
KR-finBERT를 활용한 주가 예측 모델 개선

A반 1조

김상준 김현석 배수연 손나리 이진수 주성현

<https://github.com/SNU-dataproject/kr-finbert>

Contents

- | | | |
|---|-----|---------------------------|
| 1 | 서론 | Motivation |
| 2 | 본론1 | Dataset |
| 3 | 본론2 | Model |
| 4 | 결론 | Contribution / Limitation |

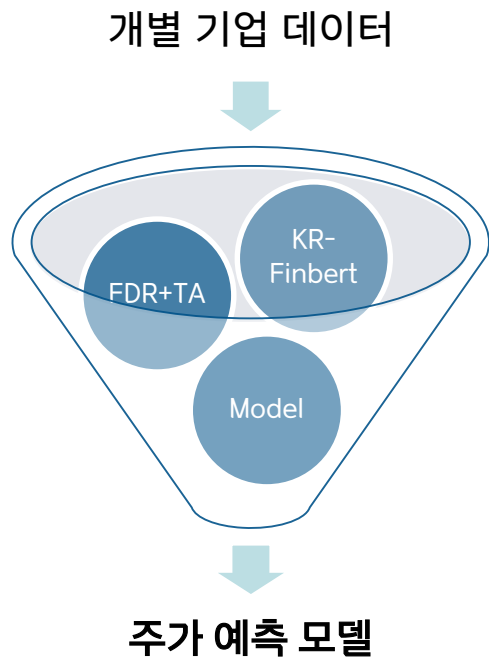
1

Motivation



1 Motivation

Overview



KR-FinBERT를 통한 감성분석과
기술 분석 지표를 활용한 **주가 예측 모델**

1 Motivation

3C분석 - ① 당행



Buisness Side

당행 시장예측시스템 내 주가예측모델에서 착안

└─> 직원이 PB 고객 상담 시 사용



Academic Side

수업에서 배운 크롤링 / 딥러닝 모델을
응용해볼 수 있을까?

1 Motivation

3C분석 - ② 산업

Summary

개인자산관리서비스 고도화



WHY?

- ✓ 자산관리서비스 수요
- ✓ 금융자산 구성 중 주식 증가
- ✓ 신뢰도 : 은행 > 빅테크

1 Motivation

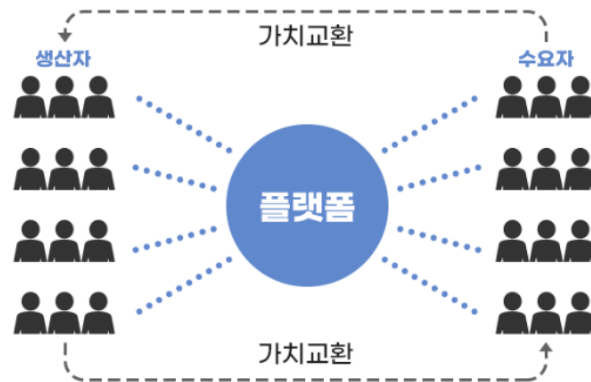
3C분석 - ② 산업(Cont'd)

“ 빅테크 등장으로 금융상품의 **생성**과 **판매** 분리 리스크 ”

-바젤은행감독위원회(BCBS)

제.판분리

플랫폼의 구조 : 양면시장(Two-sided Market)



1 Motivation

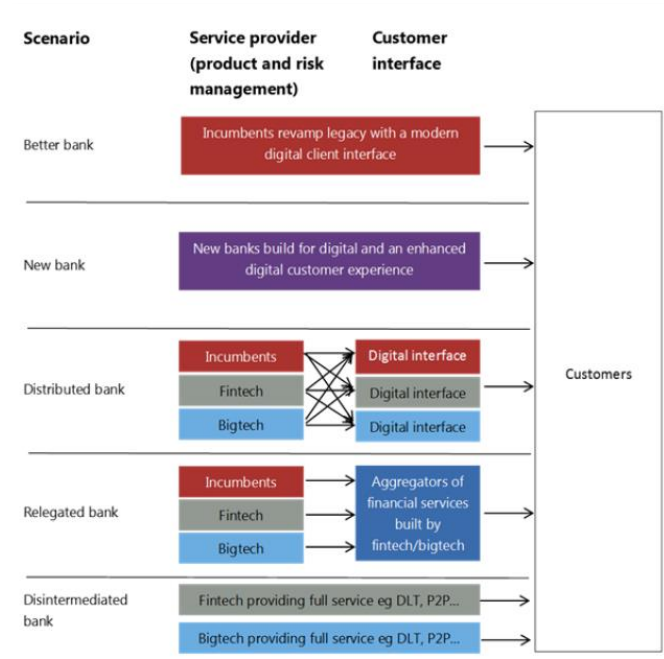
3C분석 - ② 산업(Cont'd)

기존 은행들의 전략적 성장 분야는

개인자산관리, 개인 및 중소기업 대출 관련

비즈니스 확대 등의 사업

-McKinsey, Future of Asia: The future of financial services



-바젤은행감독위원회, 2017

1 Motivation

3C분석 - ③ 고객

- ✓ 자산관리 서비스 수요 “40세 이하 초고액자산가의 약 40%는 자산관리 서비스에 비용 지불 의사...”

