

# 주식데이터와 머신러닝을 활용한 로보어드바이저 서비스



상장마켓

ALL

KOSPI200

KOSPI

KOSDAQ

KONEX

단일선택

업종구분

ALL

농업, 임업 및 어업

광업

제조업

전기, 가스, 증기 및 공기 조절 공급업

수도, 하수 및 폐기물 처리, 원료 재생업

건설업

도매 및 소매업

운수 및 창고업

숙박 및 음식점업

정보통신업

금융 및 보험업

부동산업

전문, 과학 및 기술 서비스업

사업시설 관리, 사업 지원 및 임대 서비스업

교육 서비스업

예술, 스포츠 및 여가관련 서비스업

협회 및 단체, 수리 및 기타 개인 서비스업

다중선택

조회기간

날짜선택

시작현금

숫자입력

07/01/2020 - 10/01/2020

100000000

익절률

%입력

손절률

%입력

7

3

최소 보유 기간

숫자입력

최대 보유 종목수

숫자입력

10

10

매수 조건 설정

거래기법

조건입력

숫자입력

종합점수

내림차순

추가

매도 조건 설정

거래기법

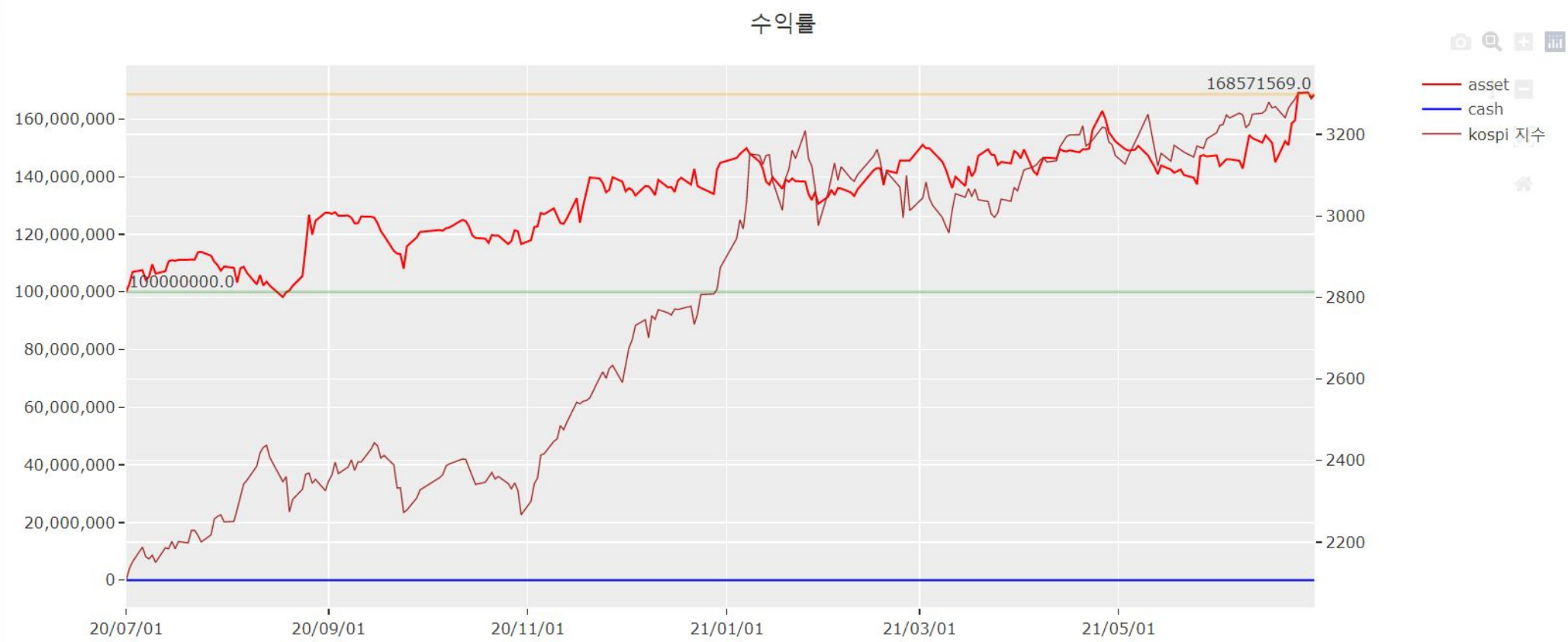
조건입력

숫자입력

추가

새로고침

Backtest



결과 내역				
백테스트 기간	누적수익률	일평균수익률	매수조건	매도조건
2020-07-01 ~ 2021-06-30	68.57 %	0.28 %	TOTAL_SCORE sort False, TOTAL_SCORE > 80	TOTAL_SCORE < 50

거래 내역						
날짜	매수/매도	종목 (종목코드)	거래수량	매수단가 매수금액	매도단가 매도금액	수익률 실현손익
2020-07-01	buy	코디엠 (224060)	35,842	279 원 9,999,918 원		
2020-07-01	buy	슈팩스비앤피 (058530)	59,171	169 원 9,999,899 원		
2020-07-01	buy	GV (045890)	12,484	801 원 9,999,684 원		
2020-07-01	buy	한샘 (009240)	110	90,800 원 9,988,000 원		
2020-07-01	buy	스킨앤스킨 (159910)	33,905	295 원 10,001,975 원		
2020-07-01	buy	케이피에스 (256940)	743	13,450 원 9,993,350 원		
2020-07-01	buy	신원 (009270)	6,272	1,595 원 10,003,840 원		
2020-07-01	buy	일신석재 (007110)	4,330	2,310 원 10,002,300 원		

현재자산

종목 ↑↓	수량 ↑↓	매입단가 현재가 ↑↓	매입금액 평가금액 ↑↓	평가손익 손익률 ↑↓
