

# An Investigation into the Impact of Pretrained Language Models on Online News Sentiment and its Influence on Investor-Specific Trading Intensity

## 사전학습모델을 활용한 온라인 뉴스 감성이 투자자별 거래강도에 미치는 영향 분석에 관한 연구

Sungho Park<sup>1</sup>, Heewon Lee<sup>2</sup>, Youngbuk Jo<sup>3</sup>, Kangbae Lee<sup>4</sup>

박성호<sup>1</sup>, 이희원<sup>2</sup>, 조용복<sup>3</sup>, 이강배<sup>4</sup>

<sup>1</sup> Post-Doctor, MIS Department, Donga-University, Korea, [psh2975@dau.ac.kr](mailto:psh2975@dau.ac.kr)

<sup>2</sup> Senior Researcher, AI Researching Department, EMRO, co., Ltd., Korea, [slcw@naver.com](mailto:slcw@naver.com)

<sup>3</sup> Professor, MIS Department, Donga-University, Korea, [ybcho@dau.ac.kr](mailto:ybcho@dau.ac.kr)

<sup>4</sup> Professor, MIS Department, Donga-University, Korea, [Kanglee@dau.ac.kr](mailto:Kanglee@dau.ac.kr)

Corresponding author: Kangbae Lee

**Abstract:** This study delves into the intricate relationship between online news sentiment and the trading intensity of various types of investors, while empirically investigating the divergent trading behaviors exhibited by these investor groups. To gauge the trading intensity of each investor, the study employs the net purchase price as a proxy variable, which serves as the dependent variable. As the independent variable, the study leverages the emotional content of online news articles. The research methodology encompasses a thorough preprocessing phase for the text data extracted from online news sources in the Korean language. This preprocessing stage is pivotal in ensuring the data's quality and consistency. Subsequently, the study utilizes KB-BERT, a state-of-the-art pre-trained model, to extract sentiment or sensibility from the online news articles. This advanced natural language processing technique enables a nuanced understanding of the emotional content, providing valuable insights into the affective dimension of the news. The findings of this research substantiate that individual investors are more susceptible to the emotional sway of online news than institutional investors and foreign investors. This empirical evidence reinforces existing research outcomes that highlight the irrational behaviors often exhibited by individual investors in response to emotional stimuli from news sources. In essence, the study underscores the pivotal role of online news sentiment in shaping investor decision-making processes and market dynamics, shedding light on the nuanced interplay between information, emotions, and financial markets. These insights contribute to a deeper understanding of market behavior and can inform strategies for more informed and rational investing.

**Keywords:** Online New Sentiment, Pre-Trained Model, BERT, KOPSI, Net Buying Amount by Investor Type, Trading Intensity

**요약:** 본 연구는 온라인 뉴스 감성이 투자자별 거래강도에 미치는 영향을 분석하여, 투자자별 거래행태의 차이를 실증적으로 검증한다. 투자자별 거래강도를 측정하기 위한 대용변수로

Received: June 27, 2023; 1<sup>st</sup> Review Result: August 03, 2023; 2<sup>nd</sup> Review Result: September 06, 2023  
Accepted: September 25, 2023

투자자별 순매수 대금을 종속변수로 사용하였다. 독립변수는 온라인 뉴스의 감성을 사용하였다. 온라인 뉴스의 텍스트데이터에 대한 한글 전처리 프로세스를 수행 후 사전학습 모델인 KB-BERT를 활용하여 온라인 뉴스의 감성을 추출하였다. 실증분석 결과 개인은 온라인 뉴스 감성에 기관과 외국인보다 큰 영향을 받음을 확인하였다. 또한, 변동성 지수에 기관과 외국인은 음의 영향을 받았으며, 개인투자자는 양의 영향을 받음을 확인하였다. 이는 기관과 외국인은 시장심리를 객관적으로 활용하였으나, 개인은 시장심리를 역방향으로 활용한 것으로 개인의 역투자 행태를 설명할 수 있다. 연구결과를 통해 개인은 기관과 외국인보다 온라인 뉴스 감성에 큰 영향을 받는 것을 확인하여 개인투자자의 비이성적 행태와 관련한 연구결과를 실증적으로 뒷받침한다.

