

# 텍스트정보처리론 3 팀 기말 레포트

불면증 보조제 관련 정보 취합 및 요약 제공 서비스

김정현, 이상은, 전유진, 최민준

## Introduction

매슬로우의 5 단계의 욕구 중 첫 번째 단계는 '생리적 욕구'이다<sup>1</sup>. 이것은 우리의 생존에서 가장 기본적인 요소인 식욕, 수면욕 등을 뜻한다. 이처럼 깊은 수면은 우리의 삶에서 중요하며, 질이 좋은 수면을 취하는 것은 우리가 정신적, 신체적으로 회복할 수 있도록 하며 일상을 살아갈 수 있도록 도와준다<sup>2</sup>. 하지만, 전세계 인구 중 30%는 불면증으로 시달리고 있으며 이로 인한 다양한 건강 문제를 겪고 있다<sup>3</sup>. 또한, 국내에서는 "60 만명의 사람들이 치료가 필요할 정도의 심각한 불면증"을 앓고 있다고 한다<sup>4</sup>. 불면증은 신체 기능을 악화시키며 다양한 신체적 질환을 야기하고 우울증과 같은 다양한 건강 문제로 이어질 수 있다<sup>5</sup>.

불면증 증상을 완화시키기 위해서는 다양한 방법들이 있다. 전통적으로 대대로 이어 내려오는 잠을 빨리 청할 수 있는 방법으로는 명상하기, 요가하기, 자기 전에 양을 세기 등이 있다. 또한, 최근에는 '디지털 치료제(DTx)'가 출시를 앞두고, 불면증 치료를 앱으로 처방하는 방법 또한 시장에서 활성화되어가고 있다<sup>6</sup>. 불면증을 해결할 수 있는 방법은 많다. 하지만, 이러한 불면증 증상을 완화시킬 수 있는 다양한 방법들이 개발되었음에도 불구하고, 현재 세계 불면증 보조제 시장 매출은 약 \$7.6 million USD(약 9.8 조 원)에 달한다고한다<sup>7</sup>. 보조제 시장에 대한 수치는 불면증 문제의 심각성과 더불어 보조제에 대한 높은 수요를 보여주고 있다.

---

<sup>1</sup> McLeod, Saul. "Maslow's Hierarchy of Needs." Simply Psychology, 21 May 2018.

<sup>2</sup> 광성희(2018). 불면증 수용과정에 관한 내러티브 탐구. 미간행 박사학위 논문. 대전대학교 대학원

<sup>3</sup> Roth, Thomas. "Insomnia: Definition, prevalence, etiology, and consequences." *Journal of Clinical Sleep Medicine*, vol. 3, no. 5 suppl, 15 Aug. 2007, <https://doi.org/10.5664/jcsm.26929>.

<sup>4</sup> 서민지, <https://medigatenews.com/news/1501114092>, 8 July 2021

<sup>5</sup> 손진욱, <http://www.dshosp.co.kr/ds/health200/post/10144>,

<sup>6</sup> 옥기원, <https://www.hani.co.kr/arti/economy/it/1075379.html>, 12 January 2023

<sup>7</sup> Future Market Insights, Sleep Supplement Market Outlook for 2024 to 2034, <https://www.futuremarketinsights.com/reports/sleep-supplement-market#:~:text=Sleep%20Supplement%20Market%20Outlook%20for,rising%20in%20the%20global%20market.>

레딧(Reddit)은 사용자들이 콘텐츠를 제출하고 투표하며 토론할 수 있는 대규모 커뮤니티 기반의 소셜 뉴스 웹사이트이다<sup>8</sup>. 2022 년 11 월 자료에 따르면, 미국에서 레딧은 사용자들이 4 번째로 자주 방문한 사이트이고, 세계적으로는 사용자들이 6 위로 자주 방문하는 웹사이트이다. 사용자들은 링크, 텍스트 게시물, 이미지 등 다양한 형태의 콘텐츠를 'Sub-Reddit'이라고 불리는 주제별 포럼에 게시할 수 있다.

r/insomnia 는 불면증을 주제로 한 서브레딧으로, 불면증을 겪는 사용자들이 모여 자신의 경험을 공유하고, 다양한 해결 방법에 대해 논의하는 커뮤니티이다. r/insomnia 에서는 실제로 불면증을 겪고 있는 사용자들이 자신의 문제를 공유하고, 다양한 치료법과 댓글을 통해 이용자가 직접 복용한 보조제에 대한 피드백을 공유한다. r/insomnia 를 통해 불면증 해결 방법에 접근하고 이는 보조제의 효과에 대한 다양한 관점을 제공한다. 실질적인 사람들의 후기를 볼 수 있다는 점이 매력적으로 느껴져 r/insomnia 이용자의 후기를 조사 대상으로 선정하였다.

