

2021년 제 23회 한국경영학회 융합학술대회

# 감성분석과 LSTM을 이용한 테마형(2차전지) 상장지수펀드(ETF) 가격 예측

2021. 8. 18

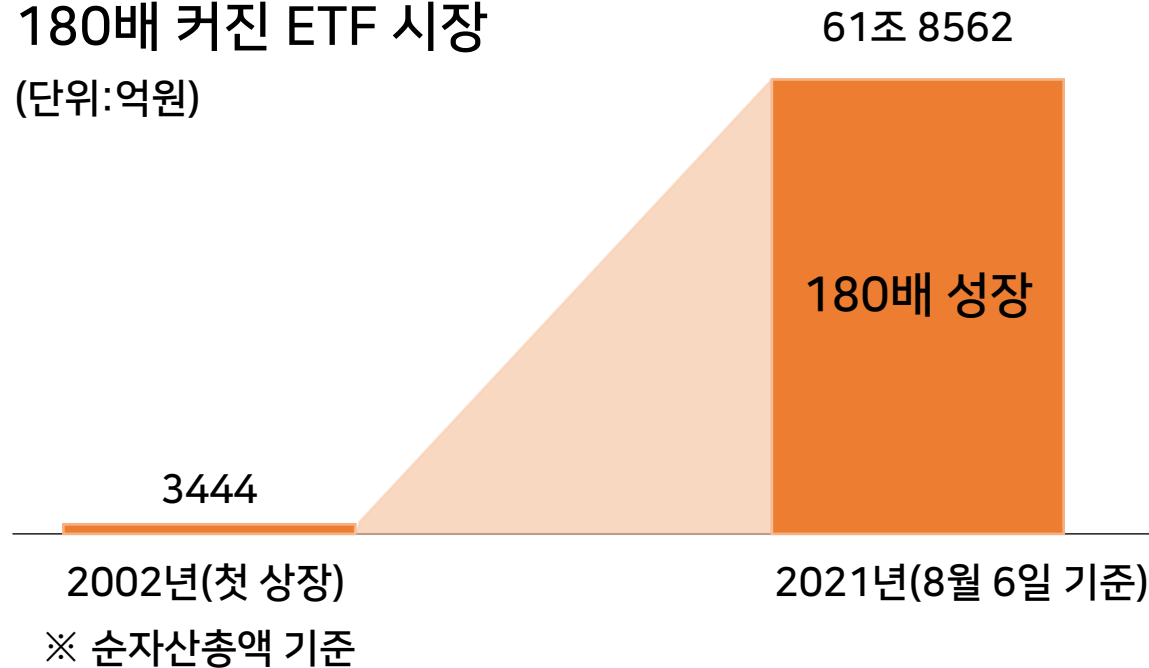
연세대학교 미래캠퍼스 경제학과 하지민  
연세대학교 미래캠퍼스 경영학부 이종혁  
연세대학교 미래캠퍼스 경영학부 신태수

# 목차

1. 서론
2. 선행연구
3. 연구모형
4. 결과 및 분석
5. 결론

# 1. 서론

## 180배 커진 ETF 시장 (단위: 억원)



참고자료: 이태훈, 『국내 상장 ETF 500개 돌파』, 한국경제, 2021.08.09.

참고자료: 강신애, 『테마형ETF에 주목해야 하는 3가지 이유』, 대한금융신문, 2021.02.01.

초기 국내 ETF는 주식형 상품만 있었지만 2009년 자본시장법 시행 후 채권, 금, 원유 등에 투자하는 ETF가 나오며 시장의 양적·질적 수준이 크게 높아짐

2020년부터는 BBIG(배터리·바이오·인터넷·게임), 미래 차, 신재생에너지 등 특정 업종과 테마에 투자하는 ETF가 인기

# 1. 서론 - 연구 목적

- 그 이유는 테마형 ETF는 종합주가지수와 연계되는 인덱스형 ETF 비해 장기적으로 더 높은 수익률을 기대할 수 있음
- 개인들이 신기술, 신제품, 새로운 트렌드를 수용하는 기간은 점차 빨라지고 있으며, 이것이 테마 ETF들의 수익률 달성 시기도 단축시킬 가능성이 높음
- 그동안 ETF에 대한 높은 관심에도 불구하고 현재 다른 금융상품과는 달리 특히 테마형 ETF에 대한 가격 예측에 대한 연구는 상대적으로 미진한 실정
- KOSPI 지수와 연계되는 인덱스형 ETF와는 달리 테마형 ETF는 뉴스 데이터, 원자재 데이터, 구성 종목 등 더 다양한 데이터와 쉽게 연계된다는 특성을 고려한 모형 개발 필요

