

## A Study on the Matching Rate of Basic Care Products and Reviews through Text Mining

Ji-Eun Kim\*, Young-Suk Chung\*\*, Seong-Hyun Park\*\*\*

\*Student, Dept. Information Technology Convergence, Kongju National University, Cheonan, Korea

\*\*Lecturer, Dept. of Computer Engineering, Kongju National University, Chungnam, Korea

\*\*\*Lecturer, Dept. of Computer Science & Engineering, Kongju National University, Cheonan, Korea

### [Abstract]

As mask wearing has become commonplace during the COVID-19 pandemic, skin problems such as "maskne" have occurred. The demand for basic cosmetics and skincare products has increased due to problems such as increased temperature and humidity inside the mask, and accumulation of waste, sebum, and dead skin cells. However, the number of consumers who have difficulty purchasing cosmetics due to the large number of types of basic cosmetics has begun to increase. To help consumers in these difficulties, this paper analyzed the matching rate between product reviews and product details pages on basic cosmetics sites. By analyzing data collected through crawling and text mining techniques, we investigated the consistency between actual consumer experience and product information provided by manufacturers. According to the analysis, the high matching rate between the two information was confirmed with 94.21% oily skin, 90.35% dry skin, 97.68% complex, 95.37% dry skin, 99.24% sensitive skin, 94.17% sensitive skin, 90.01% calming, 85.79% moisturizing, 89.26% moisturizing, 98.07% oil oil, 82.42% water, 92.66% repurchase, and 94.21% recommended. This means that the manufacturer is accurately promoting the effectiveness of the product, indicating that consumers are actually experiencing the expected results after purchasing the product.

▶ **Key words:** Cosmetics, Web Crawling, Text Mining, Review, Selenium

### [요약]

COVID-19 팬데믹 동안 마스크 착용이 일상화되면서 '마스크네'와 같은 피부 문제가 발생하였다. 마스크 내부의 온도와 습도가 높아져 노폐물, 피지, 각질 등이 쌓이면서 발생하는 트러블로, 기초 화장품과 피부 관리 제품의 수요를 증가시켰다. 그러나 많은 기초 화장품 종류로化妆품을 사기 어려워하는 소비자들이 늘어나기 시작하였다. 본 논문은 이런 어려움을 겪는 소비자들을 돕고자 기초 화장품 사이트의 제품 리뷰와 제품 상세 페이지 간의 일치율을 분석하였다. 크롤링과 텍스트 마이닝 기법을 통해 수집된 데이터를 분석하여 실제 소비자 경험과 제조사가 제공하는 제품 정보 간의 일치율을 조사하였다. 분석 결과, 지성 94.21%, 건성 90.35%, 복합성 97.68%, 속건조 95.37%, 예민 99.24%, 건조 84.17%, 민감 90.01%, 진정 85.79%, 보습 89.26%, 유분기 98.07%, 수분 82.42%, 재구매 92.66%, 추천 94.21%로 결과가 나왔으며 두 정보 간의 높은 일치율을 확인할 수 있었다. 이는 제조사가 제품의 효과를 정확하게 홍보하고 있음은 의미하고 소비자들이 제품 구매 후 기대했던 결과를 실제로 경험하고 있음을 보여준다.

▶ **주제어:** 화장품, 웹 크롤링, 텍스트 마이닝, 리뷰, 셀레니움

- 
- First Author: Ji-Eun Kim, Corresponding Author: Young-Suk Chung
  - \*Ji-Eun Kim (rlawldms0615@naver.com), Dept. Information Technology Convergence, Kongju National University
  - \*\*Young-Suk Chung (merope@kongju.ac.kr), Dept. of Computer Engineering, Kongju National University
  - \*\*\*Seong-Hyun Park (a94270816@gmail.com), Dept. of Computer Science & Engineering, Kongju National University
  - Received: 2024. 08. 30, Revised: 2024. 12. 19, Accepted: 2025. 01. 17.

