

# 데이터사이언스

이한준 교수

# 논문 연구 참고

- 참고) 논문 읽기에 도움되는 도구 <https://www.themoonlight.io/ko>

The screenshot displays the Moonlight website interface. At the top, there is a navigation bar with links for Home, Features, Download, Pricing, FAQ, and Blog. A language selector shows 'KO' and a '로그인' (Login) button. The main content area features a purple background with text about Normalized Binary Cross Entropy (NBCE). A pink box highlights the NBCE formula: 
$$NBCE = \frac{-\frac{1}{N} \sum_{i=1}^N y_i \log p_i + (1 - y_i) \log (1 - p_i)}{-(p \log p + (1 - p) \log (1 - p))}$$
. An AI-powered 'IMAGE EXPLANATION' overlay is visible, explaining the formula in Korean. It states: 'NBCE 공식을 설명합니다. Normalized Binary Cross Entropy (NBCE) 공식은 다음과 같습니다. NBCE = ... 이 공식을 구성하는 각 항목에 대해 설명하겠습니다: N: 데이터 포인트의 총 수입니다. 테스트 데이터의 샘플 수를 나타냅니다. y\_i: 실제 레이블입니다. 각 데이터 포인트에 대해 클릭 여부(0 또는 1)를 나타냅니다. p\_i: 해당 데이터 포인트에 대한 예측 확률입니다. 이 확률은 데이터가 클릭될 확률을 의미합니다. log: 로그 함수입니다. 확률의 로그를 취하여 계산의 안정성을 높입니다.' The bottom of the page includes the Moonlight logo, the tagline '논문을 함께 읽는 AI 동료', and installation instructions for Chrome and Edge. At the very bottom, a dashed box contains the text: '문라이트는 누구나 논문을 업로드하여 무료로 체험할 수 있습니다.'

문라이트는 강력한 AI PDF 리더로, 논문과 대화하고, 요약할 수 있도록 도와줍니다.  
연구에 필요한 PDF 문서를 이해, 관리하는데 걸리는 시간을 크게 절약할 수 있습니다.

문라이트는 누구나 논문을 업로드하여 무료로 체험할 수 있습니다.



1



2



3

독창성

독창성

Search Helper X

## Abstract



Keeping up with the rapid growth of academic literature is an ongoing challenge for researchers. To address this, we present **Moonlight: AI Colleague for Research Papers**, an intelligent PDF reader designed to aid comprehension as well as navigation through academic papers. Moonlight simplifies interactions with scholarly content through its **AI Colleague** and **Navigation** features. **AI Colleague** offers a suite of tools, including text and image explanations, contextual translations, auto-highlighting, and conversational chat—like an insightful colleague. **Navigation** includes smart citation cards, section previews, and external link previews. Like moonlight on a dark night, Moonlight aims to brighten the path to deeper research, transforming how people engage with academic papers. Visit Moonlight at <https://www.themoonlight.io/>.

## 1 Introduction



In recent years, the academic world has witnessed an unprecedented surge in research output, driven in part by generative AI tools like

ChatGPT (Achiam et al., 2023). Thousands of papers are published daily across various disciplines, offering immense potential but also posing significant challenges for researchers. Keeping up with these advances, deciphering methodologies, and extracting key insights in dense academic texts are all while maintaining productivity is a daunting task for many scholars.

To address these challenges, we introduce **Moonlight**, an AI-powered PDF reader designed to define how researchers engage with academic texts. Existing tools such as traditional readers, GPT-based chat windows, and reference managers typically address isolated aspects of the research workflow, often leading users to switch between multiple applications. Moonlight integrates these functions into a single platform (Fig. 1) as a comprehensive research assistant. Leveraging recent advances in AI, including large language models (Achiam et al., 2023; Hurst et al., 2023) and computer vision (Zhao et al., 2023), Moonlight provides a