## 2. 선행연구

### - 딥러닝 이용한 주가지수, ETF 예측모형에 대한 선행연구

연구자	데이터셋	분석모델	예측유형	예측성과
이중석, 신오순(2021)	TIGER 방송통신 ETF 등 20 종목	랜덤 포레스트 기법	상승 보합 하락 예측	2개월 예측 평균 41.8% 가까운 미래 예측 정확도 높음
황희수(2019)	일별 KOSPI 종가, KOSPI 연동 ETF	신경회로망	상승 하락 예측	KOSPI 주가 MAPE가 0.628 ~ 0.648 범위로 효과적인 예측 가능
김홍지 등(2018)	KOSPI 상위 30종목 중 18개 종목	스태킹 기반 앙상블 모델 (ANN, CNN, MLP)	가격 예측	제안한 방법이 기존 방법에 비해 MAPE가 8.74%에서 3.35%로 감소
박재연(2016)	KOSPI 지수	SVM, 라쏘 회귀분석, 인공신경망	상승 하락 예측	SVM과 인공신경망 학습 모델의 경우 모두 0.5 보다 높았음

## 2. 선행연구

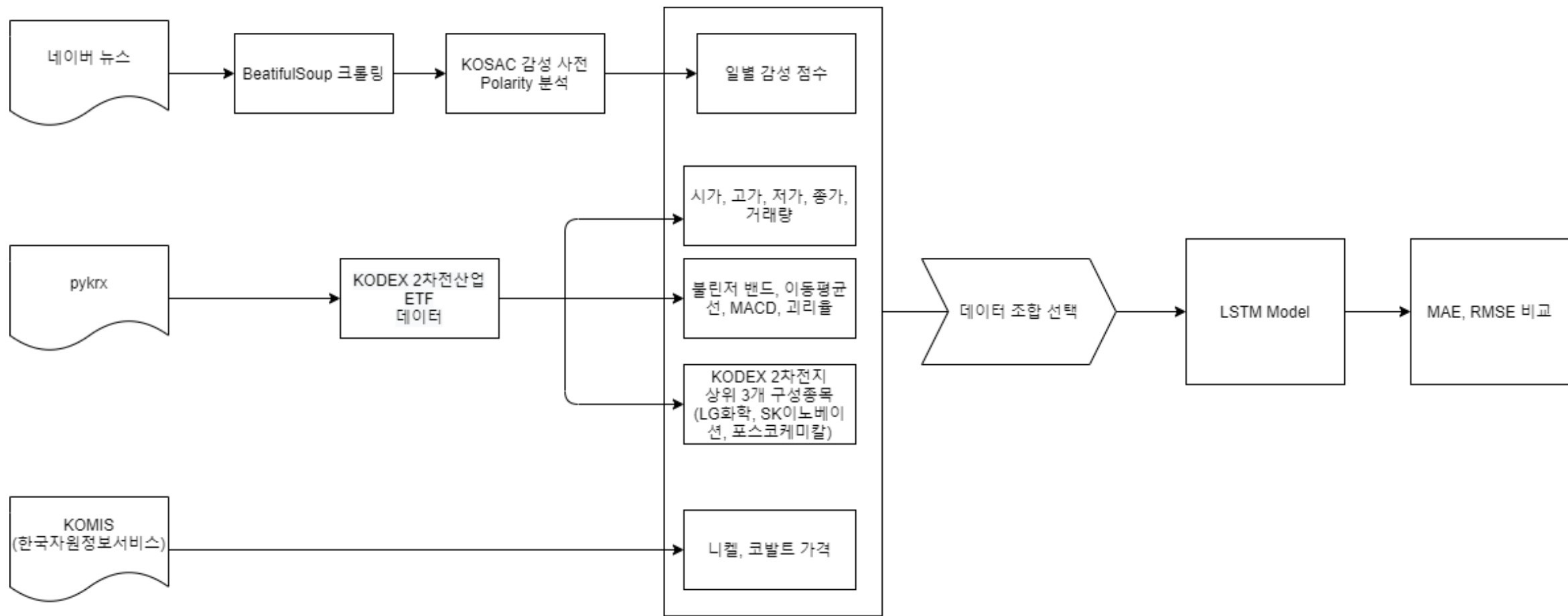
### - 뉴스 감성 분석과 주가 예측에 대한 선행연구

연구자	데이터셋	분석모델	예측유형	예측성과
Jang, Choi, Lee(2020)	다우존스지수	LSTM, BERT	지수 예측	거시경제지표와 BERT 뉴스 감성 정보를 조합한 실험이 가장 높은 정확도 보임
박민수(2020)	글로벌 자동차 제조사 중 6개사 주가	선형 회귀와 랜덤 포레스트	가격 예측	오차 기준 최소 0.08%에서 최대 0.88%의 높은 예측 성능을 보여줌