-글로벌 IT컨설팅기업 캡제미니(Capgemini)의 '2020년 세계 부 보고서'

- ✓ 금융자산 구성 중 주식의 증가

- ✓ 신뢰도 : 은행 > 빅테크

한국 가계 금융자산구성

단위 %

연말	현금·예금	금융투자상품				보험·연금	기타
		주식	채권	펀드	소계		
2015	42.8	16.5	5.7	3.5	25.7	31.1	0.5
2016	43.7	15.7	5.1	3.2	24.0	31.8	0.5
2017	43.1	17.4	4.2	3.0	24.7	31.8	0.5
2018	44.3	15.2	4.2	2.9	22.3	32.9	0.5
2019	44.8	15.3	3.6	2.8	21.7	32.8	0.6
2020	43.4	19.4	3.4	2.4	25.2	30.8	0.7

* 자료 : 금융자산부채잔액표(한국은행, '21.4.8)

Respondents' trust in counterparties to safeguard their data

Graph 1

Trust varies by counterparty...

... and by gender, ethnicity and age²

Concerns when sharing data



2

Dataset



2 Dataset

['Open', 'High', 'Low', 'Close', 'Volume', 'Change',

} 주가 데이터

'Rn', 'MRn', 'Volatility10', 'Volatility30', 'Volatility60', 'Volatility80',
'Volatility120', 'Volatility180', 'Volatility360', 'MA3', 'MA10', 'MA20',
'MA60', 'MA120', 'MA180', 'MA360', 'ATR', 'MACD-line', 'MACD-hist',
'MACD-sign', 'PSAR', 'RSI', 'ADX', 'OBV', 'OSCILLATOR',

} 기술적 지표

'negative', 'neutral', 'positive', 'sentiment', 'sentiment_to_num']

} 감성 지표

총 37개의 컬럼 사용

2 Dataset

Step1. FinanceDataReader & Technical Analysis Library

기술적 지표

지표 종류	의미	종류
추세 지표	주가가 진행하는 방향인 추세를 알아보는 지표 현재의 추세가 일정 기간 지속된다고 가정하고 이를 이용해 이익을 내고자 할 때 이용	- 이동평균선 - 추세선 - MACD (Moving Average Convergence Divergence) - DMI (Directional Movement Index)
모멘텀 지표	현재 주가의 일정 기간 이전의 추가 차이 또는 비율을 나타냄 현재의 추가 추세가 속도를 더하고 있는지, 끌어 들고 있는지를 판단하여 추가 추세와의 진행과 변곡점을 찾는 지표	- 이력도 - 스토캐스틱 (Stochastic) - CCI (Commodity Channel index) - 투자심리선
변동성 지표	주가의 방향이 아닌 변동성을 기준으로 시작할 한 지표 매매 타이밍을 잡거나 단기매매에 활용	- 볼린저 밴드 (Bollinger Band) - Envelope - Parabolic SAR (Stop And Reverse) - ATR
시장강도 지표	주가의 추세나 변동성이 얼마나 강한지 나타내는 지표 주로 거래량을 이용하여 분석한다.	- OBV (On Balance Volume) - MF (Money Flow Index) - VO (Volume Oscillator) - VR (Volume Ratio)
기타 지표	기타 지표	- 일목균형표



```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1817 entries, 2015-01-02 to 2022-05-20
Data columns (total 31 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   Open        1817 non-null   int64
 1   High        1817 non-null   int64
 2   Low         1817 non-null   int64
 3   Close       1817 non-null   int64
 4   Volume      1817 non-null   int64
 5   Change      1817 non-null   float64
 6   Volatility10 1816 non-null   float64
 7   Volatility30 1816 non-null   float64
 8   Volatility60 1816 non-null   float64
 9   Volatility80 1816 non-null   float64
10  Volatility120 1816 non-null   float64
11  Volatility180 1816 non-null   float64
12  Volatility360 1816 non-null   float64
13  MA3         1815 non-null   float64
14  MA10        1808 non-null   float64
15  MA20        1798 non-null   float64
16  MA60        1758 non-null   float64
17  MA120       1698 non-null   float64
18  MA180       1638 non-null   float64
19  MA360       1458 non-null   float64
20  ATR         1817 non-null   float64
21  MACD-line   1792 non-null   float64
22  MACD-hist   1784 non-null   float64
23  MACD-sign   1784 non-null   float64
24  PSAR        1817 non-null   float64
25  PSAR-down   907 non-null    float64
26  PSAR-up     1008 non-null   float64
27  RSI         1804 non-null   float64
28  ADX         1817 non-null   float64
29  OBV         1817 non-null   int64
30  OSCILLATOR   1784 non-null   float64
dtypes: float64(25), int64(6)
memory usage: 454.2 KB
```

주식 시장에서 사용되는 기술적 지표들 중 25개를 컬럼으로 추가

2 Dataset

Step2. NLP Raw Data Crawling

삼성전자 005930 | 코스피 | 2022.05.24 15:11 기준(장중) | [종가](#) | [기업정보](#)

66,600
전일 대비 ▼1,300 -1.91%

전일 67,900 | 고가 67,700 (상한가 88,200) | 거래량 13,764,219
시가 67,500 | 저가 66,500 (하한가 47,600) | 거래대금 921,151 백만