**핵심어:** 온라인 뉴스, 온라인 뉴스 감성, 사전학습 모델, BERT, KOSPI, 투자자별 순매수 대금, 거래강도

## 1. 서론

2019년 코로나 발생 이후 국내 주식시장에서 개인투자자의 비중은 많이 증가하였으며, 2018년부터 특히 2021년까지의 기간 개인투자자의 연간 순매수 대금은 10.93조, -5.48조, 63.92조, 76.93조로 큰 폭으로 증가하였다. 그러나 이러한 개인투자자 중 신규투자자의 60%가 손실을 기록하였다. 이에 반해 기관투자자의 순매수 대금은 순매수 대금은 -3.91조, 5.32조, -36.09조, -44.09조이며 외국인투자자의 순매수 대금은 -6.33조, 0.75조, -24.73조, -25.96조로 감소하였다. 이 과정에서 기관투자자와 외국인투자자는 수익을 기록하였으나, 개인투자자 손실을 기록했음을 설명한다. 본 논문은 코로나 19 기간동안 개인투자자의 손실을 설명하고자 한다. 개인투자자는 검색량이 높은 주식과 온라인 온라인 뉴스와 같은 접근하기 쉬운 정보에 의존하여 의사결정을 내리는 오류를 범하는 행태적 편의를 보인다[1]. 본 연구는 주식거래에서 대표적인 정보 획득 수단인 온라인 뉴스가 투자자별 거래강도에 미치는 영향의 차이를 비교 분석하여 개인투자자의 손실을 설명하고자 한다. 투자자별 거래강도를 측정하기 위해 일별 매수와 매도의 차이인 투자자별 순매수 대금을 대용변수로 사용하였다. 분석대상은 KOSPI 시장으로 한정한다. 연구 기간은 2018년 1월부터 2021년 12월까지 코스피 시장의 투자자별 순매수 거래대금, 온라인 뉴스, 경제지표, 기술적 지표를 일별로 수집하였다. 온라인 뉴스의 감성을 수치화하여 분석하기 위해 금융 분야 감성분석에 특화된 모형인 KB-BERT를 활용하였다. 수치화된 온라인 뉴스의 감성점수를 독립변수로 설정하였다. 독립변수의 영향을 효과적으로 확인하기 위해 통제변수로 경제적, 기술적 지표를 선택하여 회귀분석을 실시하였다. 연구의 주요 분석결과는 다음과 같다. 첫째, 투자자별로 온라인 뉴스 감성, 경제지표, 기술적 지표에 다른 영향을 받음을 확인하였다. 둘째, 기관투자자와 외국인투자자는 설정된 변수들에 유사한 영향을 받는 것을 확인하였다. 기관투자자와 외국인 투자자는 온라인 뉴스 감성에 양의 영향을 받았으며, 시장의 공포심리를 반영하고 있는 변동성 지수에 음의 영향을 받았다. 이를 통해 기존연구 결과인 기관투자자와 외국인투자자의 유사성을 확인하였다[1]. 셋째, 개인투자자는 기관 및 외국인투자자에 비해 온라인 뉴스의 감성에 다른 영향을 받은 것을 확인하였다. 개인투자자는 기관 및 외국인투자자와 달리 온라인 뉴스 감성에 가장 큰 영향력을

받았으며 음의 영향을 받았다. 이는 개인투자자는 온라인 뉴스라는 시장의 정보를 기관투자자와 외국인투자자에 비해 반대로 받아들였다고 해석할 수 있다. 또한 개인투자자는 시장의 심리를 대변할 수 있는 변동성 지수에 가장 큰 영향을 받았으며, 기관투자자와 외국인투자자에 비해 양의 영향을 받음을 확인했을 때, 코로나19 기간 개인투자자의 거래는 객관적 정보에 기반한 투자가 아닌, 다양한 행태적 편의에 기반한 비이성적 투자임을 설명할 수 있다. 본 연구의 구성은 다음과 같다. 1장은 서론이고 2장은 선행연구를 통해 순매수에 미치는 영향과 관련된 기존의 연구를 정리하였다. 3장은 연구자료 및 방법론을 설명하고, 4장은 연구결과를 제시한다. 마지막으로 5장은 결론으로 연구의 결론과 시사점을 제시한다.

## 2. 이론적 배경

본 장에서는 본 연구와 관련된 국내외 선행연구를 살펴본다. 경제지표와 관련된 선행연구를 분석한다. 통화량, 회사채, 다우지수 등 9개의 변수가 코스피 지수에 영향을 미치는 것을 확인하였다[2]. 국내 총생산, 무역수지, 경제성장률, 실업률 등의 경제지표와 회귀분석을 활용하여 소비자물가지수, 금리, 주가변동에 대한 영향을 확인하였다[3].

코스피 월별 수익률과 투자자별 순매수 거래대금을 상관분석과 다중회귀분석을 활용하여 투자자 집단별 상관계수를 도출하여 KOSPI와 개인투자자의 순매수는 음의 관계, 외국인투자자는 양의 상관관계, 기관투자자는 유의한 상관관계가 없음을 확인하였으며, 외국인, 기관투자자의 순매수와 KOSPI 주가 수익률 간의 관계를 검증하여 외국인투자자와 기관투자자의 순매수는 KOSPI 수익률에 영향을 주는 것을 확인하였다[4][5].