김기준(2017)	주가 데이터 & 뉴스 데이터	LSTM	가격 예측	예측 결과는 63.2%로 기존 연구보다 약 4.8% 향상된 결과
김재봉, 김형중(2017)	증권전문 사이트 'Paxnet'의 게시글	말뭉치 기반 접근법	감성 사전 구축	주식시장에 집중된 맞춤형 특화 사전을 구축함에 있어 게시글 별로 유의미한 차별성이 있음을 확인
유은지 등(2013)	주가 지수, 뉴스데이터	오피니언 마이닝	등락 예측	범용 감성사전에 비해 주제지향 감성사전을 사용한 예측 모델의 정확도가 다소 높게 나타남
안성원, 조성배(2010)	KOSPI 673 개 종목, 뉴스 데이터	텍스트 마이닝, Navie Bayes 분류, RSI 분석 기법	등락 예측	50 ~ 60%의 예측성공률

## 2. 선행연구

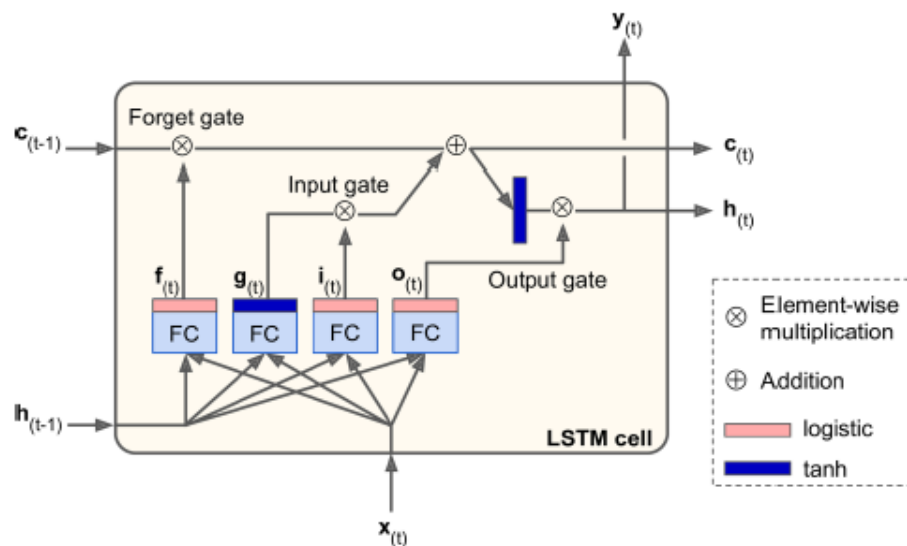
- 기존 주가지수에 대한 연구는 시계열 분석만 존재
- 테마형 ETF가 컨텍스트 데이터에 더 영향을 받을 거라 예상
- KOSPI 지수와 연계되는 인덱스형 ETF와는 달리 테마형 ETF는 뉴스 데이터, 원자재 데이터, 구성 종목 등 더 다양한 데이터와 쉽게 연계된다는 특성을 고려한 모형 개발 필요
- 본 연구는 성장가능성이 높은 종목들로 구성된 테마형 ETF인 KODEX 2차전지산업 ETF를 대상으로 정형데이터와 비정형 데이터를 함께 이용한 일별 가격을 예측하는 모형을 제시

