

The Voice of Employees:

A text mining approach to job satisfaction

Unstructured Data Analysis

14th June, 2018

고려대학교 경영대학원
경영정보학(MIS) 전공
김동원 (2017020053)
박재용 (2016020946)
정연재 (2018020058)

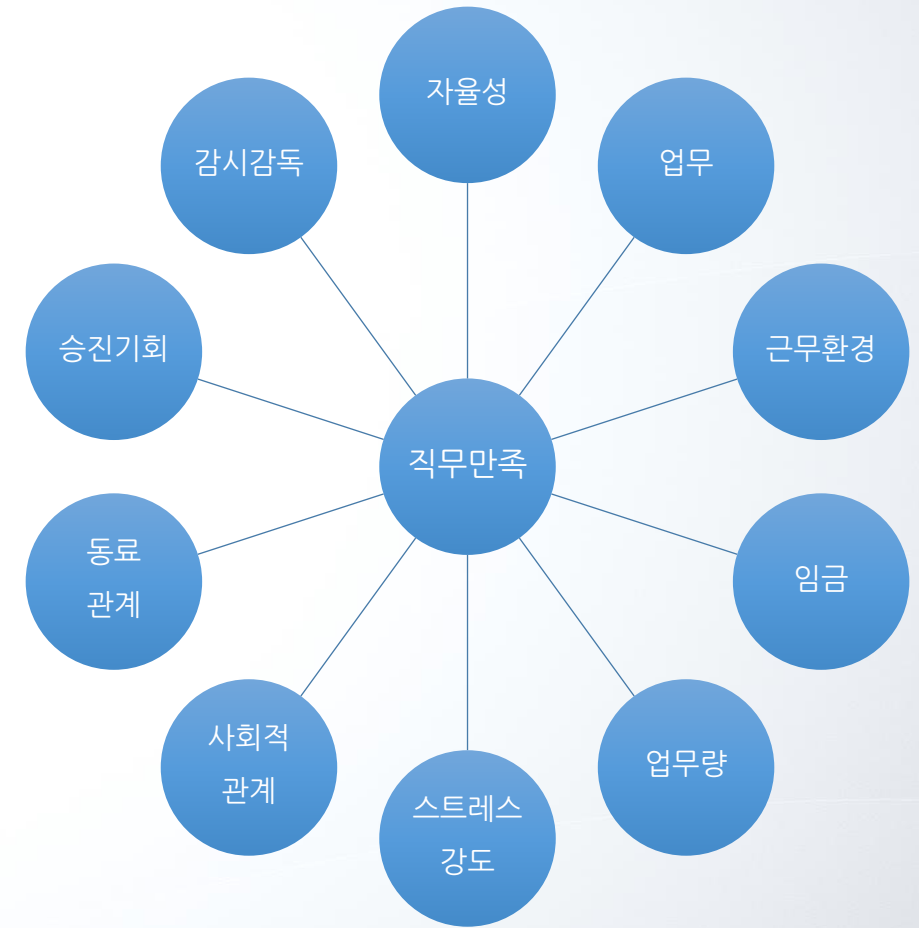
1. Introduction

문제제기

직원만족은 직원성과, 동기부여, 이직, 퇴사, 고객만족 등 다방면에 영향을 미침

온라인 리뷰는 개인이 가장 중요시여기는 요소(Feature)에 대한 경험을 담은 글

RQ: 온라인 리뷰를 통해 직원(Employee)의 직무만족도를 관리할 순 없을까?



Facets of job satisfaction
(Hackman & Oldham, 1980; Smith, Kendall, & Hulin, 1969)

1. Introduction

문제제기

RQ: 온라인 리뷰를 통해 직원(Employee)의 직무만족도를 관리할 순 없을까?

Things to do...

특정 직무만족 관련 키워드와 관련된 다른 키워드는 무엇이 있을까?

Ex) 여성직원과 관련된 만족/불만족 요소들

기업의 추천여부에 영향을 미치는 변수는 무엇일까?

Ex) 잦은 야근, 수직적 기업문화 등...

직무 만족과 관련된 토픽들은 무엇일까?

Ex) 복지, 문화, 임금 등...

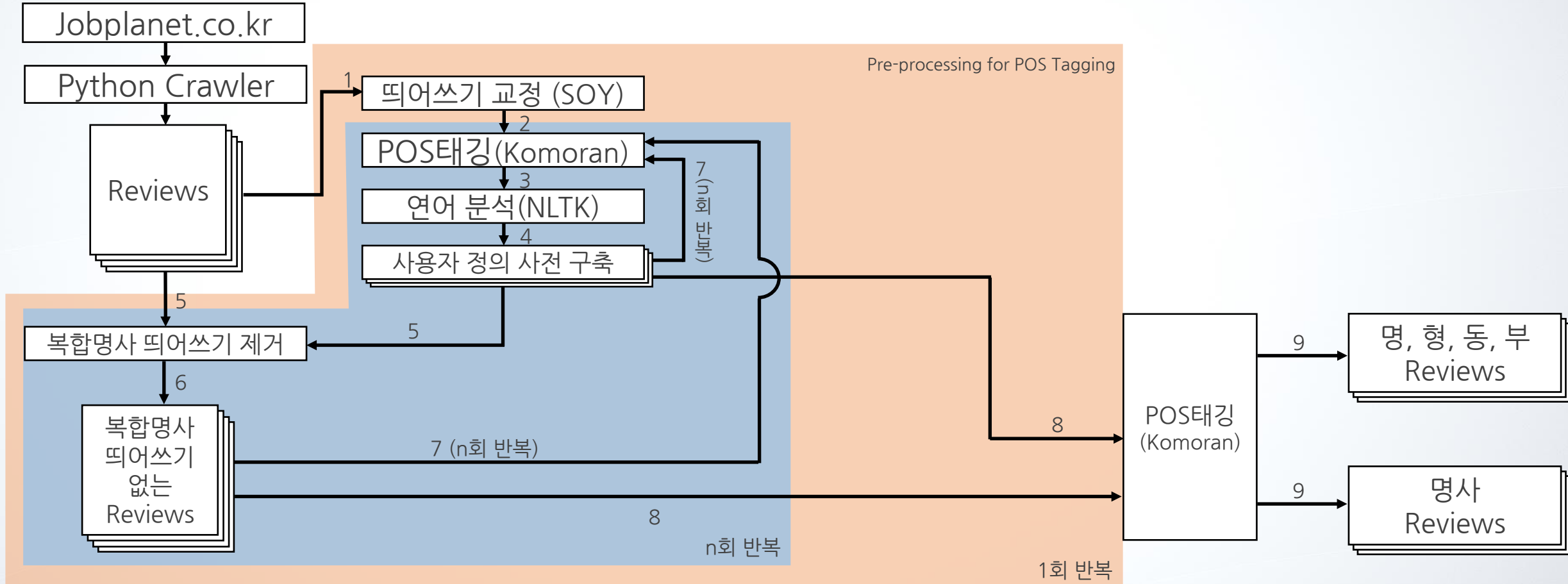
직무 만족과 관련된 토픽들의 감성은 어떠한가?

