

Analysis, prediction and visualization of stock price trends using auxiliary indicators for each stock

제안서

Project Proposal

발표자

컴퓨터학부 **배준형** (20190511)



# 목차

01 팀원, 용어 소개

02 관련 연구 및 논문 소개

- 03 기존 연구 문제 제기
  - 회사 규모 측면
  - AI에 활용된 지표 활용 측면
  - 현재 연구된 논문 측면

- 04 해결 방안 제시
  - 학문/사회적 기여도
  - 창의성
  - 난이도
- 05 기대 효과
- 06 참고 논문 및 각주



01

팀원, 용어 소개

### Team Member

- **1** 컴퓨터학부 4학년 **배준형** (20190511, Leader)
- **2** 컴퓨터학부 4학년 **서유준** (20180764)
- **3** 컴퓨터학부 4학년 **유승준** (20192425)
- **4** 컴퓨터학부 4학년 **이준혁** (20193033)

### 용어 정리

- 1. 선차트 : 시가, 고가, 저가, 종가등을 제공하는 차트
- 2. 보조지표
  - 1. MACD (moving average convergence divergence)
    - MACD는 주가 추세의 강도, 방향, 모멘텀 및
    - 지속 시간의 변화를 나타내도록 설계
    - MACD = 단기 EMA (12일) 장기 EMA (26일) ※ EMA = 지수이동평균 (Exponential Moving Average)
  - 2. 일목균형표
    - 호소다 고이치 (━圓山人) 가 제작한 차트 분석 방법
    - {9, 26, 52} 일의 최고/최저가 등을 바탕으로 주가를 평균값을 내어 기준선, 전환선, 선행 (A,B) /후행 스팬 등을 제작 가능.



# 01 팀원, 용어 소개

# 선 차트



[그림 1. 선 차트]

# 보조지표



[그림 2. 보조지표]





# 02

### 관련 연구 및 논문 소개

### 기존 논문

- 1. 인공지능 머신 러닝 기술을 이용한 주식 종목 매수/매도 **추천시스템의 분석 및 설계 (2021)** 저자 : 조병호
  - 주식종목 추천 시스템의 설계 방법과 구현 방법에 대한 내용
- 2. LSTM을 이용한 **주가 예측 모델**의 학습방법에 따른 성능분석 (2020) 저자 : 정종진, 김지연
  - 장,단기기억 DNN 을 활용한 주식 예측 알고리즘 구현 제안



그림 3. 객체지항 클래스 다이아그램 Fig. 3. Object-Oriented Class Diagn

IV. 인공지능 머신러닝 기술을 이용한 주식 종목 매수/매도 추천 시스템의 설계





본 주식 배수/배도 추원 시스템의 우수생을 알아보기 위해 그런 5의 알고리증을 착용해서 변소들보수와 파이 편으로 테스트 모르다경을 제작하였다. 과거 고자 파이 등록 대는 미국 1억 이 주식종목 배수/배도 추권 시스템 엔진을 청용한 것과 작간의 기존 기가서 알고리증을 착용 생사 배수시설과 배도시설의 수를 결과를 사람들에서 해 생사 배수시설과 배도시설의 수를 결과를 사람들에서

[그림 3. 논문 1 (조병호 저자 논문)]

#### 2. LSTM 구조와 동작

LSTM은 Fig. 1과 같이 RNN(Recurrent Neural Network)의 은닉층(Hidden Layer)에 센 상태(Cell State)를 추가해서 개선한 모델로서, 과거 학습 정보를 기억하고 새로운 학습 결과에 반영할 수 있게 된다. 따라서 LSTM은 시계열 문제 및 예측 문제에 뛰어난 성능을 밝혔하는 학습 모델이라고 할 수 있다.

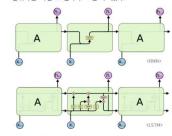


Fig. 1. Architecture of LSTM

셀의 수식은 다음과 같다.