종합정보	시세	차트	투자자별 매매동향	뉴스공시	종목분석	종목토론실	전자공시	공매도현황
종목뉴스	✓재직	✓비용						종목뉴스 인쇄
제목	정보제공	날짜						
쿠에 하얀날 재계...삼성 현대차 못대, '역대급' 투자 보따리(종합2부)	아시아경제	2022.05.24 14:55						
삼성, 미래 먹거리에 5년간 450조 쏟아붓는다...8만명 채용 [종합]	한국경제	2022.05.24 14:37						
삼성전자 5년 간 450조원 투자8만명 신규채용 관련뉴스 1건 더보기	채널24경제	2022.05.24 14:02						
삼성, 적당할 일자리 115만개 창출... 채용 안정 위해 공채 유지	조선비즈	2022.05.24 14:37						
삼성, 5년간 반도체 바이오 등에 450조원 투자...8만명 신규 채용	조선비즈	2022.05.24 14:26						
삼성, 5년간 반도체 바이오 등에 450조원 투자...8만명 신규 채용 삼성, 5년간 반도체 바이오 등에 450조원 투자...8만명 신규 채용 관련뉴스 2건 더보기	한국경제	2022.05.24 14:15						
쿠에 '반도체 조강대국' 주력수 미흡한 '삼성전자, 초유의 기업...'	파이낸셜뉴스	2022.05.24 14:20						
삼성전자, 서스텝반도체 파운드리 집중 육성... '반도체 조강대국 달성'	아시아경제	2022.05.24 14:00						
삼성, 바이오-차세대 통신 공격적 투자로 '제2반도체 신화' 구현	아시아경제	2022.05.24 14:18						
【특종주】아연전석, 삼성 반도체 바이오 등 450조 투자...반도체...	파이낸셜뉴스	2022.05.24 14:17						
삼성전자, 향후 5년간 신규 8만명 채용 '고용시장 혼동'	아시아경제	2022.05.24 14:09						
【속보】삼성, 반도체 등 핵심인력 확보는 5년간 8만명 채용 관련뉴스 1건 더보기	한국경제	2022.05.24 14:02						
삼성, 향후 5년간 450조원 투자 8만명 신규채용(상보)	아시아경제	2022.05.24 14:01						
【특종주】삼성전자 2% 가격하락...이들엔 '파란불'	조선비즈	2022.05.24 13:58						

1 2 3 4 5 6 7 8 9 10 다음 > 현재 >>

네이버 금융 뉴스 기사

N | +주가 +상전 삼성전자

장별 > 관련도순 > 최신순 > 오래된순

기간 > 전체 > 1시간 > 1일 > 1주 > 1개월 > 3개월 > 6개월 > 1년
2017.06.01. ~ 2022.02.07. >

유형 > 전체 > 포로 > 등장상 > 지면기사 > 보도자료 > 자동생성기사

언론사 > 전체 > 언론사 분류순 > 가나다순 >

기자명 > 기자명 입력 > 적용

○ 옵션 초기화 > 검색옵션 가이드

PICK 언론사가 선정한 주요기사 혹은 심층기획 기사입니다.

매일경제 PICK | A22면 TOP | 2022.02.07. | 네이버뉴스
LG연출 2조4천억 담노라 삼성 8천억 댔었다...연기금발 수급교란
불가불가 추가 추가가능 가능성 LG에너지솔루션이 수급 폭조에 힘입어 7월 9% 가가
이 급등하며 시가총액... 반면 연기금은 LG에너지솔루션 상장 후 포트폴리오 리밸...

EBN | 2022.02.07.
서울반도체, 올해 삼성전자 WOLED TV 출시 어려워-신한금융투자
투자의견은 매수 유지, 목표주가는 2만1000원이다. 7월 말전후 신한금융투자 연구원은 "최근 삼성전자의 WOL
ED 패널 구매 노이즈가 잦아지고 있다"며 "일부 보도에 따르면 패널 할당이 줄었다"고 밝혔다.

한국경제 PICK | A25면 1단 | 2022.02.06. | 네이버뉴스
"이 와중에 파업한다고"...싸늘한 삼성전자 주주들
하지만 같은 기간 삼성전자가 주가 상승률은 4.02%에 그쳤다. 노조의 행보를 악재로
판단한 투자자가 적지 않았다는 설명이다. 특히 노조가 영업이익의 25%를 성과...

네이버 뉴스 기사

분석리포트(요약)

