



ISSN: 2586-7342 © 2022 KODISA
 KJFHC website: <http://acoms.kisti.re.kr/kjfhc>
 doi: <http://dx.doi.org/10.13106/kjfhc.2023.vol9.no1.19>.

Exploring Depression Research Trends Using BERTopic and LDA

Woo-Ryeong YANG¹, Hoe-Chang YANG²

Received : November 25, 2022. Revised: December 26, 2023. Accepted : February 09, 2023.

Abstract

The purpose of this study is to explore which areas have been more interested in depression research in Korea through analysis of academic papers related to depression, and then to provide insights that can solve future depression problems. 1,032 papers searched with the keyword “depression” in scienceON were analyzed using Python 3.7 for word frequency analysis, word co-occurrence analysis, BERTopic, LDA, and OLS regression analysis. The results of word frequency and co-occurrence frequency analysis showed that related words were composed around words such as patient, disorder and symptom. As a result of topic modeling, a total of 13 topics including 'childhood depression' and 'eating anxiety' were derived. And it has been identified as a topic of interest that 'suicidal thoughts', 'treatment', 'occupational health', and 'health treatment program' were statistically significant topics, while 'child depression' and 'female treatment' were relatively less. As a result of the analysis of research trends, future research will not only study physiological and psychological factors but also social and environmental causes, as well as it was suggested that various collaborative studies of experts in academia were needed such as convergence and complex perspectives for depression relief and treatment.

Keywords: Depression, Depression Disorder, Research Trend, BERTopic, LDA

Major Classification Code: Nursing and Mental Health, Public Health, Health Policy and Economy.

1. Introduction

건강은 인간이면 누구나 갖는 욕구 중 하나이다. 국어사전에서는 건강을 정신적으로나 육체적으로 아무 탈이 없고 튼튼함, 또는 그런 상태로 정의하고 있으며, 세계보건기구(World Health Organization: WHO)는 이를 확대하여 신체적·정신적·사회적으로 완전히 안녕한 상태(WHO, 2017)로 정의하고 있다. 이는 건강에 대한 관심이 과거에는 육체적·정신적인 면에 한정되었지만, 개인이 사회생활을 통해 받는 영향이 커지고 사회가 각 개인의 건강에 대해 기대하는 것이 많아짐에 따라

(Choi et al., 2020) 건강에 대한 관심이 사회적인 면까지 확대된 것이라 할 수 있다.

정신건강(mental health)은 개인이 자신의 능력을 깨닫고, 삶의 정상적인 스트레스에 대처할 수 있으며, 생산적으로 일할 수 있고, 지역사회에 기여할 수 있는 웰빙(well-being) 상태로 정의되며, 정신건강의 수준은 다양한 생물학적, 심리적, 사회적 요인에 의해 영향을 받는다(Choi et al., 2020; Kim & Lee., 2020). 따라서 정신건강은 개인의 삶의 질을 결정하는 중요한 요인일 뿐 아니라 직업적·사회적 기능을 결정하며, 경제적으로도 큰 영향을 미치는 중요한 요인 중 하나이다(Korniloff et

1 First Author, Ph.D. Candidate, Integrated Course of Master & Doctoral, Dept. of Business Informatics, Hanyang University, Republic of Korea. Email: wooryeong325@gamil.com

2 Corresponding Author, Assistant Professor, Dept. of Distribution Management, Jangan University, Republic of Korea. Email: pricezzang@jangan.ac.kr

© Copyright: The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

al., 2012).

본 연구에서는 정신건강에 심각한 문제를 유발하는 우울증에 주목하고 있다. 우울증은 정신장애 중 가장 유병률이 높은 질환으로, 초기에 잘 대처하면 쉽게 치료할 수 있으나 적절하게 대처하지 못하는 경우 정신건강 뿐 아니라 신체 건강에 심각한 영향을 미치고 장기간의 치료를 필요로 하는 만성질환이기 때문에 사회에 큰 부담을 줄 뿐 아니라 심하면 자살이라는 심각한 결과에 이를 수 있는 질환이다(Kessler et al., 2005). 이런 문제점을 해결하기 위해 정부에서는 국민 정신건강을 증진을 위한 다양한 사업을 진행 중이지만(MOH, 2021) 그 효과는 아직 만족스럽지 않은 상황이다.

건강보험심사평가원의 「2017~2021년 우울증과 불안장애 진료현황 분석」에 따르면 2021년 우리나라 우울증 환자 수는 93만 3,481명으로 2017년 대비 35.1% 증가하였으며, 이에 따른 1인당 진료비도 56만 4,712원으로 2017년 대비 28.5% 증가하였다고 보고하였다(HIRA, 2022). 이 보도자료에 의하면 2021년 20대가 17만 7,166명으로 전체 환자의 19.0%를 차지하고 있으며 2017년 대비 2021년에 127.1%로 연평균 22.8%가 증가하였고, 2021년 기준 여성은 63만 334명으로 남성(30만 3,147명)의 2.1배인 것으로 확인되었다(HIRA, 2022). 이런 결과에 따르면 20대와 여성의 우울증 증상이 폭발적으로 늘어나고 있음을 알 수 있다.

우울증에 대한 학계의 접근방법도 다양하다. 전통적인 심리학에서는 우울증을 치료의 대상으로 접근하고 있으며 의학계의 약물치료에 대해 심리치료(psychotherapy)를 이용하여 불안, 분노, 슬픔, 무기력과 같은 다양한 형태의 부정정서와 우울증과 같은 병리적 문제를 치료할 수 있다고 주장하고 있다(Barlow, 2008; Castonguay, 2013). 이들은 치료와 관련된 많은 사례를 제시하며 심리치료가 장기적으로 단독 약물치료보다 훨씬 더 효과적이라고 보고하고 있다(Leylin & DeRubeis, 2009; Siddique, 2012). 한편, 긍정심리학자들은 지금까지의 심리치료가 대부분 개인의 결점에 중점을 두고 있기 때문에 한계가 있다고 주장하며, 강점과 긍정적으로 정서화된 감정과 같은 긍정적 측면을 통해 회복력을 발현시키고 증상을 완화할 수 있다고 주장하고 있다(Rashid & Seligman, 2018; Seligman & Csikszentmihalyi, 2014).