 $ft = \sigma(Wxh_fxt + Whh_fht-1 + bh_f)$   $it = \sigma(Wxh_fxt + Whh_iht-1 + bh_i)$   $ot = \sigma(Wxh_ixt + Whh_oht-1 + bh_o)$   $gt = tanh(Wxh_gxt + Whh_ght-1 + bh_g)$  ct = ft © ct-1 + it © gt) ht = ot © tanh(ct)

[그림 4. 논문 2 (정종진, 김지연 저자 논문)]



### 기존 연구 문제 제기

### 기존 논문에서의 문제점 1

1. 인공지능 머신 러닝 기술을 이용한 주식 종목 매수/매도 **추천시스템의 분석 및 설계 (조병호, 2021)** 

### 회사마다 데이터의 정확성을 파악하지 않았음

추천 시스템 정확도가 78.3% 라고 정확도를 소개하였지만 회사마다 어떤 지표가 높은지, 낮은지 소개를 하지 못함.

### 추첨 종목이 어떠한 이유로 선정되었는지 알 수 없음

추천 종목에 대한 신뢰도가 떨어지는 문제점 발생



#### IV. 인공지능 머신러닝 기술을 이용한 주식 종목 매수/매도 추천 시스템의 설계

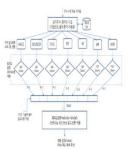


그림 4. 주식중목 매수/매도 추천 시스템 엔진



그림 5. 주식종목 매수/매도 추천을 위한 호름도

본 주식 매수/매도 추천 시스템의 우수성을 알아보기 위해 그림 5의 알고리즘을 적용해서 텐소플로우와 파이 뿐으로 테스트 프로그램을 제작하였다. 과거 고스피 총 목을 일부를 그림 4의 주식종목 매수/매도 추천 시스템 엔진을 적용한 것과 각각의 기존 7가지 알고리즘을 적용 해서 매수시점과 매도시점의 추천 결과를 시뮬레이션 해

Vol. 21, No. 4, pp.103-108, Aug. 31, 2021, pISSN 2289-0238, eISSN 2289-0246

### 본 결과 본 추천 시스템의 정확도가 평균 78.3%이고 각 주식/매도 추천 가능, 회원관리 가능, 추천종목 푸시 기

이와 같은 주식종목 매수/매도 추천 시스템의 엔진이 완성되면 추천 시스템 엔진 서버에서 추천 중목을 푸시 용음 잘 이해하도록 기술하였고 설계 방법으로는 핵심 메시지로 종목을 스마트폰 화면에 표시해주고 이미 추천 한 종목의 추천종목을 검색해볼 수 있도록 한다. 또한 추 천 종목에 대하여 소설네트워크서비스(SNS)로 글쓰기. 댓글닥기 좋아요의 기능과 회원가인 및 취리 가능 등의 스마트폰 앱의 화면 설계를 그림 6과 같이 설계하도록



그림 6. 스마트폰 앱 화면 설계 Fig. 6. Design of Screen for Smartphone App.

각의 알고리즘의 경우는 평균적으로 56.7% 정도로 본 능. 추천종목 검색 기능 및 소셜네트워크 서비스 기능 등 이 있는데 이는 객체지향 분석방법인 UML을 통해 유즈 케이스 다이아그램과 클래스 다이아그램을 통해 부석내 기술인 주식종목 매수/매도 추천 시스템의 엔진의 설계 와 주식종목의 매수/매도 추천을 위한 엔진의 동작 및 스 마트폰으로 푸시 전송되는 과정의 알고리즘 설계를 위한 플로우차트 제장 및 화여 디자인 성제 예시를 보여주었다.

이와 같은 본 논문에서 제시한 인공지능 기술을 이용 한 주식종목 매수/매도 추천시스템의 알고리즘 설계는 고전적인 주식 매수/매도 분석 알고리즘 보다 우수함이 시뮬레이션 결과로 확인되었다. 따라서 본 주식종목 매 수/매도 추천시스템 소프트웨어 분석 및 설계방법은 실 제 주식종목추천 시스템 소프트웨어를 개발하는데 있어 매우 유용하고 주요 분석 및 설계 프로세스를 세우는 기 준이 될 수 있으리라 본다. 또한 본 연구에서 제안한 주 식종목추천 시스템 분석 및 설계 방법이 나중에 실제 증 권사의 로봇어드바이저 시스템 구축이나 증권 주식종목 추천 어플을 제작하고자 할 때에 실무자들에게 어떻게 그 소프트웨어를 분석하고 설계합지에 대한 주요 가이드 라인으로 자리매김 할 수 있기를 기대한다.