기업명 삼성전자 >

조회기간 2017/06/01 ~ 2022/06/10 1주 > 조회

일자	기업명	요약	투자의견	작성주가	총가	기관명
2022/06/07	삼성전자(005...	삼성전자-Unique 합에 주목할 때, 투자의견 매수... o EUV 생태계 구축 o GAA, 3D DRAM 기반 기술 o 자체 Memory + Logic 차세대 패키징 솔루션	매수	84,000	65,500	미래에셋증권
2022/06/07	삼성전자(005...	SamsungElec-Technological differentiation co... o Creation of an EUV ecosystem o GAA and 3D DRAM-enabling technologies o In-house next-generation packaging solution...	매수	84,000	65,500	Mirae Asset S
2022/06/03	삼성전자(005...	삼성전자-P/B 1.3배 rock bottom 예상 o 2022 영업이익 14.8조원 전망 o GAA, 3D DRAM 기반 기술 o 목표주가 하향, 투자의견 'BUY' 유지	BUY	89,000	66,800	키움증권
2022/05/31	삼성전자(005...	삼성전자-실적 모멘텀 이후 주주에게 줄대는 필연적 o 스마트폰 출하 하향 조정 불구, 반도체 실적 개... o 실적 모멘텀에 기반해 안정적인 주가 흐름 예상, ...	BUY	84,000	67,400	메리츠증권
2022/05/31	삼성전자(005...	삼성전자-각종 모멘텀 이후 주주에게 줄대는 필연적 o 반도체 서버가 버림됨 o 2022년 매출액 74.6조원(-4.1% QoQ), 영업이익... o 투자의견 BUY 유지, 목표주가 100,000 원 하향	BUY	100,000	67,400	케이프투자증권
2022/05/31	삼성전자(005...	삼성전자-기술 리더십 지속 o 신주 입체화의 기술 리더십 유지 지속 o M&A 기대감 부각	BUY	87,000	67,400	NH투자증권
2022/05/30	삼성전자(005...	삼성전자-4만회 중반대이면 불발 매수 가능 o 한계선에 놓여있던 실적면 변수... o 삼성실적 악재 총류는 거의 다 나온 듯하다는 점...	중장기주...	77,000	67,700	상상인증권
2022/05/25	삼성전자(005...	삼성전자-2H22 수요가 관건 o 자율주행과 파운드리 o 2H22 수요가 관건 o 투자의견 Buy, 목표주가 87,000원 유지	BUY	87,000	66,400	이베스트투자재
2022/05/25	삼성전자(005...	SamsungElec-Demand is the key in 2H22 o Autonomous driving and foundry o Demand is the key in 2H22	BUY	87,000	66,400	eBEST Investr

증권사 리포트

2 Dataset

Step2. NLP Raw Data Crawling

2017/06/01~Latest : 삼성전자(005930), 현대자동차(005380), 카카오뱅크(323410)

✓ 네이버 금융의 뉴스기사

2022/02/08 ~ 2022/06/01

(검색 기업에 따라 제공되는 뉴스의 최초 날짜가 다름)

1	title	press	date
2	'갤럭시 Z4' 기대감 커지는데...삼성 마냥 웃을 수 없는 이유	이데일리	2022.06.01
3	"삼성·LG 제쳤다"...1분기 노트북 시장 장악한 이 기업, 어디길래?	매일경제	2022.06.01
4	올들어 이재용 삼성전자 부회장 주가평가액 1.1조원 줄었다	이데일리	2022.06.01
5	이재용 삼성전자 부회장 주가평가액 올 들어 1.1조원 증발	서울경제	2022.06.01

✓ 네이버 기사에 검색조건 추가

2017/06/01 ~ 2022/02/07

(네이버 기사는 최대 4000개의 데이터만 보여주기 때문에, 관련성을 높이면서 기간 내의 결과를 얻기 위해 +주식 +주가 +삼성 등 검색 조건 추가)

1	date	title
2	2022.02.07.	LG엔솔 2조4천억 담느라 삼전 8천억 던졌다...연기금발 수급교란
3	2022.02.07.	서울반도체, 올해 삼성전자 WOLED TV 출시 어려워-신한금융투자
4	2022.02.06.	"이 와중에 파업한다고?"...싸늘한 삼성전자 주주들
5	2022.02.04.	[차영주의 금요일장] 저가매수 VS 현금 보유...2월 증시에 대처하는 올바른 자세는?

✓ 증권사 레포트

2017/06/01 ~ Latest

1	date	title						
2	2017/07/03	Samsung Electronics-2Q17 preview: W13.9tn OP to beat expectations,2분분						
3	2017/06/29	삼성전자-이익은 증가 & 변동성은 축소 & 주식수는 감소,2분기 전세계 비금						
4	2017/06/26	삼성전자-반도체, OLED가 하반기 실적 견인,미국 기술주, 아직 고평가라고 말						
5	2017/06/22	삼성전자-모든 우려를 불식시키는 메모리 공급부족 가능성,2017년, 3D Nand						

2 Dataset

Step3. Merge, Sort, Preprocess, Combine

Preprocess

Rule

YoY or yoy or YOY	-> 전년 대비 증감율
QoQ or qoq or QOQ	-> 전분기 대비 증감율
MoM or mom or MOM	-> 전월 대비 증감율
YTD	-> 연초 누계
MTD	-> 월초 누계
분기 표현 1Q21F	-> 21년도 1분기 예측값
분기 표현 1Q21	-> 21년도 1분기
반기 표현 1H21	-> 21년도 상반기
[내용] 제거	
문장 시작의 "삼성전자-" 삭제	
다중 띄어쓰기 제거	
앞뒤 공백 제거	

Result

삼성전자-가격 협상력은 공급자 우위로
□ NAND 생산 차질로 업황 조기 개선 전망
□ 1Q22F 매출액 71.7조원(-6.3% QoQ),
영업이익 13.2조원(-4.8% QoQ) 전망
□ 투자의견 Buy, 목표주가 105,000원 유지



가격 협상력은 공급자 우위로, NAND 생산
차질로 업황 조기 개선 전망, 22년도 1분기
예측값 매출액 71.7조원(-6.3% 전분기 대비
증감율), 영업이익 13.2조원(-4.8% 전분기
대비 증감율) 전망, 투자의견 Buy, 목표주가
105,000원 유지

kr-finBERT

Negative : 0.055891796946525574
Neutral : 0.00046004404430277646
Positive : 0.943648099899292