AI와 함께 ⓘ beta

키워드 사전 ⓘ



- Moonlight AI Colleague Navigational
- Smart Citation Auto Highlight
- Translation AI Document Outlining
- PDF Reader Annotation Bookmark

3줄 요약 ⓘ



1. 🌟 AI Colleague features for simplification and engagement
2. 🗺️ Comprehensive navigation tools for efficient exploration
3. 📝 Annotation support for organized and shareable insights

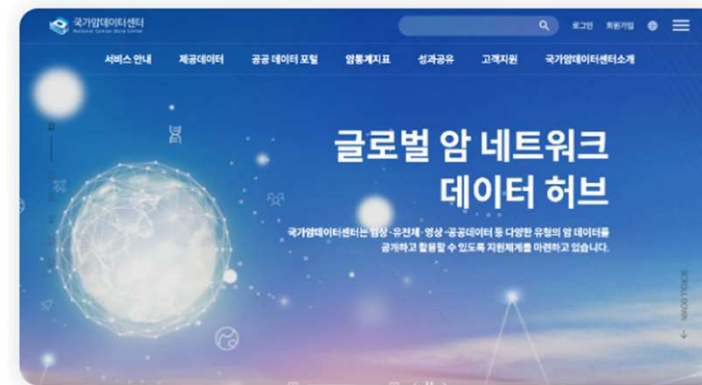
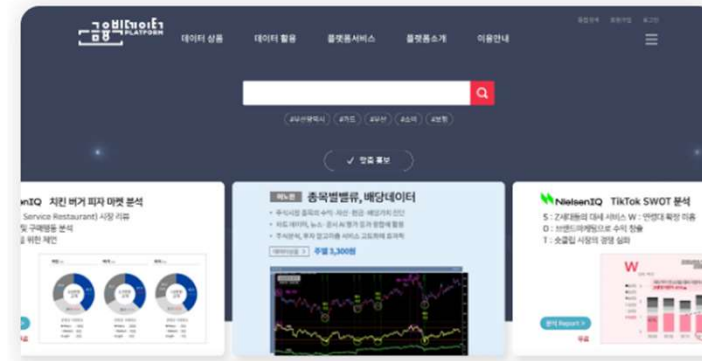
요약 ⓘ



The paper introduces Moonlight, an AI-powered PDF reader tailored for researchers, addressing the challenges presented by the exponential increase in academic literature. The authors—Taeho