Ex) 임금의 감성은 어떠한가



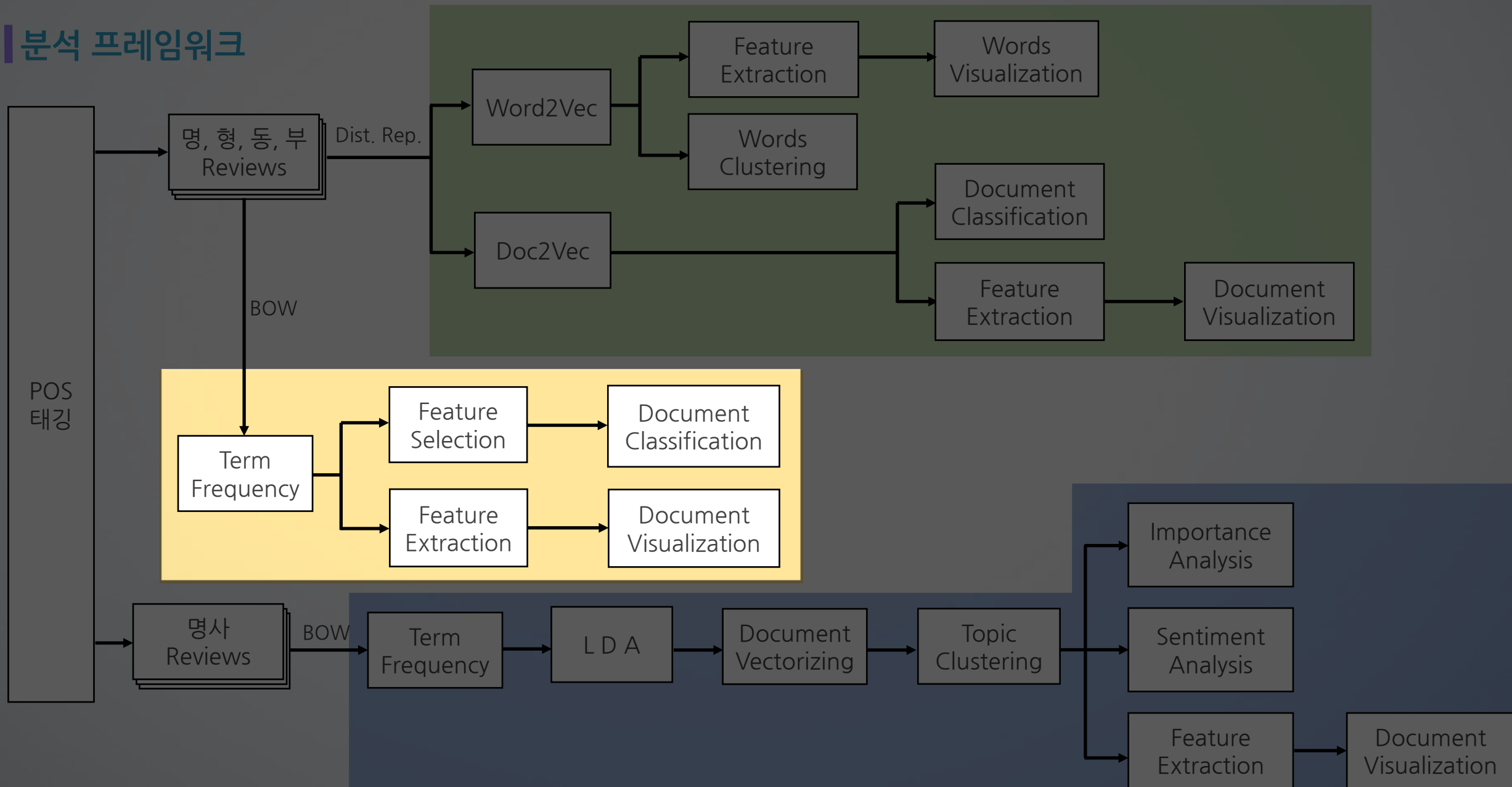
2. Preprocessing

전처리 프레임워크



3. Data Analysis

분석 프레임워크



3. Data Analysis

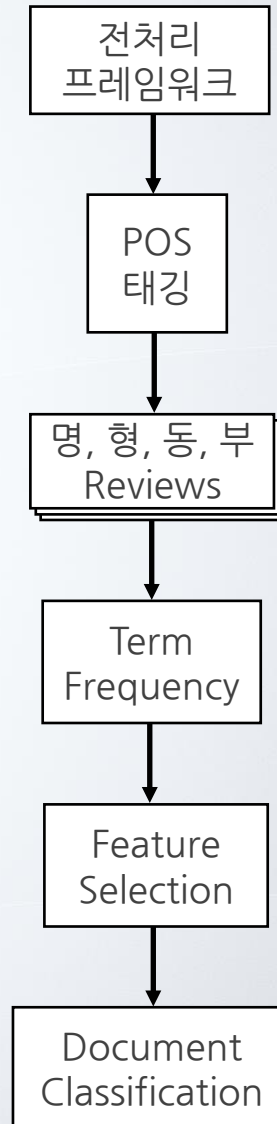
| BOW 기반 추천여부 분류 (Feature Selection)

Textual data Only

	Naïve Bayes	k-NN	C-Tree
Accuracy	0.6198	0.5823	0.5899
Precision	0.5802	0.5776	0.5649
Recall	0.8673	0.6123	0.7826
F1-Score	0.6952	0.5945	0.6562

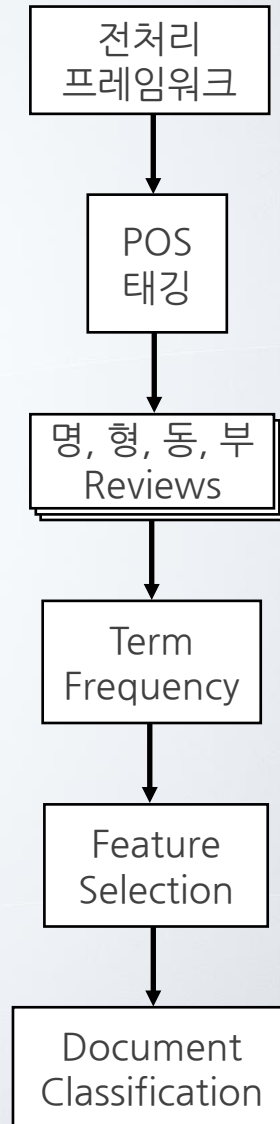
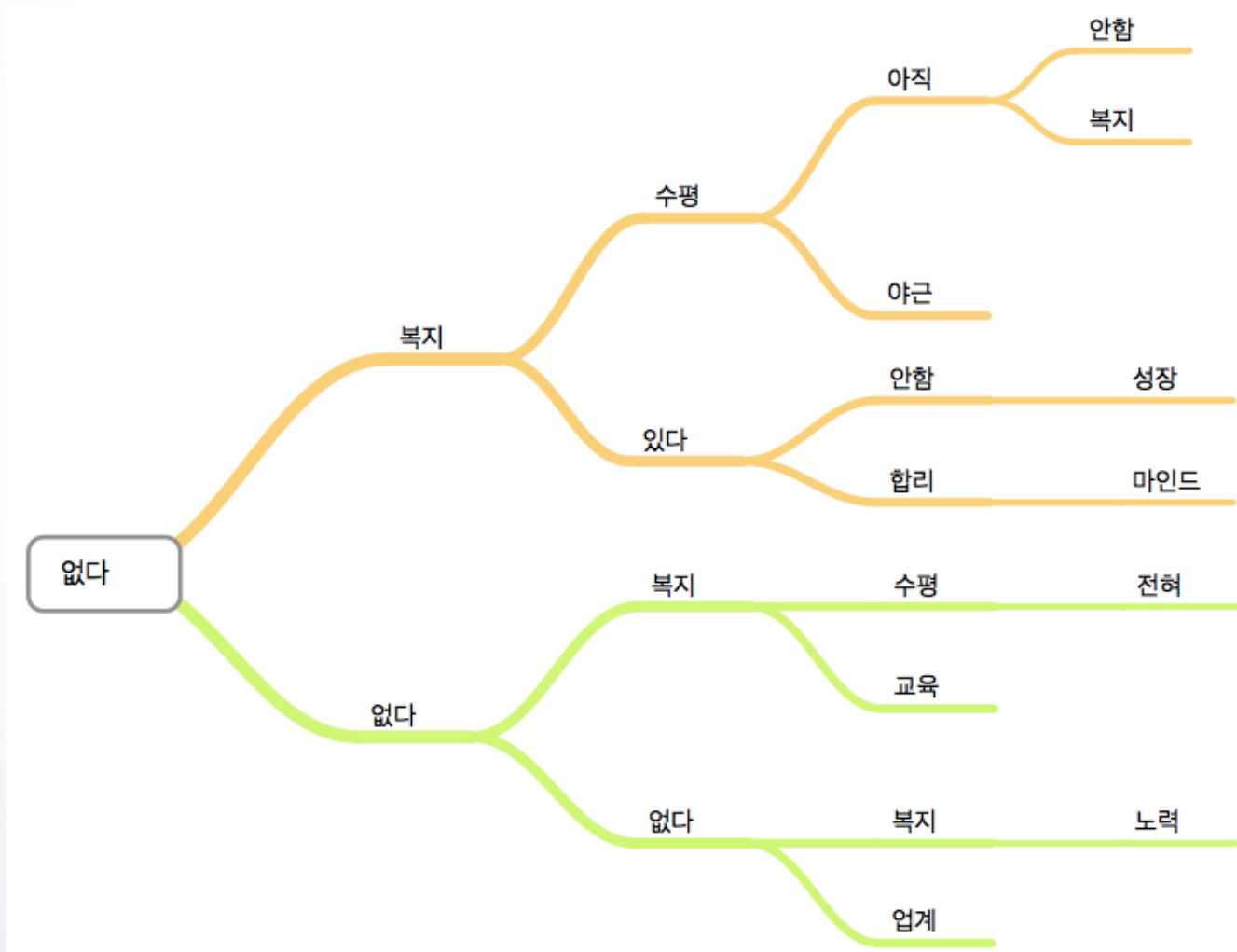
Textual & Structured data

	Naïve Bayes	k-NN	C-Tree
Accuracy	0.6379	0.6813	0.8207
Precision	0.5939	0.6677	0.8194
Recall	0.8720	0.7220	0.8227
F1-Score	0.7066	0.6938	0.8210



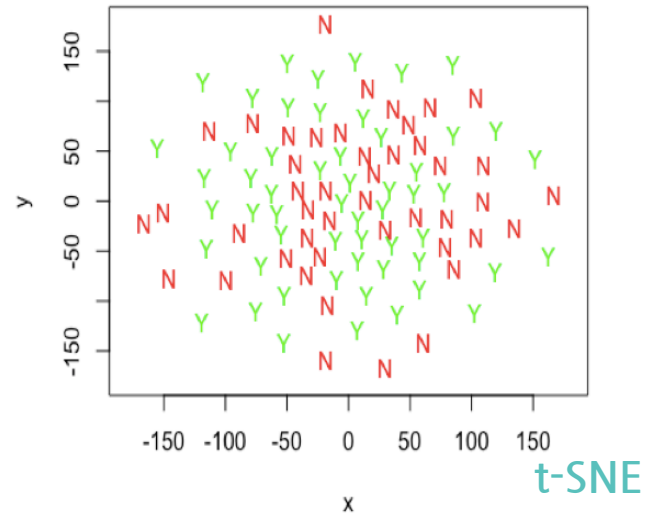
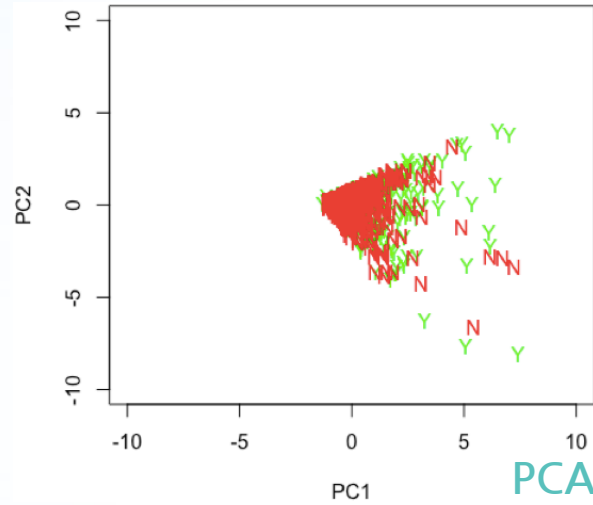
3. Data Analysis

| BOW 기반 추천여부 분류 (Feature Selection)