## I. Introduction

코로나19(COVID-19) 팬데믹 동안 마스크 착용이 일상화 되면서 많은 변화가 생겼다. 마스크는 바이러스 확산을 줄이는데 가장 중요한 역할을 하였지만 장시간 착용으로 인해 마스크로 인한 여드름이 나는 현상이 발생했고 이를 '마스크네'라는 약어도 생겨났다[1][2][3]. 이런 현상은 마스크 내부의 온도와 습도가 높은 환경이 조성되었고 노폐물과 피지, 각질 생성이 상대적으로 많아지게 되어 세균 등 미생물 증식으로 인한 트러블을 유발하여 피부 문제가 발생했다고 말할 수 있다[4]. 이러한 문제로 인해 기초 화장품과 피부 관리 제품에 대한 수요가 급증했다. 사람들이 이러한 기초 스킨케어 제품을 찾는 이유는 단순히 트러블을 완화하기 위함뿐만 아니라, 전반적인 피부 건강을 유지하고 개선하기 위함이다[5].

마스크 규제가 완화되면서 사람들은 마스크를 쓰지 않은 일상으로 돌아갔고 이에 따라 피부, 특히 외모에 대한 관심이 높아졌다. 이러한 변화는 기초 화장품 사용의 증가로 이어졌고 소비자들은 이제 마스크로 인해 발생한 피부 문제를 해결할 뿐 아니라, 피부 상태를 개선하고, 보다 건강하고 매끈한 피부를 유지하고자 하는 욕구가 늘어났다.

본 논문은 이러한 변화에 주목하여, 기초 화장품 사이트의 제품 리뷰를 분석하고 제품 상세 페이지와 비교함으로써 제품 리뷰와 상세 페이지 간의 일치율을 조사하고자 하였다. 일치율이 높을수록 제품 상세 페이지에 있는 해당 내용이 고객에게 신뢰성을 주는 역할을 한다. 크롤링을 통해 제품 리뷰는 실제 소비자들의 경험을 바탕으로 하여 해당 제품의 효과와 만족도를 반영하였고, 제품 상세 페이지는 제조사나 판매자가 제공하는 공식 정보로, 제품의 특징과 장점을 강조한 정보를 사용하였다. 텍스트 마이닝을 이용하여 시각적 자료로 이 두 정보 간의 일치율을 분석함으로써, 소비자 기대와 실제 제품 효과 간의 차이를 파악할 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 통해 화장품 사이트에서 다양한 기초 화장품과 관련된 제품 리뷰를 하는 크롤링과 리뷰의 상위 단어만 추출하는 텍스트 마이닝에 대해 설명한다. 3장에서는 수집된 리뷰와 제품 상세 페이지를 바탕으로 제품의 주요 특징과 장점, 그리고 소비자 만족도를 분석하고 일치율을 비교 분석하는 모델을 제안한다. 4장에서는 해당 모델의 결과를 나타낸다. 마지막으로 5장에서는 해당 모델의 결과를 기술하고 향후 연구 과제와 그에 따른 연구 방향에 대해 논의한다.

## II. Preliminaries

### 1. Related works

#### 1.1 Definition of Cosmetics

화장품법 제 2조 1항에 따르면 '화장품'이란 인체를 청결, 미화하여 매력을 더하고 용모를 밝게 변화시키거나 피부, 모발의 건강을 유지 또는 증진하기 위하여 인체에 사용하는 물품으로서 인체에 대한 작용이 경미한 것을 말한다. 화장품의 적용 범위는 사람의 피부와 모발에 한하며, 그 기능은 건강의 증진이나 유지를 위해 작용이 약한 수준에 제한되는 뜻으로, 치료나 치유의 개념이 아니라 건강한 인체를 유지하고 증진하는 수준의 물품이라는 뜻이다[6][7].

화장품은 일반적으로 사용 목적과 기능에 따라 여러 가지로 분류가 되는데 그 중 스킨케어 제품은 피부를 청결하게 하고, 보습하며, 보호하는 역할을 하고 자외선, 건조, 산화와 같은 유해환경으로부터 피부를 본래 가지고 있는 항상성 유지를 도와 피부의 작용이나 구조가 균형을 이루도록 하는 목적을 가진다[8].