따라서, 이번 프로젝트에서는 r/insomnia 에서 보조제에 관련된 모든 논의를 취합 및 요약하여 제공하는 서비스 개발을 목표로 두고 자하였다. 본 서비스는 레딧에 존재하는 수면제 보조제 관련된 수많은 글을 분석하여, 각 보조제의 효능, 부작용, 사용자 경험, 그리고 감정 점수 등을 정리해, 불면증을 앓고 있는 환자가 한눈에 정보를 파악하는 데에 있어서 도움을 주는 것을 주된 목표로 한다.

## 프로젝트 상세

### 1. 피쳐리스트 작성

본격적인 데이터 분석을 진행하기에 앞서 서비스의 골자를 정립하고자 피쳐리스트를 작성하였다. 핵심 기능들의 목록인 피쳐리스트를 작성함으로써 프로젝트의 목표를 구체화하여, 데이터 분석을 통해 어떤 정보를 얻고자 하는 것이며 어떤 분석법을 적용해야하는지 결정할 수 있을 것이다. 작성한 피쳐리스트는 다음과 같다.

#### ● 메인 페이지

---

<sup>8</sup> Widman, Jake . "What Is Reddit?" *Digital Trends*, 18 Dec. 2022, [www.digitaltrends.com/computing/what-is-reddit/](https://www.digitaltrends.com/computing/what-is-reddit/). Accessed 4 Feb. 2024.

- **검색** - 검색한 텍스트를 포함하는 이름을 가진 보조제들의 서브페이지 목록 띄우기
- **인기 보조제** - 전체 데이터 중 가장 많은 노출량을 가진 TOP n 개의 보조제 서브페이지 목록이 메인페이지에 상시 노출됨.

#### ● 서브 페이지

- **Overall Score** - 전체 언급된 텍스트에서 감성 점수 평균
  - 랭킹 정보 (상위 n%인자)
- **노출량**
  - 게시글 수
  - 댓글 수
- **긍정적인 키워드** - 효능, 가격, 맛등 다방면에서 해당 보조제의 긍정적인 특징으로 인식되는 키워드들 3 개
- **부정적인 키워드** - 부작용, 가격, 맛등 다방면에서 해당 보조제의 부정적인 특징으로 인식되는 키워드들 3 개
- **용량** - 언급된 최소, 최대치
- **함께 많이 언급된 보조제**
  - 감성 점수 - 두 보조제가 함께 언급된 텍스트에 한정해서 산정한 감성점수

## 2. r/insomnia 데이터 확보

프로젝트 초기 단계에서는 Python Reddit API Wrapper 인 PRAW 를 활용해 r/insomnia 의 게시글 및 댓글의 텍스트 데이터를 확보하고자 하였으나, Reddit 측의 정책 변경으로 PRAW 를 활용해 대량의 데이터를 추출하는 것이 불가능해졌다. 차선책으로 pushshift.io API 를 활용한 유저들이 공유한 2005 년~2023 년 분량의 dump file 을 확보하였다. 여기에는 insomnia 서브레딧의 69,925 개의 게시글과 433,559 개의 댓글의 정보가 포함되어 있었다. 여기서 본문의 내용이 4 글자 이하이거나, [deleted], 혹은 [removed]인 경우 제외시켜 최종적으로 46,567 개의 게시글과 403,921 개의 댓글 데이터를 확보하였다.

## 3. 보조제 목록 작성 *Appendix\_0*

리스트 작성 기준은 Google, Naver 검색, Chat GPT, Google Gemini 등에 보조제 관련해서 자주 언급되는 보조제를 기록하였다. 그 후, r/insomnia 에 작성한 리스트에 대한 보조제를 하나씩 검색해보며, 그 중 언급된 보조제 위주로 선정하였다. 하단에 보이는 표와 같이, 그에 맞게 25 개의 보조제를 선정한 후, 각 그룹에 맞게 배치하였다.

Hormonal Support	Herbal Remedies	Mineral	Amino Acid
Melatonin	Valerian Root	Magnesium	Glycine
5-HTP	Lemon Balm	Calcium	Tryptophan (or L-tryptophan)
	Lavender	Zinc	Theanine
	Passion Flower	Potassium	
	Chamomile		
	Ashwagandha		
	Kava		
	Hops		
	Skullcap		
	California Poppy		
	Magnolia Bark		
	Mulungu		
	CBD Oil		
	Black Seed Oil		
	Ashwagandha		

이를 기반으로 450,488 개의 게시글 및 댓글 데이터 중 보조제가 하나라도 언급된 데이터만 남기는 필터링을 하였고, 그 결과 약 10%에 해당하는 44,050 개의 게시글 및 댓글 데이터가 남았다. 이 과정에서 보조제별로 44,050 개의 텍스트중 게시글 노출량, 댓글 노출량과 그 합인 총 노출량의 데이터 역시 도출하였다.