명확

Negative : 0.00921998918056488
Neutral : 0.00026661704760044813
Positive : 0.9905133247375488

2 Dataset

Step3. Merge, Sort, Preprocess, Combine


Combine

date	title													
20170601	System LSI 투자 증가,삼성 파운드리 포럼 개최,17년도 2분기 영업이익 12.0조원(+22% 전분기 대비 증감율)으로 증가 예상													
20170602	모바일을 넘어 데이터 시대로,PC/휴대폰 시장 둔화에도 데이터 시장의 확대는 이익성장과 투자기회 제공,반도체 호황은 시장 기대보다 길어질 것이고, 과징													
20170607	지속되는 어닝 서프라이즈와 주주환원정책의 하모니,중장기적으로 이어질 실적 성장세,2분기 영업이익 13.6조 원 전망, 건조한 실적과 강화되는 주주환원,													
20170608	말릴 수 없는 질주,2분기 실적 기대치 상회 전망,실적 추정치 지속 상승 중,17년도년도 1분기 2분기 Preview: 반도체와 IM 실적추정치 상향조정,년도년도 1분													
20170609	17년도 2분기 실적 호조 지속 (요약),17년도 2분기 영업이익 12.9조원(+30% 전분기 대비 증감율) 예상, Samsung Electronics-Robust earnings to continue													
20170612	투자자 시선은 벌써부터 하반기로,17년도 2분기 매출 58.7조원 (16%전분기 대비 증감율), 영업이익 13.2조원 (34%전분기 대비 증감율) 전망,부문별 영업이익													
20170615	Samsung Electronics-World leading results,2분기 영업이익 13.1조원(+32.4% 전분기 대비 증감율) 전망,애플과 인텔을 사상 최초로 뛰어넘을 2분기 실적, 2													
20170616	장기적으로 IT 대형주 중 가장 양호한 주가 흐름 예상,올해 총 9.3조원에 해당하는 자사주 매입, 소각 프로그램을 실시 중,17년도 2분기 영업이익은 13.3조원													
20170620	주주환원 정책과 실적 모멘텀 지속,2분기 연결 매출액은 기존 추정치와 유사한 60.2조원,영업이익은 Galaxy S8과 스마트폰 출하량 추정치 변경을 반영하여													
20170622	모든 우려를 불식시키는 메모리 공급부족 가능성,2017년, 3D Nand의 경쟁력과 DRAM 기술력은 독보적 지위 유지,하반기에도 메모리 반도체 Tight한 수급													

Comma로 구분하며 날짜 별로 하나의 row로 연결

2 Dataset

Step4. Add kr-FinBERT value



date	title	negative	neutral	positive	sentiment	sentiment_to_num
20170622	모든 우려를 불식시키는 메모리	0.997897	0.000579	0.001524	Negative	-1
20170626	반도체, OLED가 하반기 실적 견인	0.000372	8.29E-05	0.999545	Positive	1
20170629	이익은 증가 & 변동성은 축소 &	0.688004	0.200839	0.111158	Negative	-1
20170703	Samsung Electronics-17년도 2분	0.000255	0.9994	0.000345	Neutral	0
20170704	17년도 2분기 Preview: 영업이익	0.074701	0.012141	0.913158	Positive	1
20170705	주가가 올랐어도 여전히 싸다,2분	0.002253	0.001028	0.996718	Positive	1
20170706	Earnings Season을 화려하게 개	3.7E-05	0.999838	0.000125	Neutral	0

sentiment	sentiment_to_num
negative	-1
neutral	0
positive	1

각각 row에 kr-finBERT 모델 적용하여 가장 높은 값으로 감성 결정

2 Dataset

Step5. Add kr-FinBERT value to FinanceDataReader

결측치 처리

negative	neutral	positive	sentiment
0.003329	0.071755	0.924916	1
0.011197	0.002902	0.985901	1
0	0	0	
0.000138	0.000351	0.999511	1
3.66E-05	8.24E-05	0.999881	1
4.13E-05	8.48E-05	0.999874	1
0.001564	0.995416	0.00302	0



negative	neutral	positive	sentiment
0.003329	0.071755	0.924916	1
0.011197	0.002902	0.985901	1
0.005667	0.001626	0.992706	1
0.000138	0.000351	0.999511	1
3.66E-05	8.24E-05	0.999881	1
4.13E-05	8.48E-05	0.999874	1
0.001564	0.995416	0.00302	0

