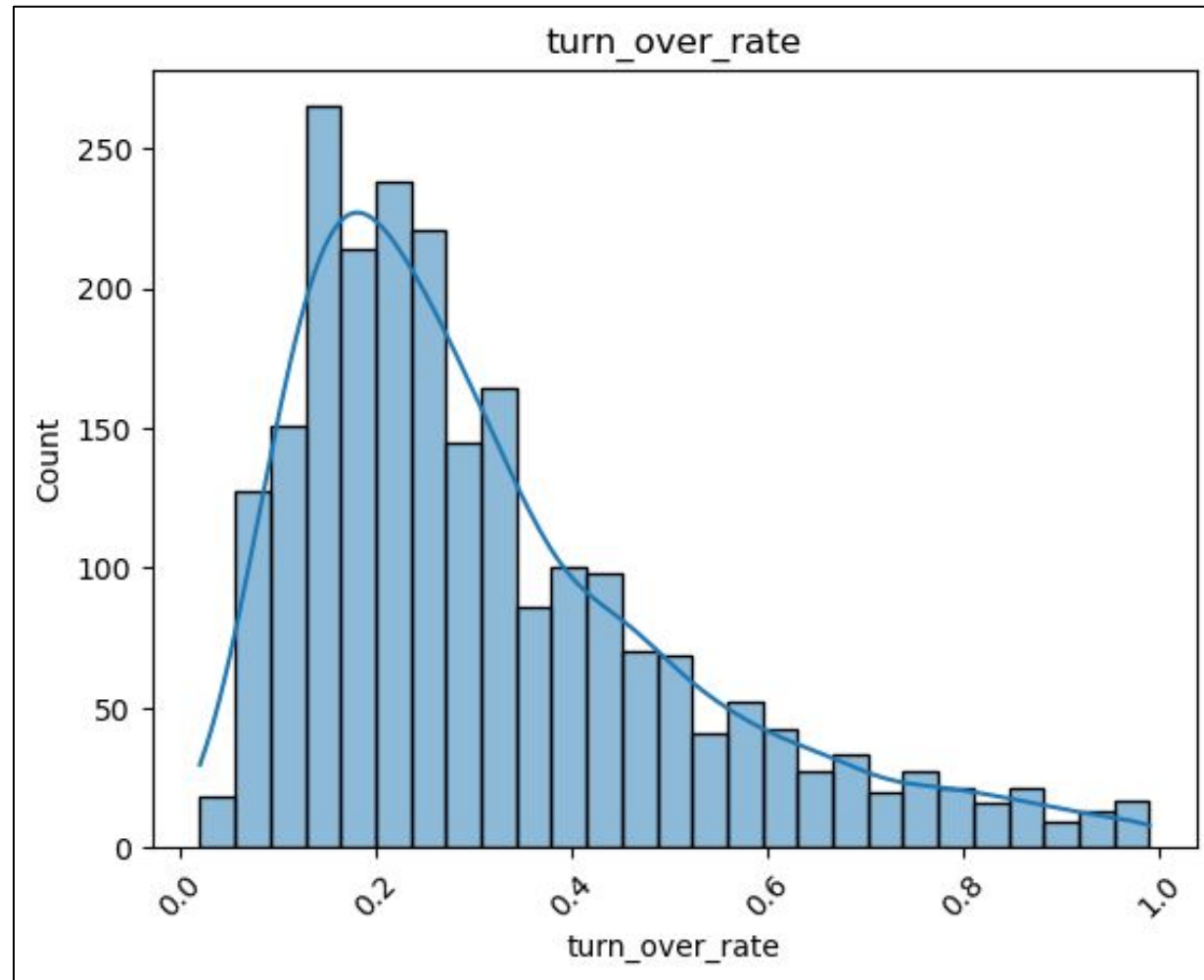