#### 4. 텍스트 전처리 *Appendix\_1*

텍스트 전처리의 첫 단계는 공백과 개행문자의 정리이다. 텍스트 데이터는 작성 환경에 따라 불필요한 공백과 개행문자가 포함될 수 있다. 이를 제거함으로써 텍스트의 가독성을 높이고, 후속 처리 단계에서의 일관성을 확보할 수 있다. 정규 표현식(re) 모듈을 이용하였다.

다음으로, URL 과 이메일 주소를 제거하는 작업을 수행하고, 텍스트를 모두 소문자로 변환하였다. 예를 들어, 'Apple'과 'apple'을 같은 단어로 처리하기 위해 모든 텍스트를 소문자로 변환하였다. 이를 통해 분석의 일관성을 유지하고, 단어 빈도 분석 등의 작업에서 정확도를 높일 수 있다.

텍스트의 토큰화 과정에서는 nltk 의 'word\_tokenize' 함수를 이용하여 문장을 구성하는 각 단어를 개별적인 단위로 분리하였다. 토큰화는 텍스트 분석의 기본 단계로, 각 단어를 개별적으로 분석할 수 있게 한다.

불용어 및 비문자의 제거는 텍스트의 중요한 정보를 남기고, 불필요한 정보를 배제하는 과정이다. 불용어는 조사와 같이 자주 등장하지만 텍스트의 주요 의미에 큰 영향을 미치지 않는 단어들이다. 불용어 리스트는 nltk의 'stopwords' 리스트를 이용하였다. 비문자는 텍스트의 분석에서 필요하지 않은 특수문자 등을 포함한다. 그러나 숫자와 단위는 복용량 추출을 위해 보존되었다.

마지막으로, nltk의 'WordNetLemmatizer' 함수를 이용하여 단어들을 원형(lemma)로 복원하는 과정을 거쳤다. 원형 복원은 단어의 형태소 분석을 통해 다양한 형태의 단어를 그 기본 형태로 되돌리는 작업이다. 예를 들어, 'running', 'ran', 'runs' 등을 'run'으로 통일함으로써 단어의 일관성을 유지하고, 분석의 정확도를 높일 수 있다.

이와 같은 전처리 과정을 거친 후, 게시글 파일에는 'processed\_text' 컬럼에, 댓글 파일에는 'processed\_comment' 컬럼에 각각 전처리된 텍스트가 저장되었다. 이를 통해 각 텍스트의 분석이 보다 용이해졌다.

## 5. 복용량 정보 추출 *Appendix\_3.3*

### ● 리스트에 언급된 보조제의 복용량 추출

전처리된 텍스트에서 사용된 보조제 리스트에 있는 보조제와 복용량이 언급된 경우에는 용량 정보를 추출하였다. 복용량 정보는 'mg', 'ml', 'pills', 'capsules', 'capsule', 'tablets', 'tablet', 'oz', 'mcg'와 같이 하나 이상의 숫자 뒤에 오는 패턴을 만족하는 텍스트를 찾고, 보조제 이름이 해당 텍스트에서 언급된 경우 보조제 정보와 용량 정보를 매핑하였다.

이 방법을 통해 텍스트에 등장하는 숫자 중 용량과 관련이 없는 숫자나 보조제 리스트에 없는 약물 등의 용량 정보를 제외하고 보조제에 대한 복용량만을 추출할 수 있었다.

## ● 복용량이 여러 번 등장하는 경우의 복용량 추출

그러나 한 게시글에 여러 가지 보조제와 여러 개의 용량 정보가 함께 포함되어 있는 경우, 각각의 용량이 어떤 보조제를 가리키는지 특정할 수 없는 문제가 있었다. 이를 해결하기 위해 보조제 리스트에 있는 모든 보조제 이름을 순회하면서 리스트에 해당 보조제 이름이 포함되어 있다면 그 위치를 저장하고, 언급된 용량의 위치를 저장하여 각각의 용량과 가장 가까운 하나의 보조제를 매핑하였다.