바로 위의 값과
null이 아닌 아래 값의
평균을 이용

크롤링한 데이터의 감성분석 결과와 FinanceDataReader를 날짜 별로 연결하고, 결측치 처리

3

Model



3 Model



KR-Finbert

금융데이터만으로 학습시킨 BERT 한국어 모델



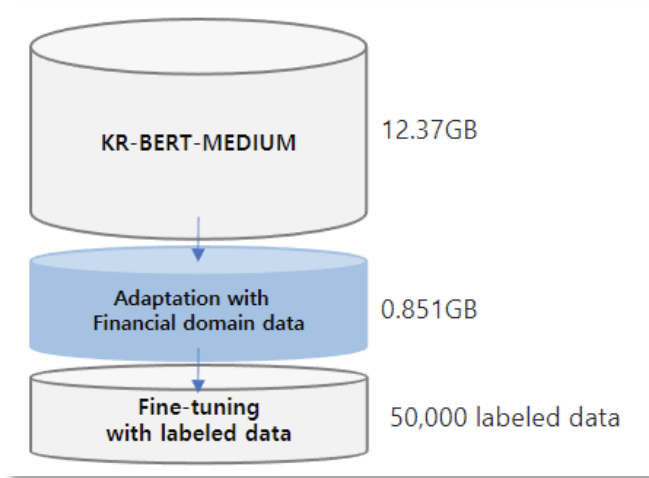
LSTM, RNN, Prophet

다양한 시계열 분석 모델

3 Model

KR-Finbert

금융데이터 만으로 학습시킨 BERT 한국어 모델

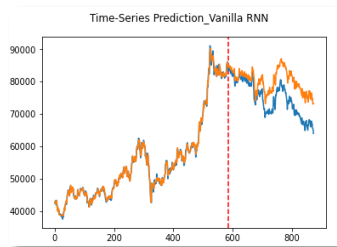


- ✓ KR-BERT-MEDIUM을 기반으로 금융 정보 학습시킨 모델
- ✓ 금융 도메인 데이터
 - 파이낸셜 타임즈 등 72개 언론사 경제 뉴스
 - 키움증권 등 16개 증권사 애널리스트 보고서
- ✓ 감성분석을 하여 Positive / Neutral / Negative 분류
 - 세 값의 합계가 1이 되도록 하는 0~1 사이의 값

3 Model

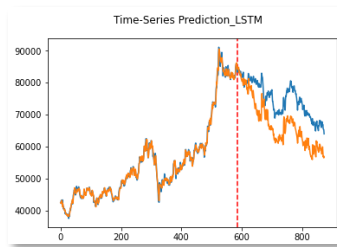
모델링 과정

RNN



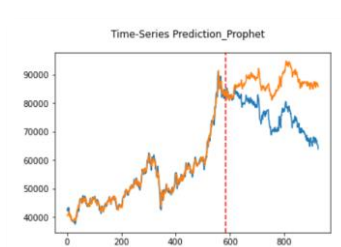
MSE: 12,943,624

LSTM



MSE: 23,056,754

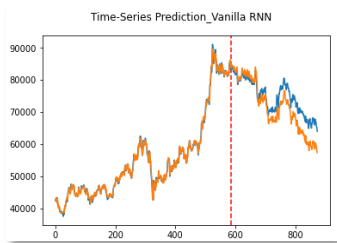
Neural Prophet



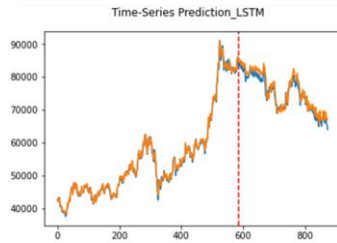
MSE: 188,622,756

Base Model

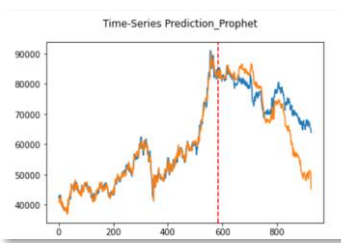
Base Model
+
FinBERT



MSE: 6,606,130



MSE: 2,446,035

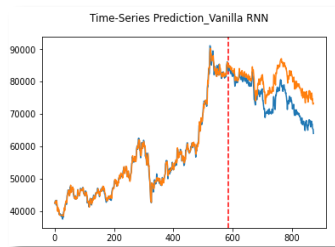


MSE: 70,610,409

3 Model

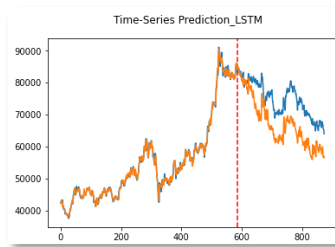
모델링 과정

RNN



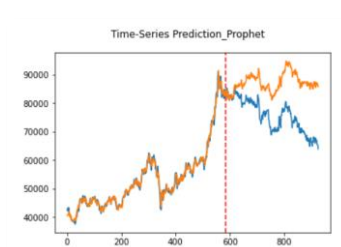
MSE: 12,943,624

LSTM



MSE: 23,056,754

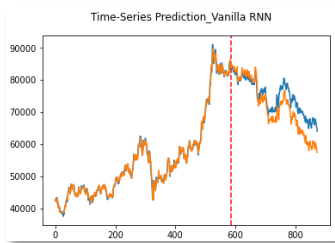
Neural Prophet



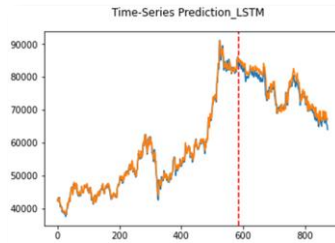
MSE: 188,622,756

Base Model

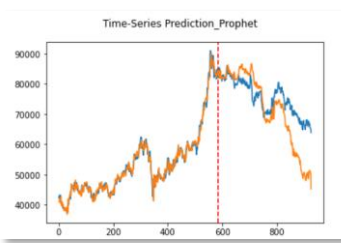
Base Model
+
FinBERT



MSE: 6,606,130



MSE: 2,446,035



MSE: 70,610,409

3 Model

MSE t검정

- 변수1 : Base Model
- 변수2 : Base Model + FinBERT

RNN

t-검정: 이분산 가정 두 집단		
	변수 1	변수 2
평균	33781214.73	6798337.981
분산	4.34189E+15	4.17564E+13
관측수	19	20
가설 평균차	0	
자유도	18	
t 통계량	1.776850323	
P(T<=t) 단측 검정	0.0462481	
t 기각치 단측 검정	1.734063607	
P(T<=t) 양측 검정	0.092496201	
t 기각치 양측 검정	2.10092204	

