

Stock Market Prediction Using Sentiment on YouTube Channels

Su-Ji Cho · Cheol-Won Yang · Ki-Kwang Lee[†]

School of Business Administration, Dankook University

유튜브 주식채널의 감성을 활용한 코스피 수익률 등락 예측

조수지 · 양철원 · 이기광[†]

단국대학교 경영학부

Recently in Korea, YouTube stock channels increased rapidly due to the high social interest in the stock market during the COVID-19 period. Accordingly, the role of new media channels such as YouTube is attracting attention in the process of generating and disseminating market information. Nevertheless, prior studies on the market forecasting power of YouTube stock channels remain insignificant. In this study, the market forecasting power of the information from the YouTube stock channel was examined and compared with traditional news media. To measure information from each YouTube stock channel and news media, positive and negative opinions were extracted. As a result of the analysis, opinion in channels operated by media outlets were found to be leading indicators of KOSPI market returns among YouTube stock channels. The prediction accuracy by using logistic regression model show 74%. On the other hand, Sampro TV, a popular YouTube stock channel, and the traditional news media simply reported the market situation of the day or instead showed a tendency to lag behind the market. This study is differentiated from previous studies in that it verified the market predictive power of the information provided by the YouTube stock channel, which has recently shown a growing trend in Korea. In the future, the results of advanced analysis can be confirmed by expanding the research results for individual stocks.

Keywords : Text Mining; YouTube; Classification; Sentiment Analysis; Stock Prediction

1. 서 론

주식시장에서 정보가 가격에 반영되는 과정은 중요한 연구주제이며, 정보에 대한 투자자들의 반응, 그리고 주식 수익률에의 영향에 대한 연구가 주를 이루고 있다. 한편, 시장에서 정보의 공개 및 전파는 다양한 매체를 통해 이루어 지는데, 대표적인 매체가 전자공시시스템, 애널리스트 보고서, 뉴스, 또는 소셜 미디어 등이다. 이 중 연구가 가장 활발한 매체는 뉴스이다. 다수의 선행연구는 뉴스 기사에 포함된 정보가 시장 참여자의 의견 형성이나 거래 행태, 또는 주식

가격에 중요한 영향을 미치고 있음을 보여주었다.

이렇듯 시장에서 뉴스나 소셜 미디어의 역할이 더욱 중요해지는 가운데 최근 시장정보 생성과 전달, 검색 과정에서 유튜브와 같은 새로운 미디어 채널이 주목받고 있다[1, 15]. 특히, 한국에서는 코로나19 기간 주식시장에 대한 사회적 관심이 높아지며 유튜브 주식 채널이 급속도로 증가하였다. 증권사와 경제 전문 언론사들은 자사의 유튜브 공식 채널을 통해 국내·외 시장에 대한 동향 분석이나 투자 전략을 제시한다. 기업 뿐만 아니라 주식·경제 관련 개인방송을 진행하는 영향력 있는 유튜브 인플루언서(influencer) 또한 출현하고 있다. 2019년 1월 개설한 ‘삼프로TV_경제의신과 함께’ 채널은(이하, ‘삼프로TV’) 매일 주식 개장시간 이전에 실시간 방송을 진행하며 일별 시장 동향과 경제 이슈를

Received 24 May 2023; Finally Revised 7 June 2023;
Accepted 8 June 2023

[†] Corresponding Author : kiklee@dankook.ac.kr

분석하는데, 해당 채널의 경우 2023년 5월 기준 231만명의 구독자를 보유하고 누적 조회수 약 10억 건을 달성하는 등 시장 참여자에게 상당한 영향력을 가지고 있다. 이와 관련하여 Song and Park[15]은 소비자들이 인기 유튜브 채널의 의견을 적극적으로 수용한다고 밝혔다. 실제로 2021년 6월 2일 삼프로TV의 장 시작 전 실시간 방송에서 모 증권사 애널리스트가 특정 종목을 추천한 이후, 해당 종목의 거래량이 직전 5일 평균 대비 4.5배 증가하는 모습을 보였다.

그러나 최근 유튜브 주식채널의 발달에도 불구하고 유튜브 주식채널의 시장 영향력이나 예측력에 대한 학술적 연구는 거의 이루어지고 있지 않다. 따라서 본 연구에서는 한국 시장에서 최근 새롭게 주목받고 있는 유튜브 주식채널의 시장 예측력에 대해 검증하고자 하였다.