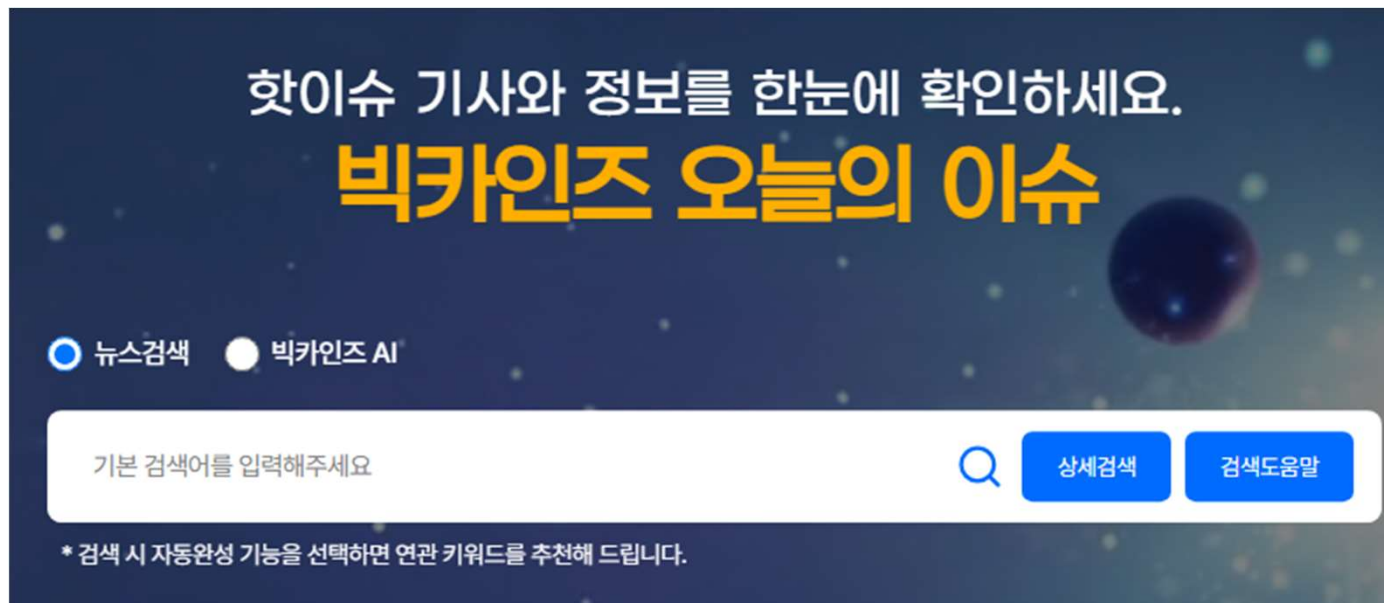
# 논문 연구 참고

- 분야별 빅데이터 플랫폼 <https://boardmix.com/kr/reviews/big-data-platforms/>



# 논문 연구 참고


- 뉴스 기사 빅데이터 플랫폼 <https://www.bigkinds.or.kr/>





# 논문 연구 참고

## • 빅카인즈 데이터 활용 연구



전체 ▾ 빅카인즈 🔍

상세검색

쉽고 빠른 AI 논 내서재 알림

주제분류 BEST 매거진 저널 + DBpia AI AI 검색 idea Citeasy 개인구독 지원사업 명지대학교 로그인 회원가입 고객센터 EN

필터

과 내재검색 본문포함 ☐ OFF

검색어 입력 🔍

선택한 항목 선택해제

이용형태

☐ 외부링크 (241)

☐ 원문이용 (197)

주제분류

☐ 사회과학 (197)

☐ 신문방송학 (57)

☐ 교육학 (33)

☐ 사회학 (18)

☐ 사회복지학 (13)

☐ 사회학 일반 (5)

more

☐ 공학 (47)

☐ 복합학 (46)

☐ 예술체육학 (38)

☐ 인문학 (29)

☐ 자연과학 (8)

☐ 농수해양학 (3)

발행연도

최근 1년 (75)

최근 3년 (245)

최근 5년 (376)

주제 추천결과 (5건)

검색 키워드 알

idea 빅카인즈 연관 리포트/소논문 주제 리포트/소논문 ▾

검색결과 (438건)

☐ 내서재 담기 ☒ 정확도순 ☐ 이용순 ☐ 최신순 20개씩 ▾

스마트팜혁신밸리 관련 사업운영실태분석: 빅카인즈 분석중심으로

그리고 빅카인즈를 활용하여 자료를 수집하고 분석하였다. 주요 결과로서 키워드 트랜드분석에서 전반기는 스마트팜혁신밸리를 통한 청년 농업인 양성정책의 홍보에 집중하였다면 후반기는 청년창업보육센터의 교육생...

#빅카인즈 #스마트팜혁신밸리 #청년 농업인

곽효연 | 조선대학교 지식경영연구원 | 기업과학신연구 | 48(1) | 2025.3 | 93-112(20pages)

학술저널 | KCI등재 | 이용수 3

빅카인즈 시스템을 활용한 UAM관련 국내 언론보도 분석연구

#빅카인즈 #언론보도 #연관어

임중훈 | 한국서비스산업학회 | 서비스산업연구 | 17(2) | 2024.11 | 81-94(14pages)

학술저널 | 이용수 36

빅카인즈 뉴스 빅데이터를 활용한 어린이 식생활 안전 관리 관련 토픽 모델링 분석: 어린이 식생활 안전관리 종합계획 시행 차수별 비교를 중심으로

박혜진, 전유화, 이승재 외 3명 | 한국식품영양과학회 | 한국식품영양과학회 학술대회발표집 | 2024.10 | 472-472(1pages)