본 연구에서는 지금까지 우울증과 관련된 학술논문 분석을 통해 우리나라 우울증 연구가 어떤 분야에 관심을 갖고 있는지를 탐색한 후 향후 우울증 문제를 해결할 수 있는 다양한 연구 방향에 대한 통찰력을 제공하고자 하는 것이 주요 목적이다. 이를 위해 본 연구에서는 scienceON(<https://scienceon.kisti.re.kr>)에서 키워드 ‘우울증’과 관련된 총

1,848편의 논문 중 중복 논문을 제외한 1,768편의 논문을 추출하였으며, 이 중에서 영문 초록이 탑재된 1,032편의 논문을 대상으로 단어빈도분석, 워드클라우드 및 단어 동시출현 빈도 분석을 실시하고 머신러닝의 비지도학습법으로 분류되는 BERTopic과 LDA를 통해 토픽을 도출한 후, 각 토픽에 대한 연도별 트렌드분석을 실시하였다.

본 연구 결과는 우울증에 관심이 있는 학자들에게 어떤 영역에 대한 연구가 수행되어야 하는지에 대한 단서의 제공과 함께 우울증 치료를 위해 어떤 접근법이 더 관심있게 연구되고 있는지와 같은 다양한 분야의 단서를 제공해줄 것으로 기대하였다.

2. Theoretical Background

2.1. 우울증(Depression or Depression Disorder)

우울증은 전 세계적으로 1억 3천만명이 넘는 사람들에게 영향을 미치는 흔한 장애로 최근 일반인을 대상으로 실시한 역학 조사에 따르면 우울증의 평생 유병률은 10~15%에 이른다(Lépine & Briley, 2011). Lépine & Briley(2011)는 우울증이 그 증상 자체로도 개인의 건강에 부정적인 영향을 미치지만 성공적으로 치료되는 경우에도 인지 장애나 사회적 기능 장애 등 잔여 증상으로 인해 상당한 고통을 유발하며, 재발 위험으로 인해 삶의 질을 크게 저하시킨다고 하였다(Lépine & Briley, 2011; Solomon et al., 2000).

우울증은 정상적인 생각과 행동을 하지 못하도록 만드는 모든 감정상태로 정의된다(Clark, 1995). 따라서 우울증이나 우울장애는 일시적으로 기분만 저하된 상태를 뜻하는 것이 아니라 여러 내외적 원인으로 인해 정상적인 활동을 방해하는 정신적인 낙담과 절망의 상태로 생각의 내용, 사고과정, 동기, 의욕, 관심, 행동, 수면, 신체활동 등 전반적인 정신기능이 저하된 상태를 뜻한다(Seung et al., 2022).

우울증의 다양한 관점은 심리학적 이론을 통해서도 확인할 수 있다. 먼저, 정신분석이론에서 Freud(1917)는 대부분 우울 증세가 배우자, 직업, 건강과 같은 자신이 아끼고 사랑하는 대상의 상실과 같은 외부적 요인에 의해 발생하며, 우울 증세를 겪게 되면 자기애의 추락과 함께 징벌에 대한 망상적 기대가 발생한다고 하였다(Freud, 1917). 정신분석이론에서는 이런 질책과 비난과 같은 징벌에 대한 망상적 기대가 실제 대상에게 적용되고 대상에 대한 자아의 복수심을 나타내는 것이라고 하였다(Freud, 1986). 대상관계이론에서 Klein(1946)은 사람은 누구나 편집-분열적 자리를 거쳐 우울적 자리라는 심

리단계로 가는 발달적 기제를 거치는데 과제 수행에 실패하게 되거나, 상실이라는 사건을 경험할 때마다 애도과정을 충분히 얻지 못하고 반복적으로 우울증을 겪는다고 하였다(Klein, 1946). 또한 대상을 사랑하거나 사랑받고자 하는 인간의 근본적인 욕구가 좌절되어 자아가 손상을 입는 경우 사람들은 나쁜 내적 대상들에게 애착과 충성심을 형성하며 이들의 나뻘을 자기 스스로 떠맡음으로써 실제 세상과 직면하지 않으려는 특징이 있다(Woods & Mitchel, 1996). 행동주의 이론에서 Maier & Seligman(1976)은 학습된 무기력을 통해 우울증이 강화물의 상실이 아니라 강화물을 통제할 수 있는 능력의 상실에 의해 발생한다고 하였으며, McMinn & Wade(1995)는 우울증에서 나타나는 불행감이나 행동의 감소가 긍정적 강화의 결핍이나 혐오스럽고 불쾌한 경험의 증가로 인한 결과이고, 자신의 행위에 대한 보상이 따르지 않는 환경적 요인이 우울증을 강화하는 요인이라고 주장하였다(McMinn & Wade, 1995). 인지이론에서 Beck(1967)은 우울증의 원인은 잘못된 추측과 가정으로 현실을 왜곡하기 때문에 발생하는 것이라 하였으며, 이런 부정확한 개념은 주로 인지발달 과정에서 나타나는 잘못된 학습에서 시작된 것이라고 하였다(Beck, 1967). 또한 그는 임의적 추론, 선택적 추상화, 과장과 축소, 과잉일반화, 흑백논리와 이분법적 사고, 개인화와 같은 잘못된 정보처리로 인해 우울증이 발생한다고 하였다(Beck, 1979).

우리 사회에서도 우울에 대한 논의는 최근 10년 동안 급속히 확산되었고, 대중매체에서도 청년 문제, 노인 문제, 자살 등 여러 사회 문제의 주요한 배경과 원인으로 우울이 지목되었다. 우울증은 다양한 정신적 상태뿐 아니라 불안장애, 강박장애, 약물 남용, 성격장애, 정신증, 인지 결핍, 치매, 파킨슨, 비만, 성기능 장애, 통증 등과 함께 나타나며, 망상과 환각이 나타나는 정신병적 우울증부터 정신병적 양상이 없는 신경증적 우울증까지를 모두 포함하는 것으로 보고되고 있다(Lee et al., 2008).

보건복지부는 2022년 6월 「코로나19 국민 정신건강 실태조사」에서 우리나라 우울 위험군은 16.9%로 2019년 3.2%의 5배가 넘는 수치로 여전히 높고 위험한 수준이라고 보도하였다(MOHW, 2022). 주요 조사 결과를 보면 연령별로는 지속해서 30대가 가장 높았으며(24.2%), 40대(17.0%), 50대(16.0%), 20대(14.3%), 60대(13.0%) 순이며, 여성이 18.6%로 남성(15.3%)보다 3.3%p 높다고 하였다(MOHW, 2022). 한편, 소득이 감소한 경우의 우울 위험군이 22.1%로 소득이 증가하거나 변화가 없는 집단(11.5%)에 비해 2배 가까이 높게 나타났으며, 1인 가구의 우울 위험군이 23.3%로 2인 이상 가구(15.6%)에 비해 높았으며, 배우자가 없는 경우가 20.6%로 기혼(14.3%)에 비해 높다

고 하였다(MOHW, 2022). 이런 결과로 볼 때 최근 우리 사회가 오랜 불경기로 인한 가처분소득 감소, 1인 가구의 증가, 이혼율 증가 등의 이유만으로도 우울 위험이 증가할 가능성이 높음을 알 수 있다.