#### References

of hmm, ann and ga for stock market forecasting. Expert Systems with Applications, 2007.

#### [그림 5. 논문 1 문제점]

### 기존 연구 문제 제기

### 기존 논문에서의 문제점 2

- 2. LSTM을 이용한 주가 예측 모델의 학습방법에 따른 성능분석 (정종진, 김지연, 2020)
- 회사 규모를 분류하지 않고 규모가 큰 대기업 만으로 학습 시행 변동성이 큰 중견, 중소기업 등에서는 해당 예측 모델이 제대로 작동하지 않을 수 있는 문제점 발생
- 모든 회사에 같은 DNN 모델을 적용하여 주가 예측 위와 같은 내용으로 각 회사마다 다른 추이를 띄고 있을 때 주식 예측이 매우 힘들어짐.



LSTM을 이용한 주가예측 모델의 학습방법에 따른 성능분석 261

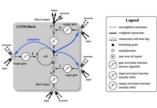


Fig. 2. Internal block architecture of LSTM

로 구성되어 있다. 입력 게이트(it ⊙ gt)는 현재의 정보 를 기억하기 위한 게이트로서, 입력(input) xt와 순환 (recurrent) ht-1을 받아서 시그모이드(σ)를 취하고 같 은 입력 xt에 대해 tanh를 취한 다음 hadamard product 연산해서 내보낸다. 망각 게이트(ft)는 과거 정 보를 잊기 위한 게이트로서, 입력 xt와 순환 ht-1을 받아 서 시그모이드를 취해서 내보낸다. 이 때 시그모이드 함 수의 출력 범위가 0 ~ 1 사이이므로 그 값이 0이면 이전 상태의 값은 잊고 1이면 이전 상태의 값을 기억하게 된 다. 각각의 게이트들은 0 ~ 1 사이의 값을 가지게 되고 셀은 연결되어 있는 게이트들의 값을 체크하면서 데이터 를 불러올지, 유지할지, 내보낼지를 결정하게 된다. 또한 게이트에 포함되는 가중치(W)와 바이어스(b)는 학습해야 할 대상이다. 본 연구에서는 딥러닝 모델인 LSTM을 이 용한 주가 예측 모델을 Keras[15]를 이용하여 구현하고 Keras 기반 LSTM 모델 적용 시에 모델 성능에 영향을 미치는 중요한 매개변수들에 대해 실험하였다.

#### 3. 주가 예측 모델 구축 및 평가

#### 3.1 데이터 수집 및 전처리

본 연구에서는 미국의 대표적인 글로벌 회사들에 대한 주가 데이터를 바탕으로 LSTM 기반의 예측 모델을 수립 하였다. 이 때 학습 및 예측하는데 있어서 필요한 주가 히스토리 데이터가 준비되어야 하는데, 이는 Yahoo API[16]를 통해 수집하였다. 원시 데이터는 날짜, 시작 가, 최고가, 최저가, 종가, 수정종가, 거래량으로 구성되 어 있다. 사용한 데이터는 2018년 11월 1일 기준 나스닥 의 상위 종목 20개 중 훈련(Training)에 10종목, 예측 (Prediction)에 5종목을 사용했으며 데이터는 2000년 1

다. 이에 따라 훈련에 사용한 종목은 애플, 어도비, 아마 존, 컴캐스트, 코스트코, CISCO, 인텔, 마이크로소프트, 엔비디아, 펩시 등이며, 예측에 사용한 종목은 구글, 페이 스북, 페이팔, 브로드컵, 퀄컴 등이다. 이들을 선택한 이 유는 동일한 주식시장의 상장종목이며 가격 단위가 미국 달러로 일정하기 때문이다. 데이터에 대한 전처리 작업으 로서 스케일링(Scaling)을 해야 하는데 그 이유는 이 회 사들의 주가 범위가 크게 다르므로 데이터의 스케일을 맞춰 가중치의 스케일도 일관성 있게 맞춰지는 효과를 위해서이다. 사용한 방식은 데이터의 최대값을 1, 최소값 을 0으로 두는 최소최대(MinMax) 스케일링 방법이다.

