

# HANA ONE-AI

주식 자동 매매 서비스

AI 주식 자동 매매 2팀

강태근 김남훈 박준하 심민정 이정민

**1. 팀 구성 및 프로젝트 소개**

**2. 시스템 구성 및 일정**

**3. 사용자 흐름 및 기능 설명**

**4. 분석 예측 모델**

# 1-1. 팀 구성

팀 원	역 할	담당 내용
강태근	팀장 / 백엔드, 인프라	소셜 로그인, 자동 매매 로직 구현, 오라클 클라우드 서버 구축
이정민	DB 구축, 예측알고리즘 구축	주식 데이터 및 공시 데이터 크롤링, 머신러닝 모델 학습
심민정	DB 모델링, 예측알고리즘 구축	DB 모델링, 딥러닝 모델 학습
김남훈	프론트, UI	매매전략 구성 및 프론트 구현
박준하	백엔드	자동매매 로직 구현, 포트폴리오 구성

# 1-1. 프로젝트 소개 : AI 자동 주식 매매

## AI 기반

## 주식추천

- ✓ 금융시장이 급변함에 따라, **다양한 요소가 주식 가치에 영향을 미침**
  - 데이터와 통계적 분석에 기반한 주식 가치를 예측하는 '과학적 방법론' 필요
- ✓ 기술적 분석과 학습을 통해 고객이 **금융 시장 '노이즈'에 휩쓸림 방지**
  - 고객이 마음 편한 투자 환경을 제공

## 자동매매

- ✓ 투자자의 전략과 실행에 있어 **정확성과 일관성을 제공**
  - 휴먼 에러(실수, 감정적 편향) 방지
- ✓ 매매 전략의 **백테스팅 및 최적화 가능**
  - 효율적인 투자 결정 도움

## 시스템

## 구현

- 1. 머신 러닝, 딥러닝 모델을 이용한 주식 종가 예측**
  - KOSPI 953개의 종목, 10년치 데이터를 이용·학습
- 2. 고객의 수익률을 만족할 포트폴리오 '자동' 구성**
  - 분산투자 이론을 활용하여 위험 최소화
- 3. 추천 시스템에 기반한 자동 매매 시스템**
  - 고객의 편의성 제공

## 2-1. 시스템 운영 방식

### 분석 예측

- ✓ 머신러닝, 딥러닝 기법을 이용한 KOSPI 10년치 주가 분석
- ✓ 학습한 모델을 활용하여 종가 예측

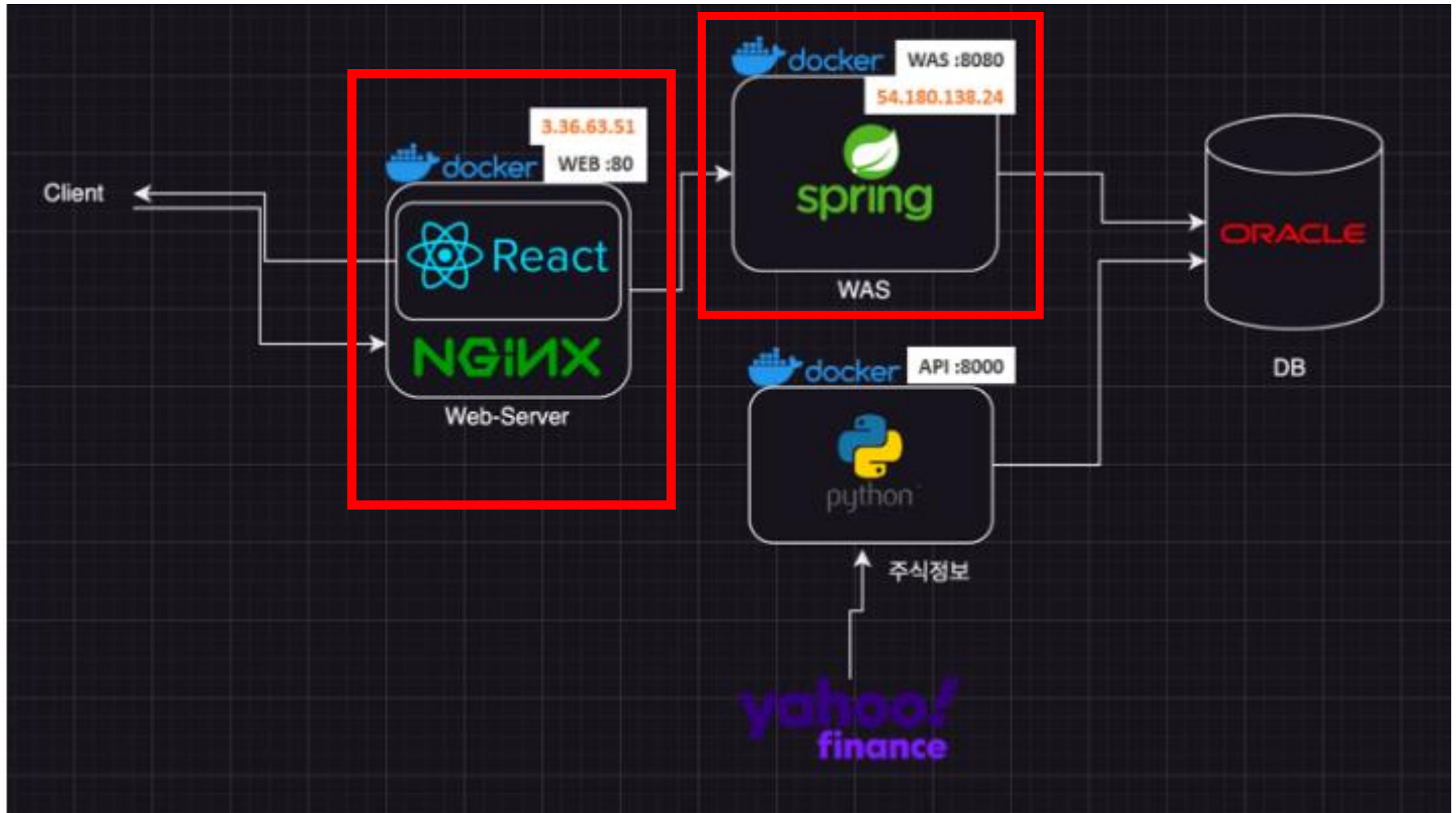
### 매매

- ✓ 전제 : 하루 24시간 기준(00시 ~ 24시)
- ✓ 당일 기준, 매수는 '전일 종가', 매도는 '당일 종가'로 실행

### 백테스팅

- ✓ 종목별 백테스팅을 통해 데이터 이상치 제거
- ✓ 백테스팅 결과를 화면에 공시

## 2-2. 시스템 설계 (아키텍처)



# 2-3. 시스템 설계 (ERD)

주식종목 KOSPI					
종목아이디	ID	VARCHAR2(30)	NOT NULL	종목코드 PK	
종목이름	NAME	VARCHAR2(255)	NOT NULL	종목이름	
주식발행수	STOCK_NO	NUMBER	NULL	주식발행수	
생성일자	CREATED_DATE	TIMESTAMP	NULL		