2.2 토픽모델링(Topic Modeling)

자연어처리(NLP: Natural language processing)는 정보 관리, 의미론적 마이닝 및 컴퓨터가 텍스트 문서에서 인간 언어 처리로부터 의미를 얻을 수 있도록 하는 컴퓨터 과학의 도전적인 과제이다(Jelodar et al., 2019). 토픽모델링은 데이터 마이닝, 잠재 데이터 발견, 데이터와 텍스트 문서 간의 관계 찾기를 위한 텍스트마이닝에서 가장 강력한 기술 중 하나이다(Yang, 2022). NMF(Non-Negative Matrix Factorization)(Févotte and Idier, 2011)나 LDA(Blei et al., 2003)는 텍스트마이닝에서 다양한 장점이 있지만 최적의 결과를 얻기 위해 토픽의 수, 불용어 목록, 형태소 분석이 필요하고 단어 간의 의미관계가 무시된다는 약점을 가지고 있기에, 자연어처리 분야에서는 텍스트 임베딩(text embedding) 기술이 보급되고 있다(Yang, 2022).

2018년 Google이 공개한 BERT는 입력데이터를 양방향에서 접근하여 문맥을 파악할 수 있고, 미세조정과 사전학습이 가능한 비지도 학습모델이다(Kim & Yang, 2022). BERTopic은 이전의 복합 토픽 모델링(Bianchi et al., 2020; Devlin et al., 2018)과는 다르게 BERT를 통해 사전학습된 언어모델 데이터를 호출해서 BoW(Bag of Word)의 요소를 제거하고 사전 학습된 언어모델을 적극적으로 활용하는 기법(Devlin et al., 2018)으로 최근 학계에서 주목을 받고 있다(Yang, 2022). 본 연구에서 활용한 BERTopic은 상황에 맞는 단어 및 문장 벡터 표현을 생성하는 것으로 이러한 벡터 표현의 의미론적 속성은 유사한 텍스트가 벡터 공간에서 가까운 방식으로 텍스트의 의미를 인코딩할 수 있도록 한다(Grootendorst, 2022). 클러스터링 기술과 TF-IDF의 클래스 기반 변형을 활용하여 일관된 주제 표현을 생성하는 BERTopic은 텍스트 임베딩 단계에서 BERT를 활용한 임베딩과 c-TF-IDF 단어 가중치를 활용한 다음, 각 도메인에 맞는 텍스트 클러스터링을 하여 텍스트에 잠재된 의미있는 주제를 찾아내는 토픽모델링 기법이다(Ko et al., 2022).

LDA(Latent Dirichlet allocation)는 정형 데이터뿐만 아니라 비정형 데이터에서도 활용도가 높아 현재까지 토픽모델을 활용한 연구에서 가장 많이 활용되고 있는 토픽모델링 기법이다(Blei et al., 2003). LDA 기반의 토픽모델링은 자연어 처리, 텍스트마이닝 및 소셜미디어 분석과 정보 검색에 적용되었으며 이산 데이터를 시연하는데 탁월할 뿐 아니라 엄청난 정보에

숨겨진 구조와 의미를 찾기 위한 효과적인 접근을 제공하는 특성이 있다(Jelodar et al., 2019; Yang, 2022). LDA는 말뭉치(corpus)의 생성확률 모델로 기본적인 아이디어는 문서가 잠재적인 토픽에 대해 무작위 혼합으로 포함되어 있으며, 토픽은 단어에 대한 분포로 구성된다는 것이다(Jelodar et al., 2019). LDA는 Blei et al. (2013)에 의해 처음 도입되었다. LDA는 단어 확률로 토픽을 나타내기 때문에 각 토픽에서 가장 높은 확률을 가진 단어는 일반적으로 LDA 토픽으로부터 단어확률이 무엇인지에 대한 좋은 아이디어를 제공하는 특성이 있다(Blei et al., 2003). LDA 매개변수(parameter)를 측정하기 위한 방법으로는 변이 방법(variational method)(Blei et al., 2003), 기대 전파(expectation propagation)와 Gibbs 샘플링(Griffiths & Steyvers, 2004)과 같은 방법들이 제안되었다.

본 연구에서는 BERTopic이 다양한 장점이 있음에도 불구하고 임베딩 특성상 기존의 평가법으로는 BERTopic의 성능을 평가하는 것이 불가능하다는 점(Bodrunova et al., 2020).에도 주목하고 있다. 실제로 Yang(2022)과 Yang & Yang(2022)의 연구에서 보면 BERTopic을 통해 도출한 토픽 수가 예상보다 훨씬 더 많아 이를 축소할 필요가 있었으며, BERTopic에서 채택되지 못한 이상치 문서들이 다수 발생하여 이를 고려하지 않는 것은 실제 연구동향을 확인하고자 하는 목적에 부합하지 않는다는 문제가 지적되었다. 따라서 본 연구에서는 Yang(2022)와 Yang & Yang(2022) 뿐 아니라 Ko et al.(2022), Kim & Yang(2022)이 적용한 것처럼 BERTopic과 LDA를 함께 활용함으로써 연구동향 분석이라는 연구목적에 적합하게 접근하고자 하였다.

Table 1: Number of publications per period

Period division	1966~1989	1990~1999	2000~2009	2010~2019	2020~	Sum
Number of publications	9	113	620	824	202	1,768

3.2. 데이터 전처리(Data Preprocessing)

데이터 전처리는 의미없는 단어나 불용어 등을 삭제함으로써 토픽을 좀 더 정교하게 도출하는 사전적 과정이다. 본 연구에서는 영문초록에 포함되어 있는 특수기호, 숫자와 함께 초록에 빈번하게 활용되고 있는 purpose, conclusion, aim, methology 등과 correlation, regression, SPSS, AMOS, coefficient 등 통계와 관련된 용어를 불용어로 제거하였으며, depression도 제거하여 우울증과 관련된 주변 단어들로 구성된 토픽을 도출하고자 하였다.

3. Research Procedure

본 연구는 우울증 연구 동향 확인을 위해 다음과 같은 연구절차를 통해 수행하였으며, Python 3.7을 활용하여 분석을 수행하였다.