본 연구는 다음과 같이 구성되어 있다. 2장에서는 선행 연구 결과를 제시하며, 3장에서는 자료와 방법론을 설명한다. 4장에서는 유튜브 주식채널의 감성을 바탕으로 KOSPI 시장지수 예측모델의 성과를 평가한다. 5장은 결론을 제시한다.

2. 미디어와 금융 시장

데이터를 통해 금융 시장의 변화를 예측하고자 하는 시도는 오래전부터 지속되어 왔다. 시장 참여자들은 크게 두 가지 방식의 분석을 통해 시장을 예측하고자 하였는데, 주식 가격의 과거 움직임을 유형화하고 미래 가격을 예측하는 기술적 분석(technical analysis)과 기업의 내재가치를 분석하고자 하는 기본적 분석(fundamental analysis)이다. 특히 기본적 분석에서는 정형 데이터(structured data), 즉, 기업의 매출, 현금흐름, 유무형 비용 등의 자료가 주식 가격 변동을 예측하기 위하여 사용되었다.

최근에는 빅데이터 기술의 활성화와 함께 전통적 정형 데이터에서 벗어나 텍스트 등의 비정형 데이터(unstructured data)를 활용하여 시장을 예측하고자 하는 시도가 이루어지고 있다. 특히, 개별 주가의 등락을 예측하기 위해 텍스트에 나타난 긍정 또는 부정 감성을 추출하는 감성분석이 주를 이룬다[8]. 소셜 미디어나 뉴스 미디어의 텍스트를 활용한 연구가 수행되었는데, 소셜 미디어는 이용자들이 관심사를 직접 검색하고, 정보를 습득하는 것 뿐만 아니라 이용자가 직접 정보를 생성할 수 있어 양방향적 특징을 가진다[14]. 또한 정보를 신속하게 전달하고 공유할 수 있다는 점에서 소셜 미디어를 통해 투자 정보를 얻고자 하는 이용자가 늘어나고 있다[5, 15]. 따라서 주식 가격 움직임에 대한 소셜 미디어와 뉴스 미디어의 예측력을 검증하는 연구가 다수 이루어졌다. Kim et al.[9]은 국내 포털사이트 주식토론방에서 형성된 여론이 1시간 이후 개별 주식 가격의 변동을

예측할 수 있는지 LSTM 기반 딥러닝 예측모델을 통해 검증하였으며, 최소 42%에서 최대 61%의 예측 정확도를 보임을 밝혔다. Jeong et al.[6]은 개별 기업의 뉴스 기사 본문 텍스트의 감성을 추출하여 기사 발행 당일 주가 등락과 일치하는지 여부를 비교하였다. 즉, 특정 종목에 대해 기사의 감성이 긍정일 경우 상승, 부정일 경우 하락을 예측한 결과 평균 56%의 예측 정확도를 보였다. Kim et al.[7]은 개별 종목과 관련한 뉴스기사 본문과 트위터 텍스트를 통해 산출한 일별 종목별 긍정지수가 익일 주가 움직임을 예측하는지 검증하였으며, 로지스틱 회귀분석 등 기계학습 예측모델의 적용한 결과 약 80% 예측 정확도를 나타내었다.

미디어를 통해 개별 기업 주가의 등락을 예측하는 연구 이외에도, 종합 주가지수의 등락을 예측하는 연구 또한 진행되었다. Kim et al.[10]은 로지스틱 회귀모델을 통해 경제 뉴스의 유형에 따라 주가지수 등락의 예측력을 비교하였으며, 시황·전망 뉴스 본문의 긍·부정 감성이 주가 변동을 설명하는 것으로 나타났다. 특히 주식시장 개장 전 발행된 시황·전망 뉴스의 당일 주가지수 등락 예측력이 74.6%로 뉴스가 시장지수 변동을 예측 가능하다고 주장하였다.

선행연구 결과는 주식 가격 움직임에 대한 소셜 미디어와 뉴스 미디어 감성의 예측력을 검증하고 있으나, 특정 콘텐츠만을 이용하거나 소수의 제한된 종목의 뉴스를 대상으로 하기 때문에 결과를 일반화하기가 어렵다. 또한 예측에 사용한 시장 자료 또한 뉴스 발행 당일 혹은 익일로 한정하였기 때문에, 미디어의 감성이 주식 가격의 선행 지표인지 여부를 판단하기 어렵다.