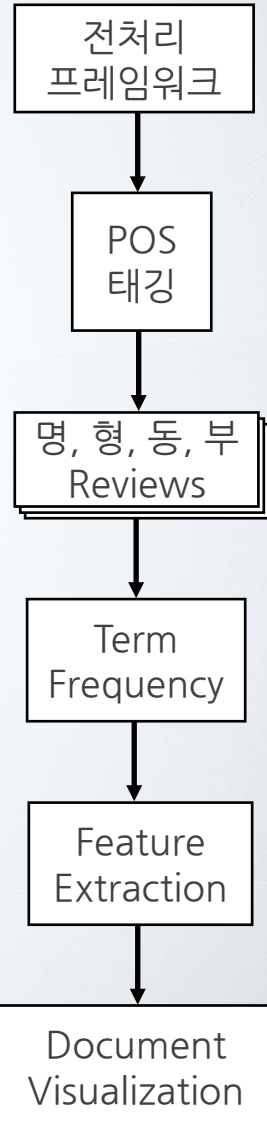
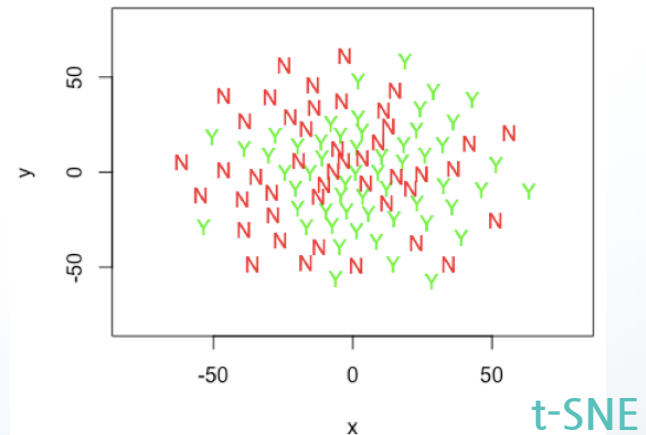
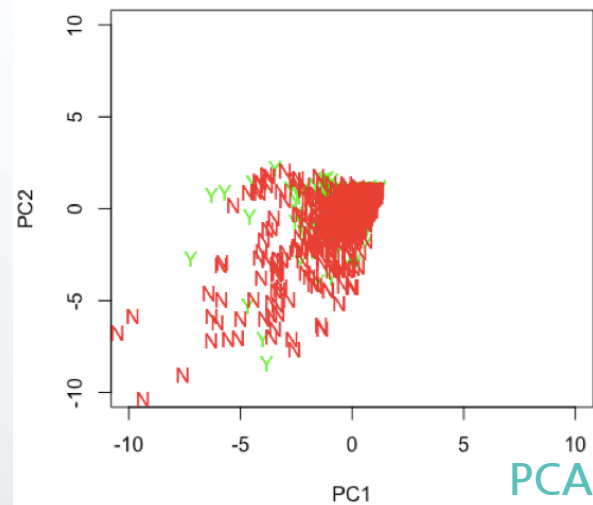


3. Data Analysis

BOW 차원 축소(추천여부-장점리뷰)

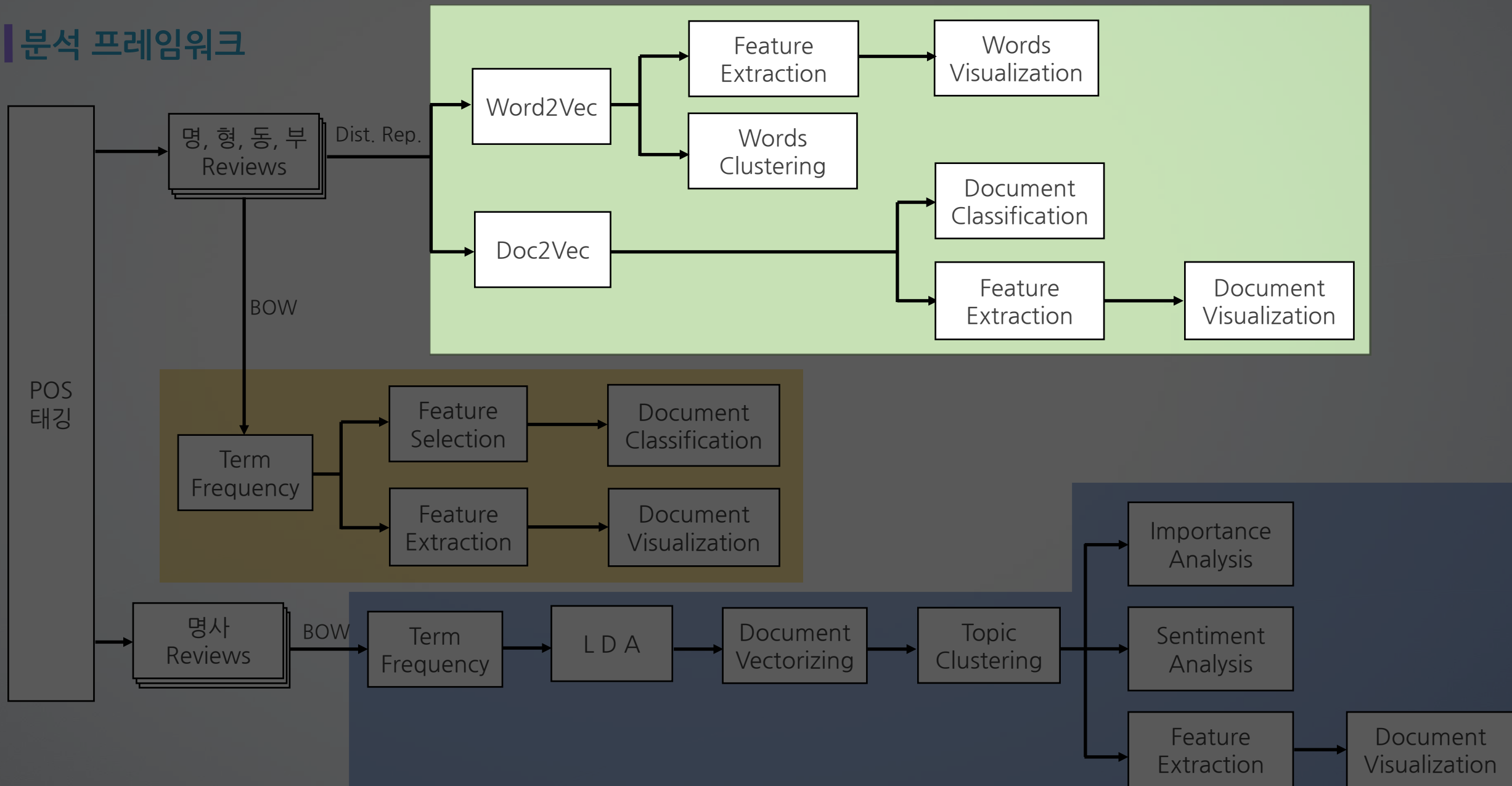


BOW 차원 축소(추천여부-단점리뷰)



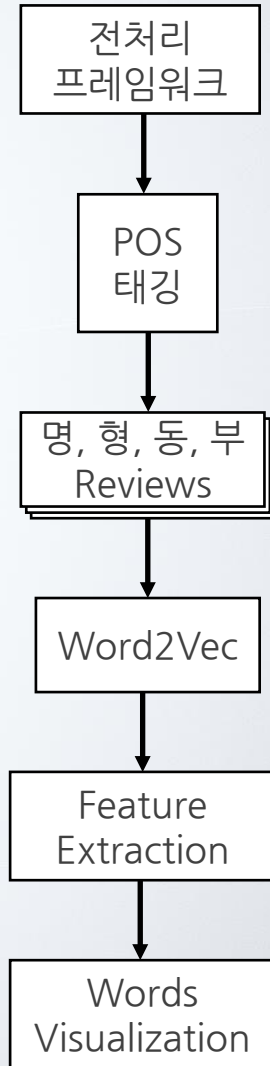
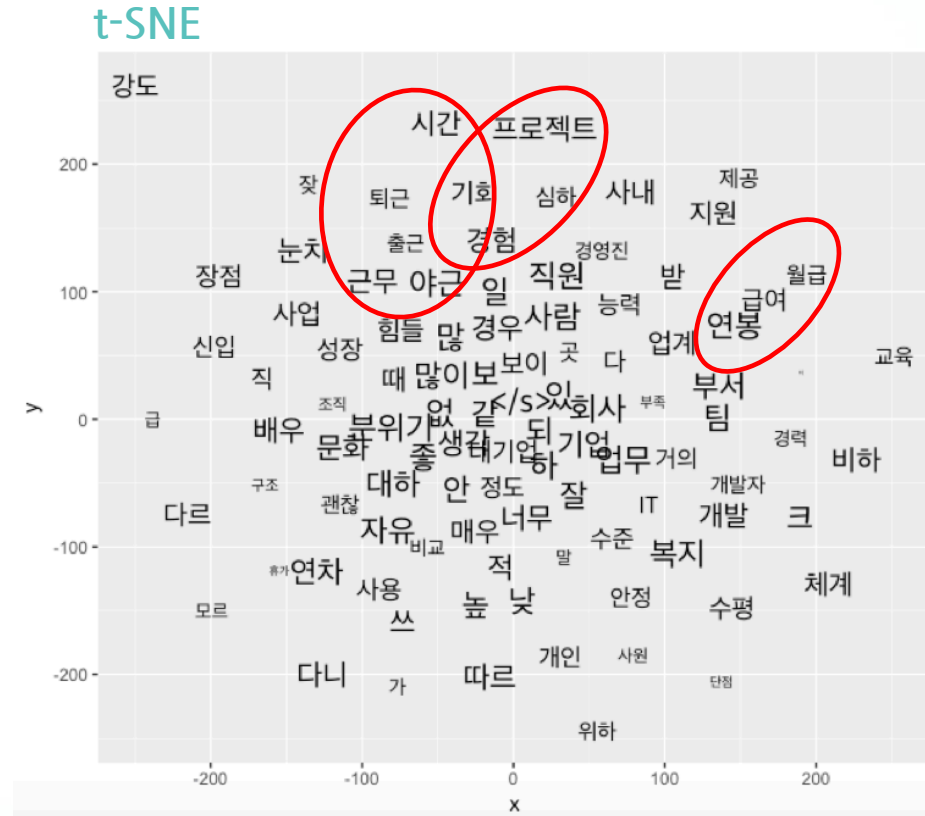
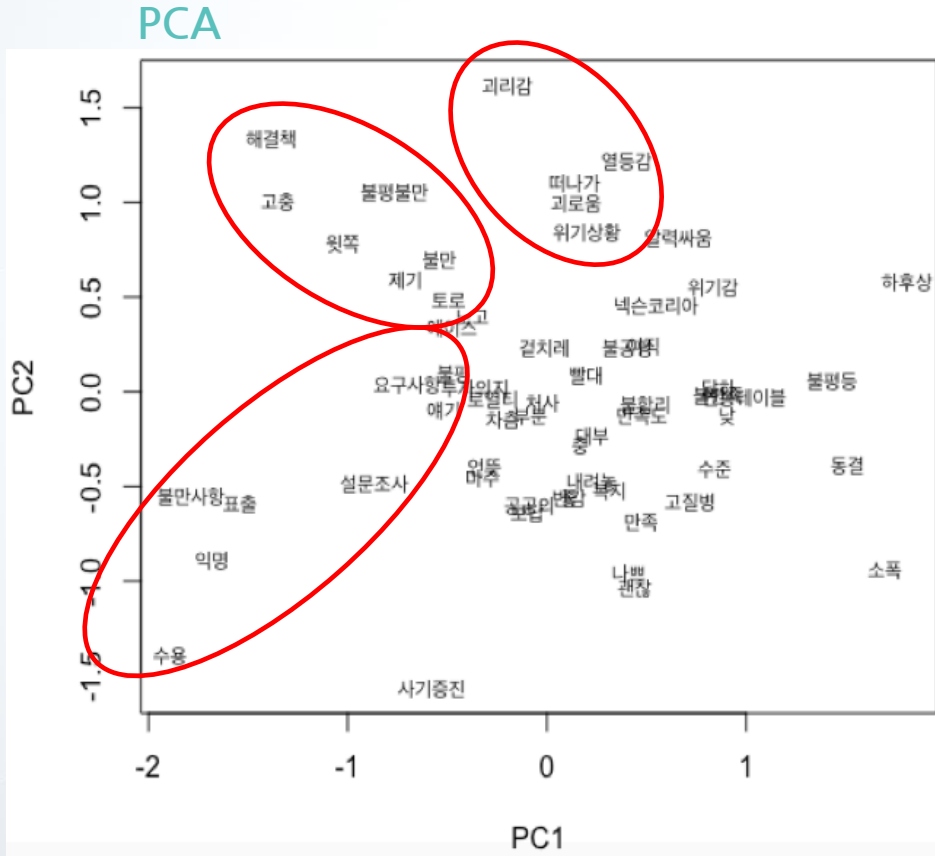
3. Data Analysis