학술대회자료 | 이용수 13

# 논문 리뷰

이한준 교수

# 리뷰 전 도움되는 지식

- 감성분석
  - 한국어 감성분석 사전 이용



github 감성분석 사전



전체 이미지 동영상 쇼핑 뉴스 짧은 동영상 도서 더보기 도구



GitHub

<https://github.com/mrlee23/KoreanSentimentAnalyzer>

**mrlee23/KoreanSentimentAnalyzer: 한국어 감성 분석기**

이 패키지는 꼬꼬마 한글 형태소 분석기와 KOSAC(Korean Sentiment Analysis Corpus)의 감성사전 데이터를 이용한 한국어 감성 분석기이다. 형태소 분석을 위해 파이썬으로 ...



GitHub

<https://github.com/semi-supervised-sentiment-analysis>

**감성사전 기반 준지도학습 감성분석**

모델 주요 기능 · make\_seeds : 단어사전 내에서 입력받은 극성 단어와 일치하는 토큰 또는 파생형 토큰을 찾는다. · fit : 극성 단어를 기반으로 감성 사전을 구축한다.



네이버 블로그

<https://m.blog.naver.com>

**[ 사전 기반 감성분석 실습(한글) ] 파이썬 python ... - 네이버 블로그**

2022. 12. 26. — 감성분석은 이미 구축되어 있는 감성어 사전에 기반하거나, 기계학습을 이용해 사전을 구축한 후 해당 사전을 이용하는 방법이 있다. 여기서는 대표적인 ...



GitHub

<https://github.com/park1200656/KnuSentiLex>

**park1200656/KnuSentiLex: KNU(케이앤유) 한국어 감성사전**

각 도메인의 감성사전을 빠르게 구축하기 위한 기초 자료로 활용하기 위해 개발되었음 ... 분석기, <https://github.com/twitter/twitter-korean-text>. About. KNU(케이앤유) ...



GitHub

[https://github.com/ukairia777/finance\\_sentiment\\_corpus](https://github.com/ukairia777/finance_sentiment_corpus)

**금융 뉴스 문장 감성 분석 데이터셋 (finance sentiment corpus)**

금융 감성 분석에 사용할 수 있는 긍정(positive), 중립(neutral), 부정(negative)이 라벨링이 된 텍스트 데이터입니다. 한국어 금융 분석 모델 개발을 위해 기존 금융 ...



velog

<https://velog.io>

**[데이터] 감성분석을 위한 준비 - 라이브러리 설치**

2024. 10. 27. — 감성사전은 이미 구축되어 있는 감성어 사전에 기반하거나, 기계학습을 이용해 사전을 구축한 후 해당 사전을 이용하는 방법이 있다. 다만, 학습을 시키는 ...



# 리뷰 전 도움되는 지식 방법 1: Rule-based (긍정/부정 단어 리스트)

- 감성분석

- 사전 이용 개념

python

 복사

 편집

```
from konlpy.tag import Okt

# 간단한 긍/ 부정 단어 목록
positive_words = ['좋다', '행복', '감사', '사랑', '축하']
negative_words = ['싫다', '화남', '짜증', '불행', '최악']

okt = Okt()

def simple_sentiment(comment):
    words = okt.morphs(comment)
    pos = sum([word in words for word in positive_words])
    neg = sum([word in words for word in negative_words])

    if pos > neg:
        return '긍정'
    elif neg > pos:
        return '부정'
    else:
        return '중립'

df['sentiment'] = df['comment'].apply(simple_sentiment)
df.to_csv('naver_comments_with_sentiment.csv', index=False)
```