### 3. 연구 모형





## 3.1 LSTM 모형



<그림 1>LSTM 구조

Equation 14-3. LSTM computations

$$i_{(t)} = \sigma(W_{xi}^T \cdot x_{(t)} + W_{hi}^T \cdot h_{(t-1)} + b_i)$$

$$f_{(t)} = \sigma(W_{xf}^T \cdot x_{(t)} + W_{hf}^T \cdot h_{(t-1)} + b_f)$$

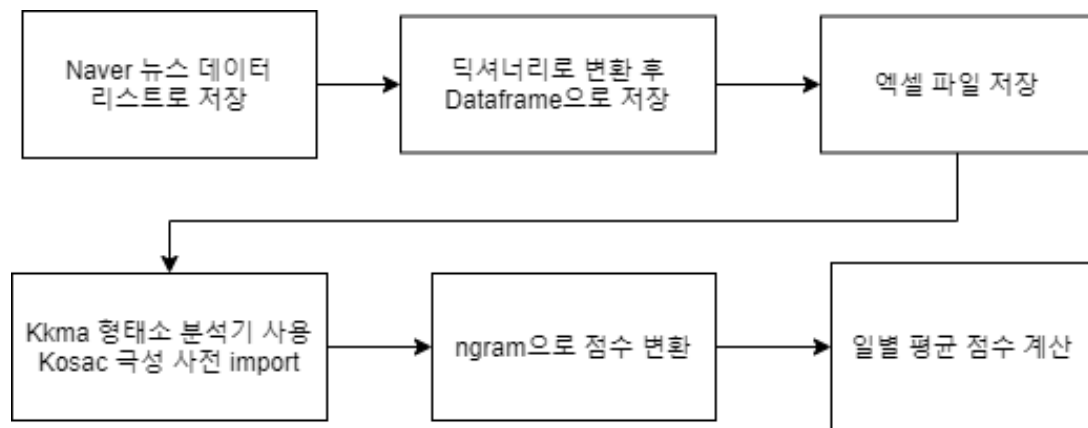
$$o_{(t)} = \sigma(W_{xo}^T \cdot x_{(t)} + W_{ho}^T \cdot h_{(t-1)} + b_o)$$

