

CHANA One-stock

1팀

AI주식자동매매

팀원

강은현 박경덕 이승희 임서영 홍윤기

목차



01 소개

팀원 및 서비스소개
시스템 아키텍처
개발일정

02 기능

ERD
흐름도
주요 기능, 모델

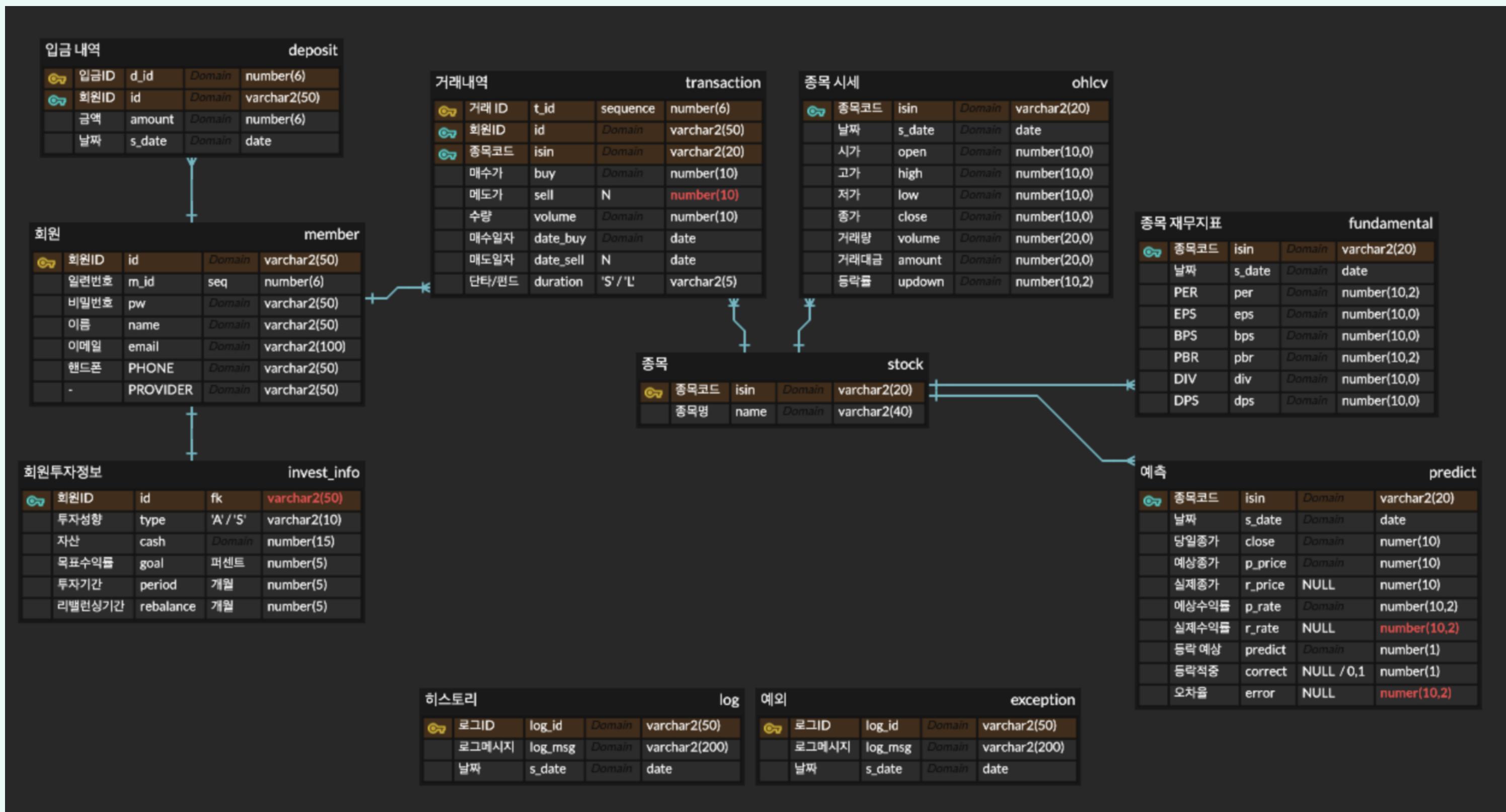
03 시연

웹
예측 모델

데이터 개선
기능 확장
UI/UX 개선

04 한계 및 고도화 방안

데이터 ERD



팀원을 소개합니다

팀원
강은현

백엔드

프로젝트 기반 구축, 소셜 로그인,
KOSPI 차트(라인차트) 구현

팀원
박경덕

인프라 설계 및 개발

RESTFUL API 개발, 데이터 트레이
이팅 기능, 검색 기능 개발, 클라우
드 서버 구축

팀원
이승희

PM & 주가 예측 모델 개발

주가 및 재무 데이터 분석, 추천 알
고리즘 연구 및 구현, 디시보드 관리

팀장
임서영

프론트엔드

웹 퍼블리싱, JS 알고리즘 개발,
email 인증, 대시보드(캔들차트),
마이페이지 개발

팀원
홍윤기

DBA & 주가 예측 모델 개발

머신러닝 모델링, 수익률 기능 개
발, DB 구축, 클라우드 서버 구축



서비스 소개

1. 코스피 953개의 종목의 10년치 데이터 전처리
2. 머신러닝 모델 개발 및 주가 예측
3. 거래량 + PER를 통해 필터링 과정
4. 최종적으로 4개의 종목 추천

개발 일정

전체 일정 계획			07/03 ~ 07/14 (총 12일)											
구분	상세 업무		07/03 ~ 07/19 (1주)						07/03 ~ 07/19 (2주)					
주요 마일스톤	분석	추가정보 예측 및 검증 모델 개발 JSON데이터 제공 API 개발, 코스피 데이터 제공 API개발	7.3	~	7.3	7.4	~	7.5	7.6	~	7.7	7.8	~	7.9
	설계	비즈니스 로직 설계 ERD 설계	7.4	~	7.5	7.5	~	7.6	7.6	~	7.7	7.7	~	7.8
	환경 구축	웹 구축, DB구축 클라우드 구축, 깃 서버 구축	7.5	~	7.6	7.6	~	7.7	7.7	~	7.8	7.8	~	7.9
주요 회의 일정	착수회의		7.10	~	7.10	7.11	~	7.11	7.11	~	7.12	7.12	~	7.13
	수시회의		7.11	~	7.11	7.12	~	7.12	7.12	~	7.13	7.13	~	7.14
분석/설계	요구사항 정의 및 분석		7.12	~	7.12	7.13	~	7.13	7.13	~	7.14	7.14	~	7.15
	기능명세서		7.13	~	7.13	7.14	~	7.14	7.14	~	7.15	7.15	~	7.16
	설계서		7.14	~	7.14	7.15	~	7.15	7.15	~	7.16	7.16	~	7.17
개발	작업명	회원 관리 - 사용자 - 등록, 조회, 수정, 삭제 회원 관리 - 사용자 - 등록, 조회, 수정, 삭제 메인페이지 - 사용자 - 추천 종목 기능 메인페이지 - 사용자 - 코스피 증시현황 메인페이지 - 사용자 - 특징 종목 리스트 정렬 마이페이지 - 사용자 - 개별 종목 수익률, 전체 포트폴리오 수익률 마이페이지 - 사용자 - 개인 정보 수정 데이터레이팅 - 사용자 - 매수/매도 기능 데이터레이팅 - 사용자 - 종목 검색 기능, 차트 조회	7.15	~	7.15	7.16	~	7.16	7.16	~	7.17	7.17	~	7.18
통합테스트	통합테스트 및 에러수정		7.16	~	7.16	7.17	~	7.17	7.17	~	7.18	7.18	~	7.19
시연	시나리오, 시연		7.17	~	7.17	7.18	~	7.18	7.18	~	7.19	7.19	~	7.20