# 리뷰 전 도움되는 지식

- 감성분석

- 사전 이용 개념

<https://github.com/park1200656/KnuSentiLex?tab=readme-ov-file>

KNU 한국어 감성사전

작성자 : 온병원, 박상민, 나철원 작성일 : 2018.05.11 소속 : 군산대학교 소프트웨어융합공학과 Data Intelligence Lab  
홈페이지 : <http://dilab.kunsan.ac.kr/>

<> 개요

- 음식, 여행지, 영화, 음악, 자동차, 스마트폰, 강의, 컴퓨터 등 특정 도메인에서 사용되는 긍부정어보다는 인간의 보편적인 기본 감정 표현을 나타내는 긍부정어로 구성된다. 보편적인 긍정 표현으로는 '감동받다', '가치 있다', '감사하다'와 보편적인 부정 표현으로는 '그저 그렇다', '도저히 ~수 없다', '열 받다' 등을 들 수 있다.
- 각 도메인의 감성사전을 빠르게 구축하기 위한 기초 자료로 활용하기 위해 개발되었음
- 본 한국어 감성사전은 다음과 같은 소스로부터 통합되어 개발되었음 (1) 국립국어원 표준국어대사전의 뜻풀이 (glosses) 분석을 통한 긍부정 추출(이 방법을 통해 대부분의 긍부정어 추출) (2) 김은영(2004)의 긍부정어 목록 (3) SentiWordNet 및 SenticNet-5.0에서 주로 사용되는 긍부정어 번역 (4) 최근 온라인에서 많이 사용되는 축약어 및 긍부정 이모티콘 목록
- 총 14,843개의 1-gram, 2-gram, 관용구, 문형, 축약어, 이모티콘 등에 대한 긍정, 중립, 부정 판별 및 정도 (degree)값 계산

<> 특징

- 표준국어대사전을 구성하는 각 단어의 뜻풀이를 분석하여 긍부정어를 추출하였음
- 1-gram, 2-gram, n-gram(어구, 문형), 축약어, 이모티콘 등의 다양한 종류의 긍부정어 포함
- 영화, 음악, 자동차 등 어떤 도메인에도 사용될 수 있는 보편적인 긍부정어로 구성

# 리뷰 전 도움되는 지식

- 감성분석
  - 최신 pre-trained model 사용
  - KoBERT

## 출력 예시

comment	sentiment	confidence
애 날기 힘든 세상임	부정	0.91
정부가 좀 더 도와줘야	중립	0.72
지원금 좋네요	긍정	0.88

여기서는 **KoBERT** 기반 감성 분석 모델 중 하나인

`beomi/kcibert-base` + 감성 분석 fine-tuning 모델을 사용할 수 있어요.

(`nlp04/kcibert-sentiment` 같은 모델도 존재)

## 예제 코드: 감성 분석 with HuggingFace KoBERT

```
python

from transformers import AutoTokenizer, AutoModelForSequenceClassification
import torch
import pandas as pd
from torch.nn.functional import softmax
from tqdm import tqdm

# 모델 로드 (감성분석 fine-tuned KoBERT)
model_name = "nlp04/kcibert-sentiment" # 또는 "beomi/kcibert-base"에 fine-tune 된 모델
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name)
model.eval() # 평가 모드

# 댓글 불러오기
df = pd.read_csv("naver_comments.csv")
comments = df['comment'].astype(str).tolist()

# 감성 분석 함수
def predict_sentiment(text):
    inputs = tokenizer(text, return_tensors="pt", truncation=True, padding=True, max_length=128)
    with torch.no_grad():
        outputs = model(**inputs)
        probs = softmax(outputs.logits, dim=-1)
        pred = torch.argmax(probs, dim=1).item()
        score = probs[0][pred].item()

    label = {0: '부정', 1: '중립', 2: '긍정'} # 모델에 따라 레이블 매핑 다를 수 있음
    return label[pred], round(score, 3)

# 전체 댓글에 대해 분석
results = []
for comment in tqdm(comments):
    label, confidence = predict_sentiment(comment)
    results.append((comment, label, confidence))

# 결과 저장
result_df = pd.DataFrame(results, columns=["comment", "sentiment", "confidence"])
result_df.to_csv("kobert_sentiment_result.csv", index=False)
```

# 논문 리뷰

- Paper 6.  
대중 매체에서 텍스트 마이닝 분석한 한국의 저출산 요인:  
네이버 뉴스 기사 댓글 중심으로
  - Dataset: 네이버 뉴스 댓글
  - Methodology: 토픽모델링, 감성분석
  - RQ: 현재 한국의 저출산 요인은 무엇인가?