JW중외제약 (001060)	325	38,150 ₩ 38,700 ₩	12,398,750 ₩ 12,577,500 ₩	178,750 ₩ +1.42 %
대원제약 (003220)	579	21,400 ₩ 22,300 ₩	12,390,600 ₩ 12,911,700 ₩	521,100 ₩ +4.04 %
디케이락 (105740)	1331	9,320 ₩ 9,520 ₩	12,404,920 ₩ 12,671,120 ₩	266,200 ₩ +2.10 %
서울리기 (043710)	10862	1,160 ₩ 1,075 ₩	12,599,920 ₩ 11,676,650 ₩	-923,270 ₩ -7.91 %
세화피앤씨 (252500)	3176	3,865 ₩ 4,030 ₩	12,275,240 ₩ 12,799,280 ₩	524,040 ₩ +4.09 %
신라섬유 (001000)	6110	2,035 ₩ 2,085 ₩	12,433,850 ₩ 12,739,350 ₩	305,500 ₩ +2.40 %
엑셀 (205100)	3276	3,785 ₩ 3,810 ₩	12,399,660 ₩ 12,481,560 ₩	81,900 ₩ +0.66 %
영우디에스피 (143540)	8235	1,530 ₩ 1,570 ₩	12,599,550 ₩ 12,928,950 ₩	329,400 ₩ +2.55 %
일신석재 (007110)	6074	2,045 ₩ 2,040 ₩	12,421,330 ₩ 12,390,960 ₩	-30,370 ₩ -0.25 %
케이엠제약 (225430)	2800	4,430 ₩ 4,475 ₩	12,404,000 ₩ 12,530,000 ₩	126,000 ₩ +1.01 %



# Team Member



김기남

Database  
백엔드



김지석

Database  
백엔드



박승원

서비스 기획  
웹 퍼블리싱



---

# CONTENTS

---

---

01. 서비스 소개

---

---

02. 상세 구현

---

---

03. Discussion

---

---

04. 시 연

---

# 01

## 서비스 소개

---

---



# Background & Problem



'주식열풍'...1분기 주식투자, 역대 최대 규모기록

단타에 빠진 '주린이' ... 상승장에도 10명중 6명 손실

지난달 개미 9조 샀지만...수익은 지지부진

삼성전자·SK하이닉스 등 매수  
반도체株 약세에 수익률 저조

기관은 삼성전자 8천억대 팔고  
삼바 등 기타 삼성株 사들여 수익

외국인 SKIET·포스코로 선방

# Robo Advisor

## 매수 / 매도

언제

무엇을

얼마나

국내 로보 어드바이저 시장 규모 전망  
(단위 : 원)