#### 1.2 Web Crawling

웹 크롤링(Web Crawling)은 자동화된 스크립트 또는 프로그램을 사용하여 웹 사이트의 콘텐츠를 수색하고 탐색하는 과정을 의미한다. 웹 크롤러(Web Crawler)는 주어진 URL을 시작으로 링크를 따라가며 웹 페이지의 데이터를 수집한다. 이 기술은 검색 엔진, 데이터 마이닝, 웹 아카이빙, 다양한 데이터 분석 작업에서 중요한 역할을 한다 [9]. 웹 크롤링은 다양한 분야에서 응용되어 사용되고 있는데 딥러닝에서는 사용될 이미지, 영상 수집 및 분류하기 위해 웹 크롤링을 활용한 연구가 진행되고 있다[10][11].

##### 1.2.1 Selenium

셀레니움은 웹의 자동화된 테스트를 돕기 위한 오픈 소스 자동화 테스트 스위트 응용 프로그램이다. 패스트 테스트 프로 웹의 자동화된 테스트 기반 소프트웨어와 매우 유사하지만, 셀레니움은 다음에 중점을 둔다. 만약 셀레니움을 사용하여 테스트를 수행한다면, 그 도구는 셀레니움 테스트라고 불린다. 셀레니움 도구는 Jason에 의해 만들어졌다[12]. 셀레니움은 다양한 플랫폼을 지원하는 자유 리눅스, 윈도우, 맥 및 기타 하드웨어 응용 프로그램이다. 셀레니움은 구글과 같은 브라우저에서 사용할 수 있다. 그중 크롬, 파이어폭스, 인터넷-프로브, 자바, C#, 파이썬 및 루비 테스트 사례 작성에 사용할 수 있으며 다양한 브라우저, 형식, 언어, 프레임워크에 호환된다. 제품 및 테스트 구성에 관계없이 셀레니움은 필요한 지원을 제공한다[13].

### 1.3 Text Mining

텍스트 마이닝(Text Mining)은 대규모 텍스트 데이터에서 유의미한 정보를 추출하고 분석하는 과정이다. 자연어 처리(NLP), 기계 학습, 통계 분석 등의 기술을 활용해 텍스트 데이터를 구조화된 형태로 변환하고, 이를 통해 패턴, 트렌드, 관계 등을 발견한다. 텍스트 마이닝 기술은 데이터의 품질이 낮거나 대량의 데이터를 실시간으로 분석하는 데 사용되며, 새로운 인사이트를 발견하거나 주요 이슈를 탐지하는 데 활발하게 연구되고 있다[14]. 텍스트 마이닝은 빅데이터 시대에서 정보의 가치를 극대화하는 중요한 도구로 자리 잡고 있다[15]. 다양한 분야에 걸쳐 진행 중인 텍스트마이닝을 통한 연구 사례에는 특정한 이벤트 발생 가능성을 파악하기 위한 방향으로 발전하고 있고 이것을 통해 북한의 도발을 예측하는 것 또한 가능할 것으로 예측된다[16].

## III. The Proposed Scheme

COVID-19 팬데믹 동안 마스크 착용이 일상화되면서 마스크로 인한 피부 문제, 즉 '마스크네'가 발생하여 기초 화장품과 피부 관리 제품의 수요가 증가했다. 타 논문에서는 딥러닝을 이용하여 제품 검색 기록과 개인 사용자의 피부 타입 등 컨텍스트 정보를 함께 반영한 제품 추천 모델 연구한 반면, 본 논문은 화장품 사이트의 제품 중 기초제품을 대상으로 소비자의 실제 리뷰와 상세 페이지를 비교 분석하여 실제 제품 효과와 소비자 기대 간의 차이를 파악하고자 한다. 크롤링과 텍스트 마이닝을 통해 수집된 데이터를 바탕으로 일치율을 분석하는 모델을 제안한다. 다음의 그림 Fig. 1.은 화장품 리뷰와 상세페이지의 일치율 모델의 프로세스이다.