포트폴리오 PORTFOLIO					
아이디	ID	NUMBER	NOT NULL	SEQUENCE	
유저아이디	MEMBER_ID	NUMBER	NOT NULL	SEQUENCE	
보유수량	AMOUNT	NUMBER	NULL	0	
평균단가	AVG_PRICE	NUMBER	NULL	0	
종목아이디	KOSPI_ID	VARCHAR2(255)	NOT NULL		
M.주지표	TRADING	NUMBER	NULL	CHECK(0, 1, 2)	
생성일자	CREATED_DATE	TIMESTAMP	NULL		
손실률	LOSS	FLOAT	NULL		
목표수익률	TARGET	FLOAT	NULL		

딥러닝예측데이터베이스 DEEPLARNING_PREDICT					
아이디	ID	NUMBER	NOT NULL	SEQUENCE	
종목아이디	CODE	VARCHAR2(30)	NULL		
생성일자	CREATED_DATE	DATE	NULL		
종목이름	NAME	VARCHAR2(100)	NULL		
전일종가	PREV_CLOSE	NUMBER	NULL		
당일종가	CLOSE	NUMBER	NULL		
예측종가	PRED_CLOSE	NUMBER	NULL		
예측수익률	PRED_RATE	NUMBER	NULL		
오차율	DIFF	NUMBER	NULL		

종목정보 STOCK					
종목정보아이디	HISTORY_ID	NUMBER	NOT NULL	SEQUENCE	
종목아이디	KOSPI_ID	VARCHAR2(30)	NOT NULL		종목코드 PK
매매거래일	TRADING_DAY	DATE	NOT NULL	SYSDATE	
시작가	OPEN	NUMBER	NULL		
최고가	HIGH	NUMBER	NULL		
최저가	LOW	NUMBER	NULL		
종가	CLOSE	NUMBER	NULL		
거래량	VOLUME	NUMBER	NULL		
시가총액	CAP	NUMBER	NULL		발행수 * 종가
동적률	FLUCTUATION	NUMBER	NULL		
생성일자	CREATED_DATE	TIMESTAMP	NULL		

회원 MEMBER					
종목아이디	ID	NUMBER	NOT NULL	SEQUENCE	
이메일	EMAIL	VARCHAR2(255)	NOT NULL	UNIQUE	email 인증 고려
비밀번호	PASSWORD	VARCHAR2(255)	NOT NULL		
잔고	BALANCE	FLOAT	NULL	1000000	계좌 대입을 생성하지 않으면 잔고 확인 필요
로그인타입	LOGIN_TYPE	NUMBER	NULL		
생성일자	CREATED_TYPE	TIMESTAMP	NULL		

머신러닝예측데이터베이스 MACHINE_PREDICT					
아이디	ID	NUMBER	NOT NULL	SEQUENCE	
종목아이디	CODE	VARCHAR2(30)	NULL		
생성일자	CREATED_DATE	DATE	NULL		
종목이름	NAME	VARCHAR2(100)	NULL		
전일종가	PREV_CLOSE	NUMBER	NULL		
당일종가	CLOSE	NUMBER	NULL		
예측종가	PRED_CLOSE	NUMBER	NULL		
예측수익률	PRED_RATE	NUMBER	NULL		
오차율	DIFF	NUMBER	NULL		

회원정보 USER_INFO					
아이디	ID	NUMBER	NOT NULL	SEQUENCE	
유저아이디	MEMBER_ID	NUMBER	NOT NULL	SEQUENCE	
유저이름	NICKNAME	VARCHAR2(20)	NULL		
생성일자	CREATED_DATE	TIMESTAMP	NULL		

거래내역 TRANSACTION					
거래아이디	TRANSACTION_ID	NUMBER	NOT NULL	SEQUENCE	추정가능성
유저아이디	MEMBER_ID	VARCHAR2(20)	NULL	SEQUENCE	
종목아이디	KOSPI_ID	VARCHAR2(30)	NULL		종목코드 PK
거래수량	AMOUNT	NUMBER	NULL		
거래일자	TRADE_DATE	DATE	NULL	SYSDATE	
거래단가	PRICE_PER_STOCK	FLOAT	NULL		
거래금액	TOTAL_PRICE	NUMBER	NULL		
생성일자	CREATED_DATE	TIMESTAMP	NULL		