## 단점 리뷰 토픽모델링 분포

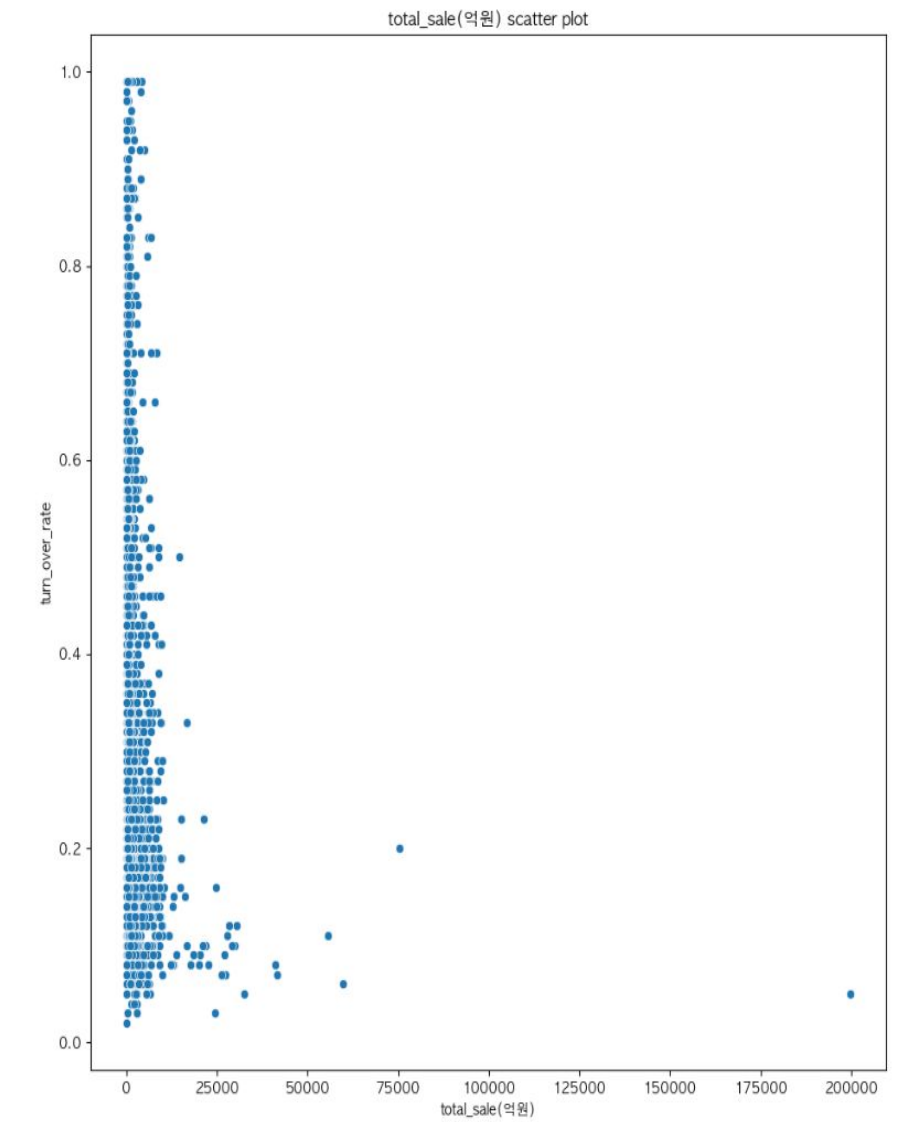