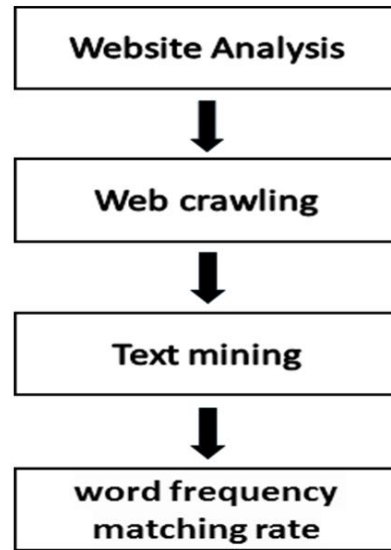


Fig. 1. System Architecture

화장품과 관련된 사이트를 선정한 기준은 분석 대상 사이트가 화장품에 대한 리뷰와 제품에 대한 상세 페이지가 제공되어야 하며 일일 방문자 수가 많은 사이트를 기준으로 하였다. 또한, 웹 크롤링이 가능한 사이트를 대상으로 하였다. 웹사이트를 분석해서 상품의 상세 페이지와 실제 사용자의 리뷰를 분석하여 웹 크롤링으로 리뷰를 추출하고 텍스트 마이닝을 통해 리뷰와 상세 페이지의 단어 빈도수의 일치율을 확인한다.

### 3.1 Cosmetics Website Analysis

셀레니움을 사용기 위해 특정 기초 상품을 선택하고 URL을 추출하여 리뷰를 모으고 수집한다. 실제 구매자들의 리뷰를 최대 20 페이지까지 수집하여 총 200가지의 리뷰들로 상품의 신뢰성을 높이도록 한다.

### 3.2 Cosmetics Web Crawling

웹크롤링을 하기 위해 파이썬을 활용하는 셀레니움 패키지를 적용하였다. 웹사이트에서 추출한 리뷰를 수집하고 수집한 리뷰를 데이터 프레임에 저장한 후 CSV 파일로 저장한다. 다음 Fig. 2.는 웹 크롤링을 위한 의사 코드이다.

```

def get_reviews(url):
    driver.get(url)

    try:
        for page_no in range(1, 11):
            WebDriverWait(driver,
                10).until(EC.element_to_be_clickable(
                    (By.ID, "reviewInfo"))).click()

    while True:
        try:
            reviews = WebDriverWait(driver,
                10).until(EC.presence_of_all_elements_located((By.CSS_SELECTOR,
                    ".txt_inner")))

            for review in reviews:

                review_text =
                    re.sub(r'^\w\s\.\.\.?\\:\:\-\(\)\[\]]',
                        '', review.text)

            df_review.loc[len(df_review)] =
                [review_text.strip()]

            break

```

Fig. 2. Static Web Crawling Algorithm Pseudocode

웹 크롤링과 데이터 처리를 위해 필요한 webdriver와 같은 여러 라이브러리를 사용한다. 객체를 통해 driver의 경로를 설정해 주고 webdriver.Chrome을 통해 드라이버를 시작한다. 주어진 URL에서 리뷰를 수집하는데 각 리뷰 페이지의 버튼을 눌러 리뷰 텍스트를 추출한다. Selenium을 사용하여 주어진 URL에서 최대 10페이지의 사용자 리뷰를 수집하는 기능을 구현한다. 각 페이지에서 리뷰 정보를 클릭하고, 해당 리뷰를 정규 표현식으로 특수 문자와 이모티콘을 제거하여 정리한 후 데이터 프레임에 저장한다. 오류 처리를 통해 리뷰 요소를 안전하게 추출하며, 반복적으로 데이터를 수집하는 구조를 갖춘다. 크롬 드라이버를 설정하여 제품 상세 페이지로 이동하게 만들었다. 제품 상세페이지는 이미지로 되어 있어 모든 이미지가 로드될 때까지 대기하고 이미지 클래스를 찾는다. 클래스를 찾으면 이미지를 다운로드 한다. 내부 모든 이미지 요소를 찾아 URL을 가져온다.

### 3.3 Review Text Mining

저장한 CSV 파일을 로드하여 텍스트 마이닝을 실행한 의사코드는 다음 Fig. 3.와 같다.

```

df = pd.read_csv(csv_file_path,
    encoding='utf-8-sig')

stop_words =
    set(stopwords.words("korean"))

def preprocess_text(text):

    text = re.sub(r"^\w\s]", "", text).lower()

    tokens = [word for word in text.split()
        if word not in stop_words]
    return tokens

df["tokens"] =
df["review"].apply(preprocess_text)

all_tokens = [token for tokens in
df["tokens"] for token in tokens]
token_counts = Counter(all_tokens)

top_20_words =
token_counts.most_common(20)
print("상위 단어:", top_20_words)

keywords = ["지성", "건설", "복합성", "속건조",
"예민", "건조", "민감", "진정", "보습", "유분기",
"수분", "재구매", "추천"]

keyword_filtered = {keyword:
    token_counts[keyword] for keyword in
    keywords if keyword in token_counts}