분석 프레임워크



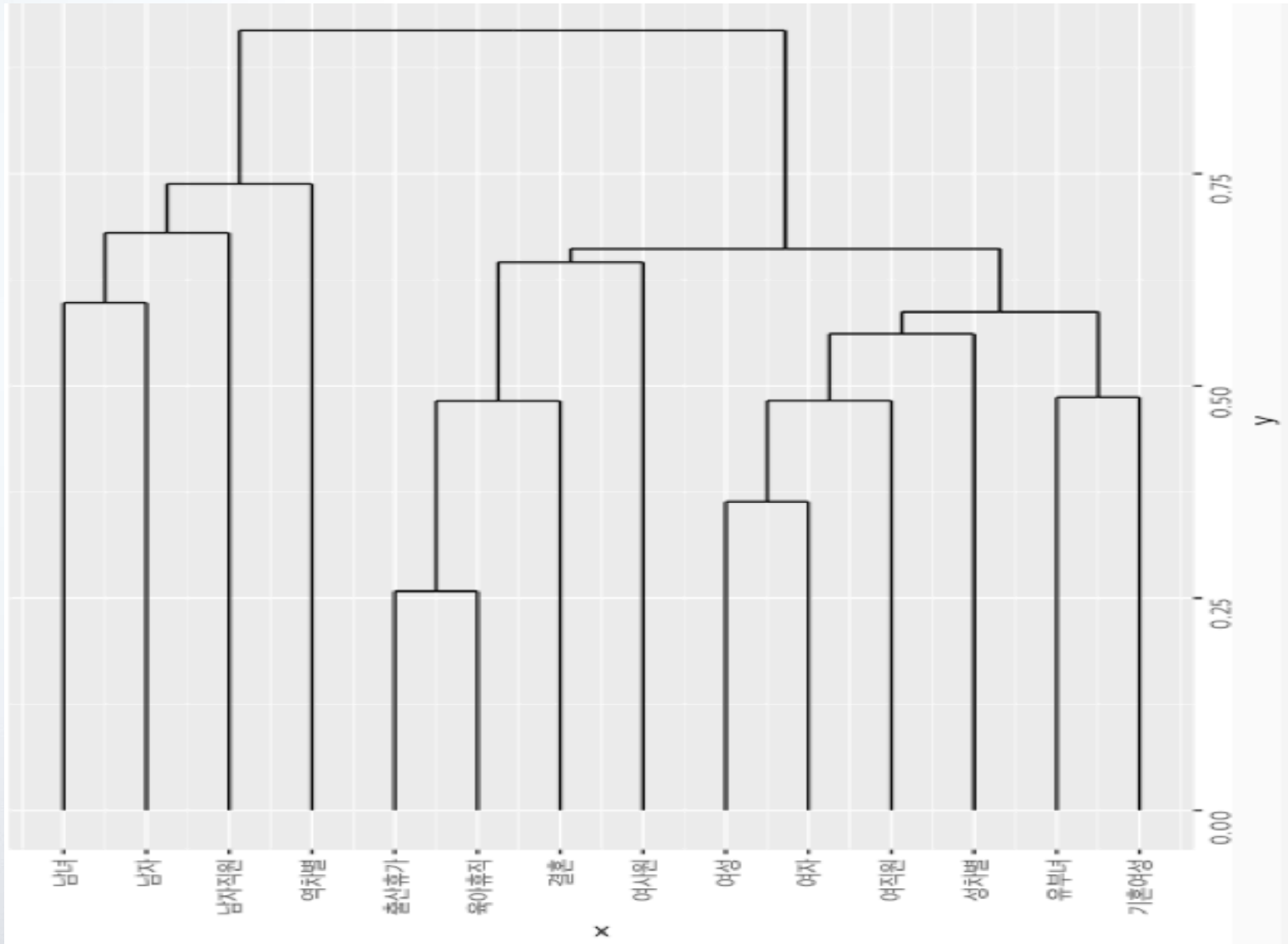
3. Data Analysis

W2V 차원 축소



3. Data Analysis

W2V 단어 클러스터링



3. Data Analysis

D2V 기반 추천여부 분류

Textual data Only

	Naïve Bayes	k-NN	C-Tree
Accuracy	0.6764	0.5457	0.5773
Precision	0.6711	0.5332	0.5752
Recall	0.6920	0.7340	0.5913
F1-Score	0.6814	0.6177	0.5831

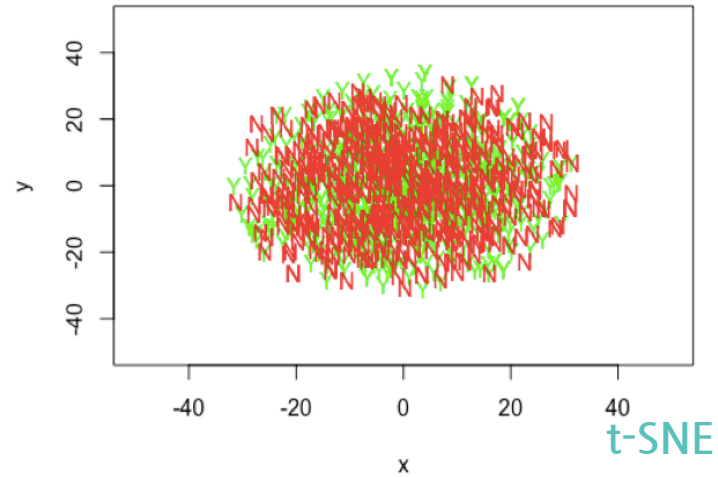
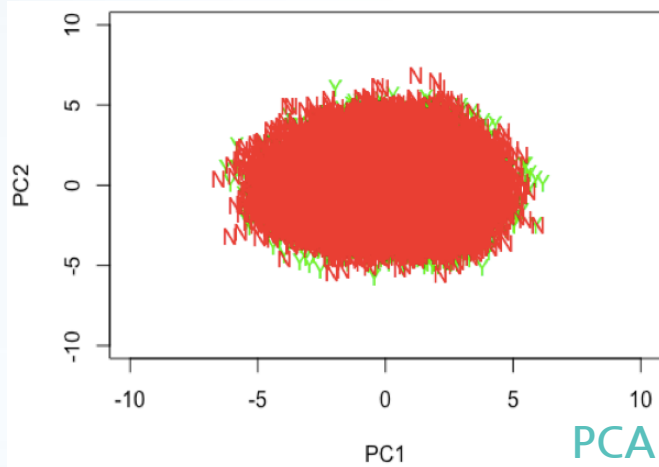
Textual & Structured data

	Naïve Bayes	k-NN	C-Tree
Accuracy	0.8286	0.6072	0.8273
Precision	0.8105	0.5762	0.8216
Recall	0.8578	0.8104	0.8362
F1-Score	0.8334	0.6735	0.8288