개인, 기관투자자의 순매수를 구분하여 순매수 압력이 코스피 200의 주가에 영향을 미친다는 것을 검증하였으며, S&P500 지수와 순매수의 압력 사이에 관계를 확인하였다[6]. 코스피, 코스닥 업종별 지수가 투자자 유형별 순매수와 기술적 지표와 펀더멘탈 지표를 사용하여 시장에 미치는 영향을 검증하였다[7][8].

유가증권 상장기업의 증권가 보고서를 분석하여 기관, 개인, 외국인 투자자별로 정보 비대칭이 발생하는 것을 확인하였으며, 모바일 이용자인 개인투자자의 주식 거래 데이터를 분석하여 개인투자자의 역투자 행태를 확인하였다[9][10]. 투자심리에 따른 수익률 동조화 효과를 분석하여 개인투자자의 매매가 수익률의 동조화를 발생시킨다고 주장하였다. 외국인, 기관투자자의 거래 데이터를 분석하여 외국인과 기관투자자는 정보거래자의 형태를 보이지만, 외국인이 정보에 있어서 우위에 있음을 확인하였다 [11][12].

온라인 뉴스를 활용한 키워드에 대한 Word2Vec을 활용하여 감성사전을 구축하였으며, 이를 활용하여 감성분석을 실시하여 온라인 뉴스의 감성이 주요한 변수임을 증명하였다[13]. 온라인 뉴스와 감성사전을 활용한 연구로 감성사전 구축을 통해 온라인 뉴스를 긍정/부정으로 분류하였으며, 이를 통해 순매수 압력과의 영향을 확인하여, 감성사전의 필요성을 확인하였다[14]. 또한 온라인 뉴스를 각 주제에 할당된 기사들을 대상으로 구축된 어휘사전을 기반으로 감성점수를 계산하여 기계학습에 적용하여 경제심리지수를 적용하였으나, 어휘사전은 한글 감성 사전에 대한 지속적인 갱신의 어려움과 단어 기반의 분석으로 전체 문맥을 파악하는데 어려움이 있음을 확인하였으며[14], 어휘사전 기반 온라인 소비자 감성지수를 개발하여 유니그램을

중심으로 한 빈도분석, 연관성 분석과 어휘자원의 검토의 필요성을 확인하였다[15]. 어휘사전 기반 감성분석의 연구를 분석한 결과 어휘사전에 대한 한계점으로 지속적인 업데이트의 필요성, 감정 어휘사전 구축에 필요한 비용, 단어 기반의 분석으로 문장과 문맥을 고려한 감성분석의 어려움을 확인하였다.

어휘사전 기반 감성분석의 어려움을 극복하기 위해 최신 기술인 사전학습 기반의 해외연구로 나스닥 지수와 아마존 뉴스 데이터를 사전학습모델인 FinBERT를 활용하여 온라인 뉴스가 주가변동에 영향이 있음을 검증하였으며[16], 금융 분야의 코퍼스를 사전학습하여 다양한 감성분석 모델과 FinBERT를 비교분석하여 FinBERT가 가장 우수한 성능을 보임을 확인하였다[17]. 국내에서는 사전학습모델을 활용한 연구로 BERT를 활용하여 뉴스 감성분석을 수행하여 거시경제지표와 조합을 통해 주가지수를 예측하여 사전학습 기반의 뉴스 감성분석의 효과성을 검증하였다[18]. 기존의 연구를 정리하면 경제지표, 기술적 지표, 온라인 뉴스가 개별적이거나 부분적으로 투자자별 거래강도에 미치는 영향은 다름을 확인할 수 있었다. 그러나 경제지표, 기술적 지표, 온라인 뉴스가 투자자별 거래강도에 미치는 영향을 종합적으로 분석하거나, 온라인 뉴스의 감성을 사전학습기반으로 계량화하여 사용한 연구는 부족한 실정이다. 본 연구는 개인투자자와 기관, 외국인 투자자의 순매수에 온라인 뉴스 영향의 차이를 분석하기 위하여 온라인 뉴스의 감성을 기존에 주로 쓰였던 어휘사전 기반의 방법이 아닌 사전학습모델을 사용하여 수치화하였으며, 선행연구에서 주로 쓰인 경제지표와 기술적 지표를 분석하였다.

### 3. 연구자료 및 방법론

#### 3.1 연구자료 및 기술적 통계

본 연구의 연구 기간은 2018년부터 2021년까지이다. 종속변수는 기관, 외국인, 개인투자자의 일별 순매수 대금을 설정하였다. 분석에 사용된 독립변수와 통제변수는 [표 1]과 같다. 온라인 뉴스는 시장의 심리를 측정할 수 있는 대용변수로 온라인 뉴스의 감성을 사용하여 독립변수로 설정하였다. 선행연구를 참고하여 통제변수로 경제지표, 기술적 지표를 각각 2개씩 설정하였다. 경제지표로 국내 경제 현황을 대표할 수 있는 실업률과 시장 금리를 반영할 수 있는 국고채 10년물을 사용하였다. 기술적 지표로는 시장의 심리를 측정할 수 있는 변동성 지수와 시장의 유동성을 측정할 수 있는 고객예탁금을 설정하였다.