시스템 아키텍쳐

Flask Rest API

PYKRX를 활용해 주가 정보 생성 및 JSON 데이터 제공

10년치 대용량 데이터 이용 예측 모델을 활용한 주가 예측 JSON 데이터 제공

주가 검증 JSON 데이터 제공

Spring

Spring Scheduler

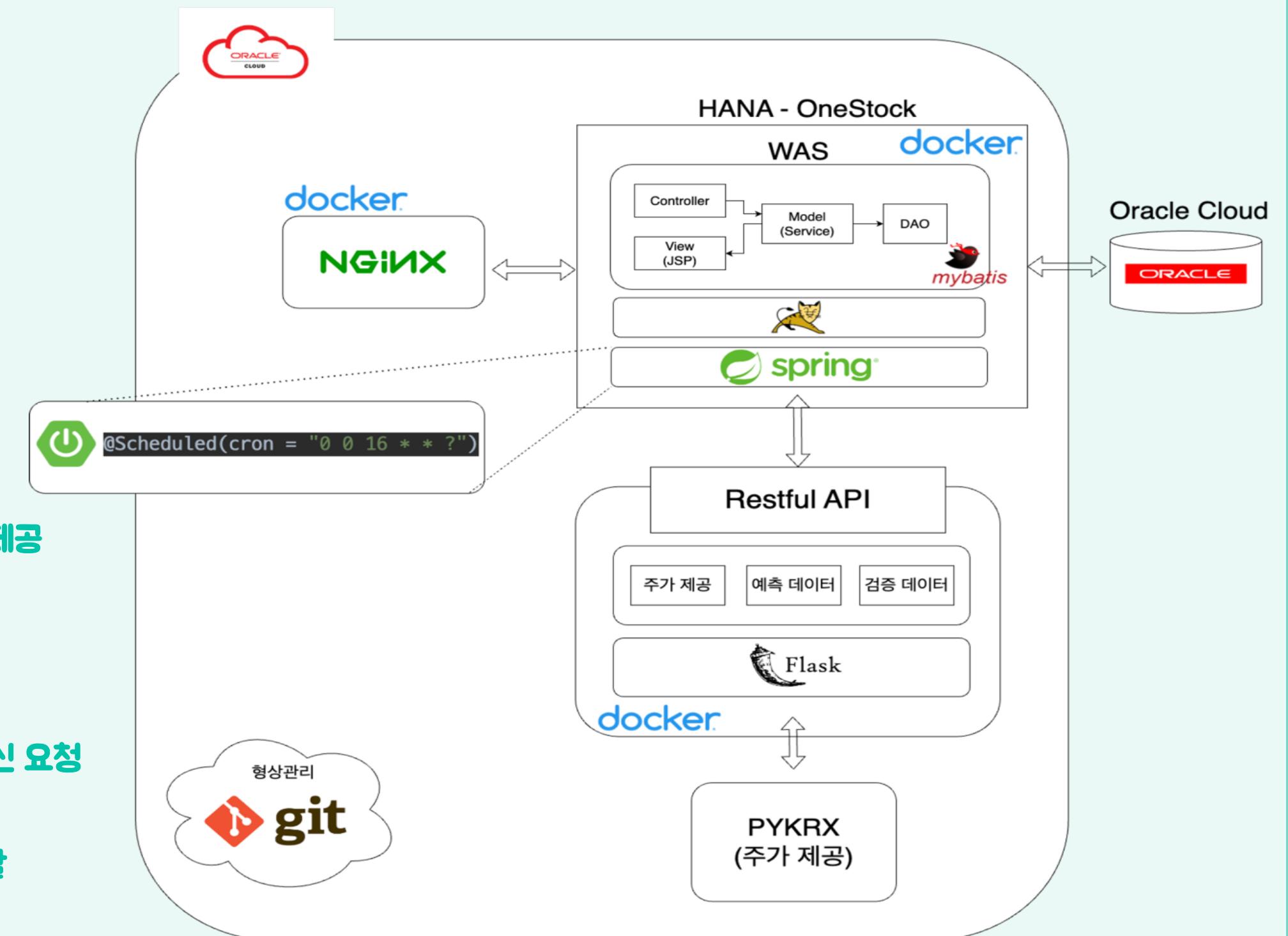
장마감 이후 Flask API에 주가 데이터 생성 요청, 예측 및 검증 데이터 생성 요청

