

<http://dx.doi.org/10.17703/JCCT.2023.9.5.509>

JCCT 2023-9-62

# 한국어 문서 요약 기법을 활용한 휘발유 재고량에 대한 미디어 분석

## Media-based Analysis of Gasoline Inventory with Korean Text Summarization

윤성연\*, 박민서\*\*

Sungyeon Yoon\*, Minseo Park\*\*

**요약** 국가 차원의 지속적인 대체 에너지 개발에도 석유 제품의 사용량은 지속적으로 증가하고 있다. 특히, 대표적인 석유 제품인 휘발유는 국제유가의 변동에 그 가격이 크게 변동한다. 주유소에서는 휘발유의 가격 변화에 대응하기 위해 휘발유 재고량을 조절한다. 따라서, 휘발유 재고량의 주요 변화 요인을 분석하여 전반적인 휘발유 소비 행태를 분석할 필요가 있다. 본 연구에서는 주유소의 휘발유 재고량 변화에 영향을 미치는 요인을 파악하기 위해 뉴스 기사를 활용한다. 첫째, 웹 크롤링을 통해 자동으로 휘발유와 관련한 기사를 수집한다. 둘째, 수집한 뉴스 기사를 KoBART(Korean Bidirectional and Auto-Regressive Transformers) 텍스트 요약 모델을 활용하여 요약한다. 셋째, 추출한 요약문을 전처리하고, N-Gram 언어 모델과 TF-IDF(Term Frequency Inverse Document Frequency)를 통해 단어 및 구 단위의 주요 요인을 도출한다. 본 연구를 통해 휘발유 소비 형태의 파악 및 예측이 가능하다.

**주요어** : 딥러닝, 한글 텍스트 요약, 자연어 처리, 휘발유 재고, 변화 요인 분석

**Abstract** Despite the continued development of alternative energies, fuel consumption is increasing. In particular, the price of gasoline fluctuates greatly according to fluctuations in international oil prices. Gas stations adjust their gasoline inventory to respond to gasoline price fluctuations. In this study, news datasets is used to analyze the gasoline consumption patterns through fluctuations of the gasoline inventory. First, collecting news datasets with web crawling. Second, summarizing news datasets using KoBART, which summarizes the Korean text datasets. Finally, preprocessing and deriving the fluctuations factors through N-Gram Language Model and TF-IDF. Through this study, it is possible to analyze and predict gasoline consumption patterns.

**Key words** : Deep Learning, KoBART, NLP, Gasoline Inventory, Analysis of Fluctuations Factors

### 1. 서 론

국가 차원의 지속적인 대체 에너지 개발에도 여전히 석유 제품이 전체 에너지 소비의 높은 비율을 차지한다

[1]. 국내 석유 제품은 대부분을 수입에 의존하고 있는 원유를 가공하여 생산하기 때문에, 국제유가의 변동은 국내 석유 제품 가격의 변동 뿐만 아니라 국내 경제에 큰 영향을 미친다.

\*정희원, 서울여자대학교 데이터사이언스학과 박사과정  
\*\*정희원, 서울여자대학교 데이터사이언스학과 교수 (교신저자)  
접수일: 2023년 7월 11일, 수정완료일: 2023년 8월 25일  
게재확정일: 2023년 9월 5일

Received: July 11, 2023 / Revised: August 25, 2023  
Accepted: September 5, 2023  
\*\*Corresponding Author: mpark@swu.ac.kr  
Dept. of Data Science, Seoul Women's Univ, Korea