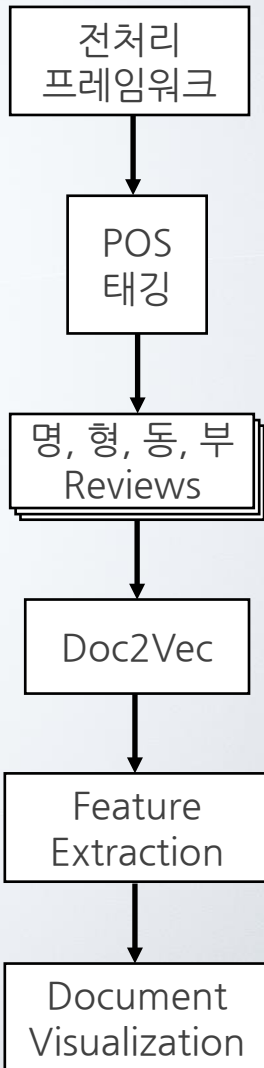
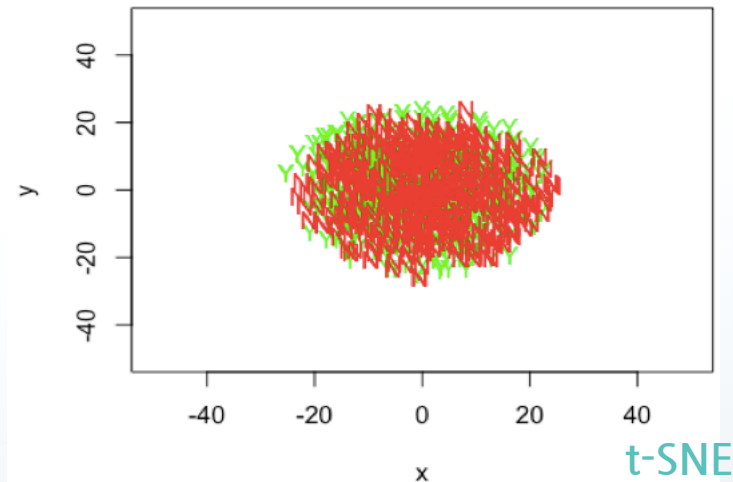
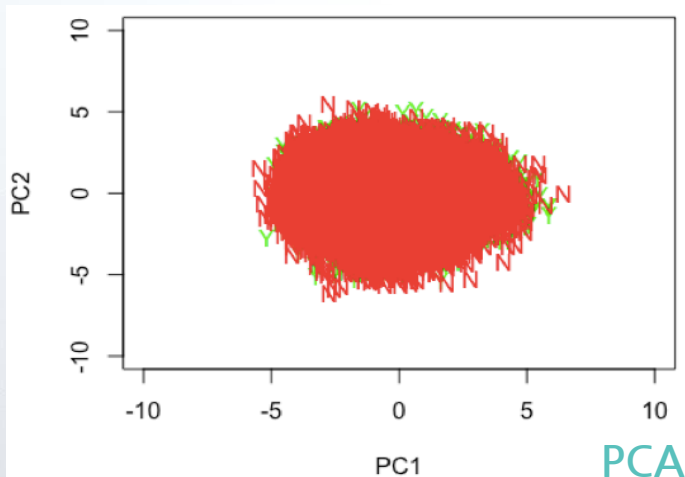


3. Data Analysis

D2V 차원 축소(추천여부-장점리뷰)

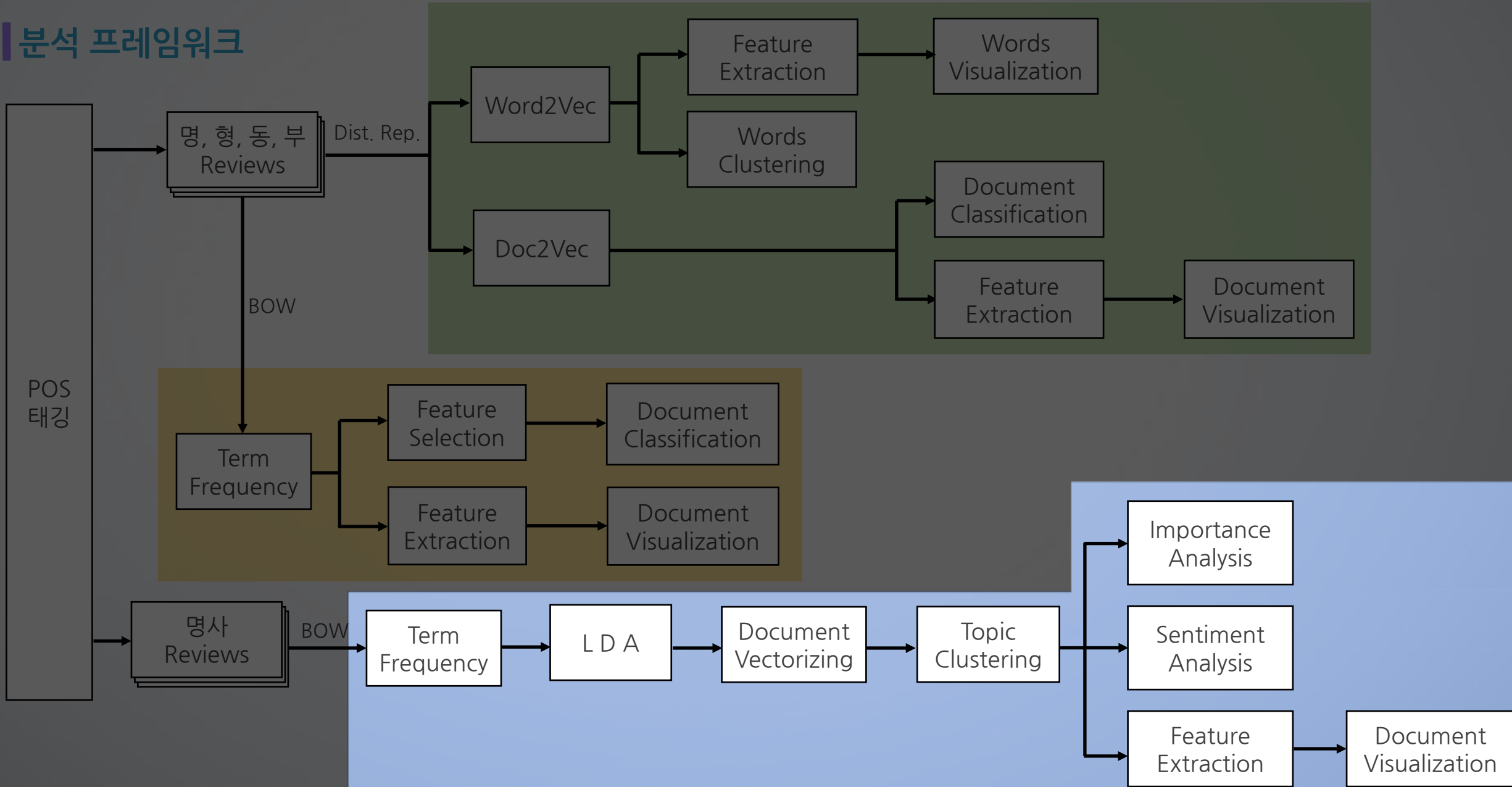


D2V 차원 축소(추천여부-단점리뷰)