## 2-4. 기술 요소

개발 언어

FrontEnd



BackEnd



DataBase

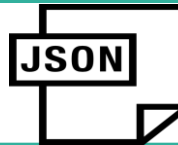


개발 S/W

인증



데이터 처리



분석·예측



UI 설계



개발 인프라

개발 인프라



서비스 인프라



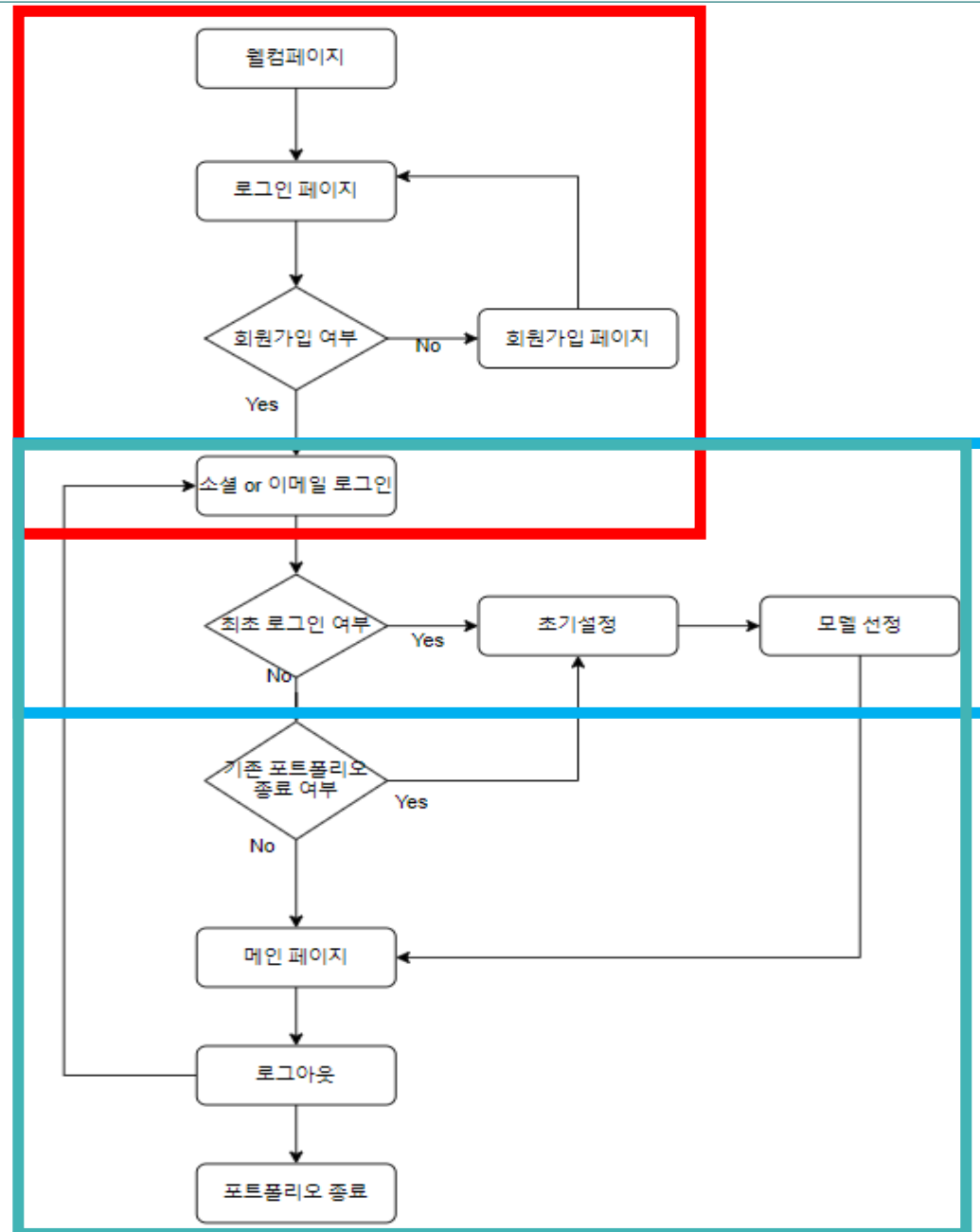


## 전체 일정 계획

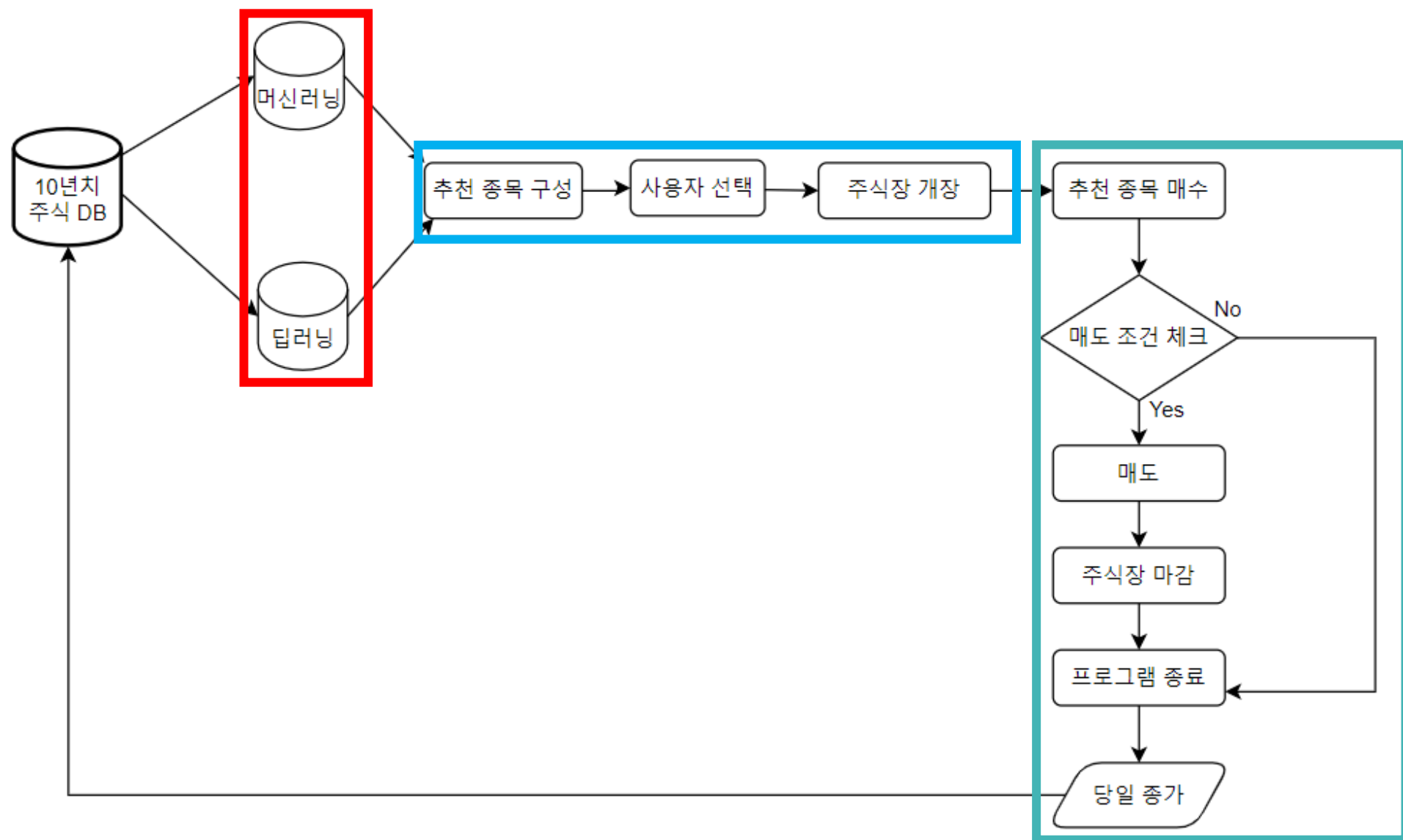
구분	상세 업무				7/3	7/4	7/5	7/6	7/7	7/8	7/9	7/10	7/11	7/12	7/13	7/14	
주요 마일스톤	분석/설계																
	개발/테스트/시연																
분석/설계	요구사항 정의 및 분석																
	기능명세서																
	설계서																
	화면설계서																
	DB설계서																
	테이블정의서																
개발	프론트																
	메인페이지																
	회원가입																
	로그인																
	대시보드 - 차트																
	백엔드																
	일반/소셜 로그인 구현																
	초기구성																
	수익률 계산 서비스																
	개별 종목 차트																
	포트폴리오 종목 리스트																
	세부 종목																
	리밸런싱																
	분석, 예측																
	알고리즘 수정																
	주식 데이터 갱신																
	종목 추천																
	인프라																
클라우드 인스턴스 생성																	
WEB, WAS 서버 구축																	
Git 서버 구축																	

# 3. 기능 설명

# 3-1. 전체 흐름도



# 3-1. 매매 흐름도



# 3-2. 메인 페이지

기업뱅킹 하나증권 은행소개 채용안내



종목

머신러닝 모델

딥러닝 모델

자동매매

The main banner features a blue gradient background. On the left, there is a cluster of logos for Samsung, Hana Securities, CJ, LG, Kakao Pay, and Hyundai AutoEver. To the right of these logos, the text '자동화를 통한 효율적인 투자 바로 시작하세요' (Start efficient investment through automation right now) is displayed. On the far right, a large green-bordered box contains the text 'AI자동주식매매' (AI Automatic Stock Trading).

KOSPI 2,628.30 ▲37.07

KOSDAQ 896.28 ▲3.21

국고채(3년) 3.6140 ▲0.0120

최다거래

급등주

급락주

# 3-2. 종목 검색

기업뱅킹   하나증권   은행소개   채용안내



종목   머신러닝 모델   딥러닝 모델   자동매매

카카오



이 페이지에서 검색할 내용을 입력하세요



종목명	종목코드	거래일자	종가	시가	고가	저가	거래량	시가총액
카카오	035720	2023. 7. 18.	51,700	52,700	52,700	51,400	1,455,102	22,968,412,093,000
카카오	035720	2023. 7. 17.	52,700	52,400	53,200	52,200	1,406,378	23,412,675,383,000
카카오	035720	2023. 7. 14.	52,800	52,300	52,800	51,600	1,900,674	23,457,101,712,000
카카오	035720	2023. 7. 13.	51,900	52,000	52,600	51,700	1,955,527	23,057,264,751,000
카카오	035720	2023. 7. 12.	51,600	50,700	51,800	50,200	1,658,795	22,923,985,764,000
카카오	035720	2023. 7. 11.	50,700	50,100	51,000	50,000	1,212,718	22,524,148,803,000
카카오	035720	2023. 7. 10.	49,850	49,850	50,300	49,450	1,068,139	22,146,525,006,500
카카오	035720	2023. 7. 7.	49,850	50,600	51,200	49,750	1,974,639	22,146,525,006,500
카카오	035720	2023. 7. 6.	50,900	48,850	51,500	48,700	3,139,685	22,613,001,461,000
카카오	035720	2023. 7. 5.	48,850	49,950	50,400	48,800	2,700,066	21,702,261,716,500
카카오	035720	2023. 7. 4.	49,950	50,900	51,000	49,900	1,445,459	22,190,951,335,500
카카오	035720	2023. 7. 3.	51,000	49,300	51,000	49,300	1,691,513	22,657,427,790,000
카카오	035720	2023. 6. 30.	49,100	48,850	49,700	48,400	1,268,925	21,813,327,539,000