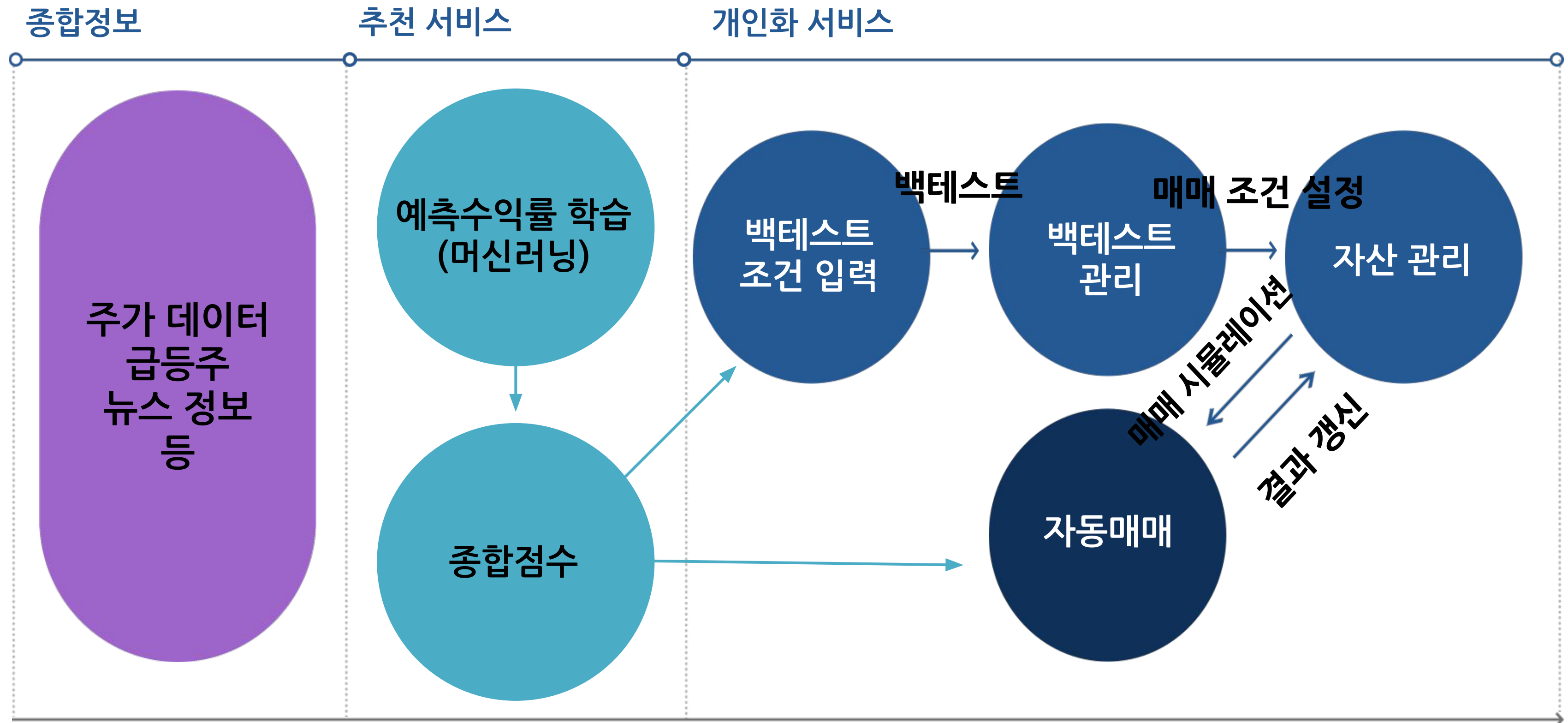
자료 : 한국과학기술정보연구원

# Our Solution

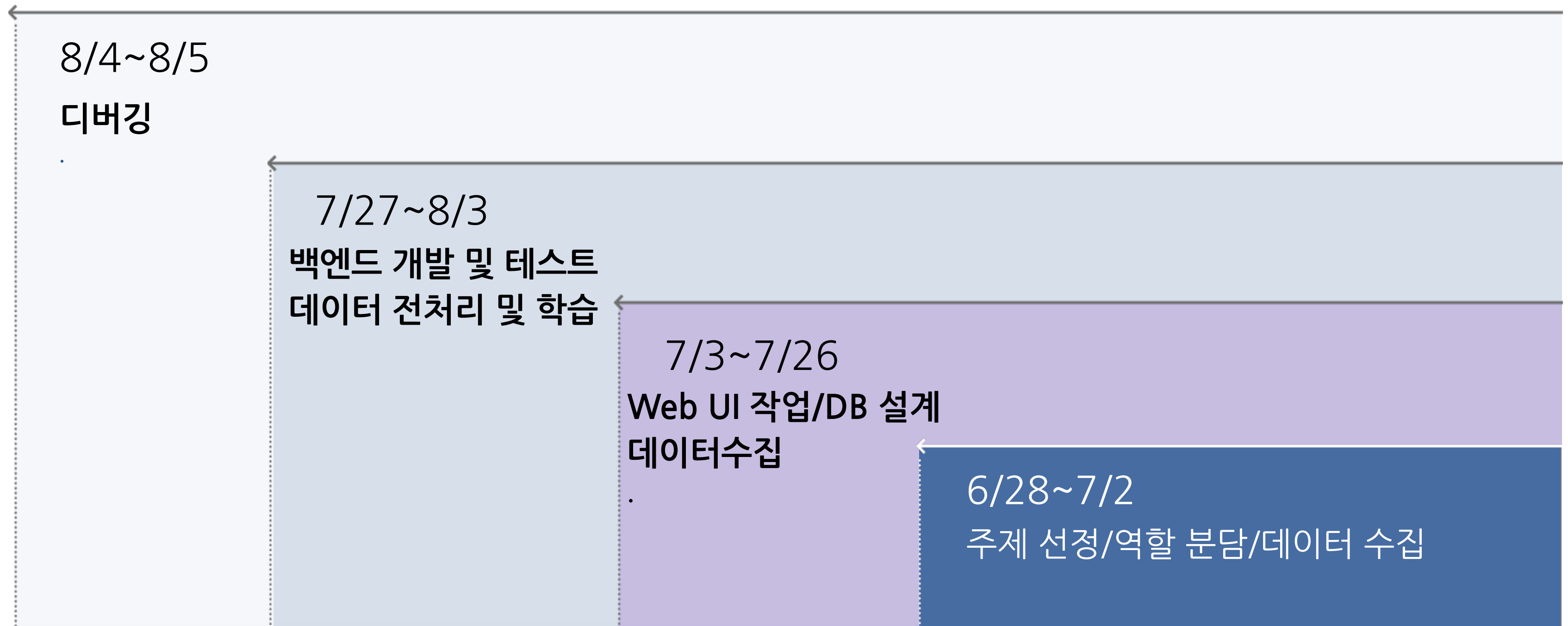
---



# 투더문 RA Service Plan



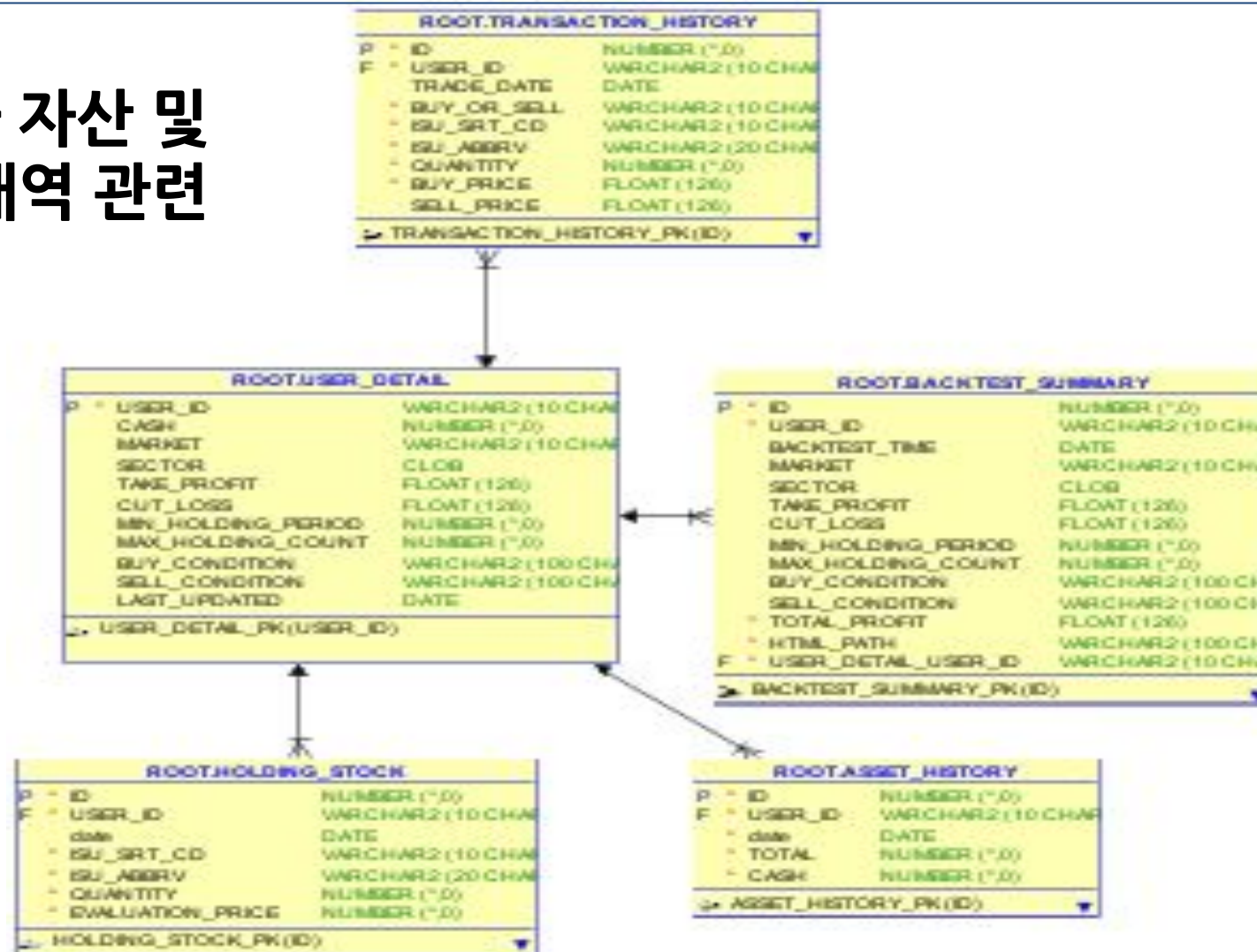
# 개발 일정





# ERD

사용자 자산 및  
거래 내역 관련



백테스트, 매매  
정보 관련



사용자 개인  
정보 관련

# 02

## 상세 구현

---

---



# 추천서비스 - 종합점수

IOWA STATE UNIVERSITY  
Digital Repository

Creative Components

Iowa State University Capstones, Theses and  
Dissertations

Fall 2019

Stock Prediction with Random Forests and Long Short-term  
Memory

Shangxuan Han

Follow this and additional works at: <https://lib.dr.iastate.edu/creativecomponents>

Part of the [Computer Engineering Commons](#)

## Recommended Citation

Han, Shangxuan, "Stock Prediction with Random Forests and Long Short-term Memory." *Components*. 393.

<https://lib.dr.iastate.edu/creativecomponents/393>

This Creative Component is brought to you for free by the Iowa State University Theses and Dissertations at Iowa State University. Components by an authorized administrator of the Iowa State University Digital Repository. please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).