# 논문 리뷰

- 연구동기
  - 저출산 문제의 심각성
- 기존 연구와의 차별성
  - 선행연구 – 저출산 현상에 대한 파급효과, 원인 규명, 정책 효과성 평가
  - 저출산에 대한 대중의 의견을 통하여 요인 규명
    - 분석 대상 데이터 측면에서 차별화

# 논문 리뷰

- 데이터 수집

- 네이버뉴스 - "저출산", "출산율" 포함 기사의 댓글 166,120건
- 저자의 석사 논문 활용

[25] 남가영, 텍스트마이닝을 활용한 저출산 요인 분석 네이버 뉴스 기사 댓글 중심으로, 전남대학교, 석사학위 논문, 2024.

- 데이터 전처리와 분석

- 명사 추출
- 축약어, 어근, 유의어, 지시어 정제
- TF-IDF
- LDA



# 논문 리뷰

- 분석 절차

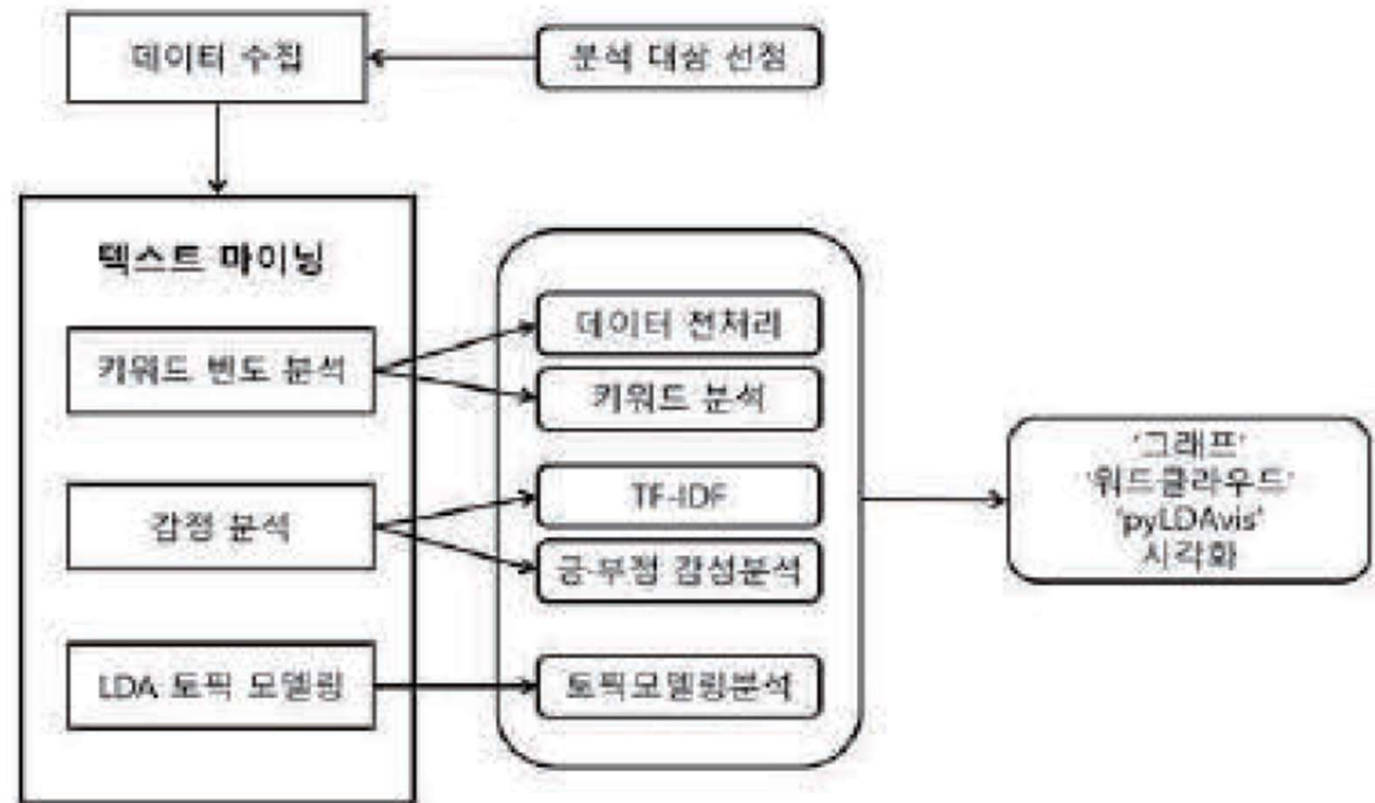


그림 3. 연구의 텍스트마이닝 분석 절차

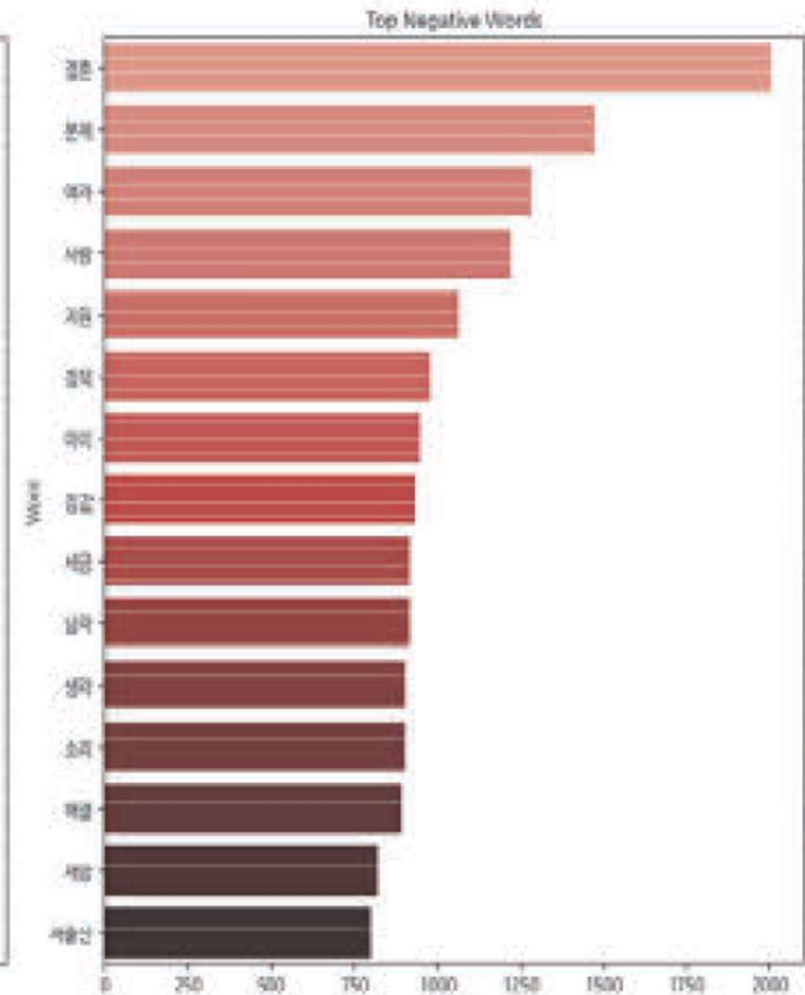
# 논문 리뷰

- 분석 결과
  - 키워드 빈도 분석

표 1. 키워드 빈도분석 상위 30개

상위 키워드					
1	결혼(18943)	11	사회(7513)	21	자식(5748)
2	아이(16852)	12	부모(7402)	22	국민(5666)
3	사람(15166)	13	국가(6750)	23	자녀(5550)
4	문제(12182)	14	저출산(6682)	24	거지(4666)
5	생각(11564)	15	여성(6440)	25	혜택(4260)