3. Data Analysis

분석 프레임워크



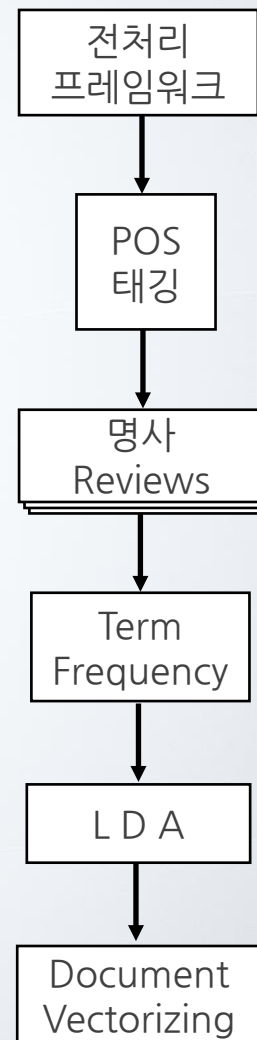
3. Data Analysis

토픽 모델링

기업문화 및 분위기

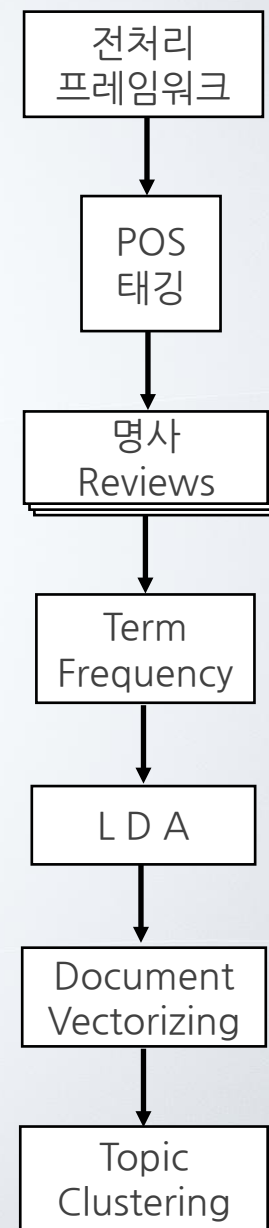
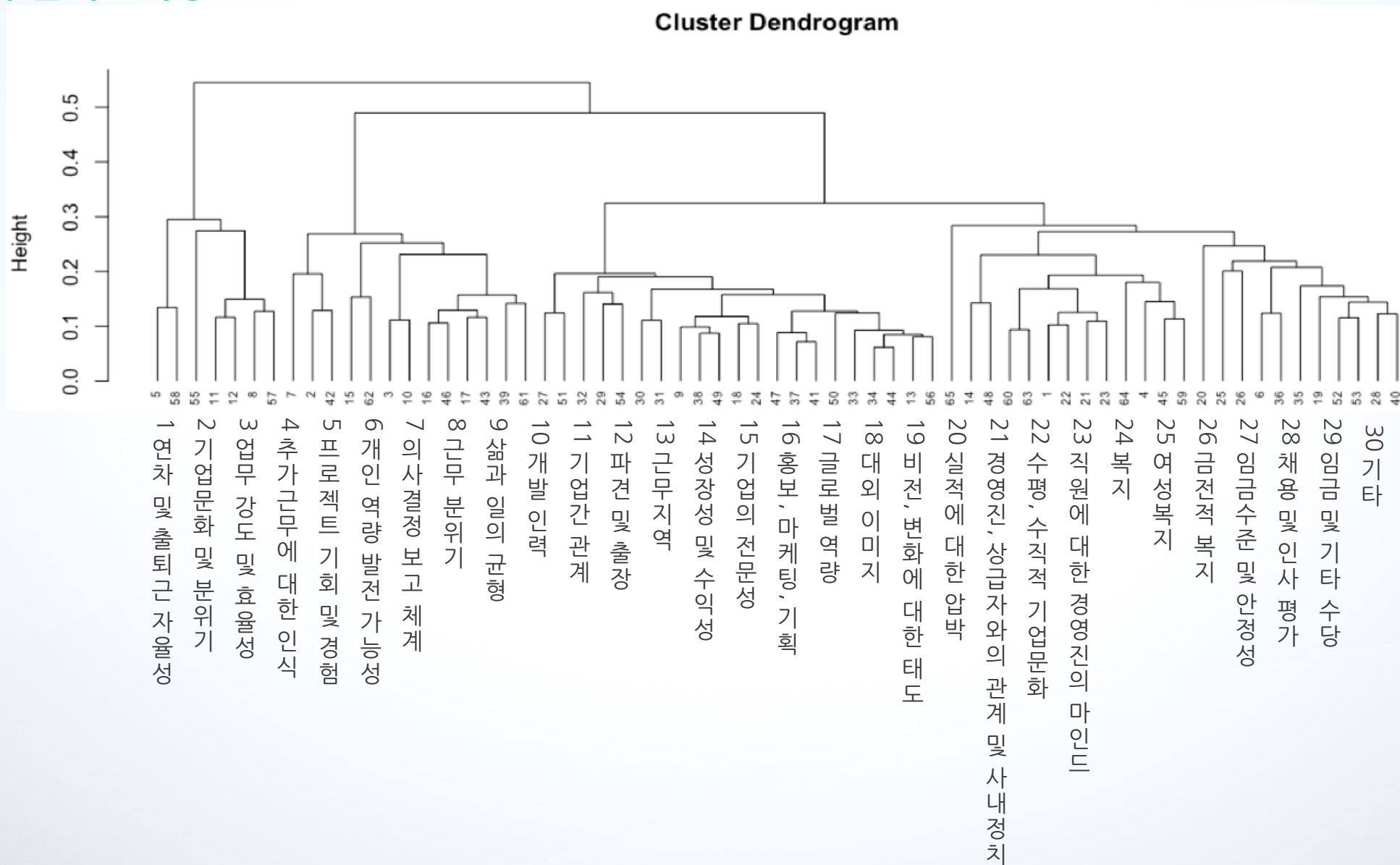
여성관련 복지

Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18	Topic 19	Topic 20	Topic 21	Topic 22
부서	분위기	경험	프로젝트	분위기	자유	직원	수평	퇴근	업무	업무	기업	경력	업계	복지	사내
업무	아근	업무	업무	자유	분위기	연봉	조직	출근	신입	보고	한국	신입	연봉	여자	복지
널	자유	분야	진행	수평	복합	복지	구조	시간	사람	작업	외국	연봉	동종	여성	분위기
아근	회식	관련	아근	문화	자율	최근	수직	아근	개인	문서	글로벌	직	복지	육아휴직	건물
차이	강요	보안	투입	사내	출퇴근	인력	문화	눈치	발전	시간	기회	사람	대비	분위기	자유
분위기	문화	기술	인력	직급	출근	사업	분위기	정시	본인	회의	영어	이직	수준	안정	카페
사람	사람	전문	개발	구조	연차	퇴사	업무	업무	단점	효율	해외	입사	급여	여직원	시설
팀	팀	지식	일정	조직문화	비교	매출	조직문화	칼퇴	성장	작성	경험	직원	평균	직원	사무실
강도	술	기회	개발자	연차	복지	사람	체계	퇴근시간	적응	체계	본사	퇴사	상위	칼퇴	식당
연봉	부서	네트워크	사람	직원	근무	경영진	의사결정	점심시간	체계	보고서	국내	채용	기업	눈치	카페테리아
특정	사내	경력	직원	비교	사람들	기존	보수	사람	생각	처리	일본	추천	비교	업무	사육
협업	수평	컨설팅	파견	사용	일하는	상태	대기업	출근시간	어려움	진행	워크	생각	업무	직장	실
커뮤니케이션	업무	개발	관리	복지	사용	구조조정	사람	주말	부분	불필요	사람	개발자	최고	생각	환경
개발	칼퇴	업체	인원	관계	사내	분위기	기업문화	반	필요	프로세스	지사	업무	규모	남자	이용
소통	퇴근	프로젝트	팀	호칭	사람	신규	기업	근무	혼자	관리	외국인	대우	업종	자유	직원
타	가족	솔루션	사업	사람	직원	규모	경직	분위기	시간	필요	장점	체계	게임	근무	헬스장
협조	비교	도움	경험	존중	시간	급여	관계	1시간	도움	의사결정	미국	적응	정도	연봉	제공
칼퇴	전체	기업	기간	기업문화	출퇴근시간	인원	스타트업	직원	반복	부서	단점	공채	타	출산휴가	밥
개인	눈치	커리어	부족	수직	부서	성장	단점	개인	노력	시스템	엔	정도	신입	보장	근무환경
생각	직원	시스템	수주	의견	팀	상황	사내	상사	전문	절차	문화	직장	경쟁사	기업	커피



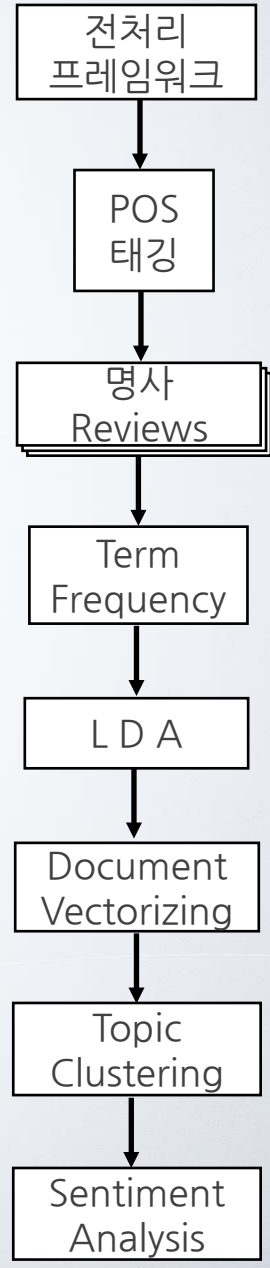
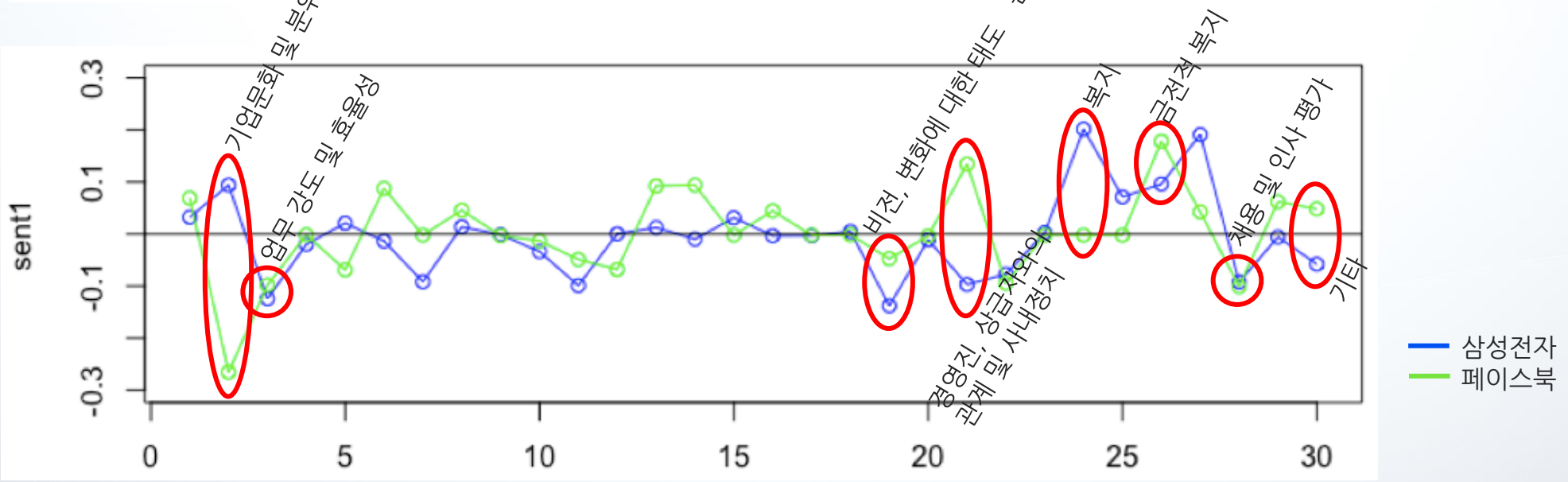
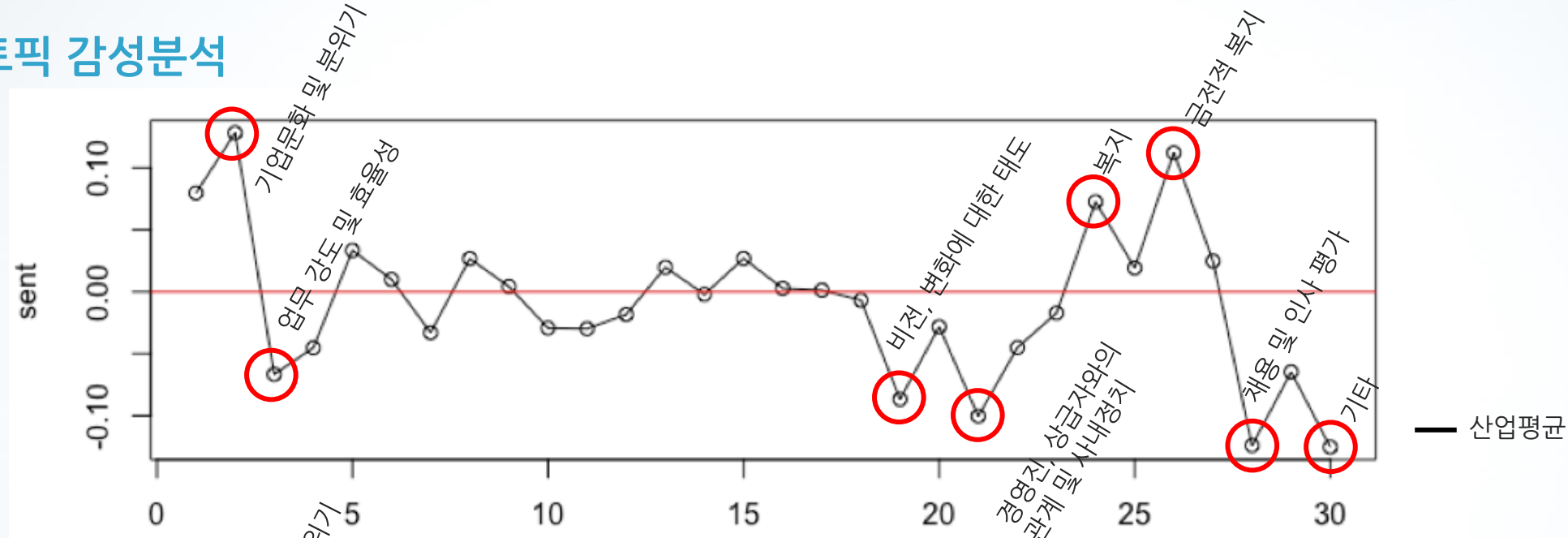
3. Data Analysis