최근 다양한 연구들에서는 국제유가의 변화와 국내 석유 제품 가격 간의 상관관계를 도출해 그 변화 추이의 차이를 파악한다[2-6]. 차경수[2]는 유종에 따라 비대칭 조정이 나타나는 기간 및 속도에 차이가 있고, 휘발유의 경우 그 속도가 상대적으로 빠른 것을 파악하였다. 신동현 등[3]은 국제유가의 변화에 따른 휘발유와 경유의 소비함수를 추정하고, 가격 비대칭성의 존재 여부를 분석하여 유종에 따라 가격 및 구조의 변화가 다르게 나타남을 파악하였다. 이처럼 유종에 따라 가격 변화 추이가 다르게 보이는 것은 해당 유종의 소비 형태에 차이가 있기 때문이다. 특히 가격의 비대칭성과 변동성이 높은 휘발유는 일반 소비자들의 자가 운전에서 주로 사용되는 연료이기 때문에, 그 변화 양상을 상세하게 파악할 필요가 있다. 김형진[4]은 2008년 5월 ~ 2022년 12월의 주간 평균 두바이유 가격, 국내 정유사의 휘발유 판매가격, 국내 주유소의 휘발유 판매가격, 환율 등 4개 데이터를 활용하여 국제유가의 변화에 대한 휘발유 가격 조정의 비대칭성에 대해 분석하였다. 국제유가의 상승으로 휘발유 가격이 상승할 경우, 정유사 및 주유소의 휘발유 재고 관리비용 또한 증가하기 때문에 휘발유 판매가격의 빠른 상승을 보이나, 반대로 국제유가의 하락은 휘발유 가격의 적극적인 조정에 큰 영향을 미치지 않는 것을 증명하였다. 차경수[5]는 2009년 2월 ~ 2020년 5월의 주간 평균의 전국 주유소와 지역별 주유소의 휘발유 평균 가격, 브랜드 별 주유소의 휘발유 평균 가격, 국제 원유 가격, 국제 휘발유 가격 등 5개 데이터를 활용하여 국제 원유 및 휘발유 가격의 변화에 따른 전국 평균, 지역 별, 브랜드 별 휘발유 가격 변화의 비대칭 정도의 차이를 분석하였다. 또한 지역 별, 브랜드 별 휘발유 가격의 비대칭성 정도를 탐색하기 위한 탐색 비용이 주유소의 휘발유 가격 조정에 영향을 미치는 것을 증명하였다. 배지영 외[6]는 1988년 1월 ~ 2012년 12월 월 말의 미국의 휘발유 현물가격과 국제 원유 가격, 국제 원유 재고량 등 3개 데이터를 활용하여 휘발유 가격의 비대칭성을 분석하였다. 전체 분석 시점을 국제유가 안정기, 국제유가 상승기, 국제적 경제위기와 연관된 5개 구간으로 구분한 뒤 국제 원유 가격 및 재고량과 휘발유 가격 변화의 비대칭성을 분석하였고, 국제 경제의 변화와 휘발유 가격 간의 관계성을 증명하였다.

이처럼 휘발유 가격 변동은 국제유가 및 휘발유 재

고 변화와 크게 연관되어 있음을 알 수 있다. 그러나 대부분의 연구가 휘발유의 가격을 통한 전반적인 소비 행태를 분석하는 데에 집중되어 있어, 수치적 해석이 불가능한 소비 행태 분석이 필요한 실정이다. 국가 에너지 수급 동향 분석에 따르면, 2021년 11월 월평균 국제 휘발유 가격이 전월 대비 26.19원 하락하고 2021년 11월 12일 시행된 정부의 유류세 인하 정책으로 20% 이상 유류세가 감소하였음에도, 실제 주유소의 휘발유의 주문량이 전월 대비 2.74% 증가에 그쳤으며, 주유소의 휘발유 평균 판매가격은 리터 당 약 25원 정도 상승하였다[7]. 이러한 현상은 2021년 12월 휘발유 주문량의 전월 대비 18.57% 상승과 주유소의 휘발유 평균 판매 가격이 전월 대비 리터 당 약 91원 하락함을 고려할 때, 주유소에서 기존 높은 가격에 주문해 두었던 휘발유 재고를 우선 소진하였다는 것을 의미한다. 따라서 국내 휘발유의 소비 동향을 파악하기 위해 주유소에서 휘발유 재고량을 결정하는 요인을 파악할 필요가 있다.

본 연구에서는 월간 휘발유 소비 동향을 파악하기 위해 휘발유 재고량 변화의 원인을 분석하고자 한다. 분석을 위해 월별 재고량이 크게 변화한 시점을 파악하고, 해당 구간의 사회적, 경제적 이슈와의 관계를 파악하기 위해 뉴스 기사를 수집한다. 뉴스 기사는 자료의 양이 많고, 빠른 시간에 다양한 정보를 알 수 있어 수치적으로 파악이 불가능한 정성적 요인을 파악할 수 있다는 장점이 있다[8]. 그러나, 뉴스 기사의 원본 문장은 동일한 내용을 강조하기 위해 중복하여 그 내용을 작성하거나, 기사 발행인의 견해가 반영되는 경우가 발생한다. 이를 그대로 활용하여 주요 이슈를 파악할 경우, 정보의 편향성 문제가 발생할 수 있다. 이에, 본 논문에서는 딥러닝 기반의 텍스트 요약 모델을 활용하여 뉴스 기사의 핵심 내용을 포함하는 요약문을 추출하고, 머신러닝 기반의 자연어 처리를 통해 휘발유의 재고량 변화의 주요 요인을 분석한다.