JSON 데이터 파싱 및 DB 업데이트

MVC 패턴을 활용한 회원관리, 주식 매매, 주식 추천 등 비즈니스 로직 개발

형상관리

오라클 클라우드 내 구축된 git server 사용



사용 API

kakao developers



데이터 분석



표현 계층



비즈니스 계층



데이터 계층

ORACLE

서버 인프라



사용 언어



컨테이너 가상화



컴퓨팅 인프라



화면 UI/UX 설계



Oracle Cloud 내 구축된 Git Server

```
teamProject-HanaOneStock — opc@instance-webproject3:~/hana-one-stock.git —...
commit 75a5367c4c7b32b34920de55e81361b20f920fb (HEAD -> main, origin/main, origin/HEAD)
)
Author: SEUNGHEE96 <dnffktmdgml@gmail.com>
Date: Thu Jul 13 09:37:38 2023 +0900

Fix : sq1<EB><AC><BB> <EC><BB><9C><EA><BB><84> <EB><B3><BB><EA><B2><BD>

commit 7a70e019f234d9fc2a9843485f0bc48a3d889f8
Merge: a14ad9 32fd1ab
Author: SEUNGHEE LEE <108311729+SEUNGHEE96@users.noreply.github.com>
Date: Thu Jul 13 09:34:16 2023 +0900

Merge pull request #40 from EunhyeonKang/main

add spring

commit 32fd16b2b425dec17535323ccc4336b62b3bbe43
Merge: e7fba21 a14ad9
Author: EunhyeonKang <rkkddmsgus1@daum.net>
Date: Thu Jul 13 09:32:44 2023 +0900

Merge branch 'main' of https://github.com/SEUNGHEE96/HANA-OneStock

commit e7fba2112bf62a01d38b7056660dbdf6cb46b
Author: EunhyeonKang <rkkddmsgus1@daum.net>
Date: Thu Jul 13 09:28:10 2023 +0900

add spring

commit a14ad95250a0eccf75bef4cebc5cbe8a9faf4f7
Merge: 661e9bf 6eeef286
Author: SEUNGHEE LEE <108311729+SEUNGHEE96@users.noreply.github.com>
Date: Thu Jul 13 09:26:22 2023 +0900

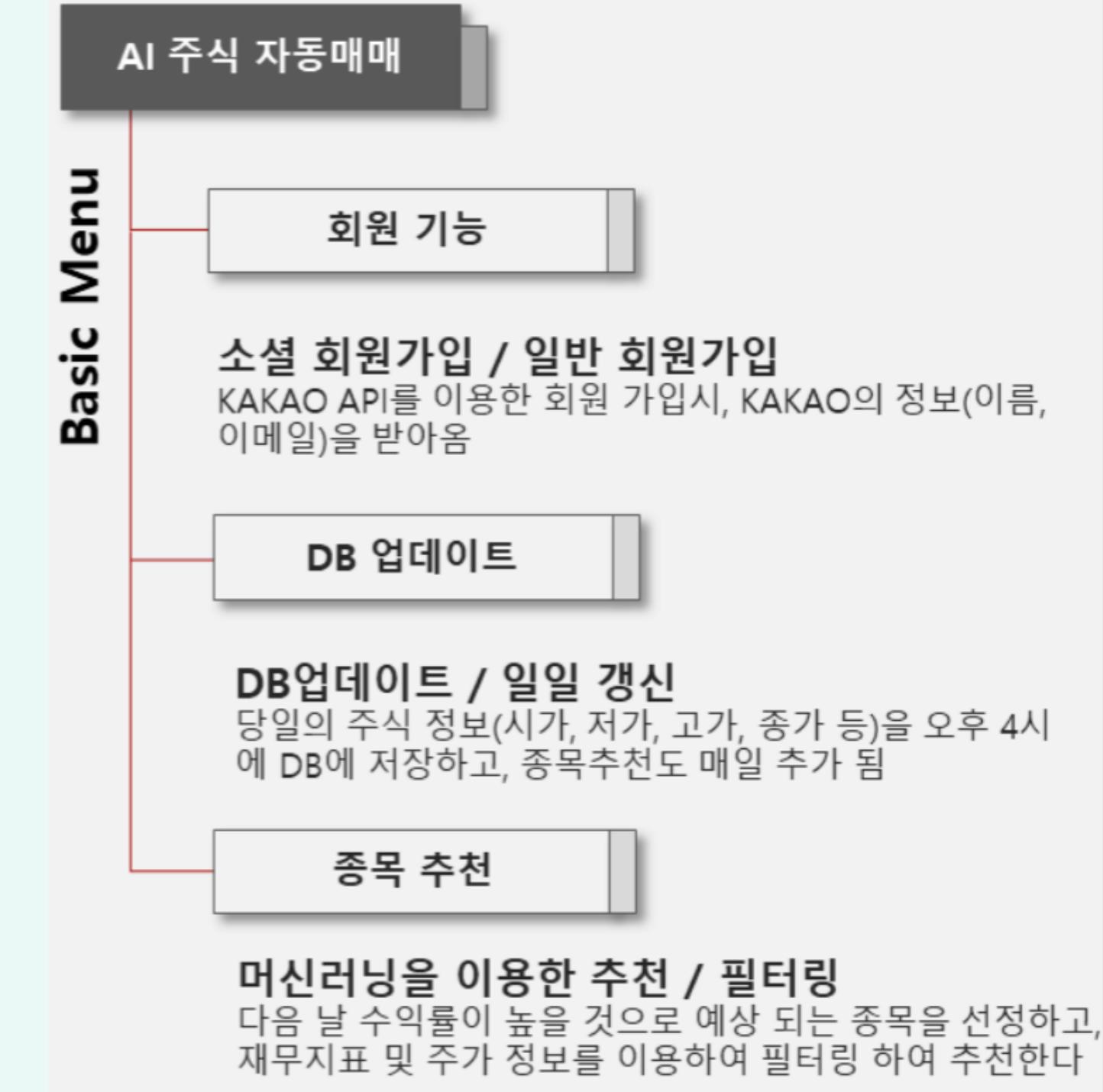
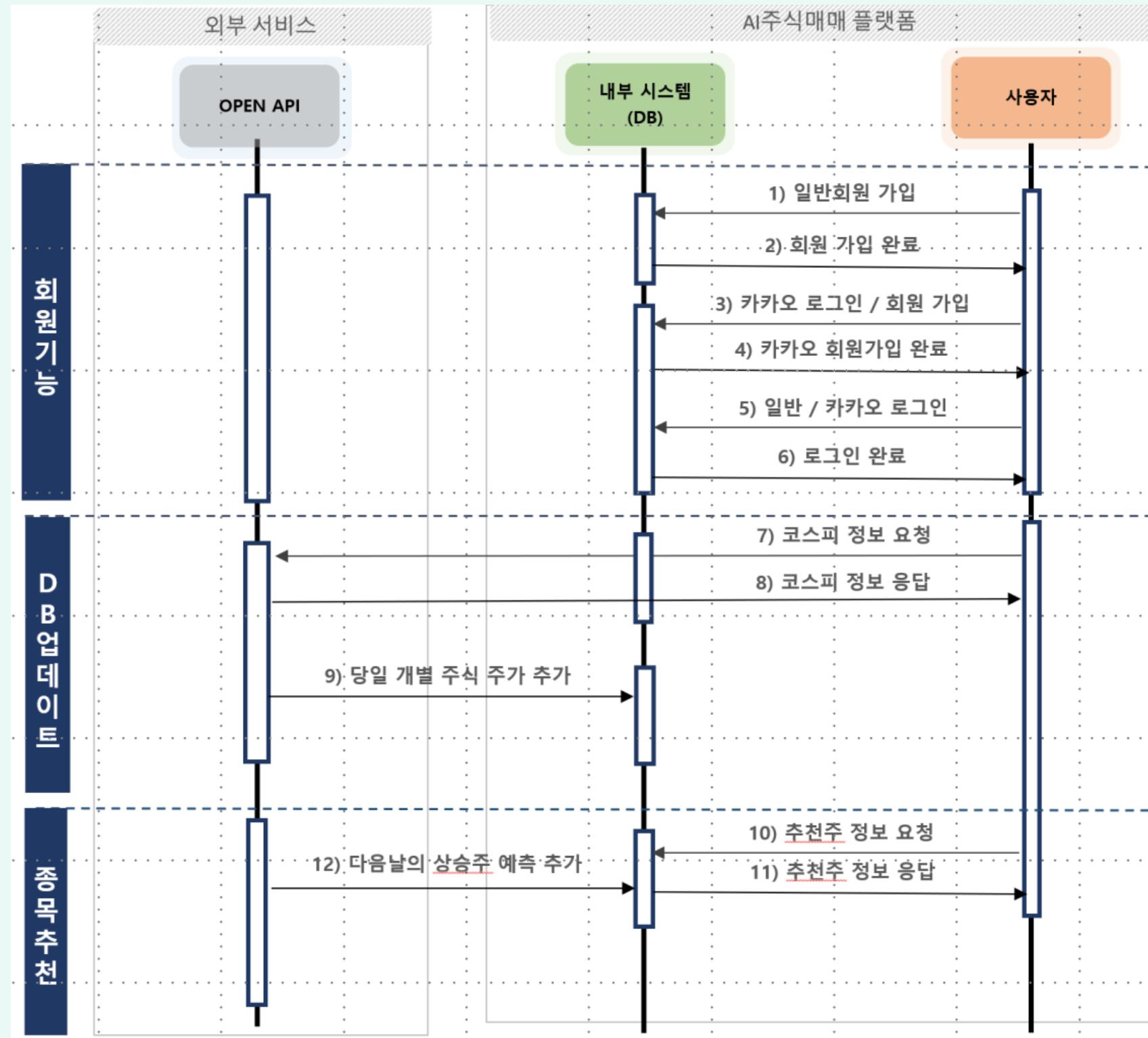
Merge pull request #39 from EunhyeonKang/main

add main page

commit 6eeef206c84e6eca03b9ef6a12f79c635595b7c1c
Author: EunhyeonKang <rkkddmsgus1@daum.net>
Date: Thu Jul 13 09:25:28 2023 +0900

add main page
:
```

