

개인프로젝트

• • •

# 자연어 처리를 통한 기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

김형준 교육생

2021/08/17

김형준

1996년생

한양대학교 ERICA (졸업예정)

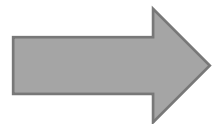
언론 정보대학  
정보사회학과 문학사

소프트웨어 융합대학  
신산업소프트웨어전공 공학사

### 기업의 평판?

**정량적 성과** + **정성적 가치**

오랜 기간 수치화된 정량적 성과를 분석한 정형 분석에서 벗어나  
**정성적 가치에 집중한 비정형데이터를 분석이 필요한 시점**



자연어 처리를 통한 기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

# 02 아이디어 소개

---

자연어 처리를 통한 기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

## 기업 평판이란 ?

1. 기업에서 평판이 중요한 이유는 평판이 전반적으로 기업의 전략적 포지셔닝과 본질적으로 연관 되어있다.
2. 좋은 평판은 대중들로부터 관심과 신뢰를 얻어, 해당기업을 지원해주고 경쟁사로부터 차별화 시켜주는 주요한 특징이 있기 때문에 평판은 **기업의 경쟁에서 우위를 창출**하게 해준다.

출처: 명성을 얻어야 부가 따른다. 저서 뉴욕 대학교 스톤 비즈니스 스쿨의 명예교수 찰스 J. 폼브런

3. 기업 평판은 정량적 성과와 정성적 가치를 통해 형성되는 대내외적 이미지이다.
4. 정성적 가치라는 말의 의미는 그 가치를 표현하고, 언어로만 표시하고 평가한다는 의미로서 그 가치의 질적인 면을 더 중요시 한다는 뜻이다.

출처 : 글 김종호 [jhkim@lgeri.co.kr](mailto:jhkim@lgeri.co.kr)