$$g_{(t)} = \tanh(W_{xg}^T \cdot x_{(t)} + W_{hg}^T \cdot h_{(t-1)} + b_g)$$

$$c_{(t)} = f_{(t)} \otimes c_{(t-1)} + i_{(t)} \otimes g_{(t)}$$

$$y_{(t)} = h_{(t)} = o_{(t)} \otimes \tanh(c_{(t)})$$

## 3.2 감성 분석



'2차전지산업'이란 키워드로 일별 최대 50개의 뉴스 데이터 크롤링을 진행

KOSAC 사전과 Kkma 형태소 분석기를 활용하여 n-gram을 계산해 뉴스 별 점수를 계산

날짜별로 점수를 평균 내어 일별 뉴스 감성 점수를 계산

## 4. 실증분석 및 결과 - 4.1 변수 개요

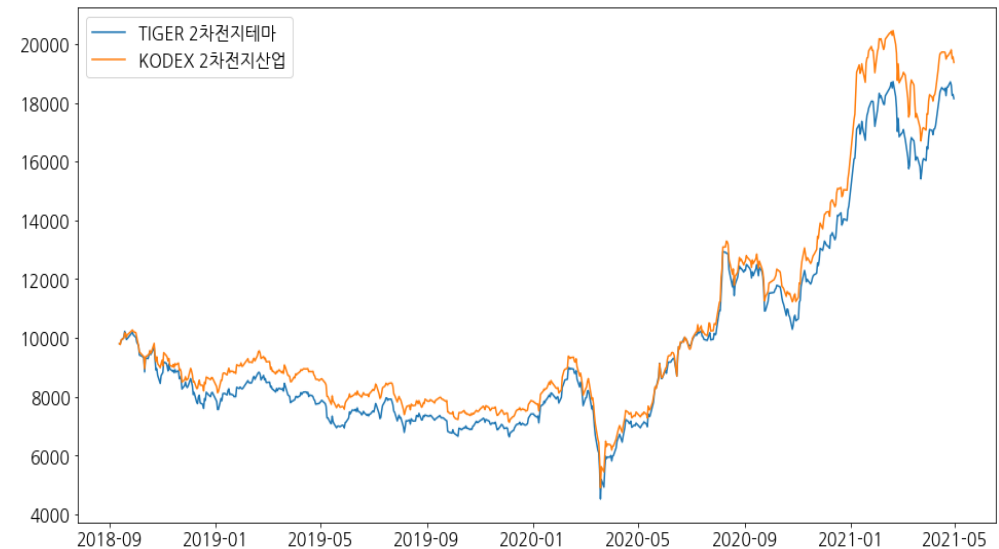
<표 3> 변수별 사용 지표

변수구분	데이터 종류	사용 지표
입력변수	기술적 지표	볼린저 밴드, 이동평균선, MACD, 괴리율
	ETF 시계열 데이터	시가, 고가, 저가, 종가
	뉴스 데이터	뉴스 감성 점수
	원자재 가격	니켈, 코발트
	구성 종목	상위 3개 구성종목 증가 데이터(LG화학, SK이노베이션, 포스코케미칼)
출력 변수		2차전지 ETF 증가 데이터