<위치를 통한 이름-용량 매핑 코드 및 주석>

```
min_distance = float('inf') #최소 거리를 무한대로 초기화

for supplement, aliases in supplements.items():
    for alias in aliases: #보조제의 각 축약어를 순회한다.
        if alias in words: #리스트에 보조제의 이름이 포함되어 있다면,
            index = words.index(alias) #보조제의 위치를 index에 저장한다.
            dosage_index = words.index(amount + ' ' + unit) #용량의 위치를 dosage_index에 저장한다. (숫자+단위로)
            distance = abs(dosage_index - index) #보조제와 용량 사이의 거리를 계산한다.
            if distance < min_distance: #현재 거리가 최소 거리보다 작다면
                min_distance = distance #최소거리를 현재거리로 업데이트한다.
                supplement_found = supplement #해당 용량과 가장 가까운 보조제를 특정한다.

if supplement_found:
    dosage_info.append((amount, unit, supplement_found))
```

## 6. 키워드 추출 및 정규화 과정 *Appendix\_3.2*

보조제별 대중들이 인식하는 서로 다른 특징들에 대한 정보를 간결하게 전달하기 위해 긍정, 부정적인 키워드를 보조제별로 도출하고자 하였다. 이를 위해 먼저 SIDER side effect resource 웹페이지를 참고하였으나, 서브레딧에서 자주 쓰이는 단어들과 달리 전문적인 의학 용어 위주의 사전이라는 점과 그 중에서도 불면증 보조제와 관련된 부작용의 키워드는 극소수라는 점 때문에 적극적으로 활용하지는 못했다. 예시로 side effects 의 첫 페이지에는 Abasia, Abdominal abscess, Abdominal adhesions, Abdominal aortic aneurysm, Abdominal aortic bruit, ... 이런식으로 수천개가 이어지는데 이 중 무엇이 불면증 커뮤니티에서 자주 언급이 될만한 부작용인지 선별하기가 어려웠다. 컴퓨팅 비용적인 측면을 포기하고 수천개의 부작용 목록을 그대로 사용한다고 하더라도 지극히 전문적인 의학 용어들의 목록을 기반으로 서브레딧의 데이터를 분석하는 것이 의미가 있는가에 대한 확신이 들지 않았다. 때문에 우리는 인터넷 서칭, Chat GPT, r/insomnia 아이볼링등을 통해 무차별적으로 긍정적이거나 부정적인 키워드들을 140 여개 수집했다. 몇개를 예시로

들자면 Garbage, Not strong, Vivid Dreams, Nightmare, Anxiety, Anxious, Improved, Works amazing, Non addictive 등이 있다.

이후 이를 비슷한 것끼리 그룹화하는 작업을 거쳤다. 예를 들어 Garbage 와 Not strong 은 Useless 로, Anxiety 와 Anxious 는 Anxiety 로 통일하고 다른 키워드들은 제거하였고, 그 결과 140 여개의 단어들은 22 개의 키워드로 압축되었다.

부정적 키워드	긍정적 키워드
Nightmare	Calming
Anxiety	Effective
Nausea	Non Addictive
Sweat	Safe
Pain	Taste Amazing
Useless	Sleepy
Paradoxical Reactions	Smells nice
No energy	Affordable
Urinate frequently	
Constipation	
Expensive	
Depression	
Taste bad	
Smells bad	

이후 서브레딧의 다양한 표현들을 한정된 키워드들 만으로 잡아내고자 WordNet 라이브러리를 활용한 동의어 정규화를 실시하였다. get\_synonyms 함수를 작성하여 WordNet 의 synset 을 통해 주어진 단어와 관련된 모든 동의어를 찾아내고, 이를 set 으로 반환하였다. 다음으로 주어진 텍스트에서 키워드를 찾기 위해

find\_keywords 함수를 작성했다. 이 함수는 각 키워드에 대해 동의어를 포함한 키워드 목록을 생성하고, 텍스트에서 해당 키워드나 동의어가 포함되어 있는지 확인한다. WordNet 을 활용한 동의어 정규화 과정을 통해 기존 14,000 여개의 키워드 검색 결과를 26,000 여개로 증가시킬 수 있었다.

## 7. 보조제간 관련성 분석 *Appendix\_3.1*

다음으로 특정 보조제를 주체로 두고, 다른 보조제들이 각각 얼마나 함께 언급되는지를 분석하였다. 이는 추후 감성 분석을 추가로 실시하기 위한 기반 데이터를 확보하기 위함으로써, 어떤 조합이 함께 복용했을때 좋은 후기를 가지고 있고, 어떤 조합이 그 반대인지에 대한 정보를 사용자에게 제공하기 위한 작업이다.

먼저 두 보조제의 조합별로 노출량을 다시 계산하고, 이를 보기 좋게 모든 보조제의 목록을 1 열에 나열한 뒤 2 열에 다른 보조제들의 노출량을 많은 순서대로 정렬하였다. 예를 들어 1 열에 Melatonin 은, 2 열에 Valerian root(2585), Magnesium(2188), Chamomile(580), ... , Black seed oil(1) 이 입력되는 식이다.