## 자연어 처리를 통한 기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

### 외부 평판 데이터

크롤링을 통한 네이버 뉴스 데이터

설 명 : 기업명을 검색한 뉴스데이터

수집 기한 : 2020년 1월 부터 2021년 8월 까지

언론진흥재단에 인정한 인터넷 신문사 기준으로  
기업당 300건 이상의 기사 수집

### 내부 평판 데이터

엘라스틱 서치를 통한 잡플래닛 데이터

설 명 : 재직자/ 퇴사자의 리뷰데이터

데이터 수 : 3만 8천건

기업의 장점, 기업의 단점, 성장 여부

외부 평판분석과 내부 평판 분석에 필요한 데이터를 수집을 통해 융합 DB 구성

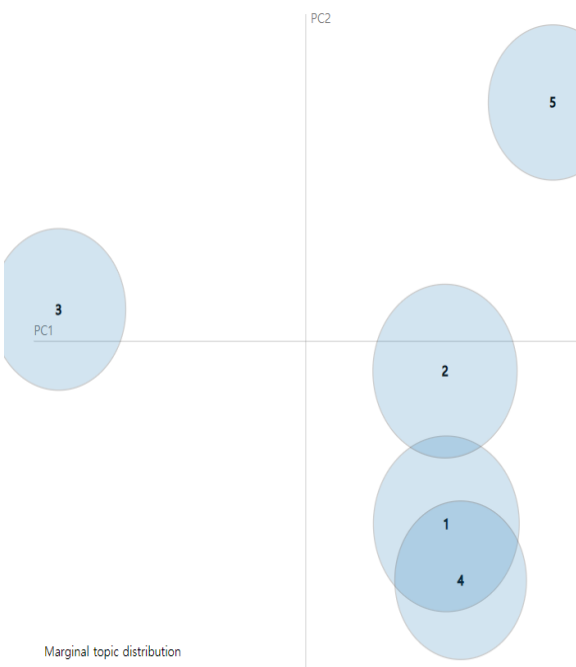
# 04 분석결과

자연어 처리를 통한기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

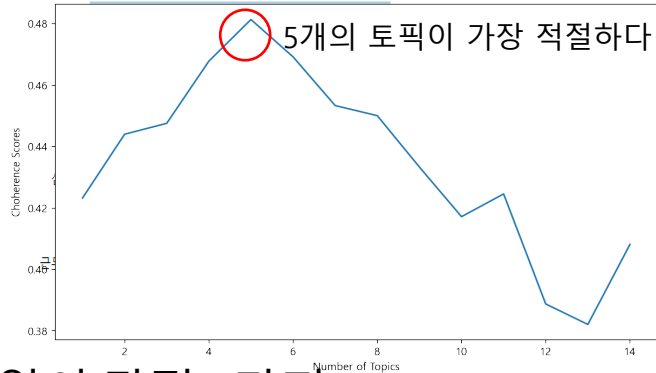
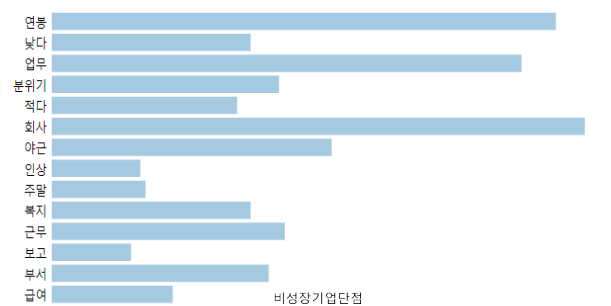
## 내부 평판 분석 토픽모델링 분류

성장기업과 비성장 기업으로 분류 후,  
기업의 장점, 기업의 단점에 대해 형태소 분석과 토픽모델링

Intertopic Distance Map (via multidimensional scaling)

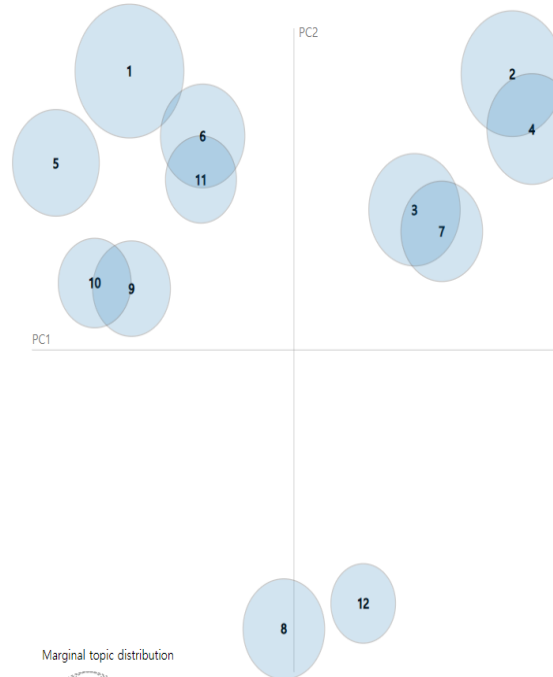


Top-30 Most Salient Terms<sup>1</sup>

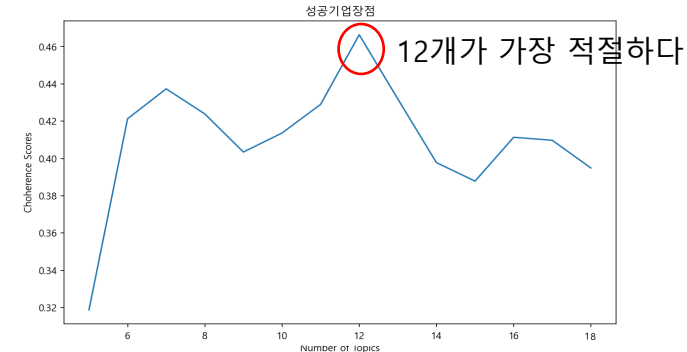
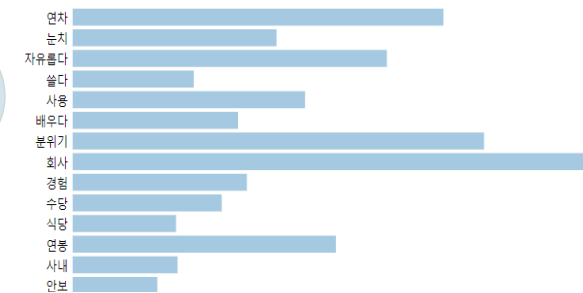