# 3-2. 백테스팅 결과 조회

기업뱅킹 하나증권 은행소개 채용안내



종목

머신러닝 모델

딥러닝 모델

자동매매

카카오



날짜

종목이름

종목코드

예측종가

실제종가

정확도

23-07-19

카카오

035720

51,250

X

X%

23-07-18

카카오

035720

53,100

51,700

97.29%

23-07-17

카카오

035720

53,300

52,700

98.86%

23-07-14

카카오

035720

52,050

52,800

98.58%

23-07-13

카카오

035720

50,750

51,900

97.78%

23-07-12

카카오

035720

50,450

51,600

97.77%

23-07-11

카카오

035720

50,150

50,700

98.92%

23-07-10

카카오

035720

50,150

49,850

99.4%

# 3-3. 자동매매 페이지 이동

## 주식 투자에 필요한 모든 것 AI 자동 매매의 기준, ONE-AI

지금 시작하기



# 3-4. 투자 프로필 설정



포트폴리오 종료

로그아웃

투자금액과 목표수익률을  
입력해주세요.











투자금액

목표수익률(%)

제출하기

# 3-5. 모델 선택

## 모델별 예측률

머신 러닝 모델	딥 러닝 모델
 <b>쏘카</b> 14020 KRW    예상가 15400KRW	 <b>와이투솔루션</b> 1056 KRW    예상가 1116KRW
 <b>신정제약우</b> 26500 KRW    예상가 29100KRW	 <b>대원강업</b> 6620 KRW    예상가 6872KRW
 <b>CJ4우(전...</b> 54100 KRW    예상가 58750KRW	 <b>DB</b> 1967 KRW    예상가 2035KRW
 <b>SGC에너지</b> 24300 KRW    예상가 26300KRW	 <b>웅진생크빅</b> 2355 KRW    예상가 2435KRW
 <b>대한해운</b> 1904 KRW    예상가 2050KRW	 <b>GS건설</b> 14120 KRW    예상가 14593KRW
<b>선택</b>	<b>선택</b>

1. 모델별/일별 예상 상승률  
Top 5 리스트 나열

2. 장 마감 후 모델들이 종목 종가를  
받아와서 분석

-> 다음날 장 시작 전 오전 6시에 예측 데이터  
결과 추출 후 예측 DB에 저장

-> 예측 데이터를 웹 화면 리스트에 반영

3. 예측 상승률이 15%이상인 경우  
이상치로 판단하여 제거

# 3-6. 마이페이지



포트폴리오 종료

로그아웃

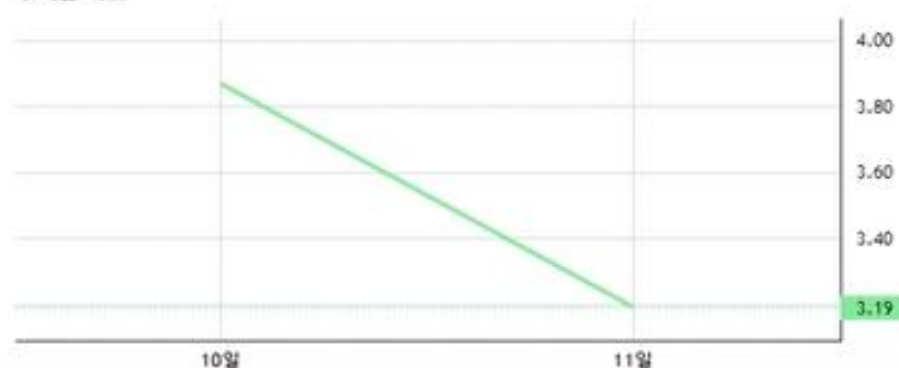


홈



거래내역

수익률 차트



sim

평가액 4953769.333333333 원

- 20.02% 쏘카
- 20.17% 대한해운
- 19.94% 신종제약우
- 19.93% 깨끗한나라우
- 19.94% 범양건영



쏘카



신종제약우

26150 KRW +-2.06%



대한해운

1930 KRW +2.01%



쏘카

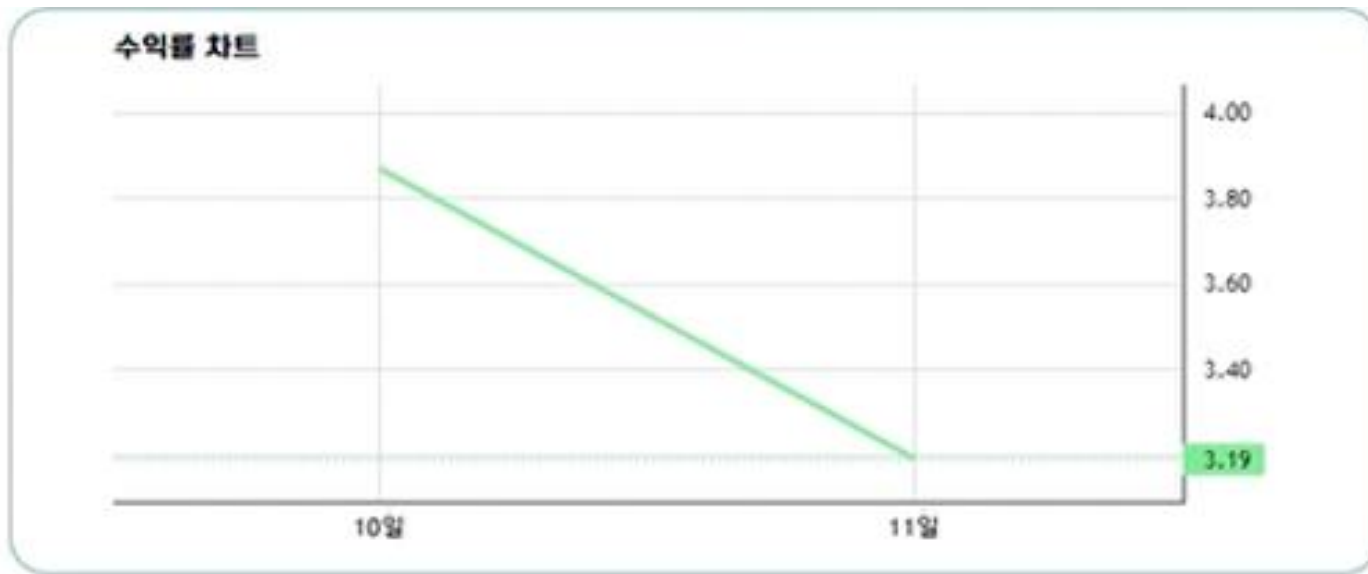
14220 KRW +-1.04%



깨끗한나라우

# 3-6. 마이페이지

## 일별 수익률 차트



1. 회원의 일별 수익률  
업데이트 및 시각화

2. 장마감 시간에 실행되는  
매도에 맞춰 차트에  
수익률을 웹 화면에 반영

3. 수익률 계산

\* 모든 종목(전일 종가)의 평가금액에서  
당일 주식 장 마감 때 매도 후  
보유 잔고로 상승률/하락률 계산

# 3-6. 마이페이지

## 회원 포트폴리오



1. 회원(포트폴리오)의 보유  
종목 비율과 평가액을 일별로  
업데이트하여 시각화
2. 추천 종목을 매수한 후 차트에  
평가금액, 보유 종목 비율을  
웹 화면에 반영