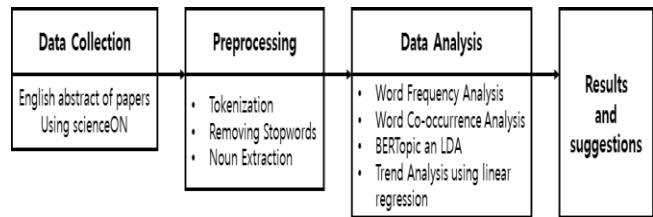


Figure 1: Research Procedure

3.1 데이터 수집(Data Collection)

2022년 10월 15일자 기준 scienceON (<https://scienceon.kisti.re.kr>)을 활용하여 “우울증”으로 검색한 결과 총 1,848편의 논문이 도출되었으며 이 중 중복 논문을 제외한 1,768편 중 영문초록 이 탑재된 1,032편의 논문을 획득하여 분석에 활용하였다. 전체 논문에 대한 기간별 발간 편수는 아래 <Table 1>과 같다. <Table 1>에서 보는 바와 같이 우울증에 대한 논문은 기간이 지날수록 늘어남을 알 수 있어 우울증이 우리 사회에 큰 문제로 작용하고 있다는 건강보험심사평가원의 보도와 보건복지부의 실태조사를 반영하고 있는 것으로 보인다.

4. Empirical Analysis Results

4.1. 단어빈도분석 결과(Results of Word Frequency Analysis)

단어 빈도분석 결과 patient(1,353), symptom(823), disorder(808), health(770), treatment(712) 등의 단어들 이 빈번하게 등장하고 있어 우울증 또는 우울장애와 관련된 치료, 우울증에 영향을 미치는 부정적인 정서 등이 연구의 관심 키워드인 것을 알 수 있었다. 빈도분석 결과 상위 20개 단어와 워드클라우드링 결과는 <Table 2>와 <Figure 2>에 제시하였다.

Table 2: Results of Word Frequency Analysis

No	Word	Freq.	No	Word	Freq.	No	Word	Freq.	No	Word	Freq.
1	patient	1353	6	method	524	11	anxiety	389	16	problem	309
2	symptom	823	7	level	521	12	woman	388	17	age	302
3	disorder	808	8	life	465	13	program	370	18	activity	287
4	health	770	9	stress	444	14	control	369	19	quality	282
5	treatment	712	10	difference	417	15	relationship	340	20	disease	280

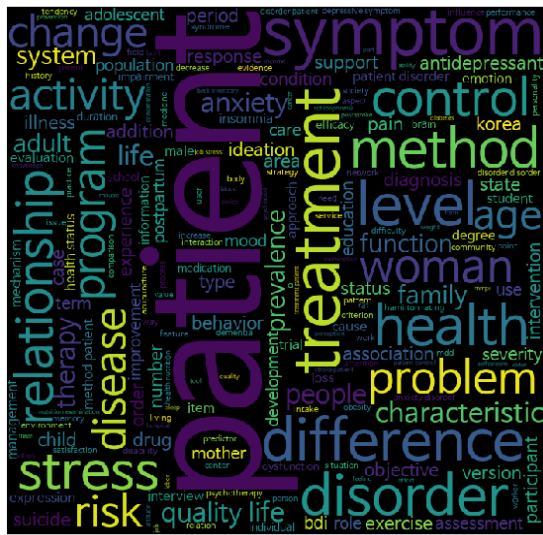


Figure 2: Result of Word Clouding

4.2 단어 동시출현 빈도분석 (Results of Word Co-occurrence Frequency Analysis)

단어 간의 관계 확인을 위해 동시출현하는 단어 빈도를 카운트하는 워드네리를 구성한 후 연결중심성 등을 고려한 단어 동시출현빈도를 분석한 결과 method-patient(205), disorder-patient(193) 등이 주로 연관되어 있으며, 전반적으로 볼 때 patient, disorder, method, symptom 등의 단어들 중심으로 연관되는 단어들 구성되어 있음을 알 수 있었다. 즉, 우울증 연구는 환자와 관련된 처방 등 방법론과 증상에 대한 연관 연구가 많은 것으로 해석할 수 있다. 동시출현 빈도 순으로 상위 10개 쌍의 단어들과 연결빈도 60회 이상으로 시각화된 결과는 <Table 3>과 <Figure 3>에 제시하였다.

Table 3: Results of Word Co-occurrence Frequency Analysis

Rank	Word	Freq.	Rank	Word	Freq.
1	method	205	6	method	144
	patient			symptom	
2	disorder	183	7	disorder	138
	patient			method	
3	patient	171	8	method	122
	symptom			treatment	
4	patient	169	9	disorder	120
	treatment			treatment	
5	method	156	10	disorder	115
	objective			symptom	

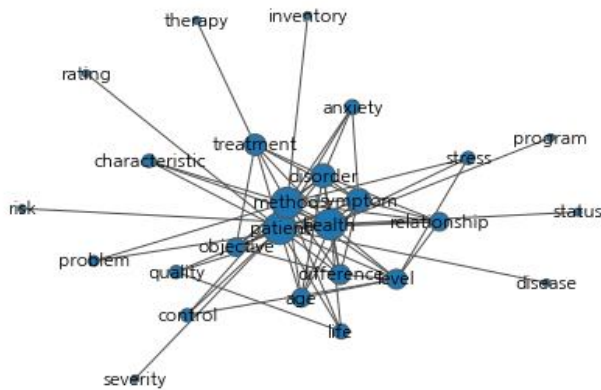


Figure 3: Visualization result for co-occurrence frequency (connection frequency = 60 times)

4.3 토픽모델링 결과(Results of Topic Modeling)

BERTopic 결과 총 20개의 토픽과 358편의 이상치 문서가 도출되었으나, 각 토픽의 편수를 30편 이상으로 제한하여 총 10개의 토픽과 616편의 이상치 문서를 추출하였다. BERTopic에서 도출된 10개 토픽은 다음과 같다. 먼저 <토픽 1>은 child, childrens, childhood, adhd 등의 단어가 연관되어 아동과 부모(또는 모)와 관련된 우울문제일 것으로 판단하여 이를 ‘아동 우울증(childhood depression)’이라고 명명하였다. <토픽 2>는 health, intake, food, nutrition, insecurity 등의 단어로 구성되어 섭식과 관련된 문제와 관련이 있다고 판단하여 이를 ‘섭식 불안(eating anxiety)’이라고 명명하였다. <토픽 3>은 rat, mouse, immobility, bdnf 등의 단어로 구성되어 동물을 이용한 스트레스 등의 실험과 관련이 있을 것으로 판단하고 이를 ‘동물 실험(animal testing)’이라고 명명하였다. <토픽 4>는 anxiety, disorder, hrv와 같은 단어로 구성되어 환자들의 증상과 관련이 있을 것으로 판단하고 이를 ‘불안 증상(anxiety symptoms)’으로 명명하였다. <토픽 5>는 antidepressant, treatment, drug 등의 단어로 구성되어 우울증에 대한 약물 치료와 관련이 있을 것으로 판단하고 이를 ‘약물 처치(medication)’라고 명명하였다. <토픽 6>은 suicide, ideation, suicidal 등의 단어로 구성되어 우울증의 대표적인 부정적 행동인 자살과 관련이 있을 것으로 판단하고 이를 ‘자살 충동(suicidal thoughts)’이라고 명명하였다. <토픽 7>은 acupuncture, treatment, method, therapy와 같은 단어로 구성되어 ‘치료와 처치(treatment)’라고 명명하였다. <토픽 8>은 job, stress, worker, labor, firefighter와 같은 단어들로 구성되어 직무와 관련된 스트레스나 증상으로 판단하고 이를 ‘직무 건강(occupational health)’이라고 명명하였다. <토픽 9>는 melancholy, novel, hamlet, desire, queer 등의 단어로 구성되어 소설 등에서