비성장 기업의 단점 5가지

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Salient Terms<sup>1</sup>



성장 기업의 장점 12가지

## 04 분석결과

## 자연어 처리를 통한기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

# 내부 평판 분석 위드 클라우드



## 비성장 기업의 단점 5가지

## 성장 기업의 장점 12가지

# 04 분석결과

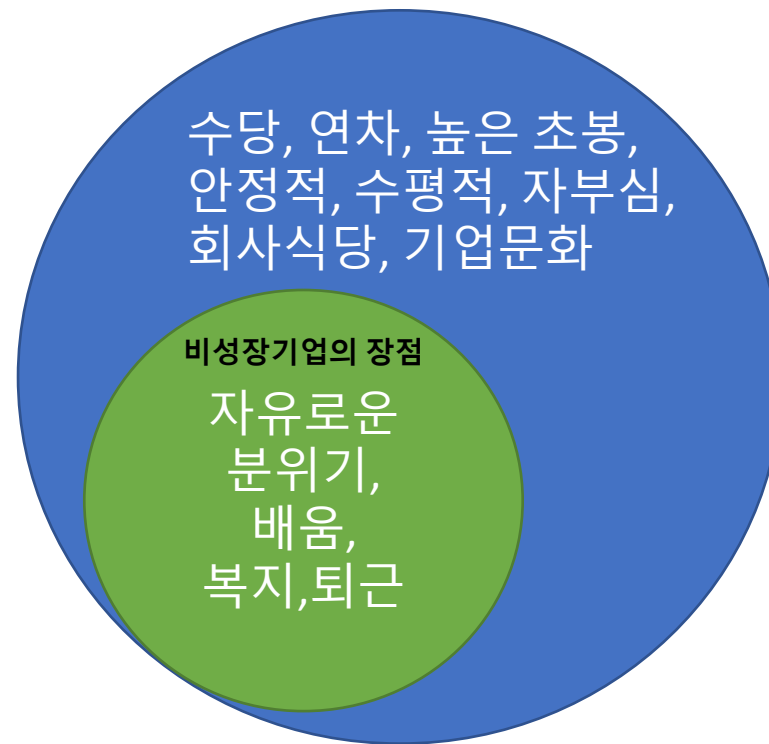
자연어 처리를 통한기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

## 내부 평판 분석 결과

비성장 기업의 단점



성장 기업의 장점



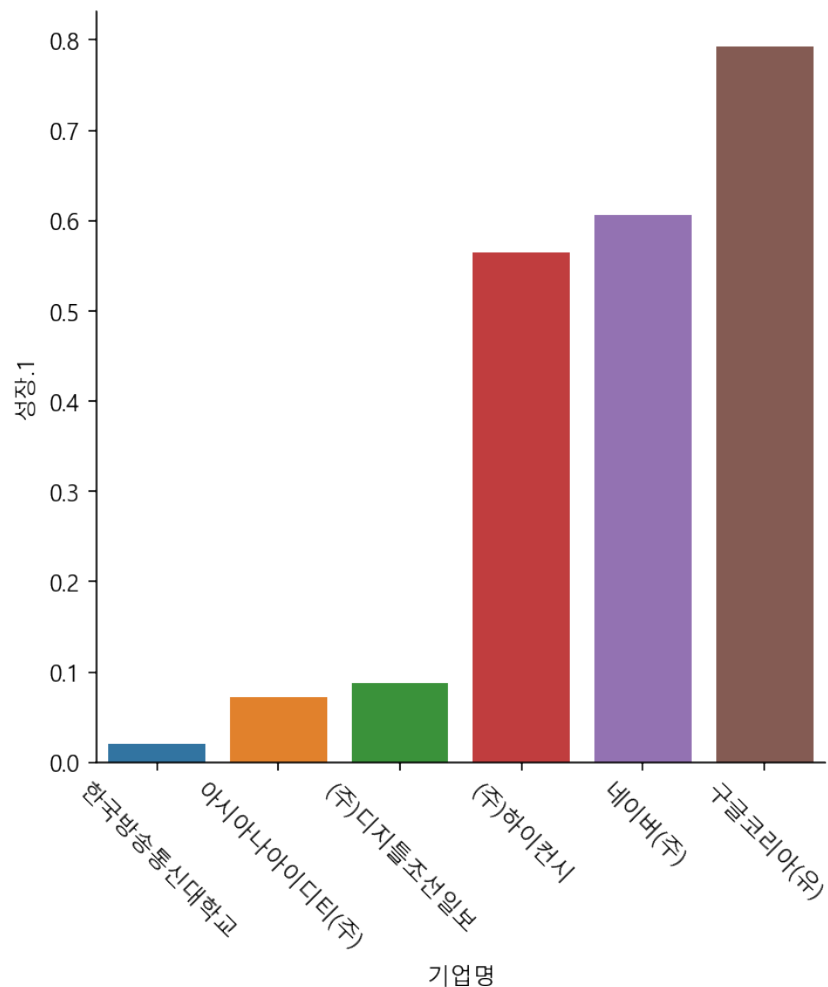
성장기업의 단점과 비성장기업의 단점은 공유하는 형태 지만, 성장기업의 장점은 비성장기업보다 많은 장점이 있었다.