LSTM

t-검정: 이분산 가정 두 집단		
	변수 1	변수 2
평균	52274963.13	17133828.33
분산	5.79905E+15	3.45494E+14
관측수	20	20
가설 평균차	0	
자유도	21	
t 통계량	2.004868129	
P(T<=t) 단측 검정	0.029019259	
t 기각치 단측 검정	1.720742903	
P(T<=t) 양측 검정	0.058038517	
t 기각치 양측 검정	2.079613845	

Neural Prophet

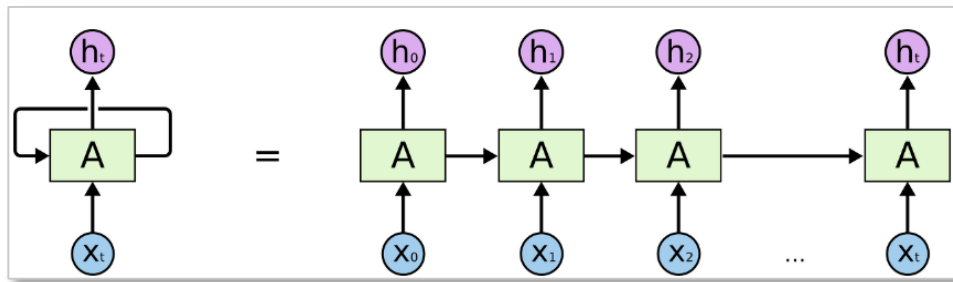
t-검정: 이분산 가정 두 집단		
	변수 1	변수 2
평균	1486447445	408587466.4
분산	4.09867E+18	4.83926E+17
관측수	20	20
가설 평균차	0	
자유도	23	
t 통계량	2.251756002	
P(T<=t) 단측 검정	0.017094347	
t 기각치 단측 검정	1.713871528	
P(T<=t) 양측 검정	0.034188694	
t 기각치 양측 검정	2.06865761	

유의수준 0.05 이하로, **FinBERT 추가 모델의 MSE가 낮음**이 통계적으로 유의

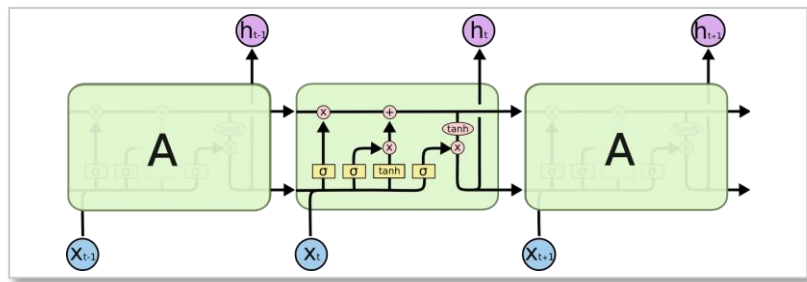
3 Model

LSTM

- ✓ 기존 RNN의 출력과 먼 위치에 있는 정보를 기억할 수 없다는 단점을 보완하여 장/단기 기억을 가능하게 설계한 신경망 구조
- ✓ 주로 시계열 처리나 자연어 처리에 사용



RNN



LSTM

3 Model

모델링 과정

Score 계산

true(내일 실제)	close(금일 실제)	pred(내일 예측값)	direction(실제 방향)	direc_pred(예측 방향)
90000	85000	80000	1 (90000 >= 85000)	0 (85000 >= 80000)