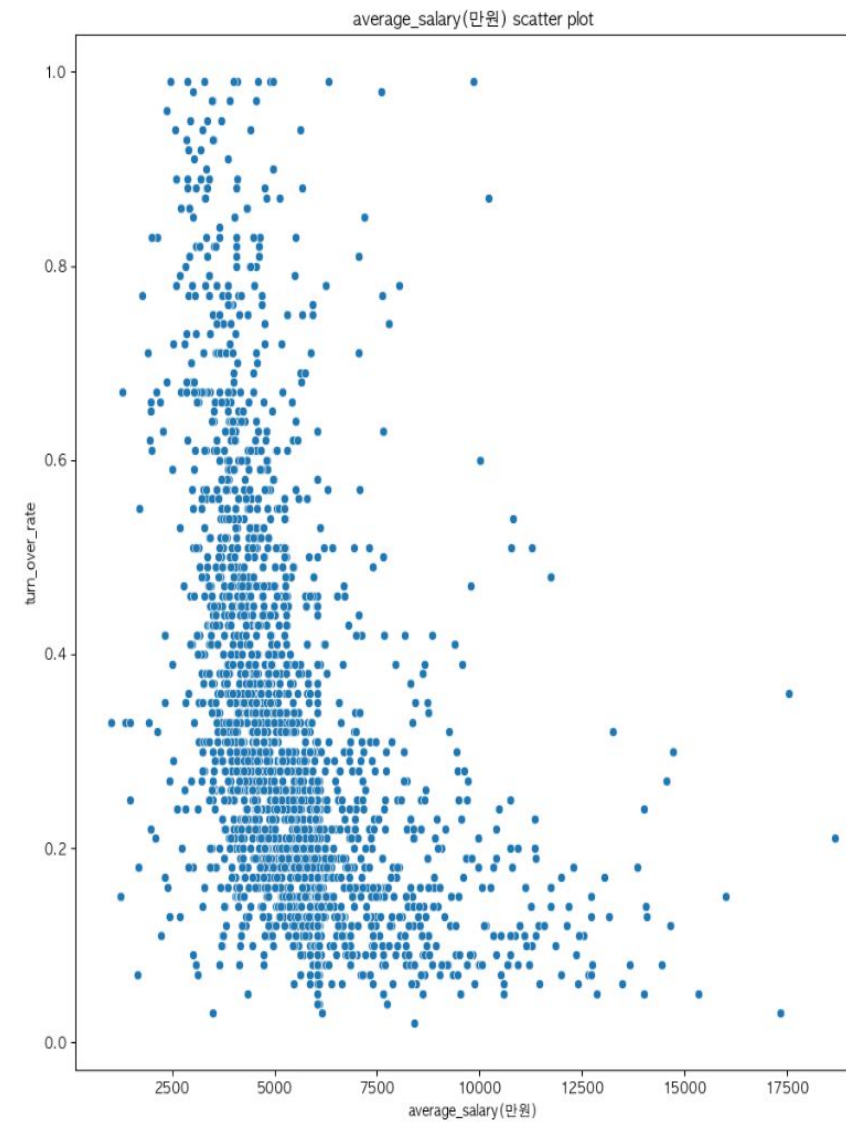