[표 1] 데이터 설명

[Table 1] Data Description

| Division            | Variable Description | Variable Name     |
|---------------------|----------------------|-------------------|
| Online              | Emotional Score      | Sentiment         |
| Economy Indicator   | Unemployment rate    | unemployment rate |
|                     | 10-year KTB          | gb10              |
| Technical Indicator | Volatility Index     | vix               |
|                     | Customer Deposit     | customer deposit  |

### 3.1.1 정형 데이터

본 연구에서 사용된 정량 데이터는 투자자별 순매수 대금, 경제지표, 기술적 지표로 구분된다. 종속변수인 투자자별 순매수 대금은 기관, 외국인, 개인투자자의 일별 순매수 대금을 사용하였다. 경제지표는 일별 데이터와 월별 데이터로 구분된다. 일별 데이터는 국고채 10년물로 국내 시장의 시장 금리를 반영하고 있다. 월별 데이터는 실업률로 월별 데이터로 구성되어 있다. 월별 데이터는 선형보간을 통해 일별 데이터로 변환하여 사용하였다. 기술적 지표는 시장의 변동성을 확인할 수 있는 VIX 지표와 시장의 유동성을 측정할 수 있는 고객예탁금을 일별 데이터 수집하여 사용하였다. 경제지표 2개, 기술적 지표 2개를 선행연구를 참조하여 선정하였다.

### 3.1.2 비정형 데이터

비정형 데이터는 2018년 ~ 2021년 사이의 온라인 뉴스 193,662건을 수집하였다. 온라인 뉴스를 수집하기 위해서 Python의 BeautifulSoup를 사용하였다. 네이버는 2022년 1분기 기준으로 64.76%의 검색 시장에서의 점유율을 확보하고 있으며, 네이버 금융 뉴스는 국내 대부분의 언론사의 뉴스를 포함하고 있다. 네이버 뉴스에서 “코스피” 키워드로 검색하여 해당 키워드가 포함된 뉴스를 수집하였다. 온라인 뉴스 데이터에는 특수문자, 숫자 등의 불필요한 문자, 중복된 데이터는 데이터 전처리 과정을 통해 제거하였다.

### 3.1.3 KB-BERT를 활용한 온라인 뉴스 감성점수

본 연구에서는 온라인 뉴스의 감성점수를 도출하기 위해 FinBERT를 활용하였다. BERT(Bidirectional Encoder Representations for Transformers)모델은 문맥기반의 자연처리 분야에서 우수한 성능을 보이고 있다. BERT는 기존의 Word2Vec과 특정 단어의 앞뒤 관계만 고려하는 것이 아니라 임의적인 마스킹을 활용하여 단어와 문장을 예측한다. 마스킹은 문장 중 특정 단어를 제외하고 주변 단어의 문맥을 고려하여 제외된 단어를 예측하는 방법이다. BERT 구조는 단방향뿐만 아니라 양방향 분석에 용이하기 때문에 기존의 임베딩 방법과 비교하면 단어 사이의 관계를 잘 반영할 수 있다. FinBERT 금융분야에 특화된 BERT로 기존의 BERT에 180만개의 뉴스기사로 구성된 금융 텍스트의 말충치를 학습한 모델이다. FinBERT는 기존의 BERT 방법론에 비하여 금융분야의 감성분류에서 정확도가 10~15% 향상되었다. 특히 최근 국민은행에서 개발된 KB-BERT가 있다. KB-BERT는 총 90GB의 코퍼스를 사용되었다. 기존의 한글관련 사전학습 모델보다 30GB 증가하였다. 전체의 40%인 40GB는 경제 관련 뉴스와 다양한 금융 문서로 구성되었다. 본 논문에서는 어휘사전을 구축하지 않고, 온라인 뉴스에 대한 감성을 수치화하기 위해 사전학습모델인 KB-BERT를 활용하였다. KB-BERT는 금융 도메인 감성 분석에 필요한 금융 시장의 낙관 및 부정적 의견에 대한 감성을 판단할 수 있는 평가하도록 설계 되어있다. KB-BERT에는 대용량의 온라인 뉴스 기사가 고성능 컴퓨팅 환경에서 학습되어 있어, 본 연구에서 사용한 금융 분야의 온라인 뉴스 감성을 추출하는데 적합하다. 이를 활용하여 본 연구에서 수집된 온라인 뉴스를 KB-BERT 감성분석 모델에 적용하여 감성점수를 추출하였다. 감성점수는 KB-BERT에서 사용된 감성 분류에 대한 확률 값을 사용하였다.