토픽 클러스터링



3. Data Analysis

토픽 감성분석



3. Data Analysis

토픽 감성분석

장점

좋은 복지 혜택 및 근무 환경을 갖추고 있음에도 내가 할 일을 스스로 찾아 할 수 있는 스타트업 같은 분위기. 지사이지만 **본사 임원진을 가까이서 만날 수 있는 기회도 잦음**.

장점

엄청난 자율성과 책임이 개인에게 주어지고 조직문화가 상하관계가 아닌 수평적임. 빠르게 성장하는 조직인만큼 빠른 결정과 판단력이 요구되고 각자개인이 알아서 일을 정해서 하는 문화. 이런회사가 있을 수 있구나 싶을 정도로 직원을 믿고 개개인의 의견을 존중해줌. 복지 최고 **경영진 역시 이런 독특한 문화를 장려하기 위해 많은 노력**. 함께 일하는 동료들로부터 배울게 많고 경영진 역시 능력자들임.

장점

직원을 직원으로 보기보단 정말 사람으로 봐준다는 느낌이 드는 게 가장 좋은 듯. 자율출퇴근, 풍부한 간식거리 제공, 업무적으로 성장할 수 있는 기회가 많은 것도 장점. 쓸데없는 데 신경 안쓰고 일에 진짜 집중하게 해 주는 곳. COO인 세릴 샌드버그가 얘기하는 Lean in (린인) 때문에 여직원들에 대한 배려가 있는듯.

장점

수평적 문화의 최고봉. open discussion을 권장하고 매우 fair하며 모든 것을 자율적으로 처리함. **사람을 가장 중시하고** collaboration을 중요한 가치로 여기며 배울 것이 많아 끊임없이 성장하고 스스로를 계발할 수 있는, 보기도문 사회적 조직.

장점

개개인이 최고의 성과를 내도록 돕고 서비스 하는 회사. 최고 복지와 함께 가족을 배려하고 돕도록 장려하는 회사. 남녀의 차별이 없는 회사 면접도 하나의 서비스인 회사. 개인의 의견이 존중받는 회사. **임원에서 사원까지 수평적인 문화**를 가진 회사. 정말 스마트하고 나이스한 사람들이 모인 회사.

장점

1. 업계 최고 수준의 글로벌 인재들과 협업하며 지적으로 자극 받고 지속적으로 성장할 수 환경
2. 경직되지 않고 유연하며 자율적이고 **수평적인 독특한 조직 문화**
3. 빠르게 성장하는 글로벌 기업에서 다양한 업무를 경험할 수 있는 기회

전처리
프레임워크

OS
환경

명사
Reviews

Term
Frequency

LDA

Document
Clustering

Topic
Clustering

Sentiment
Analysis

3. Data Analysis

토픽 감성분석

단점

너무 빠른 성장으로 인해 지속적인 변화와 공부가 필요하고, 일반적인 한국기업의 **기업문화**와는 많이 다르기 때문에 경력직의 경우 적응이 어려울 수 있음

단점

너무 자율적이다 보니 지시를 받는 것 (또는 하는 것)에 익숙하면 적응하기 어려움

단점

스스로 일을 찾아서 하는 방식이라 지시하는 방식이 익숙하면 적응하기 힘들 수 있음. 협업을 매우 중요하게 생각하므로 혼자일하는 방식을 선호할 경우 적응하기 어려울 수 있음.

단점

자율성을 바탕으로 한 성과를 중시하는 문화가 장점이자 단점으로 작용하여, 업무량이 일반 회사 대비 많은편.

단점

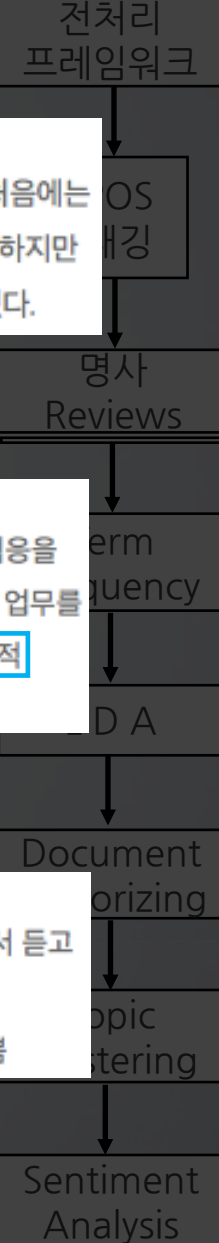
분위기가 워낙 자유롭고, 개인의 선택에 많은 것을 맡기는 기업이기 때문에 처음에는 회사가 불친절하다고 느낄 수 있다. 모든 일에 가이드를 제공하지 않기 때문. 하지만 본인이 조금 더 적극적으로 접근한다면 그만큼 무궁무진한 기회를 얻을 수 있다.

단점

기업의 빠른 발전 때문에 항상 새로운 변화에 대한 대비가 필요하고 그 변화에 적응을 하기 위해 받는 스트레스, 업무가 주어지는 것보다 스스로의 장점을 살릴수 있는 업무를 끊임없이 발견하고 연마해야 함. **한국의 기업 문화와 외국 기업문화 사이의 암묵적 문화적 갈등**

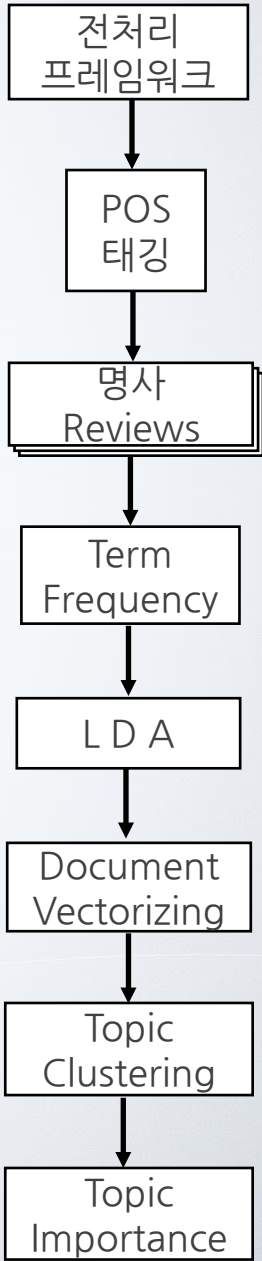
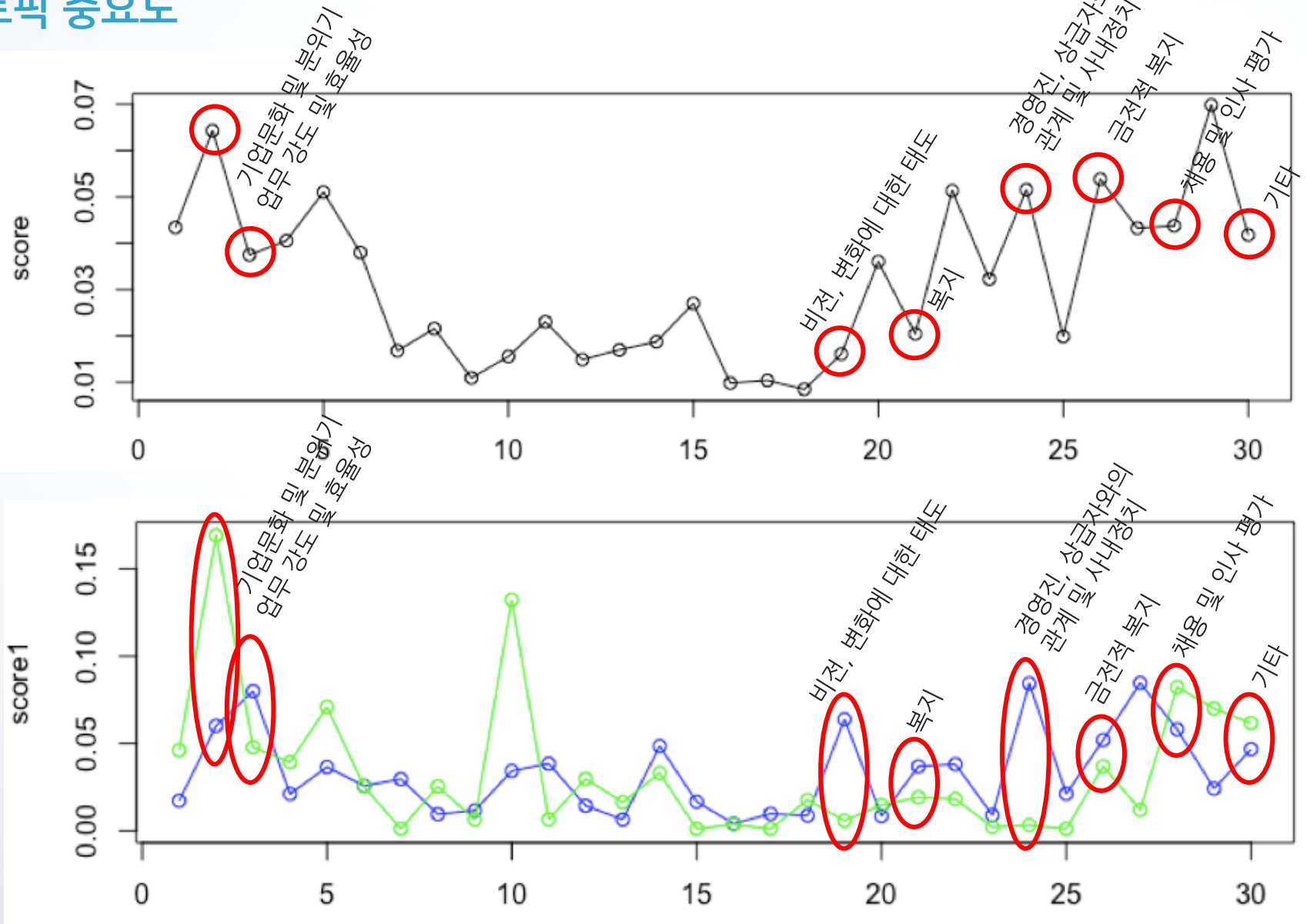
단점