# 데이터 분포 형태

## 이직률 데이터 분포



## 재무제표 분포



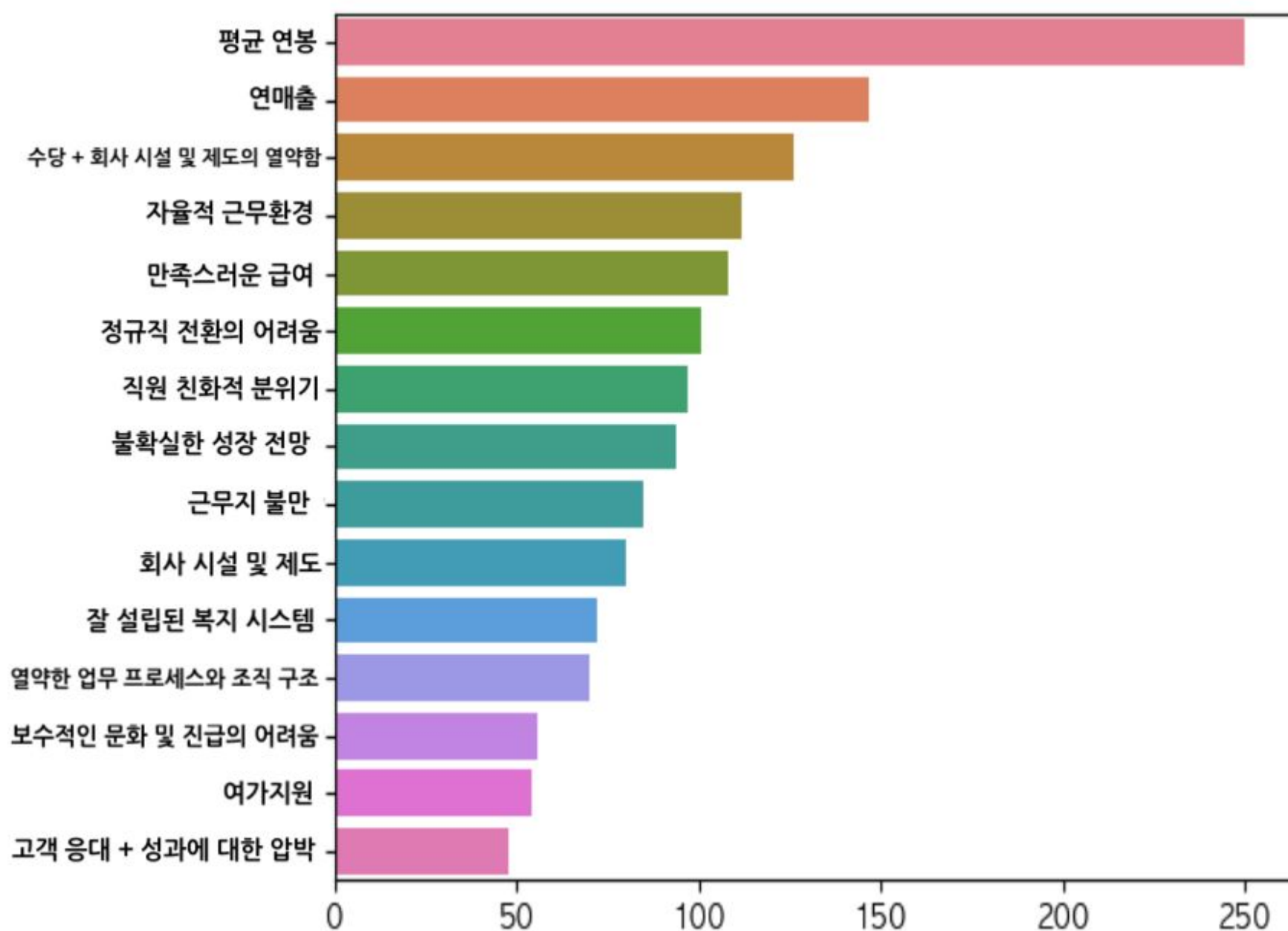
# ML 결과

MSE, RMSE, MAE를 모델 평가 지표로 사용하였습니다.