## 4. 실증분석 및 결과 - 4.1 분석대상 데이터의 기초 분석

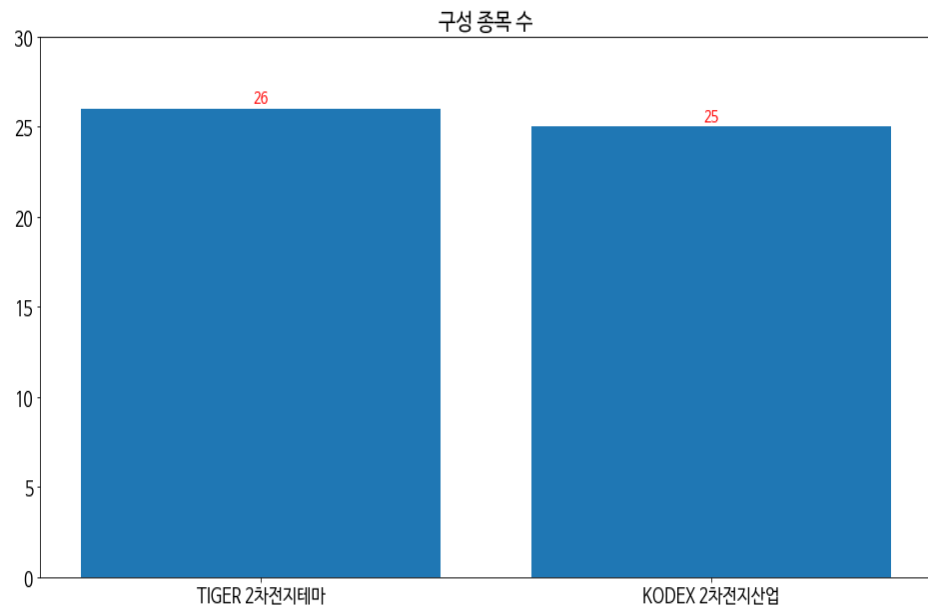
<표 4> 2차전지 테마 관련 ETF 종류

종목 코드	종목명
305540	TIGER 2차전지테마
364980	TIGER KRX2차전지K-뉴딜
2305720	KODEX 2차전지산업

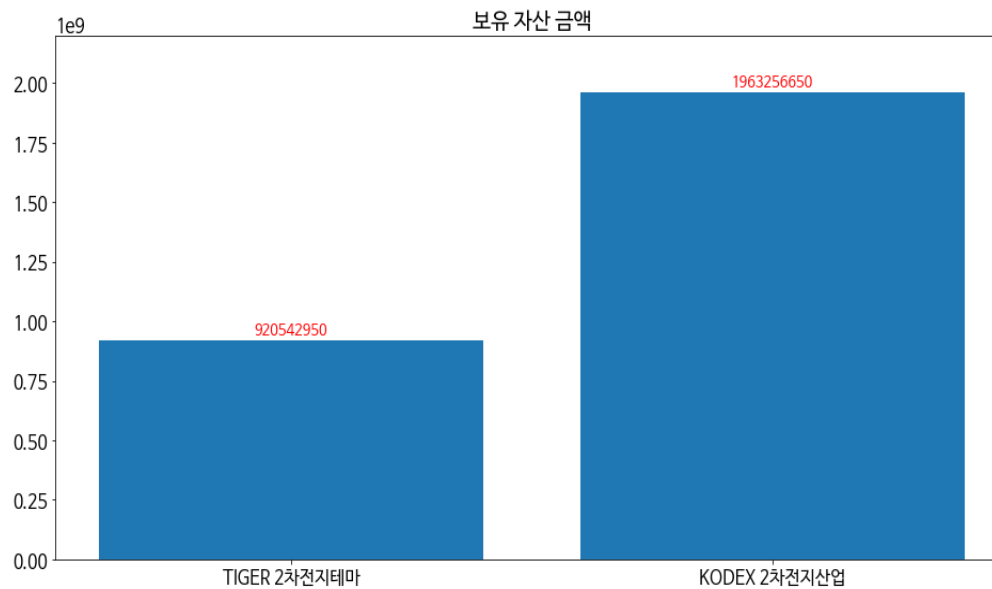


<그림 2> 2차 전지 ETF 종가 시계열 그래프

## 4. 실증분석 및 결과 - 4.1 분석대상 데이터의 기초 분석

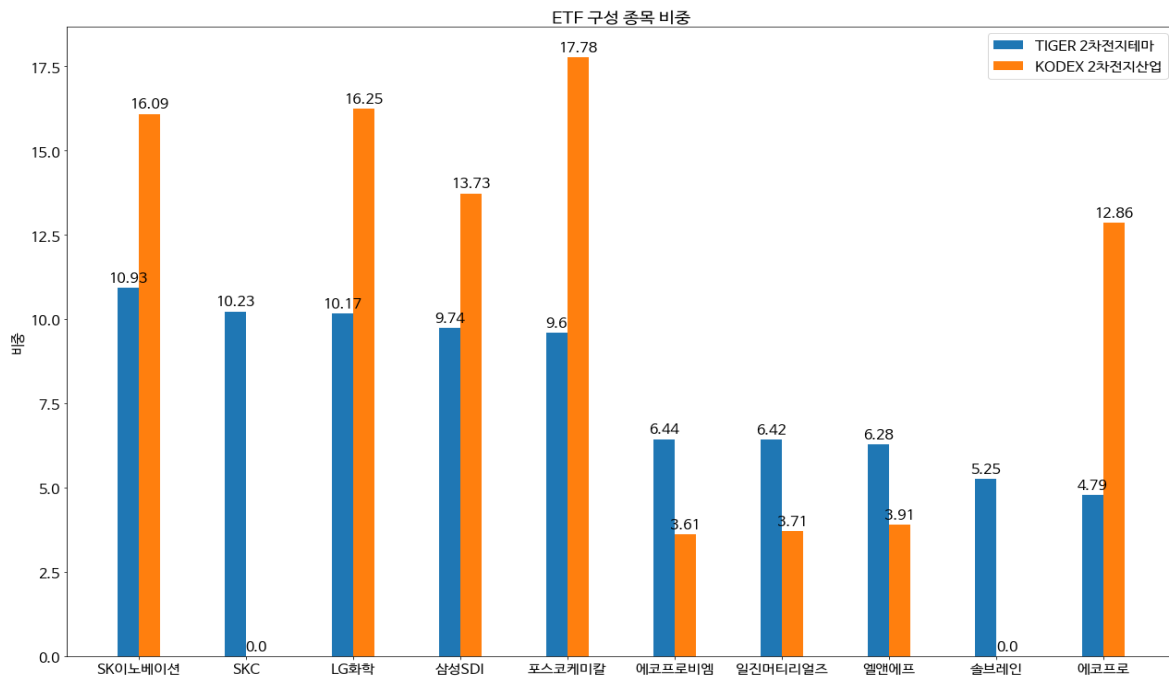


<그림 3> 2차 전지 ETF 별 구성 종목 수 비교



<그림 4> 2차 전지 ETF 별 보유 자산 금액 비교

## 4. 실증분석 및 결과 - 4.1 분석대상 데이터의 기초 분석



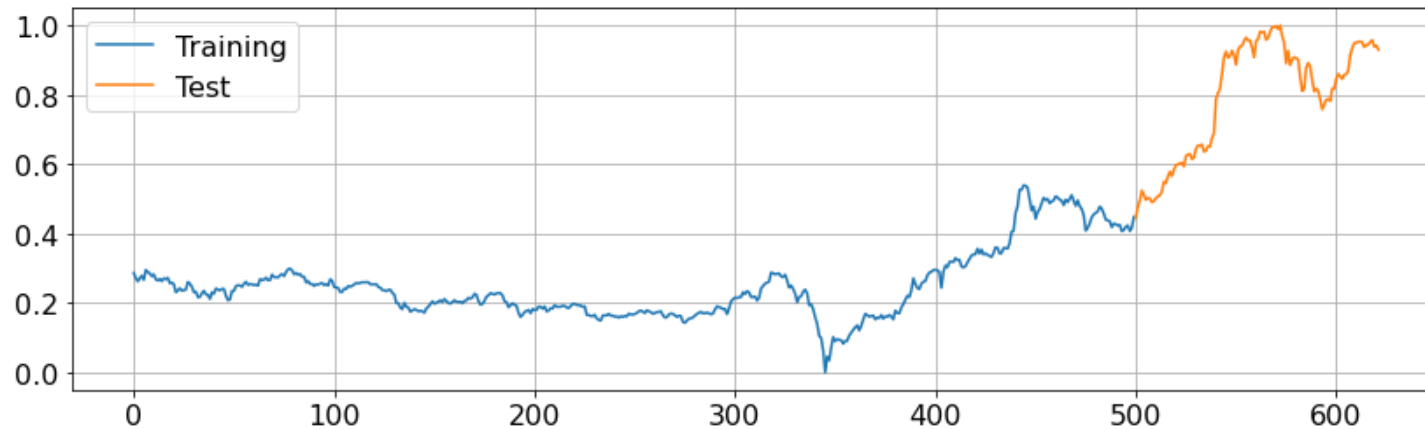
<그림 5> 2차 전지 ETF 별 상위 10 구성종목 비중

보유 자산 금액은 KODEX 2차전지산업 ETF가 TIGER 2차전지테마 보다 2배 이상 큼

TIGER 2차전지테마는 종목 비중이 골고루 분배되어 있지만, KODEX의 경우 대기업 위주로 큰 비중을 차지하고 있음

그래서 KODEX ETF가 TIGER ETF보다 더 큰 보유 자산 금액을 가지고 있는 것을 알 수 있음