```

Fig. 3. Static Text Mining Algorithm Pseudocode

저장한 CSV 데이터 파일 경로를 따라 데이터를 로드하는데 이때 인코딩은 utf-8로 지정한다. Python을 사용하여 분석에 필요한 라이브러리를 임포트 해준 후, 분석할 CSV파일을 로드한다. 불용어 목록은 한국어로 설정해야 글씨가 깨지지 않으며 텍스트 전처리 및 토큰화를 진행한다. 특수 문자 제거와 동시에 NLTK 라이브러리에서 제공하는 한국어 불용어 목록을 사용하여 토큰화 및 불용어를 제거한다. 전처리 및 토큰화 수행은 모든 리뷰에 대해 실행한다. 전체 토큰에서 빈도수를 계산하여 상위 단어를 출력하고, 특정 키워드(예: "지성", "건설")의 빈도를 필터링하여 결과를 딕셔너리 형태로 저장한다. 이 과정을 통해 리뷰 분석에서 중요한 단어와 키워드의 출현 빈도를 효과적으로 파악할 수 있다.

### IV. Evaluation

본 논문에서는 제안한 모델을 이용하여 제품 리뷰와 제품 상세 페이지 간의 일치율을 평가한다. 이를 통해 실제 소비자들의 경험과 제조사 및 판매자가 제공하는 정보 간의 일치 정도를 분석하고, 이를 바탕으로 소비자 기대와 제품 효과 간의 차이를 확인한다.

본 논문은 기초 스킨케어 제품 구입 시 일치율을 확인하기 전 상위 단어와 특정 키워드 13개를 텍스트 마이닝을 통해 데이터를 뽑아낸다. 분석에 사용된 리뷰들은 2024년 4월 28일, 29일 기준으로 200개를 선정하였다.

다음 Table 1. 은 선정된 키워드이다. 키워드 선정은 피부 타입과 주요 피부 고민을 중심으로 한 13가지 키워드를 선정하여, 일반적으로 피부 유형을 분류하는 기준과 소비자가 제품 선택 시 고려하는 주요 요소에 기반 한 것이다.

Table 1. Selected keywords

선정 키워드
지성, 건성, 복합성, 속 건조, 예민, 건조, 민감, 진정, 보습, 유분기, 수분, 재구매, 추천

결과로 상위 단어는 [('좋아요', 73), ('토리든', 35), ('같아요', 34), ('토너', 30), ('피부', 22), ('사용하기', 20), ('순하고', 20), ('수분', 19), ('좋은', 19), ('그냥', 17), ('피부에', 17), ('자극없이', 17), ('순해서', 16)]가 나왔으며 키워드 빈도는 [지성: 5, '건성': 6, '복합성': 2, '속건조': 3, '예민': 1, '여드름': 1, '진정': 4, '보습': 3, '수분': 19, '재구매': 9, '추천': 5]이 나왔으며 Fig. 4. 는 특수문자와 소문자 변환한 필터링된 특정단어 을 텍스트 마이닝으로 나타낸 그림이다.



Fig. 4. Cosmetic Review Textmining

Fig. 5.은 크롤링한 리뷰와 상세 페이지 데이터의 특정 단어 빈도수를 비교하여 그래프로 나타낸 그림이다. 상세 페이지 데이터 빈도는 지성 0, 건성 0, 복합성 0, 속건조 0, 예민 0, 건조 0, 민감 1, 진정 2, 보습 2, 유분기 0, 수분 5, 재구매 0, 추천 0, 리뷰 데이터 빈도는 지성 5, 건성 6, 복합성 2, 속건조 3, 예민 1, 건조 0, 민감 0, 진정 4, 보습 3, 유분기 0, 수분 19, 재구매 9, 추천 5로 확인 되었다. 리뷰 데이터 빈도와 상세 페이지 데이터 빈도수가 0이 된 이유는 해당 단어에 대한 데이터가 없기 때문이다.