## 3-6. 마이페이지

### 보유 종목 세부 정보



1. 회원이 보유한 종목의  
리스트 나열 (스크롤)
2. 특정 종목 클릭 시 해당 종목  
캔들차트 표시
3. 장 시작 후 추천 종목을 매수 후  
종목 리스트를 웹 화면에 반영
4. 종목의 금액 : 전일 종가  
수익률 : 전일 종가에 대한  
당일의 수익률

# 3-6. 마이페이지

## 캔들 차트



1. 회원 포트폴리오에서 보유한 종목에 대한 일봉 차트 시각화
2. 장마감 후 종목의 종가를 웹 화면에 반영

# 3-7. 거래내역



포트폴리오 종료 로그아웃



홈



거래내역

## 거래내역

당일

1주일

1개월

3개월

6개월

직접입력

전체



종목명을 입력하세요

검색

거래일자	포지션	종목코드	종목명	거래단가	거래수량	거래금액
2023. 07. 19. 오전 01:02:17	매수	45014K	코오롱모빌리티그룹우	₩8,170	244.79	₩1,999,934
2023. 07. 19. 오전 01:02:17	매수	019175	신풍제약우	₩26,150	76.48	₩1,999,952
2023. 07. 19. 오전 01:02:17	매수	003410	쌍용C&E	₩5,100	392.15	₩1,999,965
2023. 07. 19. 오전 01:02:17	매수	023530	롯데쇼핑	₩69,800	28.65	₩1,999,770
2023. 07. 19. 오전 01:02:17	매수	033180	KH 필룩스	₩442	4,524.88	₩1,999,997



## 4. 분석 예측 모델

# 4-1. 데이터 수집 FinanceDataReader

---

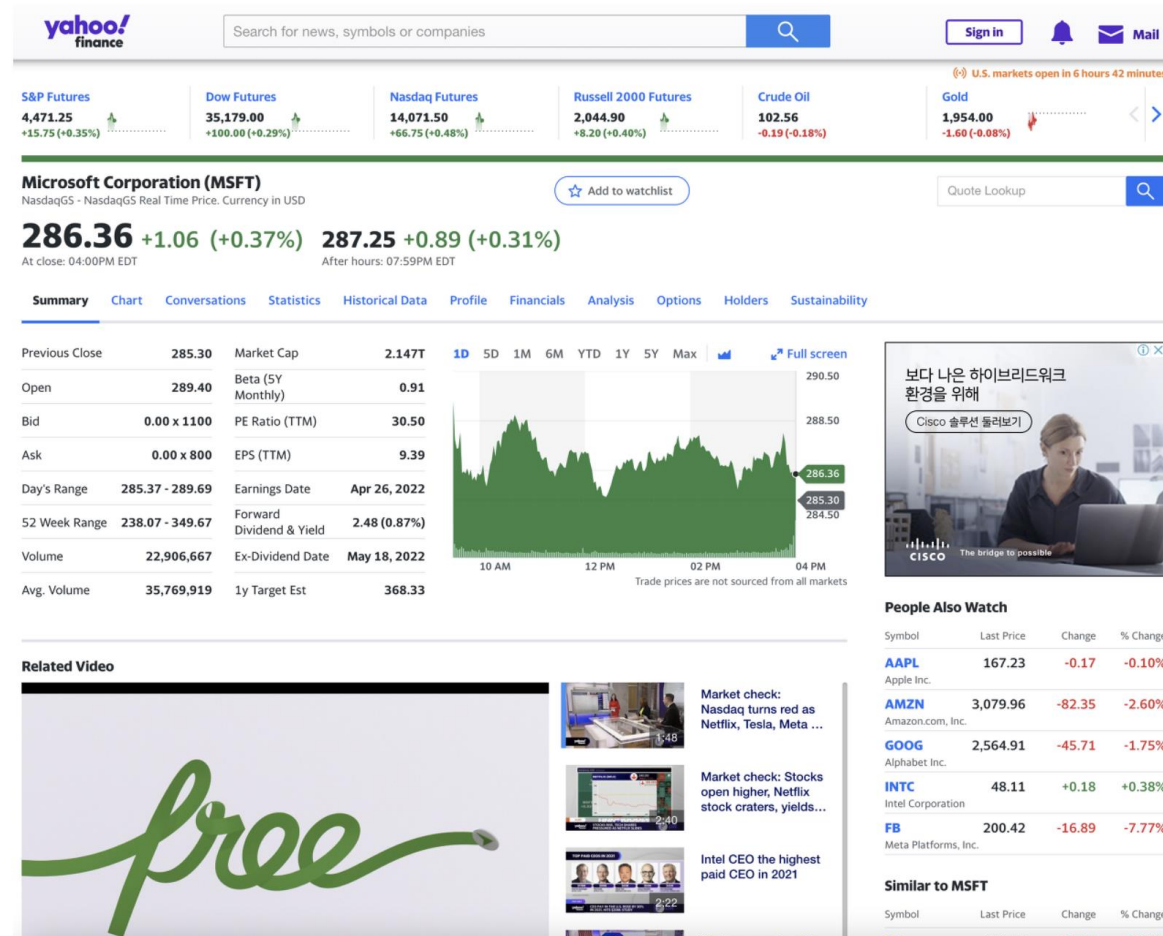


한국 주식 가격, 미국주식 가격, 지수, 환율, 암호화폐 가격, 종목 리스팅 등 금융 데이터 수집 라이브러리

## 종목 코드

- 거래소별 전체 종목코드: KRX (KOSPI, KODAQ, KONEX), NASDAQ, NYSE, AMEX, S&P 500

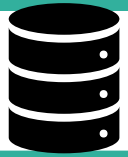
# 4-1. 데이터 수집 YahooFinance



종목코드에 해당하는 주가정보를 가져온다.