### 3.2 연구 모형

연구 모형 (1)은  $i$ 는 기관, 외국인, 개인투자자이다.  $p$ 는 국고채 10년물, 실업률이며,  $q$ 는 변동성 지수, 고객예탁금이다. 종속변수는 투자자별 일별 순매수 대금이며, 주요 독립변수인 **Sentiment**는 일별 온라인 뉴스의 감성지수 평균을 사용하였다. 통제변수로 주요 경제지표인 국고채, 실업률과 기술적 변수인 변동성 지수, 고객예탁금을 사용하였다. 국고채와 실업률은 국가경제 상황을 객관적으로 설명할 수 있는 수치로 선행연구를 참고하여 거시경제변수로 설정하였다. 또한, 시장의 공포심을 반영하는 변동성 지수와 시장의 유동성을 확인하여, 시장심리를 간접적으로 측정할 수 있는 고객예탁금을 기술적 지표로 설정하였다.

$$Y_i = B_0 + B_i \text{Senti}_i + B_{i,p} \text{Eco}_{i,p} + B_{i,q} \text{Tec}_{i,q} + \epsilon \quad (1)$$

$Y$  = 순매수 대금

$\text{Senti}$  = 온라인 뉴스 감성

$\text{Eco}$  = 경제 지표

$\text{Tec}$  = 기술적 지표

$i$  = 기관, 외국인, 개인

$p$  = 국채 10년물, 실업률

$q$  = 변동성 지수, 고객 예탁금

## 4. 분석결과

### 4.1 기술적통계

[표 2] 기술적 통계

[Table 2] Descriptive Statistics

| Category            | Variable                 | Mean        | Standard Deviation | Min        | Max         |
|---------------------|--------------------------|-------------|--------------------|------------|-------------|
| Net Purchase        | Institutional Net Buying | -47,184     | 412,725            | -3,743,204 | 1,518,870   |
|                     | Foreign Net Buying       | -65,145     | 381,572            | -2,817,359 | 1,706,880   |
|                     | Individual Net Purchases | 105,569     | 589,187            | -1,997,574 | 4,492,145   |
| Economic Indicator  | Unemployment rate        | -47184      | 412725             | -3743204   | 1518870     |
|                     | 10-year KTB              | 2440.09     | 430.15             | 1457.64    | 3305.21     |
| Technical Indicator | Volatility Index         | 108.54      | 8.01               | 95.01      | 130.92      |
|                     | Customer Deposit         | 102,180,494 | 61,681,260         | 107,122    | 166,667,690 |
| Emotional Score     | Sentiment                | -0.01       | 0.10               | -0.48      | 0.49        |

본 연구에서는 투자자별 순매수 금액의 변동이 컸던 2018년 1월부터 2021년까지 연구 기간으로 설정하여, 투자자별 순매수 금액, 온라인 뉴스, 경제지표, 기술적 지표를 수집하였다. 수집된 데이터는 연속형 데이터인 순매수 거래대금, 경제지표 기술적 지표와 텍스트데이터인 온라인 뉴스로 구분된다. 경제지표 20개, 기술적 지표 11개의 변수를 초기로 선정하였으나, 다중공선성 문제가 발생하여, 이를 해결하기 위해 변수 선택 단계선택법을 활용하였다. 본 연구에서는 투자심리를 분석하기 위해서 온라인 뉴스를 추가하였다. [표 2]는 전체 기술적 통계이다. 경제지표 2개, 기술적 지표 2개, 온라인 뉴스에서 추출한 감성은 1개, 투자자별 순매수 3개로 총 6개의 변수를 수집하였다. 온라인 뉴스의 감성을 독립변수로 활용하였으며, 경제지표인 실업률과 국채 10년물, 기술적 지표인 고객예탁금과 변동성 지수는 통제변수로 사용하였다. 거래강도를 측정하기 위한 대응변수로 투자자별 순매수 대금은 종속변수로 사용하였으며, 아래 표와 같이 투자자별로 순매수 대금에 평균과 표준편차의 차이가 있다. 특히 개인투자자의 순매수 대금의 변동은 다른 투자자 집단보다 크게 나타났다.

#### 4.2 기관투자자 회귀분석 결과

기관투자자의 거래강도에 대한 회귀분석 결과는 [표 3]과 같다. 온라인 뉴스의 감성, 고객예탁금, 실업률, 국채에 대하여 통계적으로 유의한 결과를 보였다. 기관투자자의 경우 온라인 뉴스의 감성이 긍정적일수록 거래강도는 증가하는 것을 확인할 수 있다. 고객예탁금이 증가할수록 순매수 대금은 감소하는 것을 확인하였는데 기관투자자는 시장의 유동성에 부정적으로 반응함을 확인하였다. [표 3]은 온라인 뉴스가 기관투자자의 순매수 대금에 미치는 영향을 확인하기 위한 회귀분석 결과이다.