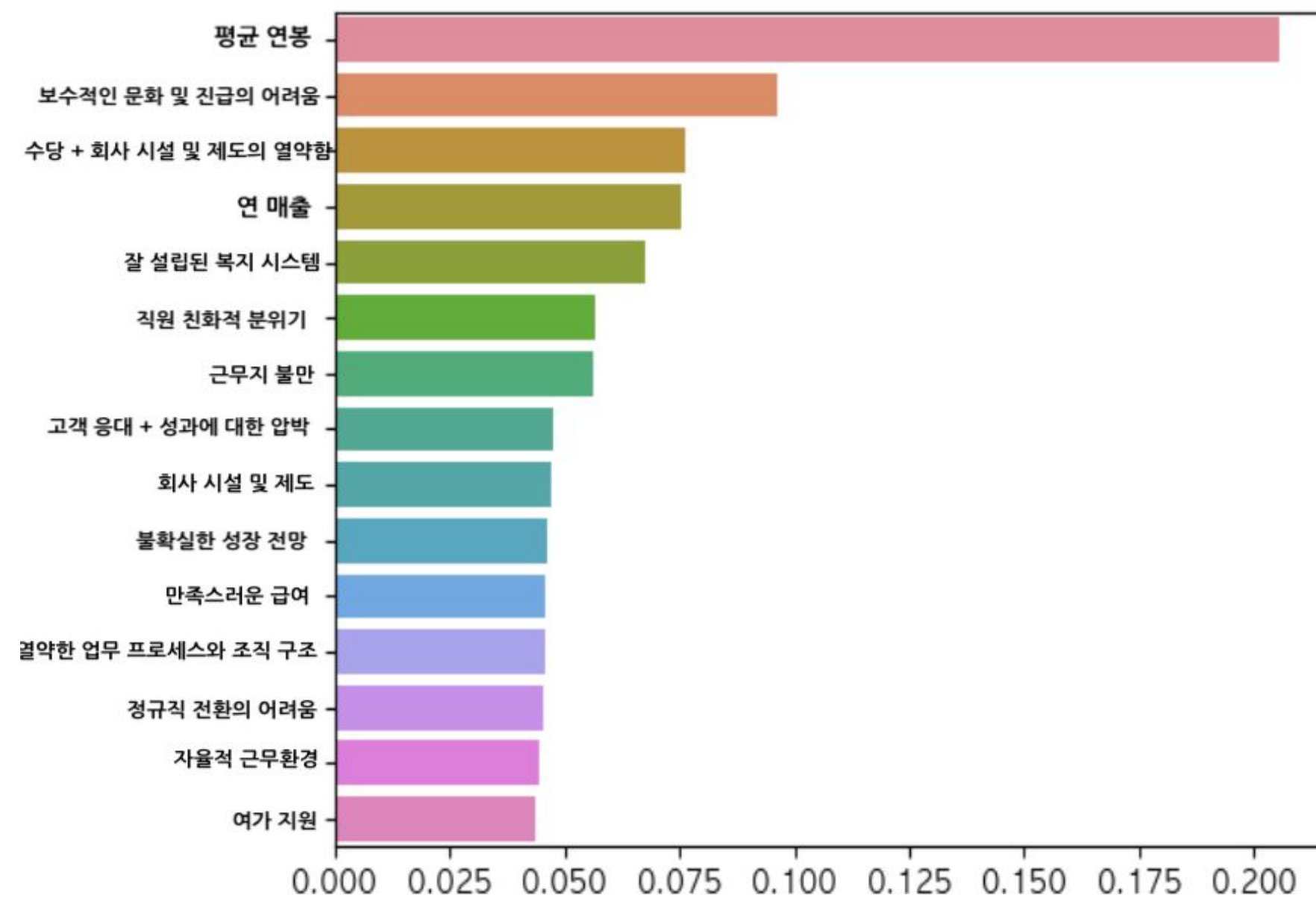
	MSE	RMSE	MAE
LR	0.0135	0.116	0.088
Ridge	0.0135	0.116	0.088
Lasso	0.0137	0.117	0.091
XGB	0.0112	0.106	0.081
LGBM	0.0108	0.104	0.079
GB	0.0121	0.110	0.085
RF	0.0121	0.110	0.086
Decision Tree	0.0121	0.110	0.084