# 4-1. 모델 생성 과정

## GOAL 1



### Data Processing

데이터 수집 및 전처리

학습 데이터 구축

## GOAL 2



### Machine Learning Deep Learning

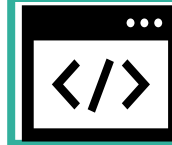
Machine learning

- Feature(여러 변수)와 Label(종가)로 분리
- 하이퍼 파라미터 튜닝을 통해 모델 성능 최적화

Deep learning

- 학습 데이터와 테스트 데이터로 분리
- 학습 데이터를 train, validation set로 분리
- Model의 input 형태에 맞게 & 전처리 및 Layer 개수와 하이퍼 파라미터 튜닝을 통해 validation loss가 작은 weight 찾기

## GOAL 3



### Full System Development

학습한 모델을 활용  
종가 예측

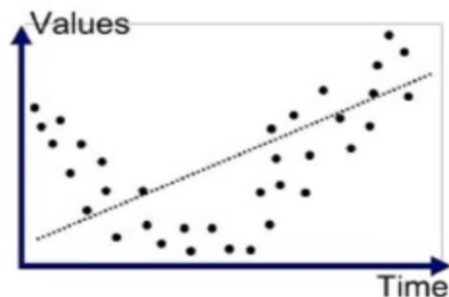
예측 종가를 이용  
주식 추천

# 머신러닝 모델

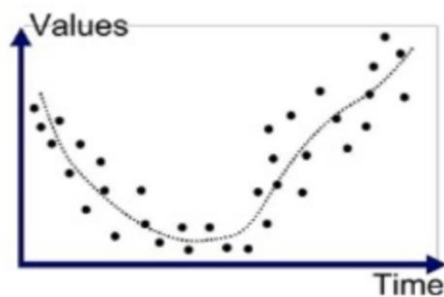
# 4-2. 과적합이란?

## 과적합 특징

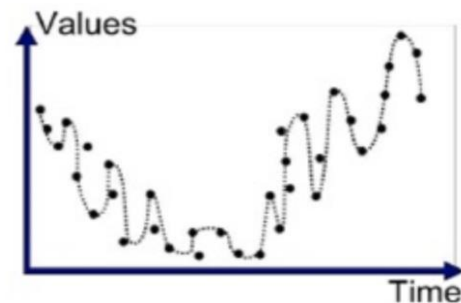
1. 모델이 훈련 데이터에 지나치게 맞추어져서 새로운 데이터에 대한 예측 능력이 떨어지는 현상
2. 훈련 데이터의 양이 적거나 데이터가 불균형한 경우
3. 모델이 너무 복잡한 경우 (예: 많은 수의 파라미터를 가진 모델)
4. 특성(feature)이 너무 많은 경우
5. 이상치(outlier)나 잡음(noise)이 있는 경우
6. 훈련 데이터와 테스트 데이터가 다른 분포를 가지는 경우



Underfitted



Good Fit/Robust

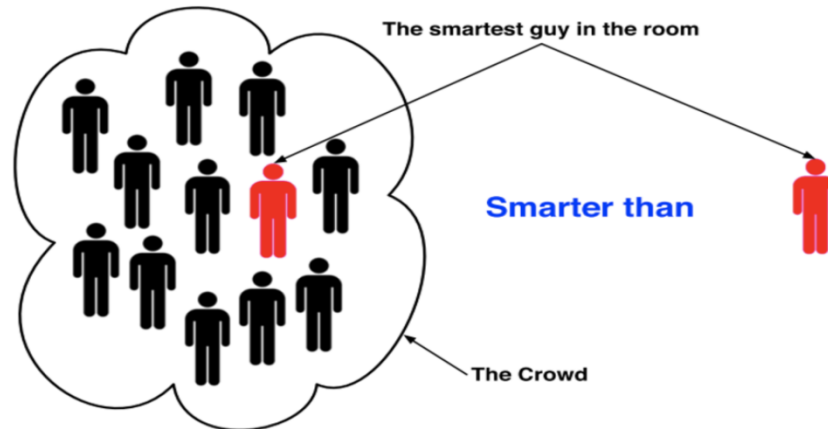


Overfitted

## 4-2. 앙상블 모델

### 앙상블 모델

1. 강력한 하나의 모델을 사용하는 대신 보다 약한 모델 여러 개를 조합하여 더 정확한 예측에 도움을 주는 방식
2. 현실세계로 예를 들면, 어려운 문제를 해결하는데 한 명의 전문가보다 여러 명의 집단지성을 이용하여 문제를 해결하는 방식



## 4-2. 랜덤 포레스트(Random Forest) 특징

### 다양성

- ✓ 각 트리가 독립적으로 학습하기 때문에 다양한 관점에서 데이터를 학습  
-> 이는 모델의 과적합을 방지하고 일반화 성능을 향상

### 간결성

- ✓ 랜덤 포레스트는 결정 트리의 간결성과 이해하기 쉬운 특성을 그대로 보유  
-> 모델의 해석력을 상승

### 확장성

- ✓ 각 트리가 독립적으로 학습하기 때문에 다양한 관점에서 데이터를 학습  
-> 병렬 처리를 활용해 대용량 데이터를 처리하는 데 효율적

### 앙상블 학습

- ✓ 여러 개의 의사 결정 트리를 결합하여 생성되는 앙상블 모델  
-> 일반적으로 단일 모델보다 더 뛰어난 성능을 제공
- ✓ 각 모델의 강점을 결합하고 약점을 상쇄하여 전체적으로 더 견고한 예측을 제공



## 4-2. 모델 학습

### 학습 1

- ✓ 시가총액, 등락률, 고가, 저가, 시가, 주식거래일, 거래량을 Feature로 설정

### 학습 2

- ✓ 주요 요소들과 실제 종가 사이의 관계를 학습하여 모델을 구축

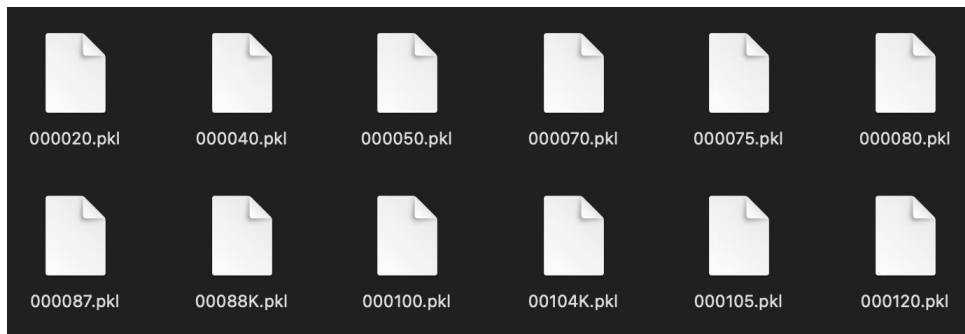
### 학습 3

- ✓ GridSearchCV를 이용하여 주어진 하이퍼파라미터에 대해 가능한 모든 조합을 시도
- ✓ 그 중에서 가장 성능이 좋은 하이퍼파라미터 조합을 찾아줌

## 4-2. 모델 저장

### 모델 저장

1. 개별 주식 종목들은 각각 독특한 특성과 도메인 특성을 가지므로, 여러 종목을 일괄적으로 예측하기 위한 단일 모델보다는, 각 종목에 대해 훈련된 모델을 사용
2. pickle 모듈을 활용하여, 각각의 학습된 모델을 파일 형태로 저장
3. 이렇게 모델을 저장하게 되면 모델의 재사용성을 높여주며, 추후에 해당 모델을 필요로 할 때 마다 학습 과정을 거칠 필요 없이 빠르게 결과를 도출



X 953

## 4-2. 모델 평가

### 모델 평가

1. 모든 종목의 오차율을 비교해보면, 50% 이상인 경우도 나온다.
2. 보통 오차율이 20% 이내에 있는 정규분포 형태
3. 이 모델이 좋다고는 할 수 없다.