[표 3] 기관투자자 회귀분석 결과

[Table 3] Regression Analysis Results for Institutional Investors

| Variable          | Institutional Investors |
|-------------------|-------------------------|
| sentiment         | 593,800(5.681)***       |
| customer deposit  | -0.3572(-5.782)***      |
| unemployment rate | -39,680(-2.678)***      |
| VIX               | -1,912.587(1.125)       |
| gb10              | 110,000(5.045)***       |

표는 온라인 뉴스가 기관투자자의 순매수 대금에 미치는 영향을 확인하기 위한 회귀분석 결과이다. 회귀모형은 다중회귀모형이다. 표의 숫자는 내림차순으로 정렬된 회귀계수이고, 괄호 안의 값은 t-값이고, \*\*\*, \*\*, \*은 각각 1%, 5%, 10% 각각 통계적 유의하다.

#### 4.3 외국인투자자 회귀분석 결과

외국인투자자의 거래강도에 대한 회귀분석 결과는 [표 4]와 같다. 외국인투자자는 기관투자자와 비슷한 결과를 나타냈다. 온라인 뉴스가 긍정적이면 순매수 대금은 증가하였다. 고객 예탁금이 증가할수록 외국인투자자의 거래강도는 감소하였다. 이를 통해 외국인투자자는 온라인 뉴스의 감성에 대칭적으로 영향을 받고, 고객 예탁금인 유동성에 부정적으로 반응하는 것을 확인하였다.

[표 4] 외국인투자자 회귀분석 결과

[Table 4] Results of Regression Analysis on Foreign Investors

| Variable          | Foreign Investors  |
|-------------------|--------------------|
| sentiment         | 679,700(7.252)***  |
| customer deposit  | -0.011(-0.011)     |
| unemployment rate | 62,300(4.689)***   |
| VIX               | -14,720(-9.958)*** |
| gb10              | -5,805.14(-0.297)  |

표는 온라인 뉴스가 기관투자자의 순매수 대금에 미치는 영향을 확인하기 위한 회귀분석 결과이다. 회귀모형은 다중회귀모형이다. 표의 숫자는 내림차순으로 정렬된 회귀계수이고, 괄호 안의 값은 t-값이고, \*\*\*, \*\*, \*은 각각 1%, 5%, 10% 각각 통계적 유의하다.

#### 4.4 개인투자자 회귀분석 결과

개인투자자의 거래강도에 대한 회귀분석 결과는 [표 5]와 같다. 개인투자자의 경우 온라인 뉴스가 긍정적일수록 거래강도는 감소하는 것으로 나타났다. 이는 개인은 기관과 외국인과 비교하면 온라인 뉴스의 감성에 대해 반대의 영향을 확인하였다. 이는 외국인투자자는 개인투자자와 다른 투자행태를 보인다는 기존의 연구를 지지하는 것을 확인하였다[12].

[표 5] 개인투자자 회귀분석 결과

[Table 5] Individual Investor Regression Analysis Results

| Variable          | Individual investor   |
|-------------------|-----------------------|
| sentiment         | -1,286,000(-8.904)*** |
| customer deposit  | 0.411(4.813)***       |
| unemployment rate | -14,420(-0.704)       |
| VIX               | 11,310(4.817)***      |
| gb10              | -118,000(-3.913)***   |

표는 온라인 뉴스가 기관투자자의 순매수 대금에 미치는 영향을 확인하기 위한 회귀분석 결과이다. 회귀모형은 다중회귀모형이다. 표의 숫자는 내림차순으로 정렬된 회귀계수이고, 괄호 안의 값은 t-값이고, \*\*\*, \*\*, \*은 각각 1%, 5%, 10% 각각 통계적 유의하다.

#### 4.5 투자자별 거래강도 비교분석

투자자별 온라인 뉴스에 대한 영향은 투자자별로 다른 [표 6]을 통해 확인할 수 있다. 주요 분석결과로는 기관과 외국인은 독립변수인 온라인 뉴스에는 양의 영향을 받으며, 변동성 지수에는 음의 영향을 받음을 확인하였다. 반면 개인투자자는 기관과 외국인과 상반된 결과로 온라인 뉴스 감성에는 음의 영향을 받았으며, 변동성 지수에는 양의 영향을 받음을 확인하였다. 이는 개인투자자는 기관투자자와 외국인투자자에 비해 정보에 비대칭 형태를 보인다는 기존의 연구를 지지하는 결과이다[12]. 또한, [표 6]의 결과는 통해 코로나19 기간에 개인투자자의 손실과 온라인 뉴스에 대한 차이와 변동성 지수에 대한 비대칭성을 통해 개인투자자를 노이즈 거래자로 구분한 기존연구를 지지한다. 투자자별 순매수 대금에 다른 영향을 미치는 것을 실증적으로 검증하였으며,

온라인 뉴스의 통계적 유의성을 확인하였다.