등장하는 우울증에 대한 연구로 판단하고 이를 ‘우울 담론(melancholy discourse)’이라고 명명하였다. <토픽 10>은 role, adult, health, elder 등의 단어로 구성되어 성인들의 다양한 증상과 관련이 있는 것으로 판단하고 이를 우울증과 관련이 있는 ‘성인 증상(adult symptoms)’으로 명명하였다.

Table 4: Results of Topic Modeling

Topic	Topic Naming	Top 10 Keywords by Topic
1	Childhood Depression	child , mother, disorder , childrens , problem , adhd , parent, childhood , symptom
2	Eating Anxiety	health , intake , food , status, prevalence, nutrition , vitamin, examination, association, insecurity
3	Animal Testing	rat , expression, cm, mouse , immobility , level, stress , control, hippocampus, bdnf
4	Anxiety Symptoms	anxiety , patient, disorder , symptom , hrv , trait, state, inventory, bdi, pressure
5	Medication	antidepressant , patient, treatment , drug , efficacy, ssri, fluoxetine, paroxetine, serotonin, inhibitor
6	Suicidal Thoughts	suicide , attempt, ideation , cholesterol, suicidal , patient , crime, level, history, risk
7	Treatment	acupuncture , treatment , medicine, patient, method, therapy , temperature, aromatherapy , part, postpartum
8	Occupational Health	job , stress , worker , labor , firefighter, work, symptom, covid19, student, health
9	Melancholy Discourse	melancholy , novel , hamlet , desire , yadam, discourse, queer , loss, ego, man
10	Adult Symptoms	satisfaction, support, role , symptom , adult , health , woman, gender, relationship, elder
11	Student Treatment Program	patient , program , control, treatment , disorder, difference, stress, method, health, student
12	Treatment in Women	disorder , symptom , woman , treatment , patient , activity, method, health, life, level
13	Health Treatment Program	patient, health , treatment , life , level, symptom, therapy , method, disorder, program

Note: Bold type indicates the keywords used to determine the topic name. adhd (attention deficit hyperactivity disorder: ADHD), bdnf (Brain-derived neurotrophic factor: BDNF), hrv (Heart Rate Variability·HRV).

616편의 이상치 문서를 대상으로 LDA를 실시한 결과 일관성(Coherence) 지수(0.2947)가 3개 토픽일 때 가장 높은 것으로 나타나 3개 토픽을 추가하였다. <토픽 11>은 patient, program, treatment, student, 등의 단어로 구성되어 학생들의 우울증 증상과 함께 이에 대한 다양한 치료프로그램에 대한 연구로 판단하고 이를 ‘학생 치료 프로그램(student treatment program)’으로 명명하였다. <토픽 12>는 disorder, symptom, woman, treatment, patient 등이 단어로 구성되어 있어 여성들이 생활속에서 겪는 우울증세에 대한 치료와 관련이 있을 것으로 판단하고 이를 ‘여성 치료(treatment in women)’로 명명하였다. <토픽 13>은 health, treatment, life, therapy, program 등의 단어로 구성되어 있어 우울증세와 관련된 건강치료와 관련이 있을 것으로 판단하고 이를 ‘건강치료 프로그램(health treatment program)’으로 명명하였다. 토픽모델링 결과와 상위 10개 토픽

키워드는 <Table 4>에 제시하였다.

4.4 각 토픽에 대한 트렌드분석 결과(Trend Analysis Results of Each Topic)

토픽모델링을 통해 특정 논문에 할당된 각 토픽의 비율을 산출하고 산출된 비율과 각 논문의 발행연도를 활용하여 각 토픽의 연도별 평균비율을 계산하면 토픽의 트렌드를 확인할 수 있는데 이런 분석방법을 동적 토픽모델링이라고 한다(Yang, 2022; Yang & Yang, 2022). 동적 토픽모델링은 시간에 따라 토픽 비율이 유동적이기 때문에 각 토픽이 시간에 따라 등장하는 것을 추적할 수 있는 장점이 있다(Blei & Lafferty, 2006; Yang, 2022). 각 토픽에 대한 트렌드 분석 결과는 <Figure 4>에 시각화하였다.