# 딥러닝 모델

# 4-3. Sequence

개념

설명

1. 순서가 있는 data

(ex) time series, text, 영상, 음성 등

2. randomforest는 inputs들이 독립적

=> 순차적으로 과거 정보를 반영할 수 있는 모델이 필요

3. time의 특징을 반영한 'Sequence model' 모델 선정

# 4-3. 딥러닝

개념

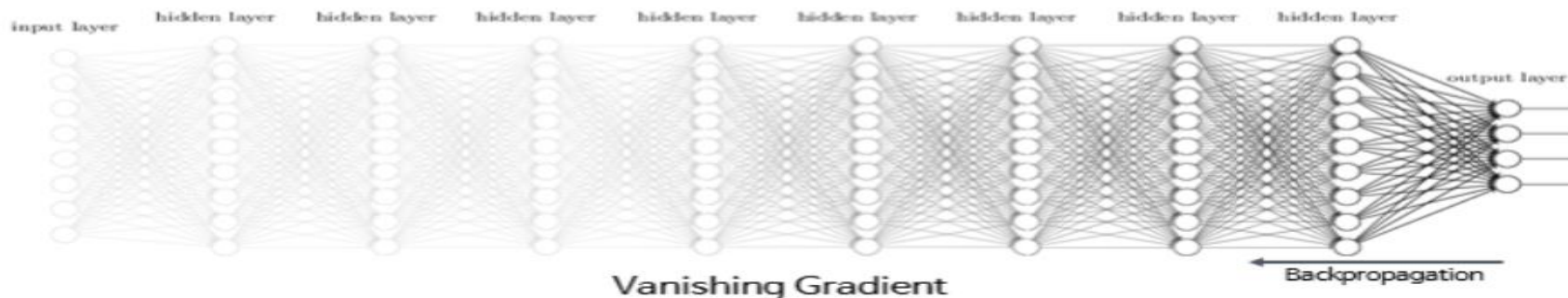
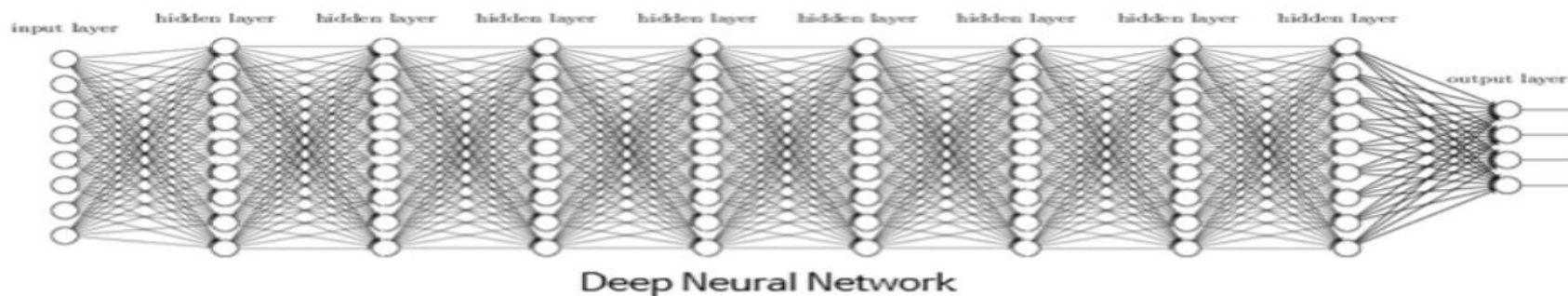
설명

## 1. 기울기 소실 문제

: layer를 많이 쌓을수록 데이터 표현력이 증가하기 때문에 학습이 잘 될 것 같지만 실제로는 layer가 많아질수록 학습이 잘 되지 않음.

## 2. Vanishing Gradient(기울기 소실)

: backpropagation(역전파)으로 출력층에서 입력층 방향으로 오차를 전파시키며 각층의 가중치를 업데이트하는데 출력층과 멀어질수록 기울기(gradients)값이 작아질 수 밖에 없음

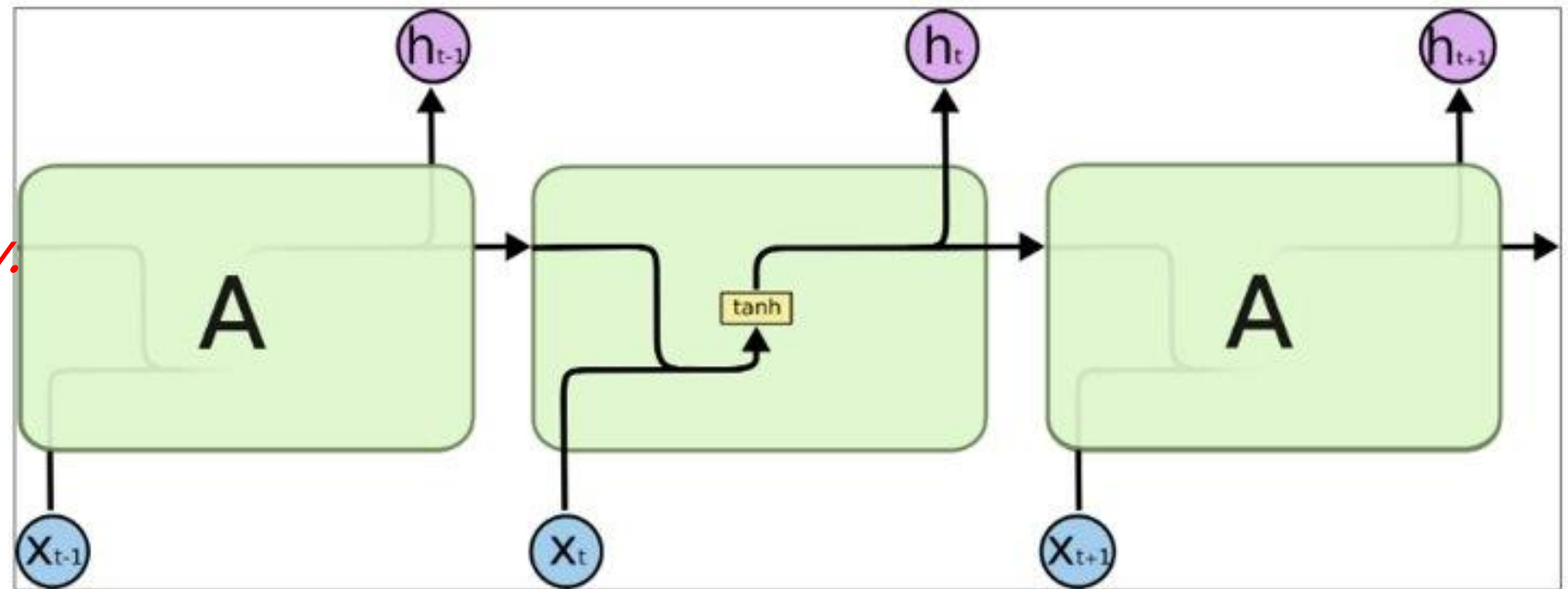


# 4-3. Recurrent Neural Network(RNN)

## 개념 설명

1. 입력정보와 그 입력정보를 사용하려는 출력지점 거리가 멀 경우,  
역전파알고리즘에 의해 기울기가 점차 줄어들어 출력과 먼 위치에 있는  
정보를 기억할 수 없고 최근의 정보를 예측에 더 많이 반영

Ex. The clouds are in the *sky*.