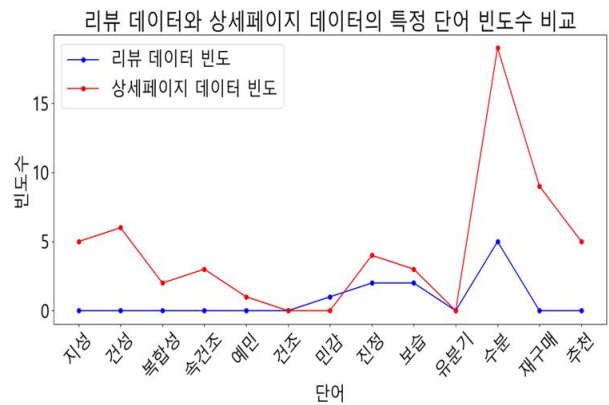


Fig. 5. Compare the frequency of specific words in reviews and detailed page data

다음 Fig. 6.은 Fig. 5.에서 나온 결과로 각 크롤링 한 리뷰와 상세 페이지 데이터의 특정 단어 일치율을 그래프로 나타낸 그림이다. 일치율은 각각 지성 94.21%, 건성 90.35%, 복합성 97.68%, 속건조 95.37%, 예민 99.24%, 건조 84.17%, 민감 90.01%, 진정 85.79%, 보습 89.26%, 유분기 98.07%, 수분 82.42%, 재구매 92.66%, 추천 94.21%로 높게 나오는 것을 확인 할 수 있다.

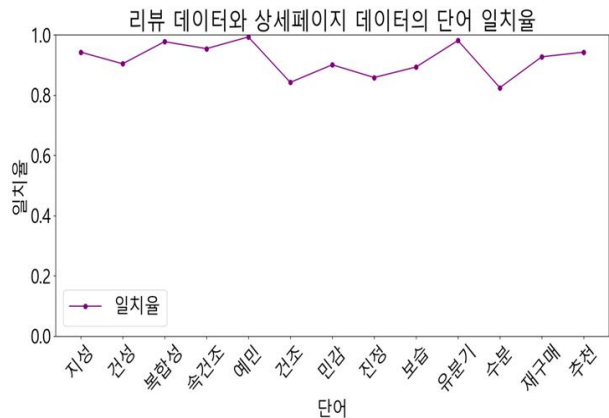


Fig. 6. Word match rate between review data and detail page data

## V. Conclusions

COVID-19 팬데믹 동안 마스크 착용이 일상화되면서 발생한 피부 문제와 이를 해결하기 위한 기초케어 화장품의 수요 증가에 주목하였다. 기초케어 화장품 제품 리뷰와 제품 상세 페이지 간의 일치율을 분석하였다.

본 논문은 텍스트 마이닝 기법을 활용하여 수집된 데이터를 통해 실제 소비자 경험과 제조사가 제공하는 제품 정보 간의 일치율을 조사한 결과, 두 정보 간의 높은 일치율을 확인할 수 있었다. 지성 94.21%, 건성 90.35%, 복합성 97.68%, 속건조 95.37%, 예민 99.24%, 건조 84.17%, 민감 90.01%, 진정 85.79%, 보습 89.26%, 유분기 98.07%, 수분 82.42%, 재구매 92.66%, 추천 94.21%이다. 높은 일치율은 제조사가 제공하는 제품 정보가 실제 소비자 경험과 일치함을 나타낸다. 이는 제조사가 제품의 효과를 정확하게 홍보하고 있음을 의미하며, 소비자들이 제품 구매 후 기대했던 결과를 실제로 경험하고 있음을 보여준다. 이로 인해 소비자들은 해당 제품의 효과와 신뢰성을 높게 평가하게 되며, 제품에 대한 신뢰도를 증대시키는 요인으로 작용한다. 제품 생산을 위한 정보로 활용할 수 있고 소비자는 제품에 대한 선택에 도움을 줄 수 있다.

그러나, 분석대상 모델이 부족하고 기초케어 제품 외에 다른 제품에 대한 연구가 필요하다. 향후 연구에서는 6000개에 도달하는 분석 제품 수와 3가지 기초 화장품 대상으로 일치율을 활용하여 소비자에게 최적의 화장품을 추천하는 모델을 개발할 예정이다.