# ML 피쳐 중요도

Light GBM



XGB





# ML 활용물 출력

## 취업 및 이직 준비 인원

```
output_user.txt
1  특징 별 상위 5개 기업: 1
2  특정 회사의 특징 백분위로 확인 : 2
3  나가기: x
4  다음의 토픽 중에서 선택해주세요: adv_topic_0~5, dadv_topic_0~6
5  회사 시설 및 제도 특징의 상위 5개의 기업은 다음과 같습니다.
6
7  유바이오로직스
8  경보제약
9  한국프라임제약
10 유진테크
11 진양제약
```

```
13 특징 별 상위 5개 기업: 1
14 특정 회사의 특징 백분위로 확인 : 2
15 나가기: x
16 프로그램에서 찾은 삼성전자 의 특징은 다음과 같습니다.
17
18 잘 설립된 복지 시스템 : 0.1669293493032455
19 해당 특징은 전체 데이터 중에서 상위 25.729% 입니다.
20
21 회사 시설 및 제도 : 0.4167281985282898
22 해당 특징은 전체 데이터 중에서 상위 23.891% 입니다.
23
24 자율적 근무환경 : 0.3976799845695495
25 해당 특징은 전체 데이터 중에서 상위 20.096% 입니다.
26
27 보수적인 문화 및 진급의 어려움 : 0.2328608185052871
28 해당 특징은 전체 데이터 중에서 상위 28.566% 입니다.
29
30 정규직 전환의 어려움 : 0.2974345982074737
31 해당 특징은 전체 데이터 중에서 상위 10.787% 입니다.
32
33 열악한 업무 프로세스와 조직구조 : 0.3217673599720001
34 해당 특징은 전체 데이터 중에서 상위 22.173% 입니다.
35
36 불확실한 성장 전망 : 0.1102671176195144
37 해당 특징은 전체 데이터 중에서 상위 54.495% 입니다.
```

## 기업 경영자

회사의 이름을 입력하세요 ::아모레퍼시픽

다음과 같은 내용을 담은 csv 파일을 업로드 하고 그 파일의 이름을 입력하세요 ::

잘 설립된 복지 시스템 : 0.6822144970627604

해당 특징은 전체 데이터 중에서 상위 6.829% 입니다.

고객 응대 + 성과에 대한 압박 : 0.6484060009802686

해당 특징은 전체 데이터 중에서 상위 2.676% 입니다.

평균 연봉 (만원) : 8.85366542803745

해당 특징은 전체 데이터 중에서 상위 99.920% 입니다.

연 매출 (억원) : 8.160803920954665

해당 특징은 전체 데이터 중에서 상위 98.243% 입니다.

LGBM 모델로 예측된 이직률 : [0.22322945]

XGB 모델로 예측된 이직률 : [0.2622404]

01 사업 개요

02 기술 개발 및 결과

# 03 데이터 분석

04 기술 활용방안

05 후행 연구 제안



# 장점 토픽 분석

## 선행 연구

유연한 직업 환경  
성과와 인정의 공평함  
잘 설립된 복지 시스템  
만족스러운 급여  
성장 기회  
자유롭게 일하는 환경  
수평 조직 문화  
직원 친화적 분위기

## 본 연구

잘 설립된 복지 시스템  
회사 시설 및 제도  
여가 지원  
자율적 근무환경  
만족스러운 급여  
직원 친화적 분위기

# 장점 토픽 분석

## Z세대 취업 위해 가장 궁금한 기업정보는 '복지'

[자료제공: 잡코리아 (Z세대 대학생 및 취업생 1,923명 조사)]