# 04 분석결과

자연어 처리를 통한기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

## 외부 평판 분석 (96개의 기업대상)



잡플래닛에서 성장/비성장 대비  
성장 가능성이 가장 높은 비율을 가진 TOP3

1. 구글코리아
2. 네이버
3. 하이컨시

잡플래닛에서 성장/비성장 대비  
성장 가능성이 가장 낮은 비율을 가진 TOP3

1. 한국방송통신대학교
2. 아이아나아이디티
3. 디지털조선일보

# 04 분석결과

자연어 처리를 통한기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

## 외부 평판 분석

서울대 언어학과에서 만든 한국어 감성어 사전을 이용해서 기업의 뉴스데이터를 분석해보자

	key	value
0	가*	1
1	가*있	1
2	가*있었	1
3	가*	1
4	가*ㄴ다*	1
...	...	...
15361	힘들ㄹ것	-1
15362	힘들ㄹ때	-1
15363	힘차	-1
15364	힘차ㄴ	-1
15365	힘차ㄴ붓	-1

15366 rows × 2 columns

이번 프로젝트에 맞게 감성어 사전을 전처리 이후  
분류된 단어 15366개를 긍정은 1, 중립은 0, 부정은 -1로 매핑하였다.

기사의 감성 분석은  $\frac{(\text{긍정} - \text{부정})}{\text{전체}}$  로 점수를 내기로 하였다.