Article

## Feature Learning for Stock Price Prediction Shows a Significant Role of Analyst Rating

Jaideep Singh and Matloob Khushi \*

School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia; [jsin9199@alumini.sydney.edu.au](mailto:jsin9199@alumini.sydney.edu.au)

\* Correspondence: [mkhushi@uni.sydney.edu.au](mailto:mkhushi@uni.sydney.edu.au)

**Abstract:** Efficient Market Hypothesis states that stock prices are a reflection of all the information present in the world and generating excess returns is not possible by merely analysing trade data which is already available to all public. Yet to further the research rejecting this idea, a rigorous literature review was conducted and a set of five technical indicators and 23 fundamental indicators was identified to establish the possibility of generating excess returns on the stock market. Leveraging these data points and various classification machine learning models, trading data of the 505 equities on the US S&P500 over the past 20 years was analysed to develop a classifier effective for our cause. From any given day, we were able to predict the direction of change in price by 1% up to 10 days in the future. The predictions had an overall accuracy of 83.62% with a precision of 85% for buy signals and a recall of 100% for sell signals. Moreover, we grouped equities by their sector and repeated the experiment to see if grouping similar assets together positively effected the results but concluded that it showed no significant improvements in the performance—rejecting the idea of sector-based analysis. Also, using feature ranking we could identify an even smaller set of 6 indicators while maintaining similar accuracies as that from the original 28 features and also uncovered the importance of buy, hold and sell analyst ratings as they came out to be the top contributors in the model. Finally, to evaluate the effectiveness of the classifier in real-life situations, it was backtested on FAANG (Facebook, Amazon, Apple, Netflix & Google) equities using a modest trading strategy where it generated high returns of above 60% over the term of the testing dataset. In conclusion, our proposed methodology with the combination of purposefully picked features shows an improvement over the previous studies, and our model predicts the direction of 1% price changes on the 10th day with high confidence and with enough buffer to even build a robotic trading system.

**Keywords:** stock index prediction; feature engineering; machine learning in stocks

## 1. Introduction

There has been extensive research to predict a firm's performance [1,2], with some of the earliest studies beginning in the 1980s. Yet, as expressed by Kiang [3], the progress in this field has been limited primarily because of lack of willingness on behalf of security analysts to share their knowledge and their inability to articulate their knowledge in unambiguous and simple terms for the domain to be taken forward. Moreover, a popular idea, the Efficient Market Hypothesis by Malkiel and Fama [4] states that prices of stocks are informationally efficient i.e., the stock prices reflect all information available out there and it is not possible to predict stock prices based on the trading data and nor is it possible to obtain excess returns by exploiting any predictability of prices. Yet, on the other hand, Cheung et al. [5] and Pesaran et al. [6] argued and established that to execute profitable