AI자동매매 처리 흐름 구성



예측모델

- 953개 종목의 10년 데이터 => 선형 회귀 분석(Scikit - Learn)
- 머신러닝 모델 1 => 다음날의 종가 예측
- 머신러닝 모델 2 => 다음날의 주가 상승 확률 예측
- 두 머신러닝 모델을 종합 => 교차 검증 후 두 모델에서 공통적으로 상승한 종목

추천모델

- 머신러닝 모델을 통해 추천된 종목
- 필터링 => 거래량 하위 10% & PER 2미만의 종목
- 예상 수익률 순 => 상위 4개 종목 추천

REST API

주가 정보 수집

사용 언어 : Python3

사용 API : PYKRX

한국거래소(KRX)에서 주가 정보를 가져오는 API

장점 : 가장 많이 사용되는 야후파이낸스에 비해 한국 주식 정보가 정확한 편

단점 : 수정주가 제공 옵션이 있으나, 반영이 안됨

REST API

데이터 처리

사용 언어 : Python3

사용 API : Pandas

Series, DataFrame을 다루는 라이브러리

장점 : SQL에서 제공하는 대부분의 기능을 사용 가능

CSV, JSON, SAV 등의 파일로 변환 용이

AI 기초지식

Artificial Intelligence

인공지능

사고나 학습등 인간이 가진
지적 능력을 컴퓨터를 통해
구현하는 기술



Machine Learning

머신러닝

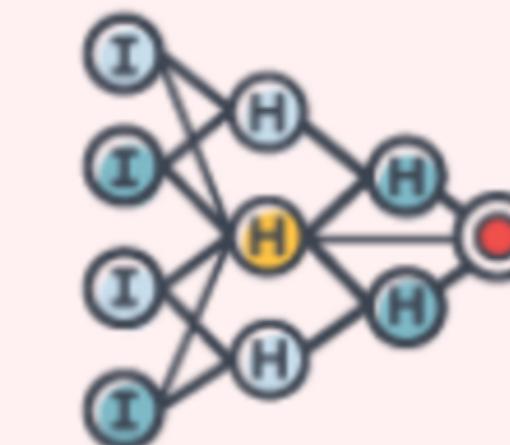
컴퓨터가 스스로 학습하여
인공지능의 성능을
향상 시키는 기술 방법



Deep Learning

딥러닝

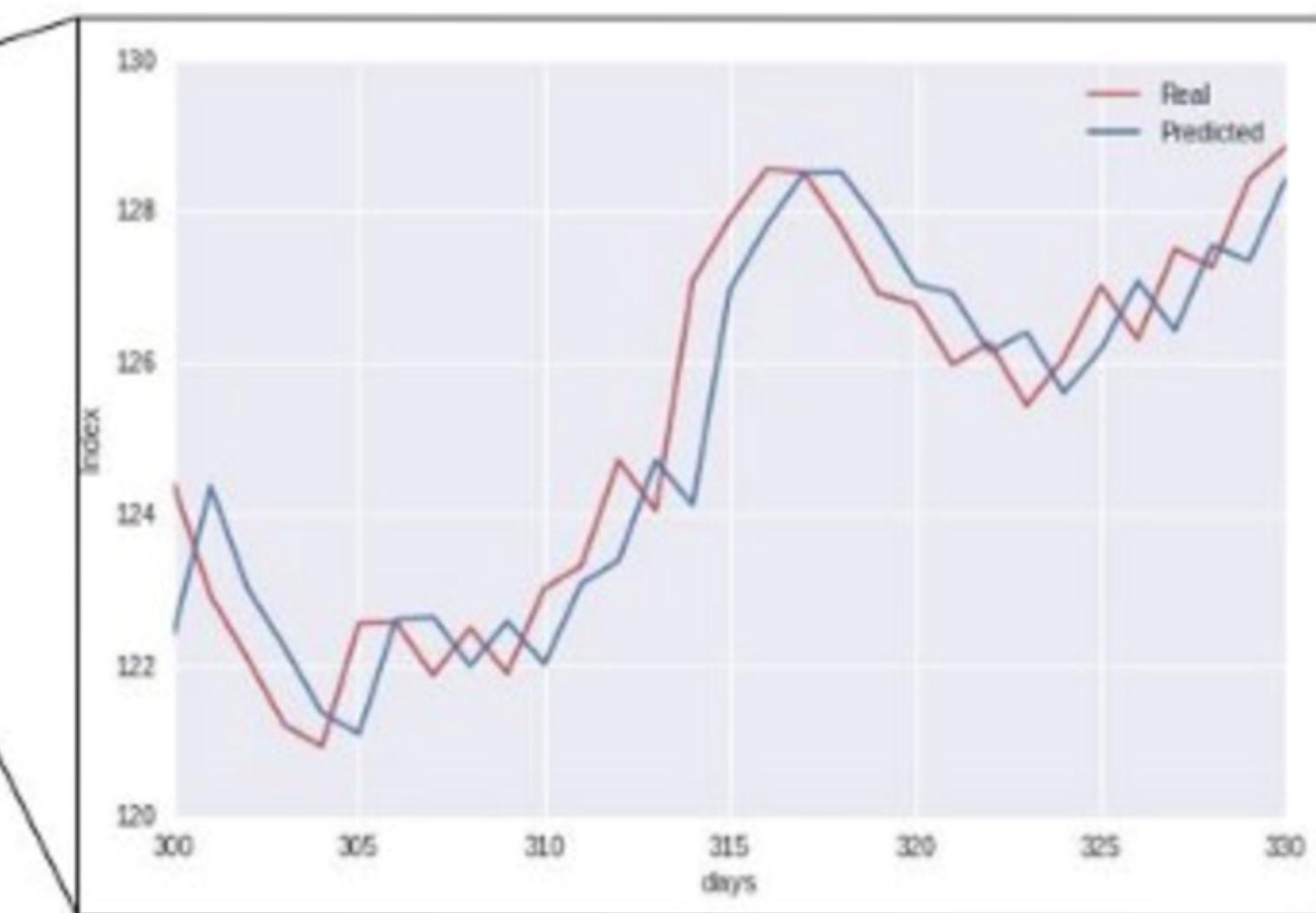
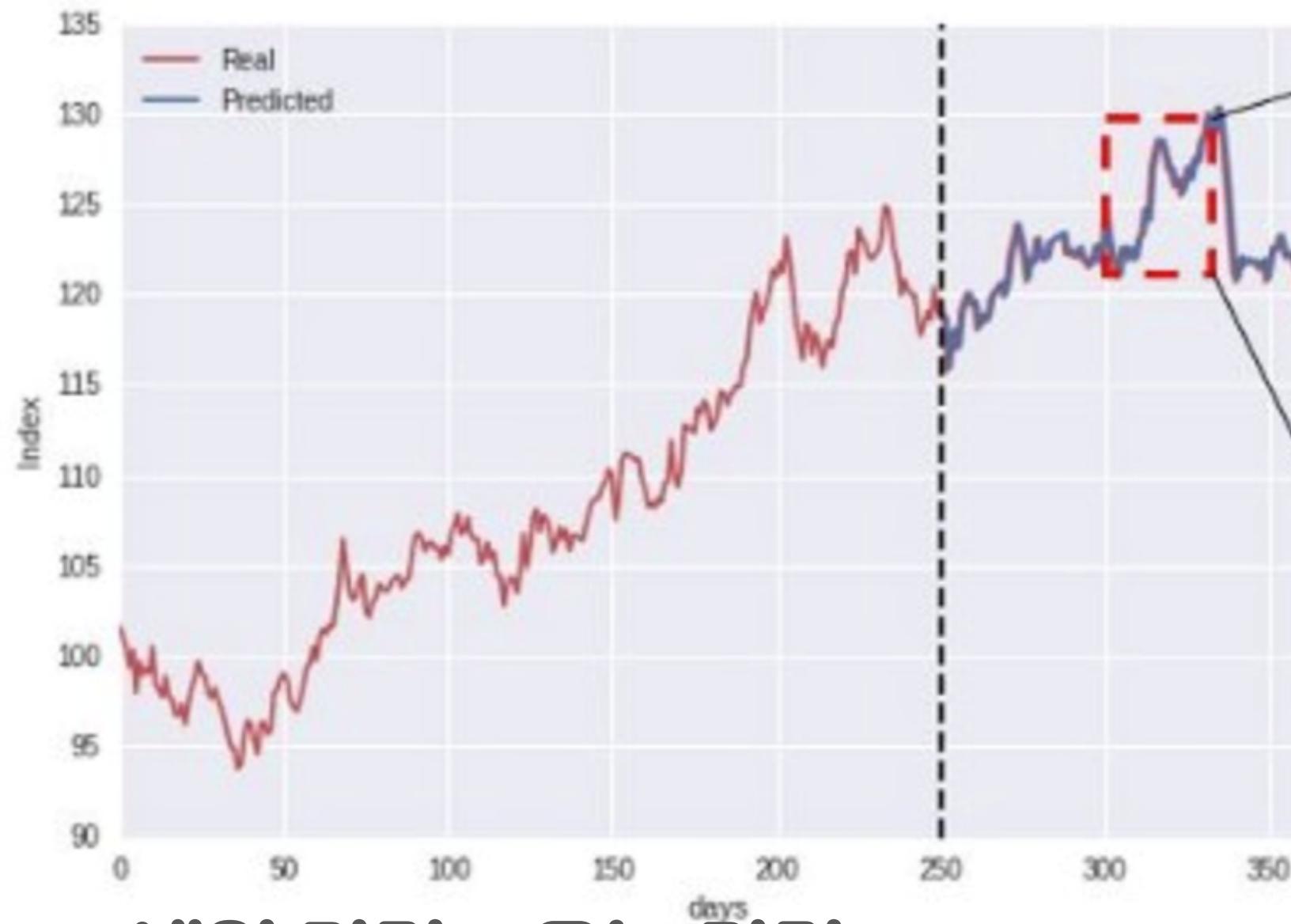
인간의 뉴런과 비슷한
인공신경망 방식으로
정보를 처리



채택하지 않은 모델

: 1. LSTM

1. LSTM은 연속된 sequence 데이터를 다룰 때 좋은 레이어
2. -> '주가'라는 숫자 패턴의 연속(수열)이므로 쓰면 좋지 않을까?



채택하지 않은 모델

: 1. LSTM

13/01/01 ~ 23/07/03 학습 -> 23/07/04 채점 -> 최상의 답안은?
다음날 재 학습을 한다고 나아질까?

a	# 23/07/07	0	2588	error
0	4460.000000	4305	3.60	
1	19340.000000	19080	1.36	
2	3850.000000	3660	5.19	