한편, 소셜미디어와 뉴스미디어 이외에도 최근에는 주식 정보 공유를 위한 창구로서 유튜브 주식 채널이 새롭게 대두되고 있다. 유튜브는 단순한 영상 공유를 넘어 현재 정보 검색을 위해 가장 널리 쓰이는 인터넷 웹사이트이다[1, 12]. 유튜브는 채널 구독(subscribe) 서비스를 통해 콘텐츠 제공자와 시청자 간 상호작용을 높이고, 좋아요(Like) 및 싫어요(Dislike)와 같은 평가 시스템을 통해 콘텐츠 제공자가 가치 있는 정보를 제공할 유인, 즉 광고비를 통한 수익 창출을 가능케 한다는 특징이 있다[11, 13]. 특히 유튜브 실시간 방송은 전통적인 대면 방식의 정보 전달효과를 유지함과 동시에 대화창 기능의 결합을 통해 영상 제공자와 시청자 간 쌍방향 상호작용을 증가시킨다[4].

유튜브의 차별화된 특성은 주식 시장의 정보 생성과 전달, 투자자 의견 형성에 있어 중요한 역할을 수행할 것으로 기대되고 있다. 그럼에도 불구하고 유튜브 주식 채널의 정보력 및 예측력에 대해서는 학술적 연구가 거의 이루어지지 않았다. 따라서 본 연구에서는 유튜브 주식채널의 예측력을 검증하고자 하였다. 구체적으로는 유튜브 주식채널의 긍정 혹은 부정 감성이 코스피 시장 수익률을 예측할

수 있는지 시점별 상관관계 분석 및 로지스틱 회귀모형을 통해 검증하고자 한다. 또한 유튜브 주식채널은 전통적인 주식정보 전달 매체인 언론기사와 차별화된 특성을 보일 것으로 기대되기 때문에, 증권 관련 언론기사와 유튜브 주식채널의 감성이 코스피 지수 수익률과 어떠한 영향관계를 가지는지 매체별 정보력을 비교·분석하고자 한다.

3. 자료 및 방법론

3.1 자료 수집 및 정제

분석을 위한 데이터는 크게 시장 데이터와 매체별 감성 추출을 위한 텍스트 및 영상 데이터를 수집하였다. 시장 데이터로서는 FnGuide를 통해 코스피200 지수의 일별 시가, 저가, 종가, 거래량 및 수익률 데이터를 수집하였다. 감성 추출을 위한 데이터는 <Table 1>과 같이 유튜브 주식채널 동영상과 증권 분야 언론 기사를 수집하였다. 유튜브 주식정보 채널은 대형 경제신문사와 일반인이 개설한 인기 유튜브 채널 간 정보력의 차이가 있을 것으로 보고 구독자 수와 영상 평균 조회수 기반 상위 채널을 선정하였다. 이에 따라 한국경제신문에서 운영하는 한경TV, 그리고 인기 채널인 삼프로TV를 선정하였다. 또한 각 유튜브 채널 특성 상 주식정보와 무관한 동영상이 다수 포함되어 있어 채널에서 정기적으로 진행하는 경제·주식 관련 실시간 방송 콘텐츠 중 평균 조회수 기준 상위 영상 클립을 수집 대상으로 하였다. 표본 기간으로는 2022년 9월 1일부터 2023년 2월 28일까지 총 6개월의 데이터를 수집하였다. 분석 대상으로서 유튜브 채널은 영상 데이터이기 때문에 감성 추출을 위해 동영상의 음성 데이터를 텍스트 형태의 스크립트(script)로 변환할 필요가 있다. 즉, 원데이터에 대한 음성 인식(Speech Recognition; Speech-to-Text; STT) 전처리가 필요하며, 본 연구에서는 Python 및 Naver CLOVA Speech 리소스를 활용하여 각 영상에서 스크립트를 추출 및 텍스트 변환하였다. 증권 언론기사는 Python

프로그래밍을 통하여 기사 제목, 출처, 작성자, 기사 본문 텍스트를 수집하였다. 이후 수집한 텍스트를 정제하고, 토큰화(tokenizing), 형태소 분석, 불용어 처리 등의 과정을 거쳐 분석에 사용할 수 있도록 전처리하였다.