내일 실제 증가는 90000으로
올랐으니 direction = 1

모델은 80000으로 내릴 것으로
예측했으니 direc_pred = 0

MSE : 4234927.5, RMSE : 2057.89404296875, SCORE : 60.9%

NOT XOR 연산

direction(실제 방향)	direc_pred(예측 방향)
0	0
0	1
1	0
1	1

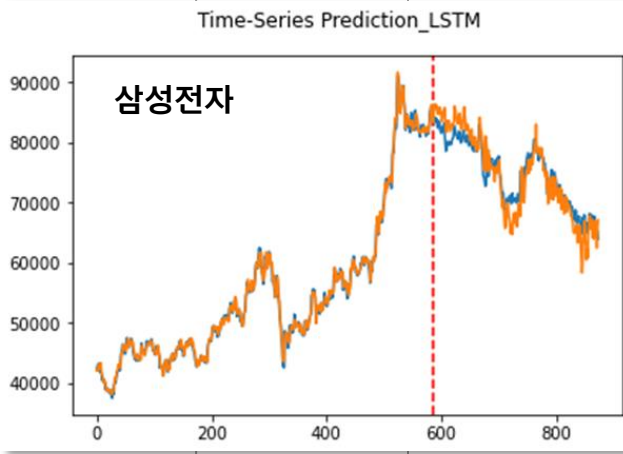
3 Model

모델링 과정

회차	column	model	optimizer	loss	epochs	learning_rate	결과(mse)	Up&Down 정확도
1	select	lstm	adam	mse	10000	0.01	81,408,112	54.6%
9	select	lstm	adamax	mse	10000	0.01	19,295,066	56.7%
12	select	lstm	adam	mse	5000	0.01	35,062,032	54.3%
20	select	lstm	adam	mse	10000	0.001	9,606,054	55.0%
25	select	lstm	adam	mse	20000	0.0005	5,515,896	57.7%
28	select	rnn	adam	mse	20000	0.001	74,880,632	51.7%
40	all	lstm	adam	mse	20000	0.0005	2,446,035	60.0%
43	all	lstm	adam	mse	20000	0.0009	4,844,823	57.6%

3 Model

모델링 과정

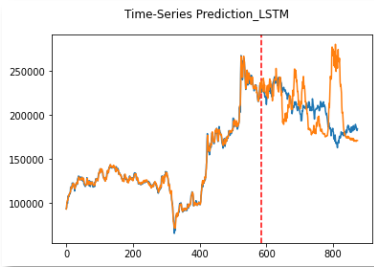
회차	column	model	optimizer	loss	epochs	learning_rate	결과(mse)	Up&Down 정확도
1	select	lstm	adam	mse	10000			
9	select	lstm	adamax	mse	10000			
12	select	lstm	adam	mse	5000			
20	select	lstm	adam	mse	10000			
25	select	lstm	adam	mse	20000			
28	select	rnn	adam	mse	20000			
40	all	lstm	adam	mse	20000	0.0005	2,446,035	60.0%
43	all	lstm	adam	mse	20000	0.0009	4,844,823	57.6%

3 Model

최종 모델

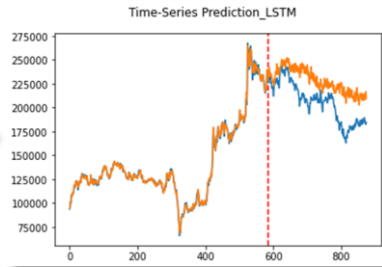
현대자동차

Base Model



MSE: 389,883,776
Up&Down 정확도: 52.2%

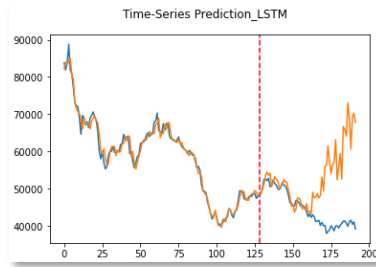
Base Model + FinBERT



MSE: 211,949,568
Up&Down 정확도: 61.6%

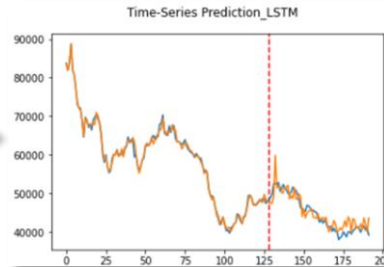
카카오뱅크

Base Model



MSE: 58,225,888
Up&Down 정확도: 56.2%

Base Model + FinBERT



MSE: 4,234,928
Up&Down 정확도: 60.9%

Base Model에 **감성지표 추가** 시 MSE 감소 및 등락 예측 정확도 향상

3 Model

최종 모델

현대자동차

Base Model



MSE: 389,883,776
Up&Down 정확도: 52.2%

Base Model + FinBERT



MSE: 211,949,568
Up&Down 정확도: 61.6%

카카오뱅크

Base Model



MSE: 58,225,888
Up&Down 정확도: 56.2%

Base Model + FinBERT



MSE: 4,234,928
Up&Down 정확도: 60.9%

소프트 코딩된 데이터 수집 및 LSTM 모델에
기업명만 변경하여 임의의 기업 주가 등락 예측 가능

Base Model에 **감성지표 추가** 시 MSE 감소 및 등락 예측 정확도 향상

4

기대효과 및 보완점



4 기대효과 및 보완점

기대효과

모델

- ✓ 기존 시장 예측 모델에 NLP를 추가함으로써 **예측력 상승**
- ✓ 증권사 리포트나 뉴스 기사 등 오피셜한 정보를 이용하여 **신뢰성 향상**

임직원

- ✓ 감성 분석 결과를 **시장 예측 모델에 활용**
- ✓ 정밀한 금융 포트폴리오 구성, PB 고객 상담에 활용, **고객 수익률 향상**
- ✓ 각 부서 의사결정 시 시황 확인용 자료로 사용

대고객

- ✓ 개인 자산관리 서비스 고도화
- ✓ 마이데이터 및 비대면 채널에 새로운 서비스를 제공함으로써 **MAU 향상**

4 기대효과 및 보완점

보완점

1. LSTM, RNN, Prophet, ARIMA 모델 외 타 모델 사용 및 앙상블 기법 도입
2. Cross Validation 등을 적용한 정확도 향상
3. 단일 기업에 대한 예측에서 산업 군에 대한 예측으로 범위 확장
4. 일 단위가 아닌 주 단위로 예측하여 추가적인 정보 제공
5. Fundamental approach를 통한 Long-term prediction 추가하여 종합적 추천
6. Transaction cost 고려

Q & A

Thank You