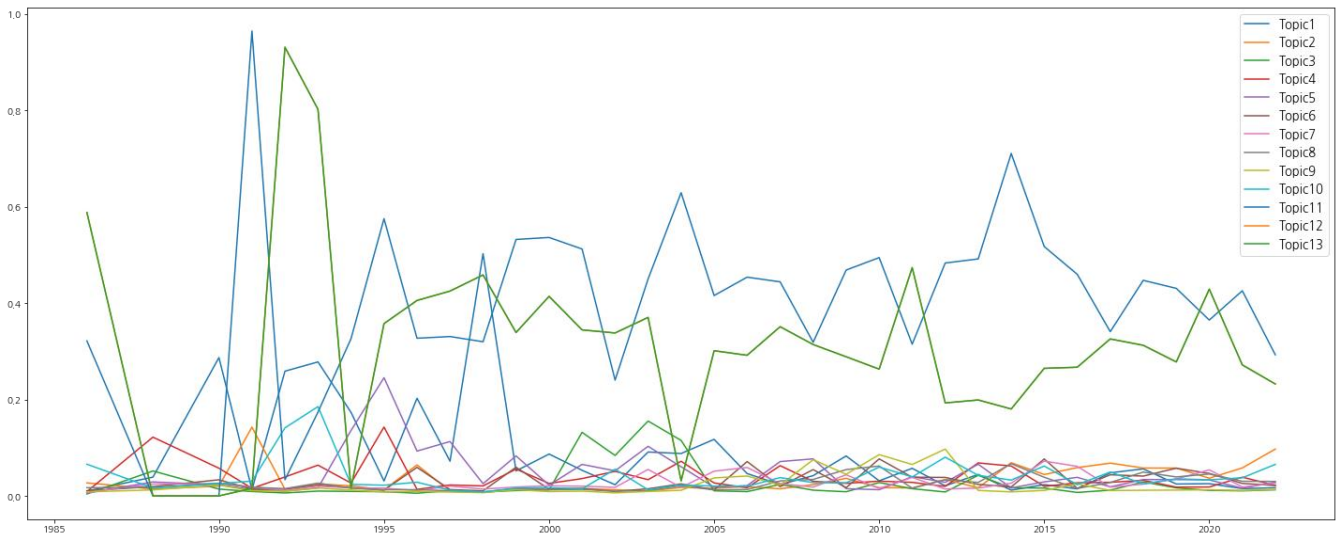


Figure 4: Visualization Results of Trend Analysis

<Figure 4>의 결과는 1990년부터 1995년 사이에 <토픽 1>의 ‘아동 우울증’과 <토픽 13>의 ‘건강치료 프로그램’과 같은 연구가 빈번하였다는 등 각 토픽이 연도별로 얼마나 유동적인지에 대해서는 확인할 수 있으나 시간의 흐름에 따라 어떤 토픽이 더 주목받고 있는지에 대해서는 확인하는데 한계가 있다(Yang, 2022; Yang & Yang, 2022).

따라서 본 연구에서는 Yang(2022)과 Yang & Yang(2022)이 적용한 분석방법처럼 독립변수를 논문의 발행년도로 설정하고, 종속변수는 해당 연도 토픽의 평균비율로 설정한 후 OLS 회귀분석을 실시하였다(Yang, 2022; Yang & Yang, 2022). 회귀분석 결과가 통계적으로 유의미한 양(+)의 계수이면 관심도가

높은 토픽(Hot topic)으로, 음(-)의 계수이면 관심도가 낮아지는 토픽(Cold topic)으로 해석할 수 있으며, 통계적으로 유의하지 않는 경우에는 중립토픽(Neutral topic)으로 결론내릴 수 있다(Griffiths & Steyvers, 2004; Yang, 2022; Yang & Yang, 2022).

연도별 토픽에 대한 회귀분석 결과 <토픽 6>의 ‘자살 충동’(coeff.=0.0007, $p<0.5$), <토픽 7>의 ‘치료와 처치’(coeff.=0.0007, $p<0.5$), <토픽 8>의 ‘직무 건강’(coeff.=0.0008, $p<0.01$)과 <토픽 13>의 ‘건강치료 프로그램’(coeff.=0.0056, $p<0.5$)이 95% 신뢰도 수준에서 상대적으로 주목받고 있는 주제로 나타났으며, 비록 95% 신뢰도 수준에서는 통계적으로 유의하지 않았지만 90% 신뢰도 수준에서 <토픽 2>의 ‘섭식

불안'(coeff.=0.0009, p=0.065)도 관심있는 주제인 것으로 나타났다. 한편 <토픽 1>의 '아동 우울증'(coeff.=-0.0045, p=0.01)은 시간이 지남에 따라 상대적으로 덜 주목받고 있는 것으로 확인되었으며 <토픽 12>의 '여성 치료'(coeff.=-0.0064, p=0.051)도 90% 신뢰도 수준에서 연구자들의 주목을 덜 받는 것을 알 수 있었다. 그러나 이런 결과는 '아동 우울증'이나 '여성 치료'가 연구자들의 관심분야에서 벗어나고 있다고 해석하기 보다는 '치료와 처치', '건강치료 프로그램'과 관련된 주제와 통합되고

있다고 해석하는 것이 바람직할 것으로 보인다. 이는 우리 사회가 어떤 세대나 계층에 한정해서 우울증세가 발현되는 것보다 상황과 환경적인 요인으로 인해 상대적 고통을 느끼는 경우에도 우울증세를 더 보인다는 보건복지부의 보도 결과와 같은 맥락으로 판단되기 때문이다.

회귀분석 결과는 < Table 5>에 제시하였으며 이를 기반으로 시간에 지남에 따라 Hot, Cold 토픽에 대한 시각화 결과는 < Figure 5>에 제시하였다.

Table 5: Results of OLS regression analysis on topics by year

Topic	Topic Naming	Coefficient	T-value	P-value	Trend
1	Childhood Depression	-0.0045	-2.864	0.007	Cold
2	Eating Anxiety	0.0009	1.906	0.065	(Marginally Hot)
3	Animal Testing	-0.0002	-0.298	0.768	-
4	Anxiety Symptoms	-0.0008	-1.680	0.102	-
5	Medication	-0.0010	-1.327	0.194	-
6	Suicidal Thoughts	0.0007	2.299	0.028	Hot
7	Treatment	0.0007	2.593	0.014	Hot
8	Occupational Health	0.0008	3.919	0.000	Hot
9	Melancholy Discourse	0.0005	1.316	0.197	-
10	Adult Symptoms	-0.0003	-0.501	0.620	-
11	Student Treatment Program	0.0012	0.407	0.687	-
12	Treatment in Women	-0.0064	-2.033	0.051	(Marginally Cold)
13	Health Treatment Program	0.0056	2.422	0.021	Hot