3.2 감성추출 알고리즘

유튜브 주식채널 스크립트와 네이버 증권 뉴스기사 본문 수집 이후, 수집한 텍스트에서 감성을 다음과 같이 추출하였다. 감성 추출에는 Cho et al.[2, 3]이 제안한 재무특화 감성사전 KOSELF(Korean Sentiment Lexicon for Finance)를 바탕으로 단어별 감성을 학습하였다. 코스피 종목에 대해 발행된 약 1만6천개의 증권사 애널리스트 보고서 본문에 대하여 긍정 또는 부정 두 가지 범주로 구분하였다. <Table 2>는 단어별 감성 학습에 사용한 학습 범주의 기준을 나타낸다. ‘긍정’ 범주는 목표주가 변화율(dTPRC)의 방향과 보고서 발행시점 이후 2일 누적초과수익률(CAR_[0,+1])의 방향이 모두 양(+)의 방향인 경우를 의미하며, 반대로 모두 음(-)의 방향인 경우 ‘부정’ 범주를 의미한다. 이후 특정 단어가 ‘부정’ 범주에서 출현한 빈도보다 ‘긍정’ 범주에서 출현한 빈도가 더 많은 경우 해당 단어를 ‘긍정’ 범주로 분류하였다.

<Table 2> Two Classes used to Train Sentiment

Criteria	Positive Sentiment	Negative Sentiment
Direction of dTPRC	Positive	Negative
Direction of CAR _[0,+1]	Positive	Negative

개별 문서의 의견변수로서 OPN은 다음과 같이 해당 문서가 긍정일 확률(POS)에서 해당 문서가 부정일 확률(NEG) 간 차이로 정의하였다.

$$OPN = \frac{POS - NEG}{LENGTH}$$

<Table 1> Data Collection for Sentiment

Criteria	YouTube		News
	Institution	Individual	
Channel Name/Source	HanKyungTV	SamproTV	Naver Finance
Channel Set Up Date	Oct. 25, 2011	Jan. 17, 2019	None
Number of Subscribers	1.05M	2.31M	None
Selected Sample	Finish up Economy and Stock Market	Market Inside	Market Condition and Outlook News
Features	Broadcasting Time	Mon. to Thu. 17:00	- Published by 12 presses - Average of 105 news posts per day
	Average Video Length	50 min.	
	Average Views	78K	

여기서 문서가 긍정(부정)일 확률은 문서 내 출현한 개별 단어의 긍정(부정) 확률을 모두 곱하여 정의하였다. 또한, OPN 변수는 각 문서의 길이에 따른 긍정·부정 확률의 편차를 보정하기 위하여 개별 문서의 길이(LENGTH)로 표준화하여 사용하였다.

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{(Precision + Recall)}$$

*TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative

3.3 시장 수익률 등락 예측 모형

본 연구에서는 유튜브 주식채널 및 뉴스기사 감성변수에 대해 주식 가격 예측력이 존재하는지 여부를 검증하기 위하여 코스피 지수 등락 예측 모형을 설계하였다. 본 연구의 예측 모형은 코스피 일수익률($Return_t$) 상승과 하락으로 구분된 아래와 같은 이진 분류(Binary Classification)를 목표로 한다. 모형은 기계학습에서 이진 분류 문제(task)에 널리 사용되는 로지스틱 회귀 모형(Logistic Regression Model)을 사용하였다.

$$y = \begin{cases} 1, & \text{if } Return_t \geq 0 \\ 0, & \text{else} \end{cases}$$

예측 모형에 대한 성과 평가를 위해 전체 데이터 중 학습 데이터와 검증 데이터를 8:2 비율로 무작위 분할하여 사용하였다. 이후 다음의 분류 모형평가 지표를 통해 각 모형의 예측 성과를 검증하였다.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} *$$

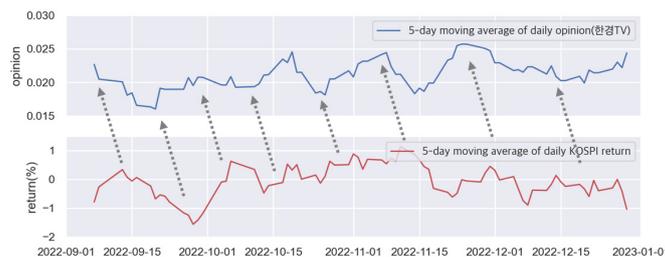
$$Precision = \frac{TP}{TP + FP}$$

여기서 ‘양성(Positive)’과 ‘음성(Negative)’은 모형이 예측한 범주를 나타내며, ‘참(True)’과 ‘거짓(False)’은 모형이 예측한 범주가 실제와 동일할지 여부를 나타낸다.