전체응답률 404.7%

**JOBKOREA**

# 장점 토픽 분석

## 선행 연구

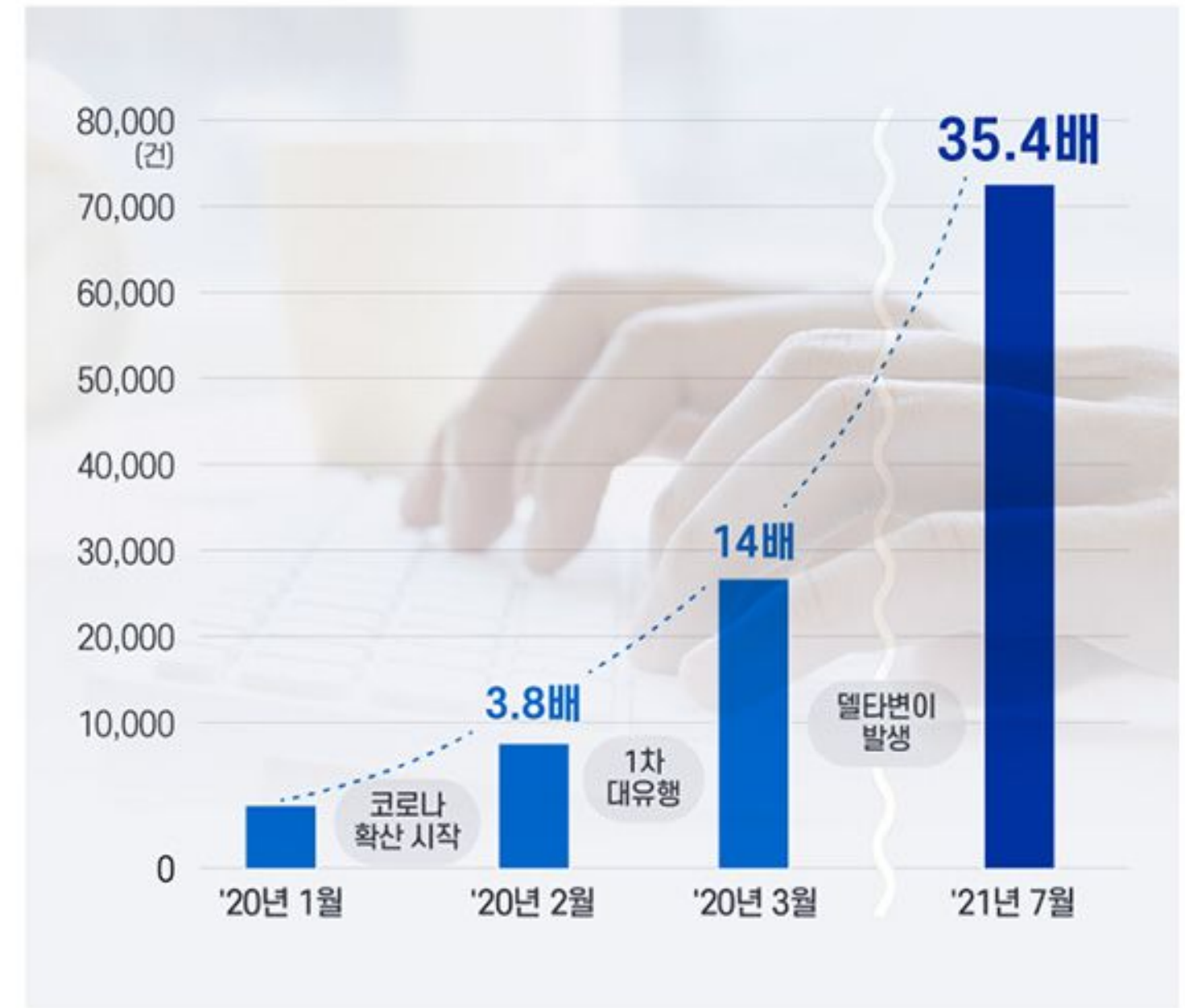
연차 (10.1%), 가능 (8.2%), 사용 (6.0%),  
야근 (5.4%), 휴가 (4.2%),  
눈치(3.0%), 칼퇴 (2.3%)

## 본 연구

재택근무 (1.3%), 재택 (0.8%),  
대기업(0.8%), 동료 (0.7%),  
포인트 (0.7%), 기회 (0.7%), 업계 (0.6%),  
사내 (0.6%), 개인 (0.5%), 교육 (0.5%)

코로나19 발생 이후 재택근무 비율 추이

(단위 : 건)