[표 6] 투자자별 회귀분석 결과

[Table 6] Regression Analysis Results for Each Investor Type

| Variable          | Institutional Investors | Foreign Investors   | Individual investor |
|-------------------|-------------------------|---------------------|---------------------|
| sentiment         | 60,570(5.681)***        | 69,550(7.252)***    | -131,400(-8.904)*** |
| customer deposit  | -85,210(-5.782)***      | -7,808(-0.011)      | 102,400(4.813)***   |
| unemployment rate | -70,010(-2.678)***      | 24,470(4.689)***    | -54,250(-0.704)     |
| VIX               | -1,157(-1.125)          | -116,000(-9.958)*** | 107,100(4.817)***   |
| gb10              | 4,285(5.045)***         | -14,420(-0.297)     | -7,769(-3.913)***   |

표는 온라인 뉴스가 기관투자자의 순매수 대금에 미치는 영향을 확인하기 위한 회귀분석 결과이다. 회귀모형은 다중회귀모형이다. 표의 숫자는 내림차순으로 정렬된 회귀계수이고, 괄호 안의 값은 t 값이고, \*\*\*, \*\*, \*은 각각 1%, 5%, 10% 각각 통계적 유의하다.

## 5. 결론

본 연구는 KOSPI 시장에서 투자자 집단별로 경제지표, 기술적 지표, 온라인 뉴스의 영향이 다름을 확인하였다. 이를 위해 투자자별 순매수 대금을 종속변수로 설정하고 독립변수로 온라인 뉴스 감성, 통제변수로 경제지표, 기술적 지표를 사용하였다. 본 논문은 투자자별 순매수에 미치는 영향의 차이를 검증하였다. 개인투자자의 순매수와 인터넷 검색량과의 상관성을 확인하였는데, 본 연구에서는 온라인 뉴스의 감성과 순매수와의 영향을 검증하였다[17]. 선행연구에 따르면 주식시장의 변동성은 거래정보에 영향을 받으며, 개인투자자의 경우 기관과 외국인 투자자보다 더 큰 영향을 받는다고 하였는데, 본 논문의 연구결과를 통해 개인투자자가 시장의 정보에 더 큰 영향을 받음을 확인하였다[18]. 기존의 연구에서는 개인투자자는 노이즈투자자로 정의하였는데 본 연구에서 확인한 개인투자자의 역투자 성향을 확인하여 연구결과를 검증하였다. 본 연구의 학문적 기여는 기존의 어휘사전기반을 구축하여 단어 기반의 감성분석이 아닌, 사전학습모델을 활용하여 온라인 뉴스의 감성을 수치화하여, 대용량의 데이터셋이 학습되어 있는 사전학습모델을 활용하여 문맥을 고려한 감성분석을 수행한 것이다. 금융분야의 온라인 뉴스는 단어만으로는 감성을 분석하기 어려움이 있기 때문에, 전체 문장의 문맥을 고려할 수 있는 사전학습모델이 적합하다. 또한, KOSPI 시장의 주요 참여자인 외국인투자자, 기관투자자, 개인투자자의 거래에 미치는 영향을 유형별로 구분하여 분석하여 개인투자자와 기관, 외국인 투자자의 순매수 대금에 온라인 뉴스라는 정보의 영향을 실증검증한 것이다. 본 연구의 실무적 기여로는 대용량의 데이터를 학습한 사전학습 모델을 통해 전이학습을 수행하면, 필요한 데이터의 양과 컴퓨팅 파워가 절약될 수 있다. 분석결과를 해석하면, 외국인과 기관투자자는 긍정적, 부정적 정보에 대칭적으로 영향을 받으며, 개인투자자는 긍정적, 부정적 정보에 대해 비대칭적 영향을 받는 경향이 있음을 확인하였다. 이는 기존의 연구에서 확인된 개인투자자의 편향 효과를 온라인 뉴스의 감성을 통해 실증한 것이다. 개인투자자는 정확한 정보를

바탕으로 이성적으로 거래하는 것이 아닌 부정확한 정보나 자신들의 경험에 의존하는 비이성적 거래를 하는 것으로 해석할 수 있다. 개인투자자의 비중이 증가하고, 이들의 비이성적 거래행태는 시장의 왜곡, 비효율성 문제를 증가시킨다. 이러한 문제해결을 위해 개인투자자의 거래행태를 이해하는 것이 선행되어야 한다. 이를 위해 본 연구는 개인투자자의 순매수에 온라인 뉴스가 미치는 영향을 분석하였는데, 이를 통해 개인투자자의 거래행태를 분석할 수 있는 기초 연구를 제공할 수 있을 것으로 기대한다.