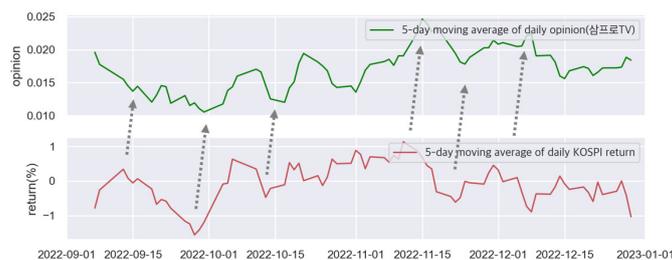
4. 결과 및 해석

4.1 시점별 상관관계 분석

먼저 각 매체의 감성이 KOSPI 가격 변동에 선행하는지 여부를 시계열 그래프를 통해 확인하였다. 본 연구에서는 비교적 짧은 기간의 표본을 사용하고 있어 추세 확인을 위해 각 변수의 5일 이동평균값을 사용하였다. <Figure 1>은 유튜브 한경TV 채널의 일별 감성변수(OPN)와 KOSPI 일수익률의 시계열 추이를 나타낸다. 이 경우 한경TV 채널의 감성이 KOSPI 일수익률에 약 일주일 시차를 두고 선행하고 있다. 반대로 <Figure 2>에 나타낸 유튜브 삼프로TV 채널의 감성은 시장 수익률에 오히려 후행하는 경향을 보인다. <Figure 3>에서 네이버 뉴스의 감성은 별도의 시차 없이 실시간으로 감성과 수익률 간 움직임이 유사한 패턴을 보인다.



<Figure 1> Daily OPN from HanKyeongTV YouTube Channel and KOSPI Index Return



<Figure 2> Daily OPN from SamproTV YouTube Channel and KOSPI Index Return



<Figure 3> Daily OPN from Naver Finance News and KOSPI Index Return

<Table 3> Correlation Coefficient between KOSPI Index and Opinion from HanKyungTV YouTube before t-days

Index	t=0	t=1	t=2	t=3	t=4	t=5	t=6	t=7
Opening Price	0.124	0.157	0.203*	0.261**	0.325***	0.386***	0.445***	0.494***
Highest Price	0.128	0.168	0.216*	0.277**	0.342***	0.403***	0.457***	0.502***
Lowest Price	0.135	0.173	0.220*	0.281**	0.347***	0.408***	0.464***	0.511***
Closing Price	0.142	0.189*	0.237**	0.300***	0.364***	0.420***	0.470***	0.513***
Return Rate	0.083	0.234**	0.287**	0.359***	0.386***	0.345***	0.308***	0.262**
Trading Volume	-0.094	-0.086	-0.063	-0.009	0.000	-0.010	-0.059	-0.111
Index	t=8	t=9	t=10	t=11	t=12	t=13	t=14	t=15
Opening Price	0.528***	0.540***	0.532***	0.503***	0.461***	0.420***	0.383***	0.361***
Highest Price	0.527***	0.534***	0.520***	0.489***	0.446***	0.409***	0.375***	0.359***
Lowest Price	0.541***	0.549***	0.539***	0.509***	0.467***	0.426***	0.392***	0.373***
Closing Price	0.533***	0.537***	0.524***	0.492***	0.452***	0.417***	0.386***	0.371***
Return Rate	0.146	0.034	-0.065	-0.194	-0.271**	-0.286**	-0.266**	-0.200
Trading Volume	-0.209*	-0.273**	-0.367***	-0.453***	-0.537***	-0.571***	-0.614***	-0.598***