학습을 위한 데이터의 특징(Feature)을 선택할 때. 본 모델은 다음날의 수정종가를 예상하기 위한 모델이기 때 문에 시작가, 최고가, 최저가, 종가를 사용하지 않고 수정 종가를 선택하였고, 거래량은 해당 종목의 예측에 주로 쓰이는 데이터 중 하나이기 때문에 특징값으로 사용하였 다. 또한 값이 변화하는 경향을 파악하기 위해 전날 데이 터와 비교한 수정종가 변화량과 거래량 변화량을 특징값 으로 사용하였다. 그런데 거래량 특징값에 대해서는 거래 량 수치와 경향(거래량의 변화량)을 같이 학습시켜야 하 는지, 아니면 경향만 학습시키면 되는 것인지가 검토 과 정에서 문제가 되었다. Fig. 3에서 보이는 바와 같이 실 제 학습을 수행해서 모델의 성능을 평균제곱오차(RMSE) 방법으로 비교한 결과. 경향만 있는 경우가 더 성능이 좋 아서 모델에서 거래량 수치를 제외시켰다.



Fig. 3. Comparison of input feature selections

구성한 학습 데이터 셋은 학습 그리고 자체적인 평가 를 위해서 훈련(Train), 검증(Validation), 추론(Test) 데 이터 셋으로 나누게 되는데, 훈련 데이터 셋은 학습을 위 한 것이고, 검증 데이터 셋은 중간 학습 평가 및 학습률 (Learning Rate) 등의 파라미터 값을 수정하기 위한 것. 

#### [그림 5. 논문 2 문제점]

주식 별 보조지표 활용 주가 추이 분석, 예측 및 시각화

### 03

### 기존 연구 문제 제기

### 기존 논문에서의 문제점 3

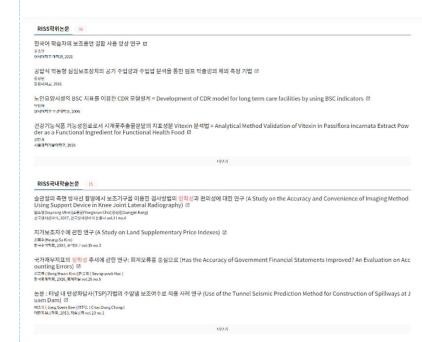
- 3. 현 시점, 보조지표를 활용한 투자방식 분석에 대한 연구진행의 부족
- 실제 분석 결과가 없어 ML, DNN 구성 과정에서
  어떤 보조지표가 주가 추이 분석에 유용한 지 알 수가 없음.

진행된 연구가 없음에 따라 유용하고 효율적인 지표를 사용하지 않고 직관적이고 추상적으로 활용 중.

-> 추가적으로 분석 DB를 구축하여 향후 연구에서 유용할 수 있도록 제공할 예정.

**X ML : Machine Learning** 

**X DNN: Deep Learning** 



[그림 6. 진행된 연구의 부족]





### 03

### 기존 연구 문제 제기

### 기존 논문에서의 문제점 3

- 2. 현 시점, 보조지표를 활용한 투자방식 분석에 대한 연구진행의 부족
- 진행된 연구가 없음에 따라 유용하고 효율적인

지표를 사용하지 않고 직관적이고 추상적으로 활용 중.

※ 해당 논문에서는 RSI, Bollinger Bands, Moving Average 등

추세, 변동성, 모멘텀 보조 지표를 활용하고 있지만 근거가 명확하지 못하다는 한계 존재.

#### 보조지표의 종류

크게 5가지로 구분한다.

구분	종류	
추세 지표	MACD, DMI, ADX, ROC 등	
모멘텀 지표	Stochastic, RSI, 이격도, 삼선전환도 등	
변동성 지표	Bollinger Band 등	
<mark>거래량 지표</mark>	<mark>거래량 이동평균선,</mark> OBV, VR 등	
기타	일목 균형표 등	

[그림 7. 보조 지표의 종류]



시도하는 논문에서도 단일 학습 모델을 통하여 결론을 도출하는 사례도 있다[13].