Academic Editor: Christos Douligoris

Received: 20 December 2020

Accepted: 9 February 2021

Published: 1 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

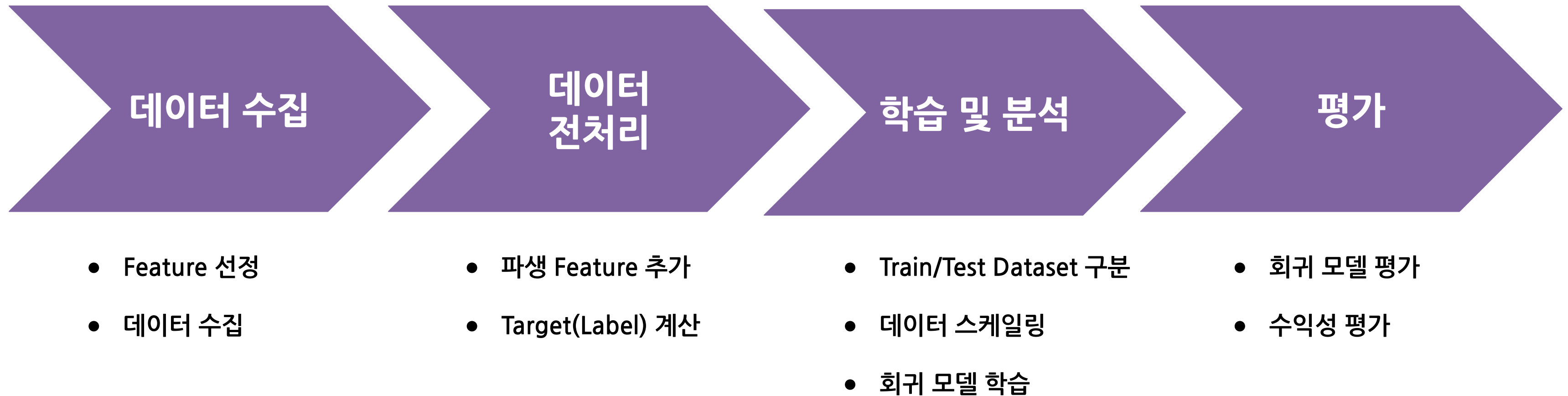
## 이전 연구들

1. 주식데이터로부터 매수/보유/매도 시그널에 대한 분류 모델 학습
2. ARIMA Model 또는 LSTM을 사용한 주가 예측

## 목 표

1. 주식 데이터와 보조 지표로부터 회귀모델을 사용하여 미래 수익률을 예측
2. 수익률을 점수화 하여 종합점수로 정의하고 주식 종목 추천의 지표로 하여 사용자에게 추천