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

시계열 추세 확인 결과 KOSPI 시장에 선행할 것으로 예상되는 유튜브 환경TV 감성에 대해서 시점별 상관분석을 진행하였다.2) <Table 3>은 KOSPI 시장변수와 유튜브 환경TV 채널 감성변수 간 시점별 상관관계를 나타내고 있다. 특히 유튜브 감성의 시장 선행 여부를 확인하기 위하여 영상 게시 시점(t=0)부터 직전 15일(t=15)까지 상관관계를 나타내었다. KOSPI 시가와 증가, 고가와 저가 모두 유튜브 환경TV 감성과 최소 8일에서 최대 10일 시차를 두고 약 50% 내외의 상관관계를 보였다. 수익률의 경우 4일 시차 구간에서 38.6%의 통계적으로 유의한 상관관계를 보였다. KOSPI 거래량의 경우 유튜브 환경TV 감성이 긍정적 일수록 통계적으로 유의하게 감소하는 경향을 보였다.

이상의 결과를 종합하면 유튜브 주식채널 중 환경TV는 KOSPI 수익률에 선행하며, 반대로 삼프로TV는 오히려 후

행하는 것으로 나타났다. 또한 네이버 시황·전망 뉴스의 감성은 KOSPI 수익률 변동과 시차 없이 동일하게 움직이는 것으로 나타났다.

4.2 예측모형 성과 평가

다음으로 각 매체의 감성이 KOSPI 수익률 변동에 대한 예측력이 있는지 예측모형을 통해 검증하였다. <Table 4>는 각 매체의 KOSPI 등락 예측에 대한 표본 외 검증(Out-of-Sample Validation) 결과를 나타낸다. 코스피 등락에 대한 분류 정확도(Accuracy)는 유튜브 환경TV에 나타난 감성을 사용하였을 때 71.4%로 가장 높았다. 유튜브 삼프로TV와 네이버 증권 뉴스기사 감성을 활용한 모형은 각각 56.5%, 60.9%의 정확도를 나타내어 상대적으로 낮았다. 예측모형이 50%의 정확도를 가질 경우 이는 무작위로 예측한 확률과 동일하기 때문에 적어도 50% 이상의 정확도를 확보하여야 한다. 이 같은 관점에서 유튜브 삼프로TV 감성을 활용한 모형의 경우 무작위 예측 결과와 크게 다르지 않았다.

2) 본문에 서술하지는 않았으나, 삼프로TV와 네이버 뉴스를 대상으로 동일한 시점에 대해 KOSPI 시장변수와 상관관계 분석을 진행한 결과 별도의 통계적 유의성을 보이지 않았다.

〈Table 4〉 Out-of-Sample Binary Classification Results

Metric	YouTube		Naver Finance News
	HanKyungTV	SamproTV	
Accuracy	0.714	0.565	0.609
Precision	0.923	0.417	0.833
Recall	0.632	0.625	0.588
F1-score	0.750	0.500	0.690

정밀도(Precision)와 민감도(Recall) 지표의 경우에도 한경TV에서 나타난 감성을 사용한 모델이 92.3%, 63.2%로 가장 우수한 성능을 나타냈다. 특히 정밀도 지표는 유튜브 한경TV 채널에서 발화자의 긍정적인 어조를 기반으로 모델이 코스피 수익률 상승을 예측할 경우, 실제로 수익률이 상승한 비율이 92.3%임을 의미한다. 주가 예측 모형의 경우 모형의 분류 성능을 평가하는 지표 중 민감도(Recall)에 비해 정밀도가 더욱 중요하다[3]. 즉, 주가 예측에서는 위양성(False Positive)으로 인한 비용이 위음성(False Negative)으로 인한 비용보다 더 크다. 따라서 본 연구의 예측 목적을 고려하였을 때, 정밀도의 측면에서 유튜브 한경TV 감성을 사용한 모형의 예측 결과가 가장 우수한 것으로 판단할 수 있다. F1-score는 정의에 따라 정밀도와 민감도의 조화평균을 의미하는데, 이 경우에도 마찬가지로 유튜브 한경TV 감성을 활용한 모형이 가장 우수하였다.

유튜브 주식채널의 주식수익률 예측력을 평가한 선행 연구는 존재하지 않기 때문에 일반적인 미디어를 활용하여 주식 수익률 예측을 목적으로 한 선행연구[6-7,9-10]와 비교하였을 때, 본 연구의 모형은 상대적으로 우수한 분류 정확도를 보이고 있다.