#### 2.4 기존 연구의 주가 예측 모델의 한계

기존 연구의 주가 예측 모델 논문은 많은 테이터를 기용함으로 과적합을 고려하지 않고 높은 학습률, 딥러닝 모델을 만든 경향이 있다. CNN, MLP 딥러닝 모델을 단독으로 사용하는 것은 과거의 학습 결과가 사라지는 현상인 장기의존성 문제 해결이 되지 않은 모델이기에 주식 테이터의 특성인 시계열 테이터 학습에 한계가 있다. 또한 모델학습률의 최도인 Loss만 측정함으로 실제로 학습이 얼마나 잘 되었는지 명시되지 않았다.

본 논문의 모델은 위와 같은 문제점을 보완하여 신뢰성 있는 학습과 과적합을 방지하기 위해 시작가, 고가, 좋가, 수정 좋가, RSI, Bollinger Bands, Moving average 20일 60일 지표 데이터를 가공해서 사용한다.

또한 관련 연구에서 문제가 될 수 있는 장기의존성 문제에서 자유로운 LSTM, GRU와 GRU 앙상블 모델을 만들어 정확도가 높은 모델을 제시한다.

[그림 8. 실제 구현된 AI 모델 예] ※ 참고논문, 변동성이 큰 주식시황에 적합한 AI 모델에 대한 연구, 신응조, 2023.06)

#### 주식 별 보조지표 활용 주가 추이 분석, 예측 및 시각화

### 03 기존 연구 문제 제기

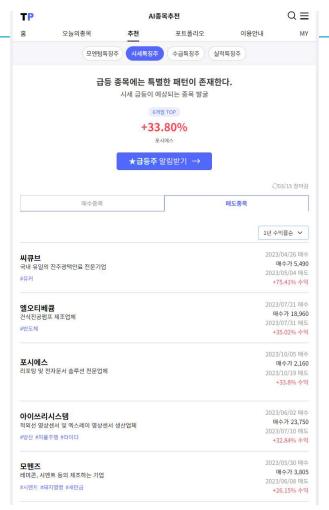
### 기존 논문에서의 문제점 4

4. 현재 주식 추천 알고리즘이 해당 주식을 왜 추천하는지 알 수 없는 시각화 문제 (+alpha)



[그림 9. AI 추천 모델 1] https://www.judal.co.kr/?view=stockList&themeldx=25



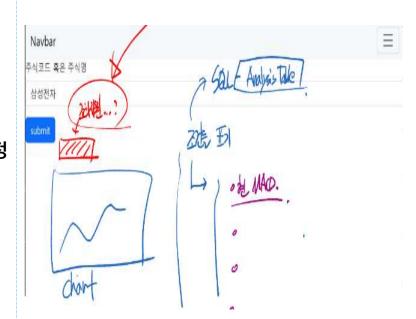


[그림 10. AI 추천 모델 2] https://m.thinkpool.com/advisor/recomm/00000054

### 04 해결 방안 제시

### 학문적, 사회적 기여

- 1. 보조지표 특성을 고려하여 계산방법 설계, 활용 방법 제시 (학문)
- 2. 보조지표가 회사 규모 별 실제 활용 가치의 파악 가능 (학문)
- 3. 부족한 보조지표 활용 방법 설계 및 제시 (학문)
  - 보조지표를 조합한 계산 방법에 따른 회사별 주식 추이 확률 제공예정
- 4. 주식 별 보조지표 계산 데이터를 DataBase로 제공하여 향후 연구에 유용하게 사용할 수 있도록 제공할 예정 (학문, 사회)
- 5. (Front-end) 실제 계산된 회사 별 확률을 시각적으로 제공하여 실 투자자에게 쉽고 빠르게 주식 추이를 직접 판단할 수 있는 Web 개발예정 (사회)



[그림 11. 구현된 Web 초안 (Prototype)]





### 04 해결 방안 제시

### 창의성

#### 1. 지표 분석 속도 향상 측면

> 일반적으로 Train Data는 .csv 파일 내지 EXCEL 파일로 제공되지만 속도 향상을 위해 MongoDB (Database) 를 사용하여 데이터 전달 예정

※ MongoDB 는 기존의 RDBMS 와 다르게 JSON 형태로 저장되며 Front-end, AI에서 데이터를 추출하는 과정이 쉽고 자유도가 높은 DataBase입니다.