논문의 구성은 다음과 같다. 제 2장에서 자연어 처리 기법과 텍스트 요약 모델을 활용한 휘발유 재고량의 변화 요인을 분석하는 방법에 대해 설명한다. 제 3장에서 연구 결과를 설명하고, 마지막으로 제 4장에서 결론을 언급한다.

## II. 연구방법

제안 방법은 데이터 수집, 뉴스 기사의 요약 문장 생성, 데이터 전처리, 자연어 처리의 네 단계로 구성된다.

### 1. 데이터 수집

본 연구는 한국석유공사 국내석유정보시스템(PEDSIS)에서 수집한 2010년 1월 ~ 2022년 12월 국내 주유소의 월평균 휘발유 주문량, 판매량의 수치 데이터를 활용한다. 국내 주유소의 월평균 휘발유 재고량은 주문량과 판매량의 차이로 산출하였다. 정부의 유류세 인하 정책의 시행 및 종료는 주유소의 휘발유 재고량 변화에 절대적이므로[9], 본 연구에서는 미디어를 활용하여 유류세 인하 이외의 요인을 파악하기 위해 유류세 인하 시점과 유류세 인하 시행 전 및 종료 후 1개월에 해당하는 2018년 10월 ~ 2019년 9월, 2021년 10월 ~ 2022년 12월을 제외한 총 129개월의 휘발유 주문량, 판매량, 재고량 데이터를 활용하였다.

뉴스 기사를 활용한 미디어 기반의 분석은 휘발유의 월평균 재고량을 기준으로 진행한다. 휘발유의 월평균 재고량을 Sturges' Formula[10] 수식을 활용해 총 9개 구간으로 구분하여 그 분포를 확인하였다. 휘발유의 재고량이 음수인 경우, 휘발유 판매량이 주문량보다 많은 것을 의미하며, 이는 주문량의 감소 혹은 판매량의 증가 때문이다. 반면, 휘발유의 재고량이 양수인 경우, 휘발유 주문량이 판매량보다 많은 것을 의미하며, 이는 주문량 증가 혹은 판매량 감소의 결과이다. 휘발유 주문량과 판매량 간의 차이가 커 휘발유 재고량이 부호와 관련 없이 큰 값을 가지는 경우의 사회적, 경제적 이슈를 모두 확인하기 위하여 하위 4개 구간인 2016년 7월, 2017년 1월, 2020년 1월, 2020년 10월과 상위 3개 구간인 2011년 7월, 2020년 5월, 2020년 11월을 포함한 총 7개 구간을 분석 구간으로 선정하였다. 또한, 뉴스 기사에 나타난 사회적, 경제적 이슈가 실제 휘발유의 재고량 변화에 영향을 주기까지의 시차를 반영하기 위하여 분석 구간 이전 2개월에 해당하는 시점을 함께 분석하였다.

분석 구간의 휘발유 재고량 변화 요인을 분석하기 위해 국가 기간 뉴스 통신사인 연합뉴스의 기사를 웹 크롤링(Web Crawling)[11] 기술을 활용해 자동으로 수집하였다. 수집한 데이터는 석유·석탄·가스, 에너지 정책, 전력, 해외 에너지 개발, 신재생 에너지, 기후 변화 등 휘발유 소비에 직·간접적인 영향을 미치는 6개 분야

로 구분된 총 18,438개 뉴스 기사의 제목, 내용, 분야, 등록일이다. 뉴스 내용의 경우, 뉴스 기사의 발행 기관(연합뉴스) 및 발행인과 관련된 개인 정보(이름, 이메일 주소 등) 등을 제거하는 과정을 수행하였다.

### 2. 딥러닝 기반의 한국어 문서 요약 기술을 활용한 기사 요약문 생성

뉴스 기사의 원본 문장에는 동일한 내용을 여러 번 중복하여 해당 내용을 강조하거나, 기사 발행인의 의견이 반영되어 작성된 사항 등이 모두 포함된다. 이를 활용해 자연어 처리 기법으로 주요 이슈를 파악할 경우, 정보의 편향성 문제가 발생할 수 있다. 이에 뉴스 기사의 핵심 내용을 추출하여 분석을 진행해야 한다.

본 연구에서는 SKT(SK telecom) Brain에서 개발한 한국어 텍스트 요약 모델인 KoBART(Korean Bidirectional and Auto-Regressive Transformers)[12]을 활용하여 뉴스 기사의 요약문을 추출하였다. KoBART는 구글에서 개발한 BERT(Bidirectional and Auto-Regressive Transformers)[13]에 한국어 위키피디아 및 뉴스 기사 등을 활용하여 한국어의 언어 특성을 추가로 학습시킨 모델이다. 수집한 총 18,438개 뉴스 기사의 뉴스 내용의 요약문을 추출하였다.