5. 결론

본 연구는 최근 주식 시장의 정보 생성 및 전달에 주목 받고 있는 유튜브 주식 채널의 시장 예측력에 대해 검증하였다. 구체적으로는 유튜브 주식채널의 각 영상에서 발화자의 긍정 혹은 부정 감성이 코스피 시장 수익률을 예측할 수 있는지 시점별 상관관계 분석 및 기계학습 예측 모델을 통해 검증하였다. 또한, 전통적으로 주식 정보 생성 및 전달 과정에서 중요한 역할을 수행하고 있는 언론기사와 유튜브 주식 채널의 정보가 차별화된 특성을 보일 것으로 보고, 증권 관련 언론기사와 유튜브 주식채널의 감성이 코스피 지수 수익률과 어떠한 영향관계를 가지는지 매체별 시장 예측력을 비교·분석하였다.

2022년 9월부터 12월까지의 표본 기간 동안 KOSPI 시장을 대상으로 매체별 감성의 시장지표 선행 여부와 지수

등락 예측력을 검증하였다. 분석 결과 유튜브 주식 채널의 경우 채널에 따라 제공하는 정보의 선·후행 여부가 상이하게 나타났다. 특히 유튜브 주식 채널 중 언론사가 운영하는 채널인 한경TV의 경우 해당 매체의 감성이 KOSPI 시장 수익률에 선행하는 것으로 나타났다. 이는 다시 말해 ‘예측(forecast)’의 관점에서 투자 가치가 있는 정보를 제공할 가능성이 있는 것으로 해석된다. 반면, 전통적인 뉴스 미디어에서 다루고 있는 시황·전망 관련 뉴스는 당일 시장에서 발생한 이슈를 실시간으로 ‘보도(broadcast)’하는 역할을 수행하는 것으로 나타났다. 이는 시황·전망 관련 뉴스가 당일 시장 수익률 변동을 예측한다는 선행연구 결과와도 일치한다[10]. 또한 인기 유튜브 주식 채널 ‘삼프로TV’ 감성의 경우 시장 수익률에 후행, 즉 사후적으로 반응하는 경향을 보였다.

본 연구는 최근 급격한 성장세를 보이고 있는 유튜브 주식 채널 정보의 시장 예측력을 검증하였다는 점에서 기존 소셜 미디어 및 뉴스 미디어의 예측력을 검증한 연구와 차별점을 가진다. 또한, 유튜브 주식 채널 정보와 전통적 주식 정보 전달 채널인 뉴스 미디어의 예측력을 비교·분석하였다는 점에서 기존의 연구결과와 차별화된다. 마지막으로, 본 연구에서 사용한 감성 추출 알고리즘은 별도의 수동 라벨링 과정을 필요로 하지 않기 때문에 학습 데이터의 규모와 무관하게 효율적으로 단어의 감성을 학습할 수 있다.

그럼에도 불구하고 본 연구는 다음과 같은 한계점을 가진다. 첫째, 본 연구는 개별 종목의 수익률이 아닌 코스피 지수 수익률을 예측 대상으로 하였기 때문에 개별 종목에 대한 매체별 예측력을 측정하기 어렵다. 둘째, 본 연구에서 사용한 표본 분석 대상과 기간이 다소 제한적이다. 유튜브 주식 채널의 경우 각 언론사, 증권사, 또는 유명 인플루언서 등 각기 다른 배경의 주식 정보 제공 채널이 존재하나, 본 연구에서는 상위 구독자 기준 두 개의 주식 정보 채널만을 검증하였기 때문에 분석 결과의 일반화에 한계가 있다.

향후 다음과 같은 후속 연구를 통해 본 연구결과를 확장할 수 있을 것으로 기대된다. 첫째, 향후 표본 수집 대상과 기간을 확장하여 개별 종목에 대한 매체별 예측력을 비교·분석할 수 있다. 둘째, 매체별 시장 예측력을 검증하기 위하여 랜덤 포레스트, 서포트 벡터 머신, 인공신경망 등 비선형 분류 모형을 적용함으로써 예측 모형 간 비교·분석과 더불어 고도화된 결과를 얻을 수 있을 것으로 기대된다.