#### 2. 보조지표를 활용하는 계산 수식 설계

- > 보조지표, 선 차트를 조합하여 근거를 통해 가설한 수식 설계 및 프로그램으로 구현
- 3. **데이터 시각화** (\* 논문 완성 후 추가 목표)
  - > Front-end 분 들의 노력을 통해 최대한 쉽게 시각화 할 예정.



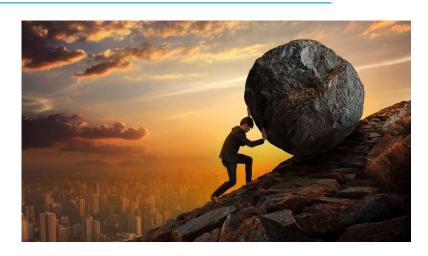


### 04 해결 방안 제시

### 난이도

- 1. 일 별 데이터에 대한 필요한 정보를 추출하는 과정 : 3/5
  - 어떻게 주식 데이터를 가져올 것인가?
- 2. 보조지표 계산방법을 세우고 직접 계산하는 과정 : 5/5
  - 보조지표 계산 방식에 대한 설계 과정 생각 및 가설 설정 과정
- 3. 해당 데이터를 MongoDB 에 올리고 가공하는 과정 설계 : 4/5
  - 실제로 데이터={"Crawling 데이터", "분석데이터", ""} 등을 DB로 넣어주고 가져와야 함
  - Python을 이용하여 MongoDB 를 자유롭게 사용할 수 있는 Driver 제작 과정 필요
- 4. 회사 규모 별 보조지표 별 추이 정확성 확률 계산 알고리즘 수립 및 논문 제작 : 5/5
  - 데이터 분석 프로그램 (python, c, c++ 등) 을 제작 해야함
  - 데이터 시각화와 분석 필요
- 5. 구성된 DB를 통해 사용자에게 시각화하는 과정: 5/5
  - Javascript (Spring, Node.js, ..) 등을 활용하여 DB에 올라간 데이터를 가공하여 사용할 수 있어야함.
  - Front-end 설계과정에서 DB 운용방법 및 설계과정을 생각 해야함.





#### 주식 별 보조지표 활용 주가 추이 분석, 예측 및 시각화

### 05 참고 논문 및 자료

### 참고 논문

1. 인공지능 머신 러닝 기술을 이용한 주식 종목 매수/매도 **추천시스템의 분석 및 설계 (2021)** 저자 : 조병호

2. LSTM을 이용한 **주가 예측 모델**의 학습방법에 따른 성능분석 (2020) 저자 : 정종진, 김지연

3. 변동성이 큰 주식시황에 적합한 AI 모델에 대한 연구 (2023.06) 저자: 신응조

4. 재무비율과 기술적 분석을 통한 AI 주식 트레이딩 알고리즘 모델링 저자: 정해성 등 5명 (2019 대학생 프로젝트 경진대회)

### 참고한 자료들

- 1. PPT Template
  - 국도리 템플릿 https://m.blog.naver.com/doll\_qls/221265211071
- 2. 용어 정리
  - 네이버 사전
  - 보조지표의 본질과 종류, 활용법 https://mcguffin.tistory.com/118

### 이미지 출처

1. 돌 굴리는 이미지

https://www.philmckinney.com/embrace-the-hard-the-secret-to-a-fulfilling-life/



### 끝까지 들어주셔서 감사합니다.

Presenter 컴퓨터학부 4학년 **배준형** 

Project Team **Member** 컴퓨터학부 4학년 **서유준** 컴퓨터학부 4학년 **유승준** 컴퓨터학부 4학년 **이준혁** 