3	170200.000000	166800	2.04	
4	6980.000000	6670	4.65	
...
948	5731.889083	6020	4.79	
949	3135.000000	3060	2.45	
950	8050.000000	9960	19.18	
951	5590.000000	5410	3.33	
952	2295.000000	2130	7.75	

b	# 23/07/06	0	2587	error
0	4460.000000	4275	4.33	
1	19340.000000	19080	1.36	
2	3850.000000	3715	3.63	
3	170200.000000	166600	2.16	
4	6980.000000	6730	3.71	
...
948	5731.889083	6160	6.95	
949	3135.000000	3070	2.12	
950	8050.000000	8050	0.00	
951	5590.000000	5530	1.08	
952	2295.000000	2210	3.85	

c	# 23/07/05	0	2586	error
0	4460.000000	4390	1.59	
1	19340.000000	19320	0.10	
2	3850.000000	3800	1.32	
3	170200.000000	167800	1.43	
4	6980.000000	6860	1.75	
...
948	5731.889083	6340	9.59	
949	3135.000000	3080	1.79	
950	8050.000000	8050	0.00	
951	5590.000000	5670	1.41	
952	2295.000000	2185	5.03	

d	# 23/07/04	0	2585	error
0	4460.000000	4460	0.00	
1	19340.000000	19340	0.00	
2	3850.000000	3850	0.00	
3	170200.000000	170200	0.00	
4	6980.000000	6980	0.00	
...
948	5731.889083	6300	9.02	
949	3135.000000	3135	0.00	
950	8050.000000	8050	0.00	
951	5590.000000	5590	0.00	
952	2295.000000	2295	0.00	

953 rows × 3 columns

채택하지 않은 모델

: 2. 변동률 예측 모델

기호지세, 파죽지세 => 추세가 중요하지 않을까???
가격은 변동률이라는 변화 추세의 결과일 뿐!!