하지만 이것만으로는 보조제별 연관성을 보여주는 데이터라고 보기 어려웠다. 왜냐하면 보조제별로 절대적인 노출량 자체가 크게 차이가 났기 때문이다. 예를 들어 Magnesium 은 무려 2188 개의 텍스트에서 Melatonin 과 함께 언급되었지만, 애초에 Magnesium 이 한번이라도 언급된 텍스트는 6857 개로, 모든 Magnesium 텍스트중 약 31.9%가 Melatonin 과 함께 언급된 것이다. 반면에 580 번 함께 언급된 Chamomile 은 절대적인 노출량이 1431 개에 불과해 모든 Chamomile 텍스트중 약 40.5%가 Melatonin 과 함께 언급되었다. 분석의 목적상 절대적인 텍스트 개수보다는, 상대적인 비율이 대중의 인식을 더욱 대표한다고 판단, 모든 보조제들의 함께 언급된 횟수에 각자의 총 노출량을 나눈 뒤 해당 수치가 높은 순으로 나열하도록 하였다. 그 결과 Melatonin 은 Valerian root(2585/3271), 5-HTP(189/376), Passion flower(102/224), Lemon balm(220/484), Chamomile(580/1431), ... , Vitamin d(62/550), Black seed oil(1/30) 의 결과가 나왔다.

## 8. 감성 분석 *Appendix\_4*



## ● 감성분석에 사용한 모델 설명

Hugging Face 는 자연어 처리(NLP) 분야에서 널리 사용되는 AI 및 머신러닝 툴을 제공하는 플랫폼이다. 특히, Transformers 라이브러리는 다양한 NLP 작업을 위해 설계된 강력한 도구 모음으로, 자연어 이해(NLU)와 생성(NLG) 작업을 쉽게 수행할 수 있도록 돕는다. 본 보고서에서는 이 라이브러리 중 "bert-base-multilingual-uncased-sentiment" 모델을 사용하였다.

"bert-base-multilingual-uncased-sentiment" 모델은 다국어 감정 분석을 위해 설계된 모델이다. 이 모델은 상품에 대한 리뷰의 긍정, 부정 척도를 계산하여 1 점에서 5 점 사이의 별점으로 예측한다. 모델 훈련에 사용된 영어 리뷰는 총 15 만 개로 대규모의 데이터셋을 이미 확보하고 있다.

모델의 성능 평가는 5000 개의 제품 리뷰를 사용하여 감성 분석을 수행하고, 모델이 예측한 별점과 실제 사용자가 매긴 별점을 비교하는 방식으로 이루어졌다고 한다. 평가 결과, 모델이 예측한 별점이 실제 별점과 정확히 일치하는 비율은 67%였으며, 최대 1 점 차이가 나는 비율은 95%였다. 이 결과는 모델이 제품 리뷰의 감성을 어느 정도 정확하게 예측할 수 있음을 보여준다.

## ● 보조제 별 분석

분석하고자 하는 보조제의 이름이 언급된 모든 게시글과 댓글을 전처리한 후, 각각의 텍스트에 대해서 감성 분석을 수행하였다. 각각의 게시글과 댓글에 대한 예측 감성 점수와 예측의 신뢰성을 나타내는 확신 점수를 결과값으로 받았다. 그리고 같은 보조제에 대한 글들끼리 모아서 감성 점수의 평균을 도출하였다.

**\* 도출 결과 예시: 2 stars (약간 부정적) / 0.843938577( 예측에 84%의 확신을 가짐 )**

## ● 조합별 분석

다음으로, 함께 복용하면 더 효과가 좋아지거나 반대로 효과가 감소하는 보조제의 조합을 알아보기 위해, 2 개 이상의 보조제가 함께 언급된 텍스트에 대해 감성 분석을 수행하였다. 이 결과를 특정 단독 보조제가 언급된 모든 텍스트의 감성 분석 결과와 비교하여 점수 차이를 도출하였다.

점수 차이가 양수인 경우, 함께 복용했을 때 더 효과가 좋은 조합으로 해석할 수 있다. 반대로 음수인 경우, 함께 복용할 때 오히려 효과가 반감되는 조합으로 해석할 수 있다. 예를 들어, 멜라토닌 단독 사용의 평균 점수가 3.5 이고, 멜라토닌과 마그네슘 조합의 평균 점수가 4.5 라면, 점수 차이는 1 이 된다. 이는 멜라토닌만 복용했을 때보다 멜라토닌과 마그네슘을 함께 복용했을 때 평균적으로 더 긍정적인 경험을 했음을 의미한다.

단독 보조제를 감성 분석할 때와 같은 모델을 사용하여 각 리뷰 텍스트에 대해 1 부터 5 까지의 별점으로 감성 점수를 부여하고, 특정 보조제들 간의 조합별로 감성 점수의 평균을 도출하는 작업이 이루어졌다. 이때 하나의 텍스트에서 두 개보다 많은 보조제가 언급되는 경우, 여러 보조제의 조합이 하나의 행으로 묶여서 분석 결과가 도출되는 문제점이 있었다. 하지만 분석의 정확성을 위해서는 보조제의 조합이 반드시 일대일로 이루어져야 한다고 판단하였다. 그래서 일차적으로 추출한 여러 보조제의 조합이 함께 언급된 텍스트의 감성 점수 데이터를 일대일 조합으로 확장하여 저장하는 작업을 거쳤다. 이 과정을 통해 각 보조제 조합을 개별 행으로 확장함으로써, 단일 보조제와 조합 보조제 간의 감성 점수를 정확하게 비교할 수 있게 되었다.