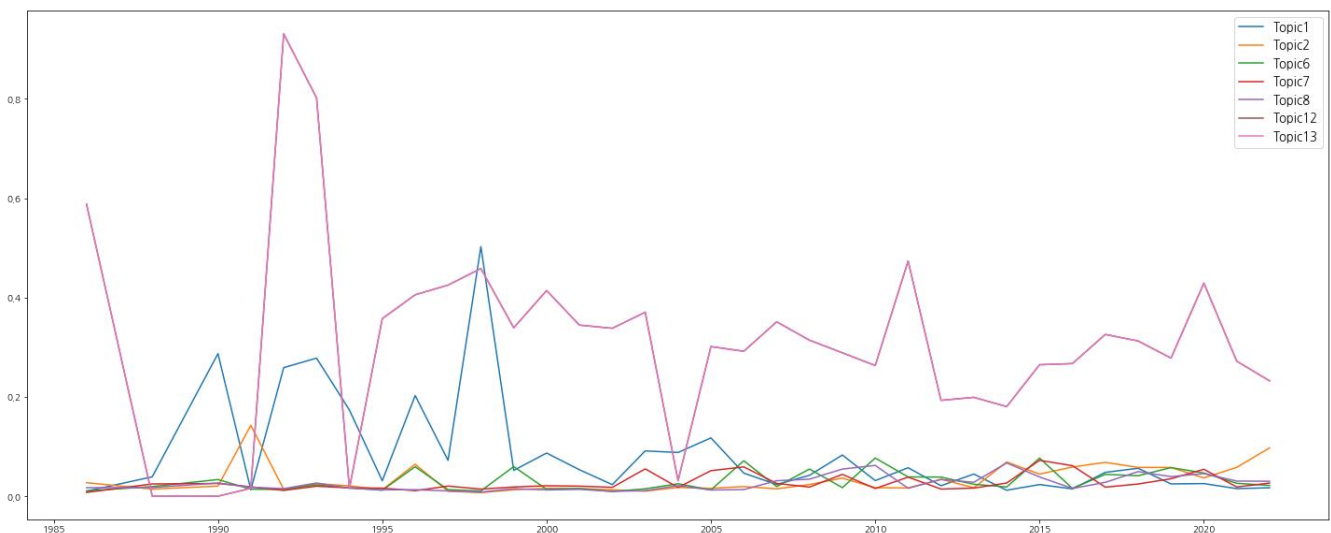


Figure 5: Visualization Results of Hot & Cold Topic

5. Conclusions

본 연구의 목적은 최근 우리 사회의 부담으로 대두된 우울증과 관련된 학술논문 분석을 통해 지금까지 우리나라 우울증 연구가 어떤 분야에 더 관심을 가졌었는지를 탐색한 후 향후 우울증 문제를 해결할 수 있는 통찰력을 제공하고자 하는 것이다. 이를 위해 추출한 1,032편의 논문을 대상으로 분석을 실시한 결과와 시사점은 다음과 같다.

첫째, 단어 빈도분석 결과 patient, symptom, disorder 등의 단어가 빈번하게 등장하였으며, 동시출현 빈도분석 결과도 patient, disorder, method, symptom 등의 단어들을 중심으로 연관 단어가 구성되어 있는 것으로 나타났다. 이런 결과는 우울증 또는 우울장애와 관련된 치료와 우울증에 영향을 미치는 부정적인 정서 등이 연구의 관심키워드이며 환자의 건강을 위한 방법론, 증상, 처치와 관련된 연구가 빈번하게 이루어진 것이라 할 수 있다. 둘째, BERTopic 결과 ‘아동 우울증 (childhood depression)’, ‘섭식 불안(eating anxiety)’, ‘동물 실험 (animal testing)’과 같은 10개의 토픽이 도출되었으며, BERTopic에서 이상치로 분류된 616편을 대상으로 LDA를 실시한 결과 ‘학생 치료 프로그램(student treatment program)’, ‘여성 치료 (treatment in women)’ 등 3개의 토픽이 도출되었다. 이런 결과는 지금까지 연구자들이 우울과 관련된 다양한 분야의 연구를 수행한 것으로 판단할 수 있다. 그러나 우울증은 심리적, 사회적, 신체적, 생물학적 원인 등 다양한 것으로 알려져 있고, 이에 대한 해결방안도 약물치료, 전기경련요법 (Electroconvulsive therapy), 반복적 경두개 자기자극법 (Repetitive transcranial magnetic stimulation; rTMS), 심리요법 등의 의학적 방법뿐 아니라 심리학, 예·체능계 등의 프로그램을 활용한 다양한 방안이 제시되고 있기 때문에 연구의 다양성 차원에서 접근할 필요가 있다. 셋째, 각 토픽에 대한 트렌드 분석 결과 ‘자살 충동’, ‘치료와 처치’, ‘직무 건강’, ‘건강치료 프로그램’이 통계적으로 유의하게 주목받고 있으며, ‘아동 우울증’, ‘여성 치료’ 등이 상대적으로 덜 주목받는 것으로 나타났다. 이런 결과는 ‘아동 우울증’, ‘여성 치료’ 등이 관심주제에서 벗어났다고 판단하기보다는 통합적 관점에서 ‘치료와 처치’, ‘건강치료 프로그램’ 분야에 흡수되고 있는 것으로 해석하여야 할 것이다. 그러나 우울증 진단기준(DSM-IV)의 측정 척도를 검토해보면 기분, 관심, 불면증, 피곤, 낮은 자존감과 같은 생리·심리적 요인뿐 아니라 사건, 사고와 같은 사회·환경적 원인에 대한 연구도 필요할 것으로 판단된다. 또한, 치료와 처치에 있어 의학계, 정통 심리학계의 관점뿐 아니라 인지행동 미술 치료, 음악치료, 체조치료 등 예·체능계의 치료프로그램과 최

근 대두되고 있는 긍정심리학의 긍정심리치료 등 다양한 분야에서 우울증상 완화와 치료에 효과적인 방안을 제시하고 있기 때문에 우울증 완화와 치료를 위한 융·복합적 관점을 적용한 전문가들의 다양한 협업연구가 수행될 필요가 있다.

본 연구는 지금까지 우울증과 관련된 우리나라 연구자들의 관심 주제를 확인하고 향후 연구되어야 할 방향에 대한 제안을 제시하였지만 몇 가지 한계점이 있어 향후 연구에서는 이를 보완할 필요가 있다. 첫째, 토픽모델링을 수행한 많은 연구들(예: Yang, 2022, Yang & Yang, 2022)의 공통적인 한계점인 토픽명 결정의 한계를 들 수 있다. 이는 해당 분야의 전문가들이 투입됨에도 불구하고 분류된 주제의 구체적인 내용을 모두 반영할 수 없기 때문이다. 따라서 향후 연구에서는 토픽명 결정을 위한 기준과 방법에 대한 연구가 필요할 것이다. 둘째, BERTopic에서 도출된 토픽외 이상치 문서들에 대해 LDA를 통해 추가 토픽을 추출하는 것이 아직까지는 실험적이라는 점을 들 수 있다. 본 연구에서 언급한 것처럼 50% 이상의 이상치 자료를 검토하는 것이 연구동향 파악을 위해 필요한 절차라고 판단하고 있지만 향후 연구에서는 이런 방법의 효과성과 각각 단일 분석을 통한 효과성의 차이를 확인할 필요가 있다. 마지막으로 연구동향 분석이 갖는 한계점으로 동향 파악을 통해 향후 방향을 제안하기는 했지만 구체적이지 못하다는 문제를 들 수 있다. 따라서 향후 연구에서는 동향을 파악하고 인과모형을 구축하는 등 좀 더 구체적인 단서들을 제공할 수 있는 방안에 대한 접근이 필요할 것이다.