## References

- [1] B. C. Go, J. Kim, Who Drive the Rise and Fall of the Bubbles in Korean Stock Market?, *Journal of Derivatives and Quantitative Studies*, (2017), Vol.25, No.4, pp.591-622.
- [2] C. H. Ahn, C. Y. Choi, The Empirical Study of Variation of KOSPI Index & Macro Economic Variation, *International Commerce and Information Review*, (2010), Vol.12, No.4, pp.171-192.  
DOI: <https://doi.org/10.15798/kaici.12.4.201012.171>
- [3] S. J. Kim, The Macroeconomic Influence on Hospitality Firms' Stock Returns, *Korean Journal of Hospitality and Tourism*, (2012), Vol.21, No.6, pp.309-327.  
UCI: G704-001024.2012.21.6.014
- [4] M. H. Gu, Y. S. Lee, A Study on the Relationship between Investor Groups and Stock Prices, *The Korean Journal of Financial Management*, (2001), Vol.18, No.1, pp.43-66.
- [5] S. K. Kim, Y. T. Byeon, The Relation between Net Purchase of Foreign and Institution Investors and Expected Returns in the Korea Stock Market, *Management & Information Systems Review*, (2011), Vol.30, No.4, pp.23-44.  
DOI: <https://doi.org/10.29214/damis.2011.30.4.002>
- [6] D. J. Ryu, D. W. Ryu, H. J. Yang, The impact of net buying pressure on index options prices, *The Journal of Futures Markets*, (2021), Vol.41, No.1, pp.27-45.  
DOI: <https://doi.org/10.1002/fut.22158>
- [7] N. P. B. Bollen, R. E. Whaley, Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?, *The Journal of FINANCE*, (2004), Vol.59, No.2, pp.711-753.  
DOI: <https://doi.org/10.1111/j.1540-6261.2004.00647.x>
- [8] J. S. Lee, H. S. Cho, K. I. Chung, J. S. Park, Feature Selection for Stock forecasting using Multivariate Convolution Neural Network, 2020 International Conference on Information and Communication Technology Convergence (ICTC), (2020)  
DOI: <https://doi.org/10.1109/ICTC49870.2020.9289492>
- [9] K. S. Kim, J. W. Park, Informativeness of Analyst Report and Information Asymmetry among Investor Types, *Korean Journal of Financial Studies*, (2012), Vol.41, No.4, pp.547-588.  
UCI: G704-SER00000670.2012.41.4.003
- [10] M. C. Woo, W. B. Lee, The Behavior and Performance of Mobile Traders on the Korea Stock Market, *Korean Journal of Financial Studies*, (2013), Vol.42, No.1, pp.133-161.  
UCI: G704-SER00000670.2013.42.1.011
- [11] Y. H. Jeon, H. Choi, Price-Based Return Comovements and Individual Investor Trading: Evidence from Stock Splits in the Korean Stock Market, *Korean Journal of Financial Studies*, (2013), Vol.42, No.2, pp.473-420.  
UCI: G704-SER00000670.2013.42.2.001
- [12] S. H. Yoon, H. Choi, Choice of Market versus Limit Orders by Foreign Investors in the Korean Stock Market, *Korean Journal of Financial Studies*, (2014), Vol.43, No.3, pp.461-497.  
UCI: G704-SER00000670.2014.43.3.004
- [13] D. Y. Kim, Y. I. Lee, News based Stock Market Sentiment Lexicon Acquisition Using Word2Vec, *The Korea Journal*

of BigData, (2018), Vol.3, No.1, pp.13-20.

DOI: <https://doi.org/10.36498/kbigdt.2018.3.1.13>

- [14] H. J. Kim, J. H. Im, H. Y. Lee, S. Lee, Development of Economic Sentiment Index Using Online News Articles, National Accounts Review, (2019), Vol.2.
- [15] Y. S. Kim, S. G. Hong, H. J. Kang, S. R. Jung, Electronic-Composit Consumer Sentiment Index(CCSI) development by Social Bigdata Analysis, Journal of Internet Computing and Services, (2017), Vol.18, No.4, pp.121-131.  
DOI: [10.7472/jksii.2017.18.4.121](https://doi.org/10.7472/jksii.2017.18.4.121)
- [16] T. Sidogi, R. Mbuva, T. Marwal, Stock Price Prediction Using Sentiment Analysis, In 2021 IEEE International Conference on Systems, Man, and Cybernetics, (2021)  
DOI: <https://doi.org/10.1109/SMC52423.2021.9659283>
- [17] Z. Liu, D. Huang, K. Huang, Z. Li, J. Zhao, FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining, Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence Special Track on AI in FinTech, pp.4513-4519, (2021)  
DOI: <https://doi.org/10.24963/ijcai.2020/622>
- [18] E. A. Jang, H. R. Choi, H. C. Lee, Stock prediction using combination of BERT sentiment Analysis and Macro economy index, Journal of the Korea Society of Computer and Information, (2020), Vol.25, No.5 pp.47-56.  
DOI: <https://doi.org/10.9708/jksci.2020.25.05.047>