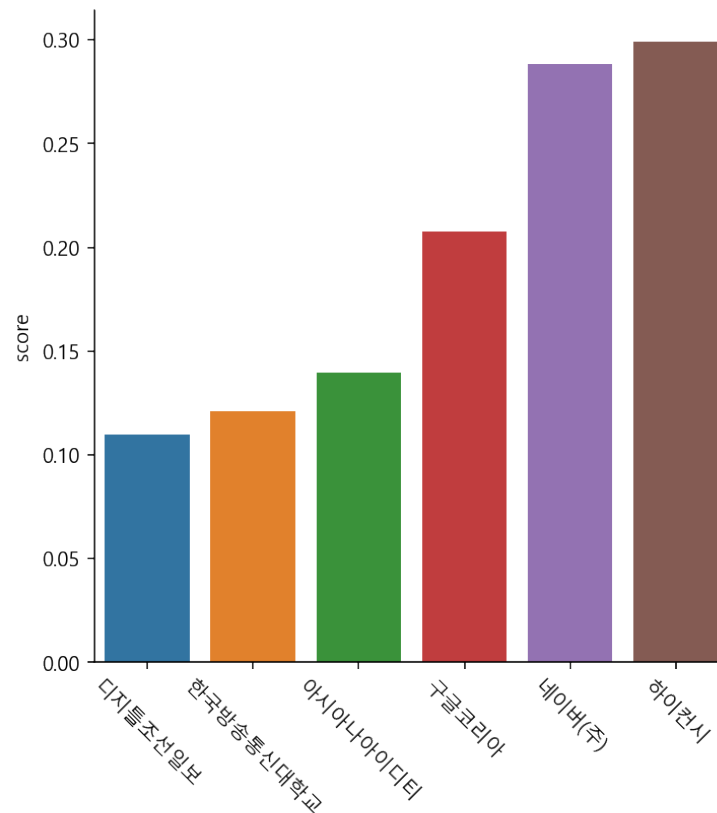
# 04 분석결과

자연어 처리를 통한기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

## 외부 평판 분석

서울대 언어학과에서 만든 한국어 감성어 사전을 이용해서 기업의 뉴스데이터를 분석해보자

	keyword	pos	neg	nue	score
0	디지털조선일보	5892	5202	596	0.110009
1	한국방송통신대학교	2521	2168	244	0.121022
2	아시아나아이디티	5157	4265	493	0.139687
3	구글코리아	17176	12284	1554	0.207842
4	네이버(주)	12911	7681	991	0.288236
5	하이컨시	3420	2003	293	0.299160



### 평판 점수별로 정렬한 결과

성장TOP3와 비성장 TO3의 비교에서

성장 기업들은 비성장 기업들보다 높은 점수를 얻은 것을 볼 수 있다.

# 04 분석결과

자연어 처리를 통한기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

## 외부 평판 분석

서울대 언어학과에서 만든 한국어 감성어 사전을 이용해서 기업의 뉴스데이터를 분석해보자

### 정규분포에 대한 가설 검정

귀무가설 : 평판score 값이 정규분포를 띌 것이다.

대립가설 : 평판score 값이 정규분포를 띄지 않을 것이다.

NormaltestResult(statistic=4.1443, pvalue=0.01259)

P.value < 0.05 , 귀무가설 기각 (대립가설 참)

평판 score 값이 정규분포를 띈다.(비모수적 검정)

### 비모수적 검정 Wilcoxon Test

귀무가설 : 내부평판의 성장여부에 따른 평판score의 유의미한 차이가 없다.

대립가설 : 내부평판의 성장여부에 따른 평판score의 유의미한 차이가 있다.

RanksumsResult(statistic=1.9639, pvalue=0.0495)

P.value < 0.05 , 귀무가설 기각 (대립가설 참)

성장여부에 따른 평판 Score의 유의미한 차이가 있다.

## 자연어 처리를 통한기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

### 내부 평판 분석 결과

기업의 단점은 비성장기업과 성장 기업이 모두 갖고 있는 사회구조적인 고질적인 문제점이 크다.

하지만, 성장 기업들은 비성장 기업보다 2배이상의 장점을 가지고 있으며 이와 같은 요소를 취한다면 비성장 기업도 충분히 성장기업으로 평가 받을 수 있다.

내부 평판이 향상 된다면 좋은 지원자와 퇴사자를 막을 수 있는 인력관리의 효율이 오를 것이다.

### 외부 평판 분석 결과

뉴스데이터를 분석을 통해 외부 데이터의 긍정/부정 감성 분석을 하였고 내부 평판의 성장여부와 외부 평판은 유의미한 차이가 있는 것을 알았다.

자연어 처리를 통한 기업의 평판 분석으로 성장기업과 비성장 기업 차이 찾기

## 기업의 평판

정량적 성과

+

정성적 가치

어떻게 보면 너무 당연한 이야기이지만

정량적 성과로만 평가 되었던 반쪽 짜리의 기업 평판에서 비정형적인 데이터 분석을 통해 이전에 없었던 기업의 평판을 알아 볼 수 있었던 계기가 된 것 같다.

시간이 너무 모자랐고 한정된 시간 내에서 크롤링을 통해 많은 기업들의 뉴스데이터를 가지고 하지 않아 아쉬운 점이 있었다.

될 수 있었다면 기업의 업종별로 분류하여 평판이 중요한 업종과 아닌 업종이 나타나는지 알아보고 싶다.

이외에도 , 이번에 진행한 기업분석 외에도 자연어처리를 통해 다른 분야에도 적용 할 수 있을 것이다.

지금까지 김형준이었습니다.