# 4-3. Long Short Term Memory(LSTM)

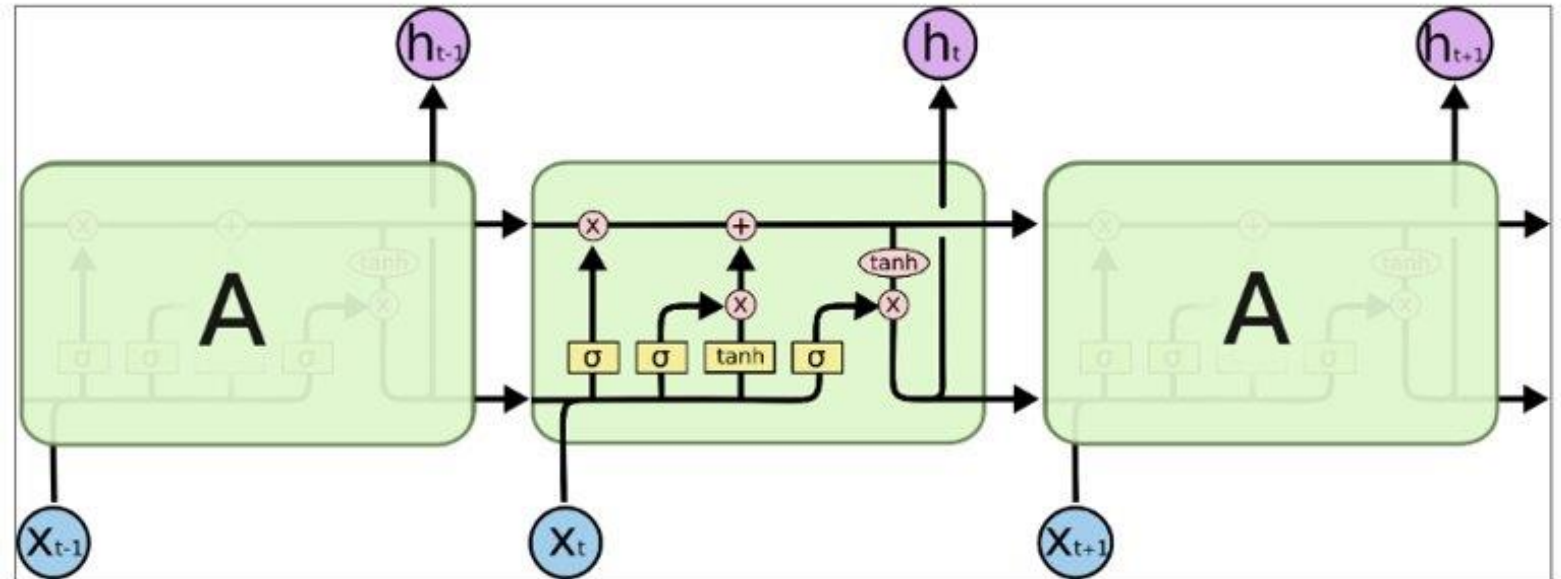
개념

설명

1. RNN이 출력과 먼 위치에 있는 정보를 기억할 수 없다는 단점을 보완하여 장/단기 기억을 가능하게 설계한 신경망의 구조

2. 주로 시계열 처리나, 자연어 처리

Ex. I grew up in France...  
I speak fluent *French*





# 4-3. LSTM 모델의 한계

## 모델 한계

### 1. 동적인 시장 조건 처리의 어려움

- 주식 시장은 복잡하고 예측하기 어려운 조건과 패턴을 가지므로, 모든 패턴을 정확하게 예측에 한계가 존재
- 과거의 패턴을 기반으로 예측을 수행하기 때문에, 시장 조건이 급격히 변하거나 예측 모델이 학습되지 않은 패턴을 마주할 경우 예측 성능이 저하

### 2. 데이터 불확실성과 노이즈

- 주식 시장은 매우 불안정하여, 예측하기 어려운 노이즈와 불확실성이 많아 모델의 성능이 불안정
- 노이즈와 잘못된 데이터가 모델에 포함되면 예측 성능이 저하

# 4-3. Time2vec

개념

설명

## 1. 주기, 비 주기 패턴을 파악

- 전체적인 시간흐름에서 주기적, 비주기적 패턴을 찾음
- 비주기적 패턴을 파악한 것이 더 좋은 성능을 냄

## 2. Time rescaling에 대해서 불변

- Time series data에서 고유값, 고유벡터를 추출하기 때문에 다른 scales(day, week 등등)로 측정되더라도, scale에 상관없이 time representation에 중요한 정보들은 불변

## 3. 여러 모델들에 쉽게 적용 가능

Learning a **Vector** Representation of Time

## 4-3. 모델 구조

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 20, 1)] input	0
t2v_1 (T2V)	(None, 20, 17)	376
lstm_1 (LSTM)	(None, 32)	6400
dense_1 (Dense)	(None, 1) output	33

	Input
	Output

[illegible]

## 4-3. 23/07/12일 종가 예측 결과

전날 종가

다음날 예측 종가    예측 등락률

CODE	CREATED_DATE	NAME	PREV_CLOSE	CLOSE	PRED_CLOSE	PRED_RATE	DIFF
900140	23/07/12	엘브이엠씨홀딩스	2620	(null)	2530	3.44	(null)
207940	23/07/12	삼성바이오로직스	723000	(null)	739301	-2.25	(null)
373220	23/07/12	LG에너지솔루션	550000	(null)	554737	-0.86	(null)
298050	23/07/12	효성첨단소재	453500	(null)	470231	-3.69	(null)
298020	23/07/12	효성티앤씨	352500	(null)	361403	-2.53	(null)
457190	23/07/12	이수스페셜티케미컬	297000	(null)	304774	-2.62	(null)
352820	23/07/12	하이브	270000	(null)	275884	-2.18	(null)
259960	23/07/12	크라프트온	182500	(null)	180517	1.09	(null)

# 4-3. 모델 학습 및 예측 프로세스

## 데이터 전처리


### ✓ 삼성전자 10년치 일봉 활용

- 삼성전자가 시총 1위로 KOSPI를 구성하는데 가장 높은 비중을 차지함으로, 해당 주가의 변동이 전반적인 KOSPI 가격 변동을 이끄므로 분석 대상으로 적합 판단

## 모 델 학 습

### ✓ Timeseries 데이터를 vector 형태로 변환한 후 lstm의 input으로 넣어 layer를 쌓으면서 과적합되지 않게 모델 구축

### ✓ GridSearchCV를 이용하여 주어진 하이퍼 파라미터에 대해 가능한 모든 조합을 시도하고, 그 중에서 가장 validation loss가 적은 하이퍼파라미터 조합을 찾아 weight 저장

 checkpoint\_best.hdf5

## 예측

### ✓ 학습된 weight를 불러와 최신 20일치 종가 데이터를 전처리하여 input으로 넣으면 21일째(다음날)의 종가 예측

## 4-3. 백트래킹

### 전체종목 평균 오차율

-> 20일치의 데이터가 존재하는 모든 종목의  
'오차율 (=  $|종가 - 종가예측| / 종가$ )'의 평균

	2023/07/04	2023/07/05	2023/07/06	2023/07/07
전체종목 평균 오차율	4.4	4.19	3.5	3.48
종목 수	947	947	947	947

THANK YOU