Supplement	Mentioned_Supplements	Text	Sentiment
Melatonin	Melatonin, Magnesium	"Melatonin and magnesium help me."	4
Vitamin C	Vitamin C, Zinc	"Vitamin C and zinc are great."	5

<확장 전 데이터 예시>

Supplement	Mentioned_Supplements	Sentiment
Melatonin	Melatonin	4
Melatonin	Magnesium	4
Vitamin C	Vitamin C	5
Vitamin C	Zinc	5

## <확장 후 데이터 예시>

	A	B	C	D	E
1	Supplement	Mentioned_Supplements	Average_Individual_Sentiment	Average_Combination_Sentiment	Difference
2	5-HTP	ashwagandha	2.223404255	2.3	0.076595745
3	5-HTP	calcium	2.223404255	2	-0.223404255
4	5-HTP	california poppy	2.223404255	2	-0.223404255
5	5-HTP	cbd oil	2.223404255	2.454545455	0.231141199
6	5-HTP	chamomile	2.223404255	2.347826087	0.124421832
7	5-HTP	glycine	2.223404255	2.222222222	-0.001182033
8	5-HTP	kava	2.223404255	1.909090909	-0.314313346
9	5-HTP	lavender	2.223404255	2.272727273	0.049323017
10	5-HTP	lemon balm	2.223404255	2.153846154	-0.069558101
11	5-HTP	magnesium	2.223404255	2.389473684	0.166069429
12	5-HTP	magnolia bark	2.223404255	2	-0.223404255
13	5-HTP	melatonin	2.223404255	2.015873016	-0.207531239
14	5-HTP	mulungu	2.223404255	3	0.776595745
15	5-HTP	passion flower	2.223404255	2.777777778	0.554373522
16	5-HTP	potassium	2.223404255	1.666666667	-0.556737589
17	5-HTP	theanine	2.223404255	1.875	-0.348404255
18	5-HTP	tryptophan	2.223404255	2.081081081	-0.142323174
19	5-HTP	valerian root	2.223404255	2	-0.223404255
20	5-HTP	vitamin d	2.223404255	2	-0.223404255

## <조합별 감성분석 결과 예시>

## 9. 최종 데이터 도출 *Appendix\_5*

보조제 별 노출량, 종합 별점, 연관 보조제 및 연관 별점, 긍/부정적 키워드 등 다양한 분석 결과값들을 서비스에 바로 적용할 수 있도록 하나의 표에 정리하는 과정을 거쳤다. 이를 통해 서비스의 어플리케이션은 이 최종 표 하나만을 가지고 있다면 피쳐리스트의 모든 기능들이 완전하게 작동할 수 있게 된 것이다. 다음은 최종

데이터 테이블의 샘플이다.