## REFERENCES

- [1] Bae Seong-eun. "The Effect of Wearing a Mask Due to COVID-19 on the Skin Condition and Skin Care Behavior of Adults." Graduate School of Oriental Medicine, Daegu Oriental Medicine University, pp1-2, 2021.
- [2] Lee-Jaeun. "Skin health care in the pandemic and untact era. Academic presentation of the Korean wellness society", pp. 465-475, 2021.
- [3] Min-Jung Kim, Min-Kyung Kim, Jeong Hee Kim, "Effects of Maskne Cosmetics to Inflammation Elicited by Volatile Organic Compounds (VOC)", vol.30, no.2, 260-269, 2024,
- [4] Kim Min Jeong, & Chae-Mok Yu. "The Effects of Long-term Mask Wearing on Aesthetic Behavior: Focusing on the COVID-19 pandemic Min-Jeong. Journal of the Korean Society of Aesthetic Arts", 21(3), pp17-35. 2020.
- [5] Kim Min Jeong, Chae-Mok Yu, "The Effects of Long-term Mask Wearing on Aesthetic Behavior: Focusing on the COVID-19 pandemic", 21(3), pp17-35.
- [6] Sun Jung Oh, & Dong-Ook Choi. "The Impact of the 2012 Amendment of the Cosmetics Law on Employment. Journal of Industrial Economics and Business", pp349-373, 2019.
- [7] Me-Kyeong Kim, & Ju-Duck Kim. A Study on Cosmetic Purchase Behavior in Adult Women. Journal of the Korean Society of Cosmetics and Beauty, 5(2), pp125-135. 2015.
- [8] Jang Min-ah. "A Behavioral Study on Consumer Choice of Functional Cosmetics Based on Baumann Skin Analysis of Korean Women in the Non-face-to-face Era." Graduate School of Konkuk University, 2022.
- [9] Olston, C., & Najork, M.. Web Crawling. Foundations and Trends in Information Retrieval, 4(3), 175-246. 2010.
- [10] Lee-juHyeok, Kim-Mi Hui, "Image Classification Model using web crawling and transfer learning", j.inst.Korean.electr. electron.eng , Vol026, No.4 , pp639~646 , December 2022
- [11] Hee-Yeon Won, Jae-Woong Kim, & Young-Suk Chung. Propose a Static Web Standard Check Model. Journal of the Korea Society of Computer and Information , 29(4), 83-89. 2024.
- [12] ArjunSatheesh, MonishaSingh"Comparative Study of Open-SourceAutomated Web TestingTools: Selenium and Sahi", Indian Journal of Science and Technology, Vol 10(13). 2017.
- [13] S. Nyamathulla , Dr. P. Ratnababu , Nazma Sultana Shaik, Bhagya Lakshmi. N. "A Review on Selenium Web Driver With Python". Annals of the Romanian Society for Cell Biology, pp. 16760-8, 2021
- [14] Feldman, R., & Sanger, J.. "The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press", pp 125-127 2006.
- [15] Soo Kyung Kim, l"Text Mining Analysis on Consumer Reviews of Hanbang Cosmetics: Focusing on Implications for the Tourism Industry", vol.21, no.4 pp.143-153, 2023
- [16] Sunkyo Cha, & Bongkyoo Yoon. "Forecasting the Level and Types of North Korea's Provocations with Text Mining. Journal of the Korean Institute of Industrial Engineers", 49(5), pp 441-447, 2023.

## Authors



Ji-Eun Kim received my bachelor's degree in chemical engineering from Cheonan Gongju National University in 2022. I am currently earning my master's degree in IT convergence from Cheonan Gongju National

University's Graduate School of Technology Convergence. She is interested in text mining, web crawling, reviews, and selenium.



Young-Suk Chung received the M. S. degree in Multimedia Engineering from Kongju national university, in 2009. Ph. D degree in Computer Engineering from Kongju national university in 2013.

He is currently an adjunct professor in Daejeon Health Sciences College. He is interested in Big data, Cloud computing, Simulation, A.I and Predictive modeling.



Seong-Hyun Park received the B. S. degree in College of Arts and Music from Chungnam National University, Korea in 2017. The M. S, and Ph. D. degrees in Computer Engineering from Kongju National

University, Korea, in 2017, 2020. He is currently a teaching in the Department of Computer Science & Engineering Kongju National University. He is interested in computer music, convergence Education. and Real Time System and Management and Clout computing and Communication