## 4. 실증분석 및 결과 - 4.2 Train Test 셋



<그림 6> 2년 반의 Kodex 2차 전지 산업 train test 데이터 셋

## 4. 실증분석 및 결과 - 4.3 연구 모형의 예측 결과

LSTM모형 유형	train		test	
	rmse	mae	rmse	mae
증가 데이터 기반	0.025239	0.016865	0.114088	0.103897
기술적 지표 기반	0.028705	0.182379	0.557953	0.159065
시가 고가 저가 증가 기반	0.023065	0.016292	0.098934	0.122237
원자재 가격 기반	0.099061	0.060783	0.641812	0.617438
ETF 상위 3 구성 종목 증가 기반	0.025612	0.018015	0.232609	0.208674
감성 점수	0.098321	0.068468	0.583326	0.559819
감성 점수+증가기반	0.024083	0.017085	0.090863	0.075886

정형 데이터 중 증가 데이터 기반 LSTM 모형이 가장 좋은 성적

추가적으로 학습을 진행한 원자재 가격과 ETF 상위 3 구성 종목 기반 모형은 예상 외로 낮은 예측력을 보임

실제 예측성과지표를 기준으로, 증가와 감성점수의 조합의 모형이 기술적 지표와 같은 정형 데이터 보다 두 배 이상 낮은 오차율을 보이고 있는 것을 확인 가능