\* 2020~2021년 재택근무 데이터 기준

자료 : 통합 인력관리 솔루션 시프티, 자료사진 이미지투데이

# 단점 토픽 분석

## 선행 연구

불확실한 성장 전망

열악한 업무 프로세스와 조직 구조

사내 정치와 불공정한 보상

낮은 연봉과 미흡한 복리후생

꼰대 문화

불합리한 업무 배분

보수적인 조직문화

## 본 연구

보수적인 문화 및 임금의 어려움

정규직 전환의 어려움

열악한 업무 프로세스와 조직구조

근무지 불만

불확실한 성장 전망

고객 응대 + 성과에 대한 압박

수당 + 회사 시설 및 제도의 열악함

# 단점 토픽 분석

## 정규직·비정규직 임금 격차

연도별 8월 기준 월평균 임금, 단위: 만 원



연합뉴스

자료: 통계청 국가통계포털(KOSIS)

원형민 기자 20231024

# 단점 토픽 분석





# 기업 특징에 대한 토픽 모델링 유효성 검증

## AMORE PACIFIC

compan y_name	잘 설립된 복지 시스템 (장점)	구내식당 + 제공 서비스 만족 (장점)	여가 지원 (장점)	자율 근무 환경 (장점)	만족스러 운 급여 (장점)	직원 친화적 분위기 (장점)	보수적인 문화 및 진급의 어려움 (단점)	정규직 전환의 어려움 (단점)	열악한 업무 프로세스와 조직 구조 (단점)	근무지 불만 (단점)	불확실한 성장 전망 (단점)	고객응대 + 성과에 대한 압박 (단점)	수당 + 회사 시설 및 제도의 열악함 (단점)
아모레 퍼시픽	0.537661	0.000000	0.0000 00	0.21750 4	0.043675	0.000 000	0.000000	0.2024 93	0.000000	0.0000 00	0.000000	0.573879	0.000000

	아모레 퍼시픽 실제 사례 조사
잘 설립된 복지 시스템	직원의 수는 4년간 20% 줄었으나, 평균 근속 연수는 9.4년에서 11.9년으로 증가 다양한 휴가 옵션 제공: 연차 15일, 리프레쉬 7일, 해피베케이션 4일 연 8회 심리 상담 전문 기관에서 상담 가능
자율근무 환경	출근 후 업무 시작 전 자유시간 보장 식사시간의 유연한 조정 탄력 근무제 시행
정규직 전환의 어려움	정규직 전환의 경우 조직 전체로 보면 50%이지만, 부서와 인사팀의 결정이 주요하다.
고객응대 + 성과에 대한 압박	협력 업체 및 해외 바이어에게 오는 연락 처리 및 수시로 진행되는 회의 참여

# ML 결과 분석

## Light GBM

평균연봉

연매출

수당 + 회사 시설 및 제도의 열악함(단점)

자율적 근무환경(장점)

만족스러운 급여(장점)

정규직 전환의 어려움(단점)

## XGB

평균연봉

보수적인 문화 및 진급의 어려움(단점)

수당 + 회사 시설 및 제도의 열악함(단점)

연매출

잘 설립된 복지 시스템(장점)

직원 친화적 분위기(장점)

01 사업 개요

02 기술 개발 및 결과

03 데이터 분석

## 04 기술 활용방안

05 후행 연구 제안

# 기술 활용 방안

## 일반 취업 및 이직 준비 인원

- 선택한 회사의 토픽 수치를 백분위로 확인 가능
- 토픽을 선택해 해당 토픽의 만족도가 높은 회사를 검색 가능

## 기업 경영자

- 인력 수급에 대한 계획과 운영 방식의 의사결정 과정에 활용 하는 서비스 제공
- 경쟁 회사와의 비교를 통해 회사 성장에 도움이 되도록 방향을 수정
- 인사 및 직원 교육에 필요한 비용 절감 가능

# 활용 예시



- 01 사업 개요
- 02 기술 개발 및 결과
- 03 데이터 분석
- 04 기술 활용방안

## 05 후행 연구 제안



# 후행 연구 제안

1



이직자 전용 추천 시스템

2

기업 별 트렌드 반영 여부에 따른  
성장 파악

3

토픽 모델링 내 주요 원인들  
해결 연구



# Q&A