### 3. 데이터 전처리

요약된 뉴스 내용을 분석하기 위해 명사 추출과 지칭어 및 유의어 지칭의 전처리(Preprocessing)를 수행하였다. 한국어는 조사, 어미 등을 명사에 붙여 쓰는 교착어이기 때문에, 이를 형태소의 단위로 쪼개는 과정이 필요하다. 본 연구에서는, 체언에 해당하는 명사를 추출하여 분석을 수행하였다. 형태소 분해 및 명사 추출에는 Okt(Open-source Korean Text Processor) 형태소 분석기[14]를 활용하였다. 이후 추출한 명사를 기준으로 '우한 폐렴'과 같이 고유한 의미를 가지나 두 개 이상의 어절로 구성된 명사의 경우 이를 지칭어로 지정하여 어절 사이의 공백을 제거하였다. 또한, '신종 코로나바이러스', '신종 코로나바이러스 감염증', '신종 바이러스 코로나' 등과 같이 동일 혹은 유사한 의미를 지니나, 뉴스 기사 발행인의 언어 습관에 따라 다르게 표현한 명사를 유의어로 지정하여 '코로나'와 같이 하나의 대표어로 변경하였다.

#### 4. 자연어 처리(Natural Language Processing)

전처리 후의 뉴스 내용은 N-Gram 언어 모델(N-Gram Language Model)[15, 16]을 통해 1~2개의 연속 단어 및 구를 토큰화(Tokenization)하고, TF-IDF(Term Frequency Inverse Document Frequency)[17, 18]를 통해 이를 벡터화(Vectorization)하여 분석 구간의 주요 이슈를 파악한다. TF-IDF는 분석 구간에 발행된 뉴스에 자주 등장하지만, 일반적으로 등장하지는 않은 특정 단어 및 구를 확인하고자 할 때 사용한다. TF-IDF 값이 클수록 뉴스 내용의 주제나 의미를 더욱 잘 내포한다는 의미이다. 추출한 TF-IDF 값의 공통점을 바탕으로 후후 재고량의 변화 추이를 예측하는 데에 활용할 수 있도록 표 1과 같이 높은 TF-IDF 결과 값을 가지는 단어 및 구를 기반으로 ‘국내 유가 정책’, ‘국제 유가 정책’, ‘전쟁·테러’, ‘국제 경제 쇼크’, ‘감염증’ 등의 5개 카테고리 분류하였다.

### III. 실험 및 결과 분석

#### 1. 주유소의 휘발유 재고량 변화 요인 분석

본 연구에서는 주유소의 휘발유 월평균 재고량이 큰 변화를 보이는 구간의 뉴스 기사를 자연어 처리하고, 그 결과를 5개의 카테고리로 분류하여 사회적, 경제적 이슈에 따른 휘발유의 소비 추이 변화를 파악하였다. 2016년 7월, 2017년 1월, 2020년 1월, 2020년 10월은 주유소의 월평균 재고량이 크게 감소한 구간이고, 2011년 7월, 2020년 5월, 2020년 11월은 주유소의 월평균 재고량이 크게 증가한 구간이다. 뉴스 기사에 나타난 사회적, 경제적 이슈가 실제 휘발유의 재고량 변화에 영향을 주기까지의 시차를 반영하기 위하여 분석 구간 이전 2개월에 해당하는 시점을 함께 확인하였다.

2016년 7월에는 판매량의 큰 증가로 인해 재고량이 크게 감소하였다. 영국의 유럽연합(EU) 탈퇴 확정으로 인한 세계 경기의 불확실성이 고조되는 와중, 미국의 원유의 공급과잉으로 인해 국제유가가 하락하는 현상을 보였다. 이와 관련하여 ‘브렉시트’, ‘탈퇴’, ‘브렉시트 탈퇴’, ‘브렉시트 우려’, ‘유럽연합 탈퇴’ 등의 단어 및 구가 주요 이슈로 도출되었다. 국제유가가 하락하였으나, 세계 경기 둔화의 영향으로 주유소에서 휘발유 주문량을 증가시키지 못하였고, 반면 저렴한 휘발유 가격에