## 5. 결론 - 연구결과 요약

- 1 작년부터 현재까지 테마형 ETF의 여러 장점과 수익성에 의해 관심이 계속 증가
- 2 이러한 테마형 ETF 특성을 살린 가격 예측 모형에 대한 연구는 부족
- 3 기존 ETF 예측 모형에선 지수 기반의 시계열 분석만 존재
- 4 더 나아가 테마형 ETF에 대한 컨텍스트 정보를 활용해서 감성 정보를 통해 예측모형을 만들고자 함
- 5 이를 딥러닝 모델로 결합하였고 그 결과 인덱스 펀드의 정형 데이터와 테마형 ETF의 특징을 살릴 감성분석을 합한 것이 높은 성능이 나옴

## 5. 결론 - 공헌도 및 활용 방안



본 연구를 통해 ETF의 가격을 예측하는데 뉴스 기사를 통한 감성 분석이 유효성 있고 상당한 예측력을 보여준다는 것을 알 수 있었음



연구결과를 바탕으로 ETF 매매전략으로 활용하거나 종가 데이터와 감성 분석을 적절히 사용하면 예측의 정확도를 향상시킬 수 있을 것으로 기대

## 5. 결론 - 연구 한계점 및 향후 개선 방안



일반적인 감성사전을 사용했지만 향후 ETF 전용 감성 사전을 이용할 필요 있음  
추후 정확도를 높일 가능성 존재



입력변수로서 기술적 지표 중 일부만 사용하였으며, 구성 종목도 3종목만 이용  
향후 연구에서는 본 연구에서 포함되지 않은 변수를 포함시킬 필요



비교적 최근에 만들어진 ETF라 데이터양 부족이 해결된다면 성능 향상 가능성 존재

# 참고문헌

- Jang, E., Choi, H., & Lee, H. (2020). Stock prediction using combination of BERT sentiment Analysis and Macro economy index. 한국컴퓨터정보학회논문지, 25(5), 47-56.
- Yu, E., Kim, Y., Kim, N., & Jeong, S. R. (2013). 주가지수 방향성 예측을 위한 주제지향 감성사전 구축 방안. 지능정보연구, 19(1), 95-110.
- 공인호, 정초원, 투자 초보, ETF 투자 따라잡기, MONEY 2021년 06월호, p. 18
- 김기준. "뉴스 감성 분석과 시계열 예측 기반의 주가 등락 예측." 국내석사학위논문 송실대학교 대학원, 2017. 서울
- 김재봉, 김형중. "주가지수 방향성 예측을 위한 도메인 맞춤형 감성사전 구축방안." 한국디지털콘텐츠학회 논문지 18.3 (2017): 585-592.
- 김홍지, 정지현, 고은나래, 조만재, 이기훈, "딥러닝 앙상블을 이용한 주가예측," 데이터베이스연구, 2018, pp.111 - 120 .
- 박민수. "머신러닝 기반의 온라인 미디어 감성분석을 통한 자동차 제조사 주가 해석에 관한 연구." 국내석사학위논문 서울대학교 대학원, 2020. 서울
- 박재연, 유재필, 신현준. (2016). 기술적 지표와 기계학습을 이용한 KOSPI 주가지수 예측. 정보화연구(구 정보기술아키텍처연구), 13(2), 331-340.
- 성노윤, 남기환. (2019). 시스템적인 군집 확인과 뉴스를 이용한 주가 예측. 지능정보연구, 25(3), 1-17.
- 송민중, 이미협, 김영민, 서대룡. (2016). 코스피 시장에서의 기술적 분석지표 예후 수익률 분석. Studies in Mathematical Education, 2016(3), 289-293.
- 안성원, 조성배. "뉴스 텍스트 마이닝과 시계열 분석을 이용한 주가예측." 한국정보과학회 학술발표논문집 37.1C (2010): 364-369.
- 원종민, 황현수, 정유현, 박희동. (2018). 기술적 분석지표와 딥러닝을 활용한 주가 예측 기법. 한국정보기술학회 종합학술발표논문집, (), 404-405.
- 이중석, 신오순. "머신러닝을 활용한 ETF 수익 방향 예측." 한국통신학회 학술대회논문집 . (2021): 376-377.
- 황준호, "ETF 전성시대.. 연내 500개 돌파", <아시아경제>, 2021.04.23
- 황희수. (2019). 신경회로망을 이용한 KOSPI 예측 기반의 ETF 매매. 한국융합학회논문지, 10(1), 7-12.

**감사합니다.**