References

- Barlow, S. H. (2008). Group psychotherapy specialty practice. *Professional Psychology: Research and Practice*, 39(2), 240-244.
- Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York: Hoeber Medical Division.
- Beck, A. T. (1979). *Cognitive therapy and the emotional disorders*. Penguin.
- Bianchi, F., Terragni, S., Hovy, D., Nozza, D., & Fersini, E. (2020). Cross-lingual contextualized topic models with zero-shot learning. arXiv preprint arXiv:2004.07737.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Blei, D. M., & Lafferty, J. D. (2006, June). *Dynamic topic models*. In Proceedings of the 23rd international conference on Machine learning (pp. 113-120).
- Bodrunova, S. S., Orekhov, A. V., Blekanov, I. S., Lyudkevich, N. S., & Tarasov, N. A. (2020). Topic detection based on sentence embeddings and agglomerative clustering with markov moment. *Future Internet*, 12(9), 144. <https://doi.org/10.3390/fi12090144>
- Castonguay, L. G. (2013). Psychotherapy outcome: An issue worth

- re-visiting 50 years later. *Psychotherapy*, 50(1), 52-67
- Choi, S. S., Choi, M. O., Kim, K. M., Lee, M. K., Park, E. J., & Choi, E. J. (2020). *Mental Health*. 2ed. Seoul: Hakjisa.
- Clark, D. C. (1995). *Epidemiology, assessment, and management of suicide in depressed patients*. In E. E. Beckham & W. R. Leber (Eds), *Handbook of depression*(2nd.), PP.526-538). New York: Guilford Press. Roy.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Févotte, C., & Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β -divergence. *Neural computation*, 23(9), 2421-2456.
- Freud, S. (1917). *Mourning and melancholia*. The standard edition of the complete psychological works of Sigmund Freud, 14(1914-1916), 237-258.
- Freud, S. (1986). Vue d'ensemble des névroses de transfert. *Œuvres complètes*, 13, 281-302.
- Griffiths, T. L., & Steyvers, M. (2004). *Finding scientific topics*, in Proceedings of the National Academy of Sciences, 101 (Supplement 1), 5228-5235. <https://doi.org/10.1073/pnas.0307752101>.
- Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. arXiv preprint arXiv:2203.05794.
- HIRA (2022). *Analysis of treatment status for depression and anxiety disorders in the last 5 years (2017-2021)*. Wonju: HIRA(Health Insurance Review and Assessment Service). <https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA020041000100&brdScnBltno=4&brdBltNo=10627&pageIndex=1>
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent dirichlet allocation(LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169-15211.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry*, 62(6), 593-602.
- Kim, J. A., & Lee, S. Y. (2020). N-3 and N-6 Fatty Acid Intakes and Depressive Symptoms in Korean Female Adults Population Using Data from the 2014-2018 Korea National Health and Nutrition Examination Survey. *Korean Journal of Health Promotion*, 20(3), 125-134.
- Kim, S. W., & Yang, K. D. (2022). Topic model augmentation and extension method using LDA and BERTopic. *Korean Society for Information Management*, 39(3), 99-132.
- Klein, M. (1946). Notes on some schizoid mechanisms. *International Journal of Psycho-Analysis*, 27, 99-110.
- Ko, Y. S., Lee, S. B., Cha, M. J., Kim, S. D., Lee, J. H., Ham, J. Y., & Song, M. (2022). Topic modeling insomnia social media corpus using BERTopic and building automatic deep learning classification model. *Korean Society for Information Management*, 39(2), 111-129.
- Korniloff, K., Vanhala, M., Kautiainen, H., Koponen, H., Peltonen, M., Mäntyselkä, P., Oksa, H., Kampman, O., & Häkkinen, A. (2012). Lifetime leisure-time physical activity and the risk of depressive symptoms at the ages of 65-74 years: the FIN-D2D survey. *Preventive medicine*, 54(5), 313-315.
- Lee, S. H., Kim, B. R., & Kim, C. H. (2008). What is Depression? From the Viewpoint of Modern Psychiatry. *Journal of the Korean Neuropsychiatric Association*, 47(1), 3-10.
- Leykin, Y., & DeRubeis, R. J. (2009). Allegiance in psychotherapy outcome research: Separating association from bias. *Clinical Psychology: Science and Practice*, 16(1), 54-65.
- Lépine, J. P., & Briley, M. (2011). The increasing burden of depression. *Neuropsychiatric disease and treatment*, 7(Suppl 1), 3-7.
- Maier, S. F., & Seligman, M. E. (1976). Learned helplessness: theory and evidence. *Journal of experimental psychology: general*, 105(1), 3-46.
- McMinn, M. R., & Wade, N. G. (1995). Beliefs about the prevalence of dissociative identity disorder, sexual abuse, and ritual abuse among religious and nonreligious therapists. *Professional Psychology: Research and Practice*, 26(3), 257-261.
- MOHW(2021). *2021 Mental Health Project Information*. Sejong: Ministry of Health and Welfare.
- MOHW(2022). *Announcement of the results of the 「Corona 19 National Mental Health Survey」 in the second quarter of 2022*. Press reference dated August 10, 2022. Sejong: Ministry of Health and Welfare.
- Rashid, T., & Seligman, M. P. (2018). *Positive psychotherapy: Clinician manual*. New York: Oxford University Press.
- Seligman, M. E., & Csikszentmihalyi, M. (2014). *Positive psychology: An introduction*. In *Flow and the foundations of positive psychology* (pp. 279-298). Springer, Dordrecht.
- Seung, H. B., Kwon, H. J., & Kim, S. H. (2022). Effectiveness and Safety of Traditional East Asian Herbal Medicine as Monotherapy for Major Depressive Disorder: A Systematic Review and Meta-Analysis. *Journal of Oriental Neuropsychiatry*, 33(1), 79-111.
- Siddique, S. (2012). Storymaking: In-between anthropological enquiry and Transactional Analysis Psychotherapy. *European Journal of Psychotherapy & Counselling*, 14(3), 249-259.
- Solomon, D. A., Keller, M. B., Leon, A. C., Mueller, T. I., Lavori, P. W., Shea, M. T., ...Endicott, J. (2000). Multiple recurrences of major depressive disorder. *American Journal of Psychiatry*, 157(2), 229-233.
- WHO (2017). *Depression and other common mental disorders Global Health Estimates [Internet]*. WHO 2017 [cited in December 18]. Available from: <https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-ME-R-2017.2-eng.pdf?sequence=1&isAllowed=y>.
- Woods, N. F., & Mitchell, E. S. (1996). Patterns of depressed mood in midlife women: observations from the Seattle Midlife Women's Health Study. *Research in nursing & health*, 19(2), 111-123.
- Yang, H. C. (2022). Analysis of distribution industry research trends using BERTopic and LDA. *Journal of Creativity and Innovation (JCI)*, 15(4), 71-103.
- Yang, W. R., & Yang, H. C. (2022). Topic modeling analysis of social media marketing using BERTopic and LDA. *Journal of Industrial Distribution & Business*, 13(9), 39-52.