Acknowledgement

This paper was supported by the research fund of the National Research Foundation of Korea (NRF-2019S1A5 A2A03038389).

References

- [1] Cheng X., Dale, C., and Liu, J., Statistic and Social Network of YouTube Videos, *16th International Workshop on Quality of Service*, 2008, pp. 229-238.
- [2] Cho, S.J., Kim, H.K., and Yang, C.W., Building the Korean Sentiment Lexicon for Finance (KOSELF), *The Korean Journal of Financial Studies*, 2021, Vol. 50, No. 2, pp. 135-170.
- [3] Cho, S.J., Lee, K.K., and Yang, C.W., Developing the Automated Sentiment Learning Algorithm to Build the Korean Sentiment Lexicon for Finance, *Journal of Korean Society of Industrial and Systems Engineering*, 2023, Vol. 46, No. 1, pp. 32-41.
- [4] Cho, Y. and Lim, S., Psychological Effects of Interactivity for Internet Live Broadcasting Viewers-Moderating Role of User Motivations on Parasocial Interaction, Social Presence, and Flow, *Korean Journal of Broadcasting & Telecommunications Research*, 2019, Vol. 105, pp. 82-117.
- [5] Hsieh, J.K., Hsieh, Y.C., Chiu, H.C., and Feng, Y.C. Post-adoption Switching Behavior for Online Service Substitutes: A Perspective of the Push-pull-mooring Framework, *Computers in Human Behavior*, 2012, Vol. 28, No. 5, pp. 1912-1920.
- [6] Jeong, J.S., Kim, D.S. and Kim, J.W., Influence Analysis of Internet Buzz to Corporate Performance: Individual Stock Price Prediction Using Sentiment Analysis of Online News, *Journal of Intelligence and Information Systems*, 2015, Vol. 21, No. 4, pp. 37-51.
- [7] Kim, D., Park, J., and Choi, J., A Comparative Study between Stock Price Prediction Models Using Sentiment Analysis and Machine Learning Based on SNS and News Articles, *Journal of Information Technology Services*, 2014, Vol. 13, No. 3, pp. 221-233.
- [8] Kim, H.S. and Kim, C.S., An Analysis for IT Proposal Evaluation Results using Big Data-based Opinion Mining, *Journal of Korean Society of Industrial and Systems Engineering*, 2018, Vol. 41, No. 1, pp. 1-10.
- [9] Kim, M., Ryu, J., Cha, D., and Sim, M.K., Stock Price Prediction Using Sentiment Analysis: From “Stock Discussion Room” in Naver, *Journal of Society for e-Business Studies*, 2020, Vol. 25, No. 4, pp. 61-75.
- [10] Kim, Y., Kim, N., and Jeong, S.R., Stock-Index Invest Model Using News Big Data Opinion Mining, *Journal of Intelligence and Information Systems*, 2012, Vol. 18, No. 2, pp. 143-156.
- [11] Krishna, A., Zambreno, J., and Krishnan, S., Polarity Trend Analysis of Public Sentiment on YouTube, In *Proceedings of 19th International Conference on Management of Data (COMAD)*, Dec, 2013, Ahmedabad, India.
- [12] Nasmedia, 2019 Internet User Survey, 2019, pp.1-51.
- [13] Radwan, S., A most unusual tool for trading, YouTube: A study that evaluates the possibility of using YouTube to make abnormal returns on small-cap stocks, Bachelors Thesis, Lund, Sweden, LUND University, 2021, pp. 1-36.
- [14] Song J.E. and Jang, W.H., Developing the Korean Wave through Encouraging the Participation of YouTube users: The Case Study of the Korean Wave Youth Fans in Hong Kong, *The Korea Contents Society*, 2013, Vol. 13, No. 4, pp. 155-169.
- [15] Song, H.Y. and Park, H.W., Comparison of popular YouTube video scripts and commentary networks in the economic sector: focusing on Sampro TV channels, *Journal of Korean Data Analysis Society*, 2022, Vol. 24, No. 2, pp. 843-859.

ORCID

Su-Ji Cho | <https://orcid.org/0000-0003-1511-5348>Cheol-Won Yang | <https://orcid.org/0000-0001-9023-5089>Ki-Kwang Lee | <http://orcid.org/0000-0003-2291-8376>