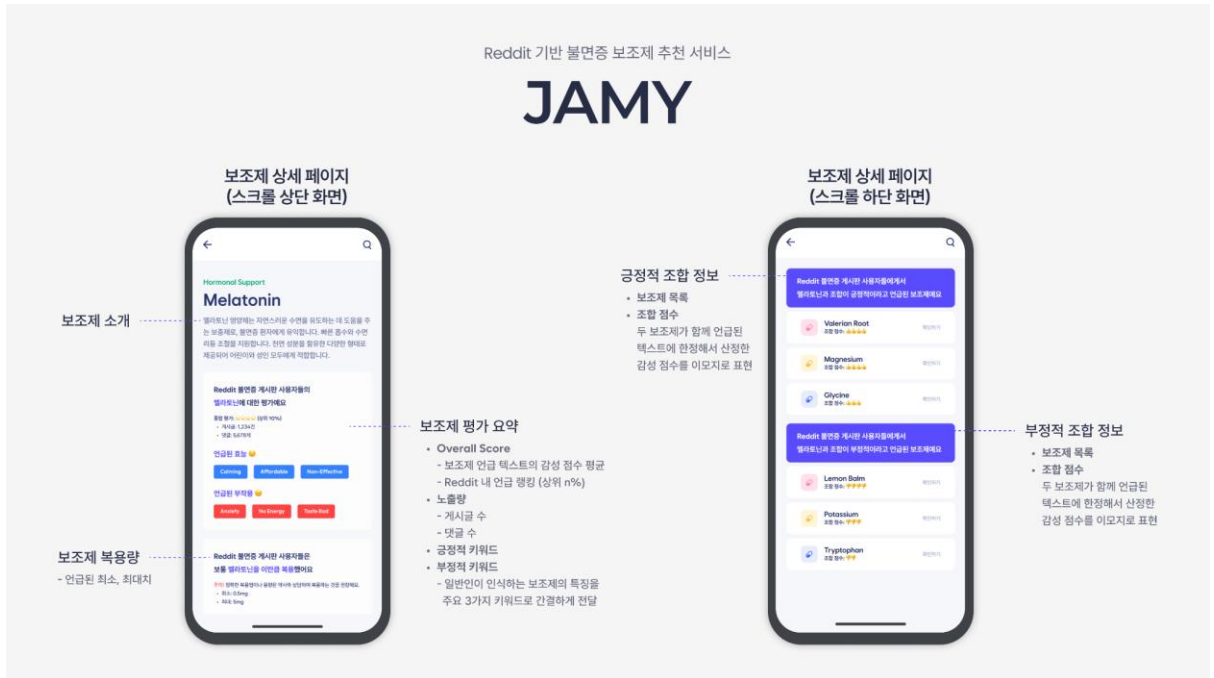
Supplement	Submission_Count	Comment_Count	Total_Count	Sentiment_Rounded	First_Related_Supplement	First_Sentiment_Difference	Second_Related_Supplement	Second_Sentiment_Difference	Third_Related_Supplement	Third_Sentiment_Difference	First_Neg_Keyword	Second_Neg_Keyword	Third_Neg_Keyword	First_Pos_Keyword	Second_Pos_Keyword	Third_Pos_Keyword
Melatonin	6087	18500	24587	1.5	Valerian root(2585/3271)	1.904762	5-HTP(189/376)	0.666667	Passion flower(102/224)	-0.333333	Nightmare(732/1046)	Useless(234/374)	Pain(3609/6829)	Non Addictive(35/62)	Sleepy(1369/2583)	Effective(5332/10254)
5-HTP	81	295	376	5	Tryptophan(40/587)	1.904762	Mulungu(3/58)	-1.09524	Glycine(18/441)	1.904762	Expensive(12/537)	Nausea(4/265)	Useless(5/374)	Sleepy(28/2583)	Effective(107/10254)	Safe(91/9402)
Valerian Root	683	2588	3271	5	Skullcap(88/150)	-0.34524	Passion flower(111/224)	0.666667	poppy(20/42)	1.904762	Taste bad(2/13)	Expensive(45/537)	Constipation(9/109)	Sleepy(251/2583)	Non Addictive(6/62)	Affordable(3/35)
Lemon Balm	79	405	484	2.75	Magnolia bark(35/83)	-2.09524	California poppy(11/42)	-0.333333	Passion flower(50/224)	-0.34524	Urine frequently(1/10)	Taste bad(1/13)	Expensive(13/537)	Affordable(2/35)	Non Addictive(1/62)	Calming(120/9004)
Lavender	193	738	931	1	Chamomile(194/1431)	0.154762	Lemon balm(49/484)	-0.333333	poppy(4/42)	-2.09524	Urine frequently(1/10)	Depression(318/9765)	Nausea(7/265)	Sleepy(92/2583)	Calming(251/9004)	Safe(241/9402)
Passion Flower	42	182	224	3.25	California poppy(11/42)	0.904762	Skullcap(24/150)	0.166667	Mulungu(7/58)	0.154762	Depression(224/9765)	Expensive(6/537)	Constipation(1/109)	Affordable(1/35)	Sleepy(19/2583)	Safe(57/9402)
Chamomile	297	1134	1431	4	California poppy(12/42)	-0.92857	Passion flower(58/224)	-0.333333	Skullcap(38/150)	0.904762	Urine frequently(1/10)	Constipation(6/109)	Nausea(13/265)	Taste Amazing(1/4)	Sleepy(153/2583)	Non Addictive(3/62)
Ashwagandha	104	406	510	2.1667	Magnolia bark(7/83)	-1.09524	Lemon balm(37/484)	-0.333333	Passion flower(17/224)	-0.92857	Anxiety(146/6879)	Constipation(2/109)	Sweat(12/782)	Safe(144/9402)	Effective(147/10254)	Calming(127/9004)
Kava	94	453	547	2	Amanita(13/59)	-0.59524	Mulungu(10/58)	-0.083333	poppy(7/42)	-1.09524	Taste bad(1/13)	Expensive(14/537)	Constipation(2/109)	Non Addictive(3/62)	Safe(137/9402)	Effective(148/10254)
Hops	1	6	7	2.5		-2.09524		-0.333333		-0.59524	Expensive(1/537)	Sweat(1/782)	Anxiety(4/6879)	Sleepy(1/2583)	Calming(1/9004)	Safe(1/9402)
Skullcap	13	137	150	1	California poppy(8/42)	0.904762	Magnolia bark(9/83)	0.066667	Passion flower(24/224)	-2.09524	Depression(68/9765)	Useless(2/374)	Anxiety(31/6879)	Affordable(2/35)	Sleepy(14/2583)	Effective(49/10254)
California Poppy	8	34	42	4	Black seed oil(2/30)	-1.09524	Magnolia bark(5/83)	0.904762	Skullcap(8/150)	0.904762	Constipation(1/109)	Nausea(1/265)	Depression(23/9765)	Safe(11/9402)	Calming(8/9004)	Effective(8/10254)