자율성 - 직장생활 초년 시작은 조금 어려울 수 있음. 교육은 스스로 알아서 찾아서 듣고 배워야함. 너무 자유로워서 나중에 다른 기업으로 옮기면 적응이 어려울 수 있음
업무강도 - 일인당 담당 영역이 많고 항상 변해서 계속 배워야하고 일이 많고 바쁨



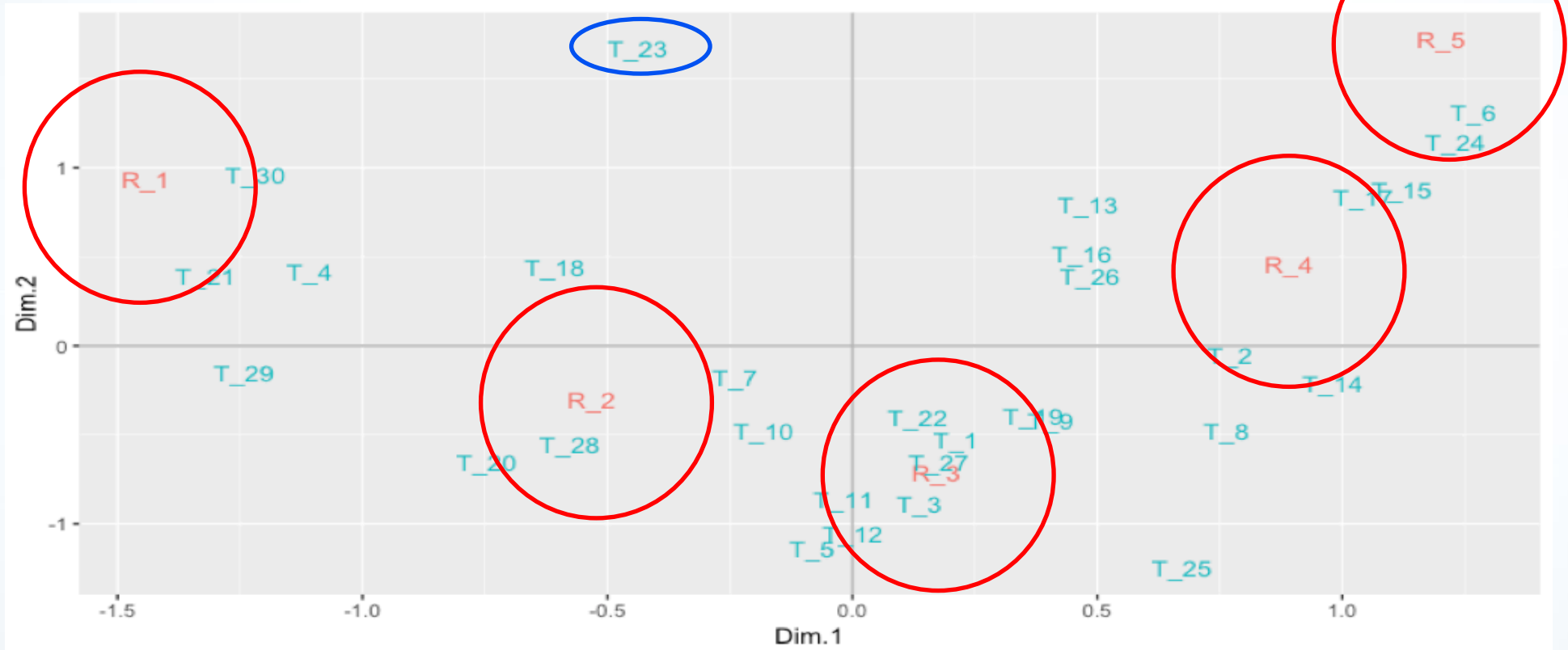
3. Data Analysis

토픽 중요도

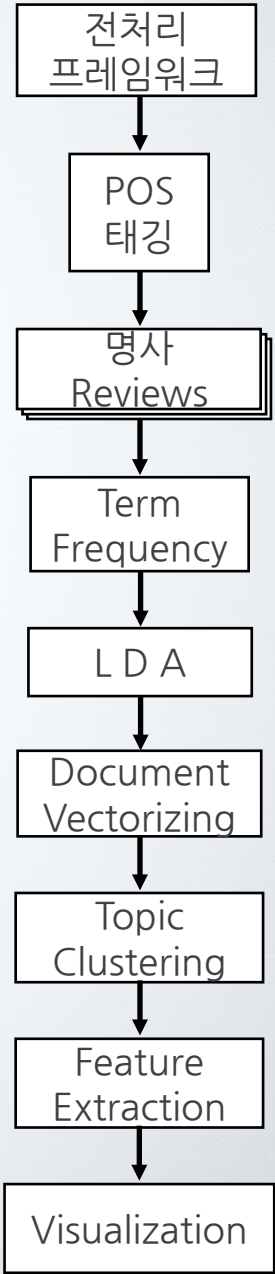


3. Data Analysis

토픽 차원 축소 (Corresponding Analysis)



1 연차 및 출퇴근 자율성	7 의사결정 보고 체계	13 근무지역	19 비전, 변화에 대한 태도	25 여성복지
2 기업문화 및 분위기	8 근무 분위기	14 성장성 및 수익성	20 실적에 대한 압박	26 금전적 복지
3 업무 강도 및 효율성	9 삶과 일의 균형	15 기업의 전문성	21 경영진, 상급자와의 관계 및 사내정치	27 임금수준 및 안정성
4 추가근무에 대한 인식	10 개발 인력	16 홍보, 마케팅, 기획	22 수평, 수직적 기업문화	28 채용 및 인사 평가
5 프로젝트 기회 및 경험	11 기업간 관계	17 글로벌 역량	23 직원에 대한 경영진의 마인드	29 임금 및 기타 수당
6 개인 역량 발전 가능성	12 파견 및 출장	18 대외 이미지	24 복지	30 기타



4. Conclusion

| Summary of Results

직무 만족(추천여부)과 관련된 키워드를 발견 (BOW-Classification)

특정 키워드와 유사한 만족관련 키워드 발견 (W2V-Clustering)

만족 관련 토픽 발견, 군집화, 감성분석 및 중요도 분석 (LDA)

기업 평점에 영향을 주는 토픽들을 발견 (Corresponding Analysis)

| Business Implications

각 기업별 직원 만족과 관련된 토픽차이를 살펴 볼 수 있음(지역별, 직무별 등의 차이 분석도 가능)
(additional idea:경쟁사와의 비교를 통한 이직 방지 등 가능)

직무 만족을 위한 정책의 실효성을 재 검토 할 수 있음
(자율성이 반드시 직무만족도에 긍정적으로 영향을 미치지 않음)

4. Conclusion

Contributions

형태소 분석 시 복합명사와 미등록단어를 보다 효과적으로 분석가능한 프레임워크 제안

재화 및 서비스 관련 온라인 리뷰가 아닌 직원들의 직무만족 리뷰를 분석

추후 직무 만족 관련 연구에서 고려될 수 있는 직무 만족 토픽을 발견 (서베이 작성 시 활용 가능한 키워드)

Limitations

한글 데이터 전처리의 한계 (형태소 분석 등에서 한계)

더욱 다양하고 많은 데이터로 더욱 다양한 분석 가능

리뷰가 10개 이하의 기업제거, 총 등장 빈도수가 10이하인 단어 제거(추후 중요한 만족 관련 토픽이 될 수 있음)

THANK YOU

Unstructured Data Analysis

고려대학교 경영대학원
경영정보학(MIS) 전공
김동원 (2017020053)
박재용 (2016020946)
정연재 (2018020058)