The screenshot shows a software window titled "[0162] 특이증목 - 상한가/하한가". At the top, there are several tabs and filters: 1. 고가/신저가 (highlighted in red), 2. 상한가/하한가 (highlighted in red), 3. 조회 (highlighted in red). Below these are more filter options: 전체, 코스피, 코스닥, 상한, 상승, 보합, 하락, 하한, 전일상한, 전일하한, 등락률순, 증목코드순, 연속횟수순(상위100개). The main table lists 10 stocks with columns: 종목명, 현재가, 전일대비, 등락률, 거래, 5. 전일거래량, 매도잔량, 매도호가, 매수호가, 매수잔량, 횟수. The stocks listed are: 제넨바이오, 필로시스헬스, 중앙오션, 시스웍, 캠온, 유나이티드제, 진원생명과학, 마이크로디지.

종목명	현재가	전일대비	등락률	거래	5. 전일거래량	매도잔량	매도호가	매수호가	매수잔량	횟수
제넨바이오	5,050 ↑ 1,165 +29.99	73,158,899	73,292,194	0	5,050	1,003,932	1			
필로시스헬스	4,905 ↑ 1,130 +29.93	8,554,128	47,775,484	0	4,905	2,483,338	5			
중앙오션	1,455 ↑ 335 +29.91	4,213,156	797,725	0	1,455	2,327,949	1			
시스웍	7,470 ↑ 1,720 +29.91	446,679	100,296	0	7,470	4,359,296	4			
캠온	3,325 ↑ 765 +29.88	66,264,930	30,137,317	0	3,325	831,462	1			
유나이티드제	51,800 ↑ 11,900 +29.82	7,064,988	6,714,312	0	51,800	98,922	1			
진원생명과학	18,300 ↑ 4,200 +29.79	27,214,066	60,526,984	0	18,300	882,935	1			
마이크로디지	19,000 ↑ 4,350 +29.69	3,813,981	70,585	0	19,000	90,110	1			

상한가 하한가란?

하루에 변할 수 있는 주가엔 한계가 있음 => 30%

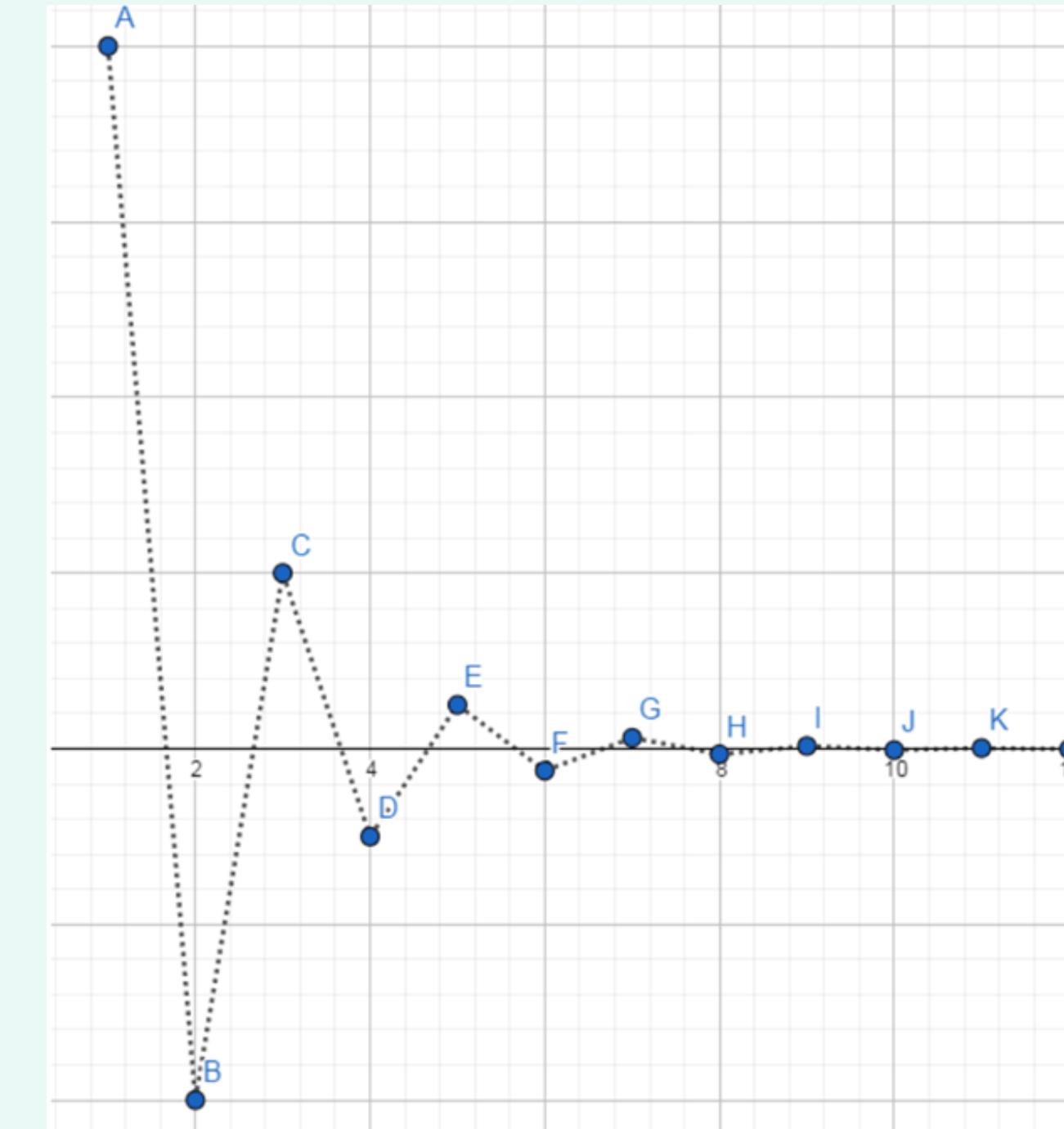
주가 변동률의 범위 : -30 % ~ 30%

즉 $-0.3 < P < 0.3$ 범위의 예측 결과가 나온다.

채택하지 않은 모델

: 2. 변동률 예측 모델

	예측	실제
0	0.040523	-0.19
1	0.018659	-0.68
2	0.005334	-1.94
3	-0.032473	-1.54
4	-0.023461	-0.17
...
2854	0.034698	-3.28
2855	-0.074102	2.54
2856	0.012022	0.75
2857	0.052563	-3.62
2858	-0.078216	-1.88



(진동발산)

0에 수렴하는 형태가 되어 유의미한 변동률을 도출하지 못함

채택하지 않은 모델

: 3. 그 외 모델

**과하게 많은 데이터 수, feature의 수
=> 과적합, 과도한 Training 시간, 인프라의 열악**

차트 온오프라인



프로젝트 타당성

차트 매매법 => 종목 시세 정보를 시각화한 차트의 추세를 보고 투자

AI 매매법 => 종목 시세 정보의 수치의 패턴을 분석하여 투자

과연 옳은 방법이라 할 수 있나?