## 10. 앱 UI/UX 디자인

<앱 메인 페이지>



<앱 상세 페이지>



## 결론

이번 프로젝트는 Reddit의 불면증 게시판, r/insomnia 서브레딧 사용자들의 수면 보조제 관련 텍스트 데이터를 수집하고 분석하여, 각 보조제별 정보를 빠르고 명확하게 제공하는 서비스 'JAMY'를 제안한다. 본 서비스는 보조제의 실제 효과와 주의점에 대한 사용자 후기를 한눈에 파악할 수 있도록 제공함으로써, 사용자가 불면증 보조제를 선택하는 데 있어 유용하게 참고할 수 있다.

### 1. 프로젝트 의의

- 사용자의 수면 보조제에 대한 정보 접근성을 향상시키고, 시간을 절약한다.  
기존에는 불면증 보조제에 대한 정보를 얻기 위해 여러 웹사이트를 직접 검색하고, 수많은 리뷰를 읽어야 했다. 하지만, 본 서비스를 통해 정보를 한눈에 파악함으로써, 개인의 불면증 증상을 완화하기 위한 가장 적합한 보조제 선택에 큰 도움을 얻을 수 있다.
- 궁극적으로 본 서비스는 불면증으로 어려움을 겪는 사람들이 더 나은 삶을 영위할 수 있도록 돕는다. 양질의 삶 뿐만 아니라 생존에 있어 인간에게 수면은 빼놓을 수 없는 중요한 요소이며, 우리 서비스는 사용자들의 현명하고 건강한 선택에 효과적인 도구가 될 것이다.

### 2. 한계

본 프로젝트는 Reddit 의 불면증 게시판, r/insomnia 서브레딧의 데이터만을 수집하여 분석하였으므로, 특정 연령층이나 문화권의 의견이 대표될 수 있어 분석의 객관성이 한계로 지적될 수 있다. 또한, 일부 보조제의 낮은 언급 빈도로 데이터 확보가 충분하지 못하였다. 이는 분석 결과의 신뢰성을 저해할 수 있어, 다양한 데이터를 포함하여 더욱 포괄적인 분석을 수행할 필요가 있다. 마지막으로, 의학적 자문 없이 진행된 프로젝트기 때문에 분석 결과의 타당성을 검증하고 보다 안전한 정보를 제공해야 한다.

결론적으로, 본 연구는 사용자 중심의 수면 보조제 정보를 종합적으로 제공하여 사용자의 더 나은 선택을 돕는다는 점에서 중요한 의의를 가진다. 하지만, 앞서 언급한 한계점을 극복하고 서비스를 발전시키기 위한 지속적인 노력이 필요하다.

## 출처

### [국내 자료]

곽성희(2018). 불면증 수용과정에 관한 내러티브 탐구. 미간행 박사학위 논문. 대전대학교 대학원

서민지, <https://medigatenews.com/news/1501114092>, 8 July 2021

손진욱, <http://www.dshosp.co.kr/ds/health200/post/10144>,

옥기원, <https://www.hani.co.kr/arti/economy/it/1075379.html>, 12 January 2023

### [해외 자료]

Future Market Insights, Sleep Supplement Market Outlook for 2024 to 2034, <https://www.futuremarketinsights.com/reports/sleep-supplement-market#:~:text=Sleep%20Supplement%20Market%20Outlook%20for,rising%20in%20the%20global%20market>.

McLeod, Saul. "Maslow's Hierarchy of Needs." Simply Psychology, 21 May 2018.

Roth, Thomas. "Insomnia: Definition, prevalence, etiology, and consequences." *Journal of Clinical Sleep Medicine*, vol. 3, no. 5 suppl, 15 Aug. 2007, <https://doi.org/10.5664/jcsm.26929>.

Widman, Jake . "What Is Reddit?" *Digital Trends*, 18 Dec. 2022, [www.digitaltrends.com/computing/what-is-reddit/](http://www.digitaltrends.com/computing/what-is-reddit/) Accessed 4 Feb. 2024.

감성분석 모델: <https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment>

SIDER side effects resource: <http://sideeffects.embl.de/se/>