판매량이 증가하여 재고량의 큰 감소로 이어진 것으로 해석할 수 있다. 해당 단어 및 구는 ‘국제 경제 쇼크’ 카테고리로 구분하였다. 2017년 1월에는 주문량의 큰 감소로 인해 재고량이 크게 감소하였다. 원유의 지속적인 공급과잉 문제를 해결하기 위해 석유수출국기구(OPEC) 회원국과 비회원국 간의 유류 감산 합의가 이루어졌고, 이에 유가가 급등하는 현상을 보였다. 이에 ‘감산’, ‘비회원국’, ‘회원국 감산’, ‘비회원국 감산’, ‘산유국 감산’, ‘감산 합의’, ‘감산 동참’ 등의 단어 및 구가 주요 이슈로 도출되었다. 국제유가의 상승으로 인한 주유소의 휘발유 주문 비용이 증가가 휘발유 주문량 및 재고량의 감소로 이어진 것으로 해석할 수 있다. 해당 단어 및 구는 ‘국제 유가 정책’ 카테고리로 구분하였다. 2020년 1월에는 주문량 감소와 판매량 증가가 재고량의 큰 감소에 영향을 미쳤다. 2020년 1월은 신종 코로나바이러스 감염증이 전세계로 확산되기 시작한 시점이다. 또한 18개월 간 지속되었던 미-중 무역 전쟁의 첫 합의가 이루어짐에 따라 원유 수요 증가에 대한 기대로 인해 국제유가가 소폭 상승하였다. 미국과 이란의 군사 갈등 및 보복 사건도 국제유가 상승에 영향을 미쳤다. 이에 ‘우한폐렴’, ‘코로나’, ‘무역 합의’, ‘미국 이란’ 등의 단어 및 구가 주요 이슈로 도출되었다. 신종 코로나바이러스에 대한 공식적인 명칭이 지정되기 이전, 우한폐렴으로 그 명칭을 사용하였기 때문에 ‘우한폐렴’을 ‘코로나’의 유의어로 지정하지 않고 별도로 그 중요도를 확인하였다. 국제유가의 상승과 신종 감염증의 발생으로 인한 국제 경기 침체에 따라 주문량이 감소한 반면, 감염증의 위험에 대비하여 대중교통 대신 휘발유를 연료로 사용하는 일반 승용차의 통행량 증가에 따른 판매량 증가가 재고량의 하락으로 이어진 것으로 해석할 수 있다. ‘우한폐렴’, ‘코로나’는 ‘감염증’ 카테고리, ‘무역 합의’는 ‘국제 경제 쇼크’ 카테고리, ‘미국 이란’은 ‘전쟁·테러’ 카테고리로 구분하였다. 2020년 10월은 주문량의 지속적인 감소로 인해 재고량이 크게 감소하였다. 2020년은 신종 코로나바이러스가 전세계적으로 확산됨에 따라 전반적인 경제 침체가 이어졌고, 이에 국제유가가 급변하는 모습을 보였다. 이와 관련하여 ‘코로나’, ‘코로나 사태’, ‘오르내림’ 등의 단어 및 구가 주요 이슈로 도출되었다. 국제 유가의 급변에 주유소에서 휘발유 주문량을 증가시키지 못하여 재고량이 감소한 것으로 해석할 수 있다. ‘코로나’, ‘코로나 사태’는

표 1. 주유소의 휘발유 재고량 변화 요인 및 카테고리

Table 1. Fluctuation factors and categories of gasoline inventory at the gas station

|              | 연월           | N-Gram | 카테고리        |                                             |                          |                                 |                              |
|--------------|--------------|--------|-------------|---------------------------------------------|--------------------------|---------------------------------|------------------------------|
|              |              |        | 국내 유가 정책    | 국제 유가 정책                                    | 전쟁·테러                    | 국제 경제 쇼크                        | 감염증                          |
| 감소           | 2016년<br>7월  | 1-Gram |             |                                             |                          | 브렉시트, 탈퇴                        |                              |
|              |              | 2-Gram |             |                                             |                          | 브렉시트 탈퇴,<br>브렉시트 우려,<br>유럽연합 탈퇴 |                              |
|              | 2017년<br>1월  | 1-Gram |             | 감산, 비회원국                                    |                          |                                 |                              |
|              |              | 2-Gram |             | 회원국 감산, 비회원국<br>감산, 산유국 감산,<br>감산 합의, 감산 동참 |                          |                                 |                              |
|              | 2020년<br>1월  | 1-Gram |             |                                             |                          |                                 | 우한폐렴, 코로나                    |
|              |              | 2-Gram |             |                                             | 미국 이란                    | 무역 합의                           |                              |
| 2020년<br>10월 | 1-Gram       |        |             |                                             | 오르내림                     | 코로나                             |                              |
|              | 2-Gram       |        |             |                                             |                          | 코로나 사태                          |                              |
| 증가           | 2011년<br>7월  | 1-Gram | 기름값, 할인, 종료 |                                             | 리비아, 무아마르                |                                 |                              |
|              |              | 2-Gram | 기름값 할인      |                                             | 리비아 반군                   |                                 |                              |
|              | 2020년<br>5월  | 1-Gram |             |                                             | 산유국, 전쟁                  | 폭락, 급락,<br>마이너스, V자형곡선          | 코로나, 대유행                     |
|              |              | 2-Gram |             |                                             | 유가 전쟁,<br>사우디아라비아<br>러시아 |                                 | 코로나 확산,<br>코로나 사태,<br>코로나 우려 |
|              | 2020년<br>11월 | 1-Gram |             |                                             |                          |                                 | 코로나, 백신                      |
|              |              | 2-Gram |             |                                             |                          |                                 | 코로나 백신,<br>코로나 여파            |