6	여자(9893)	16	집값(6325)	26	육아(4164)
7	정책(9736)	17	교육(6141)	27	소리(4019)
8	지원(8368)	18	정부(6133)	28	이상(3949)
9	인구(7979)	19	해결(5952)	29	현실(3466)
10	남자(7745)	20	세금(5882)	30	대책(3433)

- 분석 결과
  - 감성 분석



# 논문 리뷰

- 분석 결과
  - 토픽모델링 결과

표 2. LDA 토픽 구성 단어 목록과 토픽명

	토픽명	핵심 구성 단어	비중(%)
1	여론과 대중 심리	선동, 기사, 댓글, 수준, 관심, 정신, 마인드, 생각	8.0
2	사회 계층 불평등	기득권, 정치, 정치인, 부자, 서민, 거지, 노예, 이민	9.1
3	가족생활의 경제적 부담	결혼, 일자리, 육아, 도우미, 가사, 임금, 취업, 수도권	9.4
4	젠더간 갈등	군대, 국방, 페미, 남녀, 갈등, 비혼, 연애, 대책	10.9
5	세계 인구변화	인구, 중국, 일본, 한국, 감소, 소멸, 세계, 걱정	8.9
6	가족생활의 균형	근무, 집안일, 월급, 독박, 난임, 기관, 회사	6.9
7	사회적 지원	육아, 교육, 지원, 수당, 어린이집, 맞벌이, 임신, 현실, 혜택	18.0
8	주거 및 경제적 부담	집값, 아파트, 주거, 소득, 대출, 물가, 교육비, 연금	11.2
9	출산에 대한 인식변화	세대, 문화, 시대, 혼자, 자식, 인생, 인식, 환경, 본인	17.6