프로젝트 타당성

Conclusion

코스피 지수가 확률 보행 과정의 실현값이라면 차트만으로는 예측이 어려울 수도 있겠네요. 물론, 다시 말하지만 현실에서 예측은 차트만으로는 하지 않습니다. 예를 들어 개요에 말했던 것처럼 한국이 지속적으로 발전할 거라고 생각한다면 우량주에 분산해서 넣어 놓고 주식 어플리케이션을 삭제하는 사람도 있겠죠? 이런 식으로, 주식 투자를 할 거라면 차트만 보기보다는 여러 정보를 종합해서 투자를 하는 편이 좋을 수도 있을 듯 보입니다.

결론

우선 결론을 내기 전에 알아야 할 점은 주식은 복합적인 요인들이 동시에 영향을 끼치기 때문에 단순히 특정 요소에 의해서 변화하지 않는 것을 고려해야 한다는 것입니다.

마치며

- 분류의 경우는 단순히 투자케이스 성공에 초점을 두어 이익이 보장되는 예측은 아니라고 생각된다.
- 옵션요소가 아직도 무수히 있기 때문에 (기간, 추가 할 피쳐, method ...) 좋은 모델은 나올거라 생각한다.
- 회귀의 경우는 RMSE가 적당히 있어야 손실 혹은 이익이 나타난다.
- ElasticNet으로 Regression Model을 정할 수 있다.
- 사이클을 profit의 평균지점 혹은 수익율로 모델을 평가하여 찾는 수준으로 올리면 최적을 찾을 것으로 기대된다.
- 투자가설과 규칙을 잘 정하여 투자를 해야하고 시뮬레이션 투자케이스 검증을 꼭 해봐야 하겠습니다!
- 성투하시길 바랍니다 ^~^

1. 주식가격을 sequence 데이터로 LSTM에 집어 넣는게 맞는지 고민해봅시다.

Q. 문장의 다음 단어를 예측하는 모델은 어떻게 만들죠?

1. 명사

The **police** are the official organization that is responsible for making sure that people obey the _____?

▲ 당연히 앞의 단어들을 쪽 sequence 데이터로 만들어서 LSTM에 집어 넣으면 됩니다.

LSTM은 '자료간의 순서와 연관성이 있을 것이다'라는 가정을 하고 사용하는 뉴럴네트워크 레이어니까요.

그러니 영어문장 같은 경우 앞의 단어들을 분석하면 뒤에 올 단어도 쉽게 추론할 수 있을 것입니다.

버핏의 명언



"주가 변동을 적으로 보지 말고 친구로 보라.
어리석음에 동참하지 말고 오히려 그것을 이용해서 이익을 내라."

"가격은 우리가 내는 돈이며, 가치는 그것을 통해 얻는 것이다."

"훌륭한 가격에 적당한 회사를 사는 것보다,
적당한 가격에 훌륭한 회사를 사는 것이 훨씬 낫다."

"10년 이상을 볼 것이 아니면 10분도 그 주식을 갖고 있지 말아라."

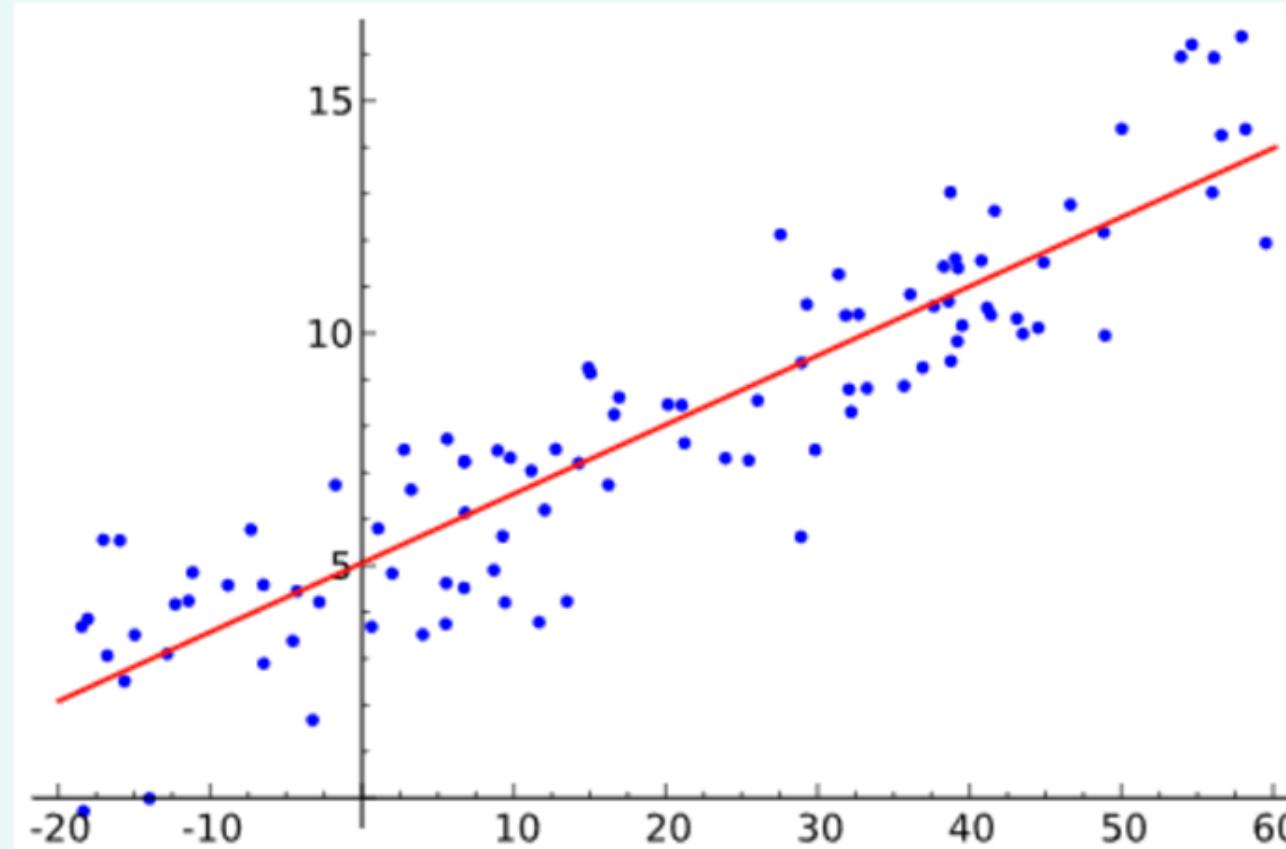
"어떤 자산에 투자하는 이유가 최근 가격이 상승해서여서는 안된다."