‘감염증’ 카테고리, ‘오르내림’은 ‘국제 경제 쇼크’ 카테고리, ‘리비아’, ‘무아마르’, ‘리비아 반군’ 등의 단어 및 구는 ‘전쟁·테러’ 카테고리, ‘산유국’, ‘전쟁’, ‘유가 전쟁’, ‘사우디아라비아 러시아’, ‘폭락’, ‘급락’, ‘마이너스’, ‘V자형곡선’ 등의 단어 및 구가 주요 이슈로 도출되었다. 저렴한 국제유가에 주유소에서 휘발유 주문량을 증가시켰기 때문에 재고량이 증가한 것으로 해석할 수 있다. ‘코로나’, ‘대유행’, ‘코로나 확산’, ‘코로나 사태’, ‘코로나 우려’, ‘산유국’, ‘전쟁’, ‘유가 전쟁’, ‘사우디아라비아 러시아’, ‘폭락’, ‘급락’, ‘마이너스’, ‘V자형곡선’ 등의 단어 및 구가 주요 이슈로 도출되었다. 저렴한 국제유가에 주유소에서 휘발유 주문량을 증가시켰기 때문에 재고량이 증가한 것으로 해석할 수 있다. ‘코로나’, ‘대유행’, ‘코로나 확산’, ‘코로나 사태’, ‘코로나 우려’는 ‘감염증’ 카테고리, ‘산유국’, ‘전쟁’, ‘유

러스가 대유행하며 전세계적으로 심각한 경제 침체가 지속됨에 따라 지속적인 석유 수요가 감소함에도 불구하고, 상대적으로 손해가 적었던 러시아가 석유 감산을 거부하며 원유 생산량 조절 합의에 실패하였다. 이에 사우디아라비아가 원유를 증산해버리면서 러시아-사우디아라비아 간에 유가 전쟁의 양상으로 번지는 상황이 발생하였다. 유가 전쟁을 통해 다량으로 생산된 원유를 소비할 수 없게 됨에 따라 지속적으로 유가가 하락하여 사상 최초로 마이너스 유가를 기록하였다. 이와 관련하여, ‘코로나’, ‘대유행’, ‘코로나 확산’, ‘코로나 사태’, ‘코로나 우려’, ‘산유국’, ‘전쟁’, ‘유가 전쟁’, ‘사우디아라비아 러시아’, ‘폭락’, ‘급락’, ‘마이너스’, ‘V자형곡선’ 등의 단어 및 구가 주요 이슈로 도출되었다. 저렴한 국제유가에 주유소에서 휘발유 주문량을 증가시켰기 때문에 재고량이 증가한 것으로 해석할 수 있다. ‘코로나’, ‘대유행’, ‘코로나 확산’, ‘코로나 사태’, ‘코로나 우려’는 ‘감염증’ 카테고리, ‘산유국’, ‘전쟁’, ‘유

표 2. 주요 단어 및 구의 TF-IDF 중요도 비교  
Table 2. Evaluation with TF-IDF of words and phrases