# 논문 리뷰

## • 분석 결과 해석

- 선행연구 FGI(Focus Group Interview) 조사 결과와의 비교

표 3. LDA 토픽 주제와 FGI 주제간의 비교

최영미외, 2019 (FGI, 38명)		본 연구
의지적 저출산	“개인이 스스로의 이익 추구에 따라 출산을 거부함으로써 출산 저하를 가져온 것”	Topic 9
숙명적 저출산	“타인이나 사회 공동체 이익에 스스로를 희생당하며 비출산을 받아들임으로써 출산 저하를 가져온 것”	Topic 9
사회적 저출산	“전 사회적 인식 변화로 인하여 사회질서가 심각하게 재적응해야 하는 상황에서 기존의 집단적 질서가 흔들리며 출산 저하를 가져온 것”	Topic 9
경제적 저출산	“국가경제에 대한 비관적 전망이나 결혼 생활을 유지하기 위한 경제적 비용 부담으로 인하여 출산 저하를 가져온 것”	Topic 3,8
정책적 저출산	“정부의 저출산 정책의 부재나 한계 혹은 정책 그 자체가 가져온 문제점으로 인하여 오히려 출산 저하를 가져온 것”	Topic 7
이타적 저출산	“인류와 세계의 복리증진을 위하여 출산을 하지 않는 것이 바람직하다는 인식하에 나타나는 비혼 및 비출산”	Topic 5
물리적 저출산	“출산 가능 인구의 신체적 어려움이나 물리적인 출산저하 요소들로 인하여 출산 가능성이 떨어져서 출산저하를 가져온 것”	Topic 6
차별적 저출산	“사회가 비혼에 의한 임신이나 장애인에 대한 출산을 암묵적으로 가로막음으로써 출산저하를 가져온 것”	-