conclusion

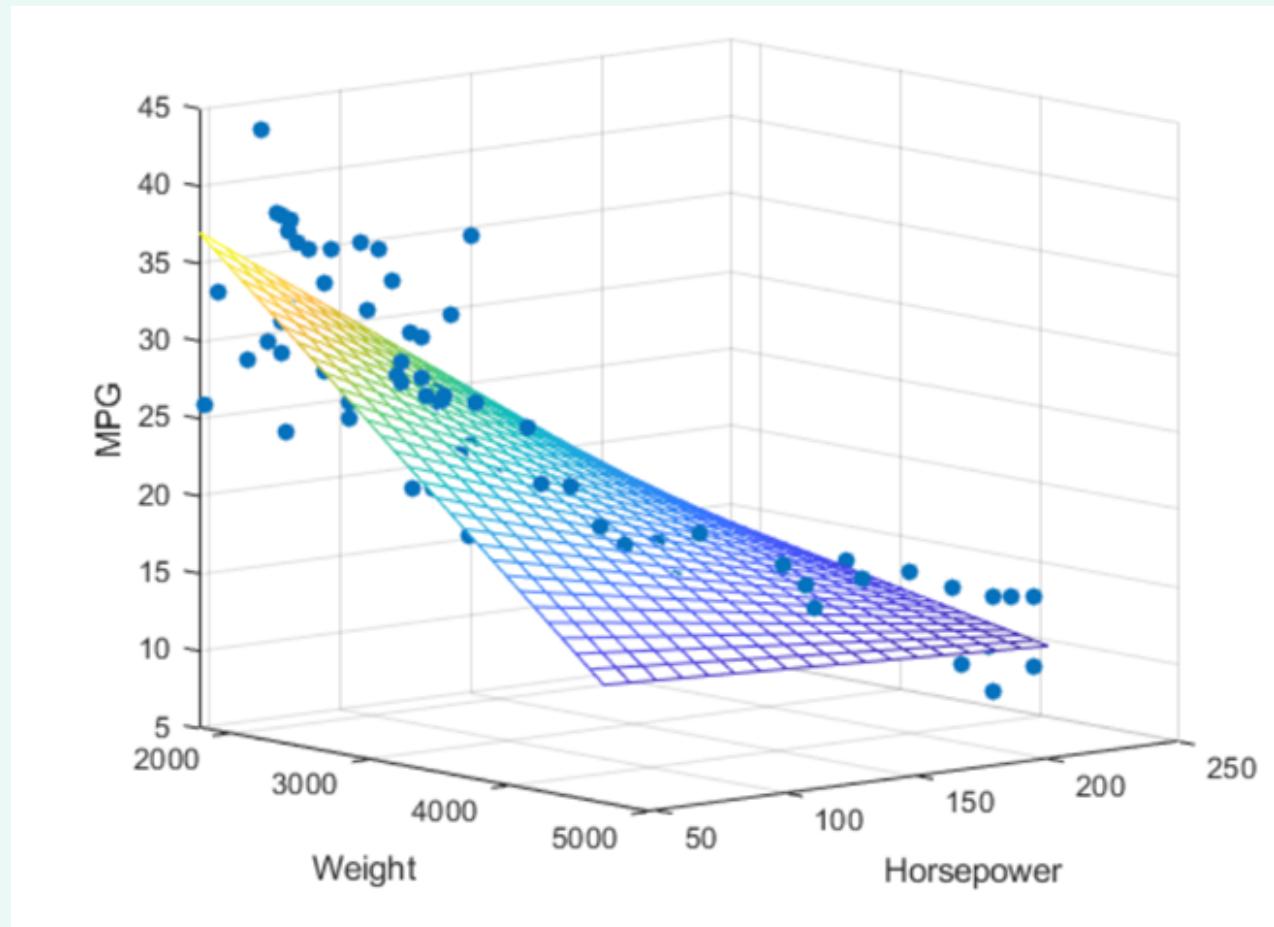


**"예측의 정확도를 확보 할 수 없다면
타당성과 경량성에 집중하자"**

선형 회귀 분석



단순 선형 회귀 분석



다중 선형 회귀 분석

장점 : 가장 직관적인 분석 기법

-> 단순하고 가벼워 서비스가 용이하다.

단점 : feature(변수)가 적으면 과소적합, 너무 많으면 과대적합의 문제가 발생

-> 단순해서 변수의 영향을 많이 받는다.

filtering

추천 주식 필터링

“거래량 + PER”

당일 매수매도를 전제로 하는
데이트레이딩에서 단기 특성에 맞추어
필터링할 조건을 선정하였음

추천 주식 필터링

: 1. 거래량



삼부토건 :

2022년 도급순위 기준 70위 업체로 75년 간의 풍부한 토목시공 경험을 토대로 항만, 댐, 도로, 지하철, 발전소 등에서 기술적 노하우와 시공실적을 보유함.

최근 윤대통령 '우크라이나 다각적 지원' 계획을 밝히자, 급등한 주식

- 데이 트레이딩의 경우 '거래량이 뒷받침 되어주어야 당일 매수 매도할 수 있는 기반이 됨'
- 또한 중장기적인 주식 거래량 분석 결과는 특정 '주식 종목 투자자들의 관심 추세'를 알려줌
- > 하위 10% 거래량 기준을 구하고, 그 기준보다 거래량이 적을 경우 추천주식에서 제외함

PER Price Earning Ration

“주가 수익 비율”

$\text{PER} = \text{1주당 가격} / \text{주당 순이익(EPS)}$

$\text{EPS} = \text{당기순이익} / \text{발행주식수}$

- ‘PER가 높으면 기업이 고평가 되어있고, PER가 낮으면 기업이 저평가 되어있다’고 볼 수 있음
- 또 다른 말로 ‘순이익으로 기업의 총 가치에 도달하는 데 걸리는 시간’으로도 볼 수 있음

추천 주식 필터링

: 2. PER

PER EPS(2023.03) ⓘ	5.64배 67,769원
추정PER EPS ⓘ	N/A N/A
PBR BPS (2023.03) ⓘ	0.72배 531,263원
배당수익률 2022.12 ⓘ	2.36%
동일업종 PER ↗	13.81배
동일업종 등락률 ↗	+1.21%

PER EPS(2023.03) ⓘ	182.87배 286원
추정PER EPS ⓘ	53.81배 972원
PBR BPS (2023.03) ⓘ	2.14배 24,405원
배당수익률 2022.12 ⓘ	0.11%
동일업종 PER ↗	77.00배
동일업종 등락률 ↗	+0.86%

오뚜기 :

필수 소비재

섹터 기준이나 비슷한 규모의 회사끼리 비교하는 것이 유의미하지만,
보통 10배 이하의 회사들은 저평가 되어 있다고 판단함.

-> 매우 보수적으로 PER가 2 미만인 회사들은 추천 주식에서 제외하였음

PER가 2 미만이라는 것은 회사의 순이익이 기업의 가치와 거의 동일하다는 것

카카오 :

기술주

FINAL

개발 고도화 방안

- 현재 종가는 수정주가가 아님
- 수정주가로 변경하면 모델의 오차율이 줄어들 것으로 기대됨



- 자산(시드머니) 충전
- 구현이 완료되지 않은 부분
추가 구현 예정

- 수정이 필요한 부분은 추가적으로 수정 예정

**THANK
YOU**