| 2017년 1월 |       |             | 2011년 7월 |      |             |
|----------|-------|-------------|----------|------|-------------|
| 단어 및 구   | 추출 문장 | TF-IDF 중요도  | 단어 및 구   | 추출문장 | TF-IDF 중요도  |
| 감산       | 요약문   | <b>0.38</b> | 기름값      | 요약문  | <b>0.15</b> |
|          | 원문    | 0.31        |          | 원문   | 0.09        |
| 비회원국     | 요약문   | <b>0.28</b> | 할인       | 요약문  | <b>0.14</b> |
|          | 원문    | 0.23        |          | 원문   | 0.08        |
| 회원국 감산   | 요약문   | <b>0.12</b> | 종료       | 요약문  | <b>0.07</b> |
|          | 원문    | 0.05        |          | 원문   | 0.01        |
| 비회원국 감산  | 요약문   | <b>0.10</b> | 리비아      | 요약문  | <b>0.09</b> |
|          | 원문    | 0.07        |          | 원문   | 0.00        |
| 산유국 감산   | 요약문   | <b>0.07</b> | 무아마르     | 요약문  | <b>0.08</b> |
|          | 원문    | 0.06        |          | 원문   | 0.01        |
| 감산 합의    | 요약문   | 0.18        | 기름값 할인   | 요약문  | <b>0.05</b> |
|          | 원문    | <b>0.20</b> |          | 원문   | <b>0.05</b> |
| 감산 동참    | 요약문   | <b>0.15</b> | 리비아 반군   | 요약문  | <b>0.03</b> |
|          | 원문    | 0.09        |          | 원문   | 0.02        |

가 전쟁', '사우디아라비아 러시아'는 '전쟁·테러' 카테고리, '폭락', '급락', '마이너스', 'V자형곡선'은 '국제 경제 쇼크' 카테고리로 구분하였다. 2020년 11월은 주문량의 큰 증가로 인해 재고량이 크게 증가하였다. 여전히 신종 코로나바이러스에 영향력이 크게 작용하나, 백신의 등장으로 일상으로의 회복이 기대되었다. 이에 따른 경제활동의 증가가 휘발유 주문량 증가에 영향을 미쳤고, 이에 재고량이 증가한 것으로 해석할 수 있다. 이와 관련하여 '코로나', '백신', '코로나 백신', '코로나 여파' 등의 단어 및 구가 주요 이슈로 도출되어 '감염증' 카테고리로 구분하였다.

## 2. 요약문을 활용한 주요 이슈 파악

비교 실험을 통해 KoBART 기반의 한국어 텍스트 요약 모델을 활용해 생성한 요약문을 통해 추출한 주요 이슈가 원 기사를 통해 추출한 주요 이슈보다 더욱 명확하고 높은 중요도를 가지는 것을 확인하였다. 표 2는 요약문과 원 기사에 나타난 주요 단어 및 구의 TF-IDF 결과 값을 비교한 표이다. 2017년 1월에는 '감산'이 요약문 0.38, 원문 0.31로 가장 높은 중요도를 보였으며, 요약문이 원문보다 더 높은 TF-IDF 결과 값을 가지는 것을 확인하였다. 주요 단어 및 구로 확인된 '비회원국', '회원국 감산', '비회원국 감산', '산유국 감산', '감산 동참', '감산 합의' 또한 원문보다 요약문에서 그

중요도를 더 높게 확인할 수 있었다. 그러나, '감산 합의'의 경우 원문의 중요도가 약 0.02 높게 나타났다. 이는 원유 감산 합의와 관련된 대다수의 기사 원문에서 원유 감산 합의의 주요 국가, 조건, 상황, 가능성 등을 상세하게 설명하며 해당 구가 반복적으로 등장했기 때문이다. 2011년 7월에는 '기름값'이 가장 높은 중요도를 보였으며, 요약문에서 약 0.06 정도 그 결과 값이 높게 나오는 것을 확인하였다. '리비아'는 원문 기사에서 결과 값 확인이 불가하였다. TF-IDF는 일반적인 단어의 중요도를 낮게 평가하는데, 주요 산유국인 리비아의 석유 생산과 관련된 기사가 다수의 월에 발행되었기 때문이다.

## IV. 결 론

본 논문에서는 사회적, 경제적 이슈에 따른 휘발유의 소비 추이를 분석하기 위해 뉴스 기사를 활용해 주유소의 휘발유 재고량 변화 요인을 분석하였다. 2010년 1월 ~ 2022년 12월의 월평균 휘발유 재고량 수치 데이터를 기반으로 휘발유의 재고량이 크게 감소하거나, 크게 증가한 구간을 분석 구간으로 선정하였다. 분석 구간에 해당하는 시점에 발행된 휘발유 소비와 관련된 뉴스 기사를 웹 크롤링(Web Crawling) 기술을 활용해 자동으로 수집하고, 이를 KoBART(Korean Bidirectional and Auto-Regressive Transformers) 한국어 텍스트 요약 모델을 활용해 주요한 내용만을 포함하도록 그 내용을 요약하였다. 이후, 요약문을 전처리하여 명사 단위로 토큰화(Tokenization)하고, N-Gram 언어 모델(N-Gram Language Model)과 TF-IDF(Term Frequency Inverse Document Frequency)를 통해 주요 단어 및 구를 도출하여 이를 카테고리화 하였다. 재고량 변화에 영향을 미치는 카테고리는 '국내 유가 정책', '국제 유가 정책', '전쟁·테러', '국제 경제 쇼크', '감염증'으로 구분하였다. 이를 통해, 카테고리에 해당하는 단어 및 구가 뉴스 기사에 등장하면 주유소 재고량이 크게 변화한다고 예측할 수 있다. 또한 요약문과 원 기사에서 도출한 주요 단어 및 구의 TF-IDF 중요도 비교를 통해 요약문에서 주요 이슈를 상대적으로 잘 파악할 수 있음을 확인하였다. 이는 원문의 불필요하거나 반복적인 서술이 제거된, 핵심적인 사항만이 요약문에 포함되기 때문이다.

텍스트 요약 모델은 다양하게 활용이 가능하다. 향후 연구로는 도출한 단어 및 구를 기반으로 텍스트 요약 모델을 활용해 문장 단위의 주요 요인을 확인하고자 한다.

## References

- [1] Korean Statistical Information Service(KOSIS), Ministry of Trade, Industry and Energy, 2021.
- [2] K. S. Cha, "A Study on the Differences of Price Asymmetry and the Price Adjustment Process between Petroleum Products," *KUKJE KYUNGJE YONGU*, Vol. 26, No. 4, 2020.
- [3] D. H. Shin and H. H. Jo, "An Analysis on the Structural Break of Asymmetric Price Effects to Transport Energy Consumption: Evidence from Gasoline and Diesel Consumption in Korea," *AKES*, Vol. 34, No. 2, 2016.
- [4] H. Kim, "An Analysis of the Asymmetry of Domestic Gasoline Price Adjustment to the Crude Oil Price Changes: Using Quantile Autoregressive Distributed Lag Model," *Environmental and Resource Economics Review*, Vol. 31, No. 4, pp. 755-775, 2022.
- [5] K. Cha, "A Study on Price Asymmetry of the Retail Gasoline Market," *Journal of Budget and Policy*, Vol. 9, No. 4, pp. 31-61, 2020.
- [6] J. Bae, S. Kim, M. Kim., S. Oh., and E. Heo, "The Asymmetric Response of Gasoline Prices to International Grude Oil Price Changes Considering Inventories," *Envioronmental and Resource Economics Review*, Vol. 22, No. 4, pp. 643-670, December 2013.
- [7] H. Jang and B. Choi, "Effects of fuel tax cut on retail prices and its implications," *Korean Energy Economic Review*, Vol. 22, No. 1, pp. 205-228, March 2023.
- [8] B. Seo, "Machine-Learning-Based News Sentiment Index (NSI) of Korea," *Bank of Korea WP*, Vol. 2022, No. 15, September, 2022.
- [9] H. S. Jang, "Effects of temporary tax cut on retail prices: Evidence from Korean gas stations", *Korean Energy Economic Review*, No. 2, 2021.
- [10] A. S. Herbert, "The Choice of a Class Interval." *American Statistical Association*, Vol. 21, No. 153, pp. 65-66, 1926.
- [11] T. Wang, J. Song, D. Son, M. Kim, D. Choi, and J. Jang, "Web crawler Improvement and Dynamic process Design and Implementation for Effective Data Collection," *JKIICE*, Vol. 26, No. 11, pp. 1729-1740, 2022.
- [12] KOBART, <https://github.com/SKT-AI/KoBART>.
- [13] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," *arXiv:1910.13461*, 2019.
- [14] Open-source Korean Text Processor, [github.com/open-korean-text/open-korean-text](https://github.com/open-korean-text/open-korean-text).
- [15] S. J. Kim, "Exploring the Possibility of Using N-gram Feature for Automatic Scoring of Argumentative Writing Task," *Writing Research*, No. 41, pp. 37-62, 2019.
- [16] H. Shin and J. Choi, "Analysis of User Reviews for Webtoon Applications Using Text Mining," *JCCT*, Vol. 8, No. 4, pp. 457 - 468, July 2022.
- [17] S. J. Lee and H. J. Kim, "Keyword Extraction from News Corpus using Modified TF-IDF," *Journal of Society for e-Business Studies*, Vol. 14, No. 4, pp. 59-73, 2009.
- [18] H. S. Lee, "Rearch of Late Adolcent Activity based on Using Big Data Analysis," *IJACT*, Vol. 10, No. 4, pp. 361 - 368, Dec. 2022.