# 추천서비스 - 종합점수

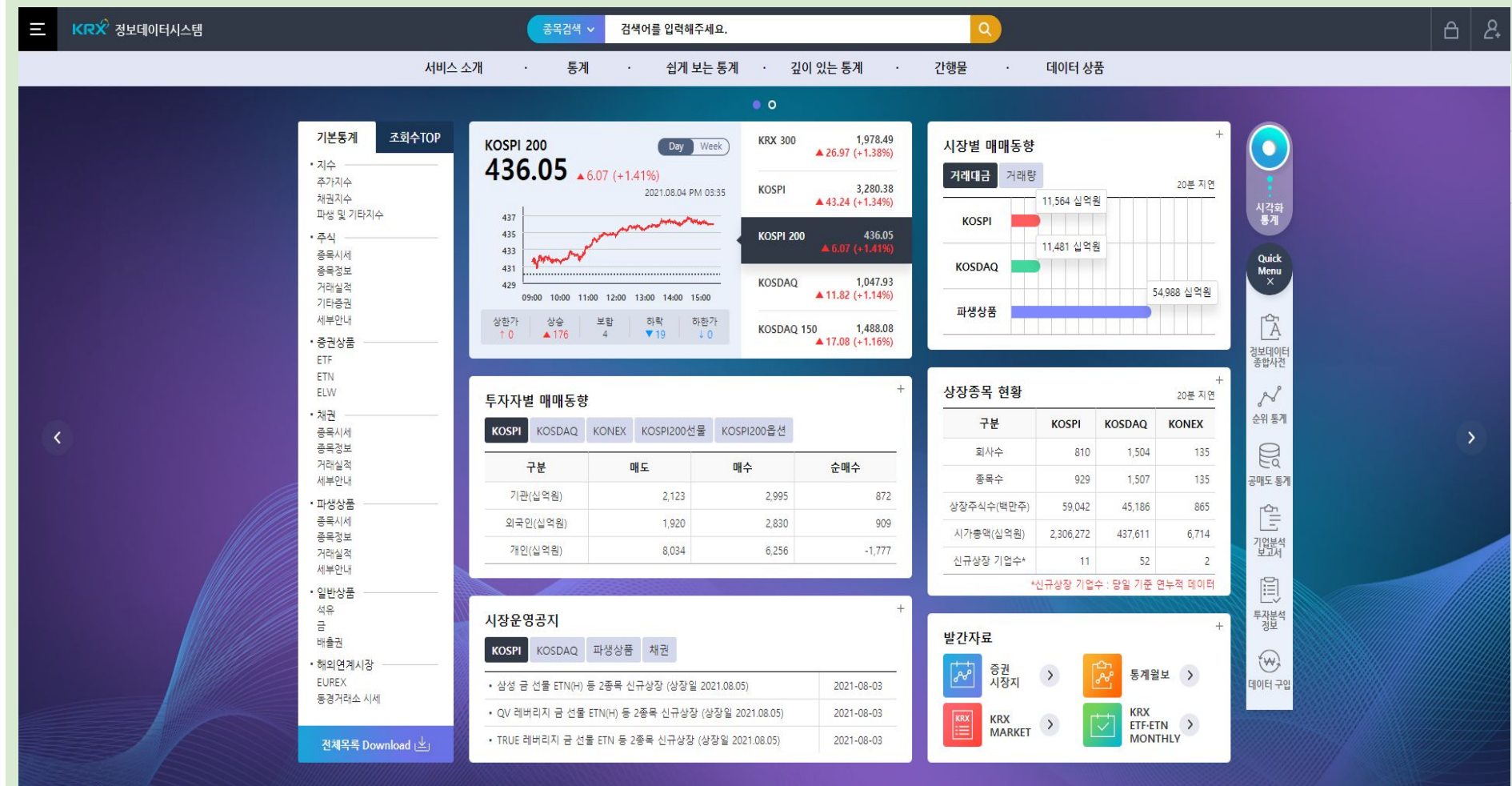


# 데이터 수집

## Feature 선정

개장일  
종목명  
종목코드  
시가총액  
EPS, PER, BPS, PBR  
종가  
거래량, 거래대금  
외국인거래량, 외국인거래대금  
외국인보유비중  
기관거래량, 기관거래대금  
공매도거래량비중, 공매도거래대금비중

### 한국거래소 KRX



기간: 2017/01/01 ~ 2021/07/19

# 데이터 전처리 - 파생 Feature 추가

## 종가

### RSI

일정 기간 동안의 가격 상승과 하락 추세에서 상승폭과 하락폭을 계산하여 그 강도를 백분율로 계산

현재 가격의 상태를 평가

일반적으로 70% 이상을 매도 시그널로 30% 이하에서 매수 시그널로 봄

### 이동평균선

종가기준으로 일정 기간 동안의 이동평균치를 나타내는 보조지표

상승과 하락 동향을 예측하고 판단하는데 사용

### 이격도

이동평균선과 종가간의 괴리

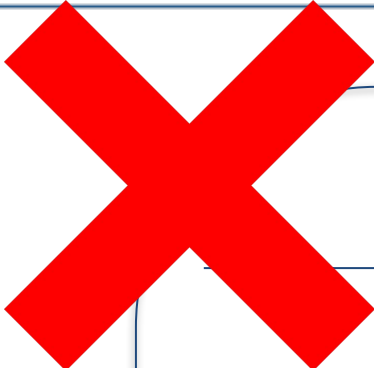
### 스토캐스틱

최근 n일간의 최고가와 최저가 범위내에서 현재 종가의 위치를 나타내는 보조지표

RSI와 마찬가지로 현재 가격의 상태를 평가



# 데이터 전처리 - Target 계산



현재 종가 ~ n일 후 종가 간 수익률

종가		
0일	10000	10 %
1일	10000	5 %
2일	11100	...
...		
n-1일	10900	
n일	11000	
n+1일	10500	

$$(n\text{일 후 종가} - \text{현재 종가}) / \text{현재종가} \times 100$$



매매 알고리즘에 따른 수익률

종가		
0일	10000	10.6 %
1일	10000	10.8 %
2일	11100	...
...		
n-1일	10900	(익절: 7 %인 경우)
n일	11000	※시간 기회비용 고려(0.2)

m일에 익절%를 넘긴 경우  
 $(m\text{일 후 종가} - \text{현재 종가}) / \text{현재종가} \times 100 - 0.2 \times m$

n일까지 익절%를 넘기지 못한 경우  
 $(n\text{일 후 종가} - \text{현재 종가}) / \text{현재종가} \times 100 - 0.2 \times n$

# 데이터 학습

## Train Dataset

2017/01/01 ~ 2020/06/30  
3년 6개월

전처리 및 결측치 제거 후  
1876773 rows × 48 columns

## Test Dataset

2020/07/01 ~ 2021/06/30  
1년

전처리 및 결측치 제거 후  
624974 rows × 48 columns

## Data Scaling - Standard Scaler

### 1차 학습 모델

Linear Regression, Lasso Regression, Ridge Regression

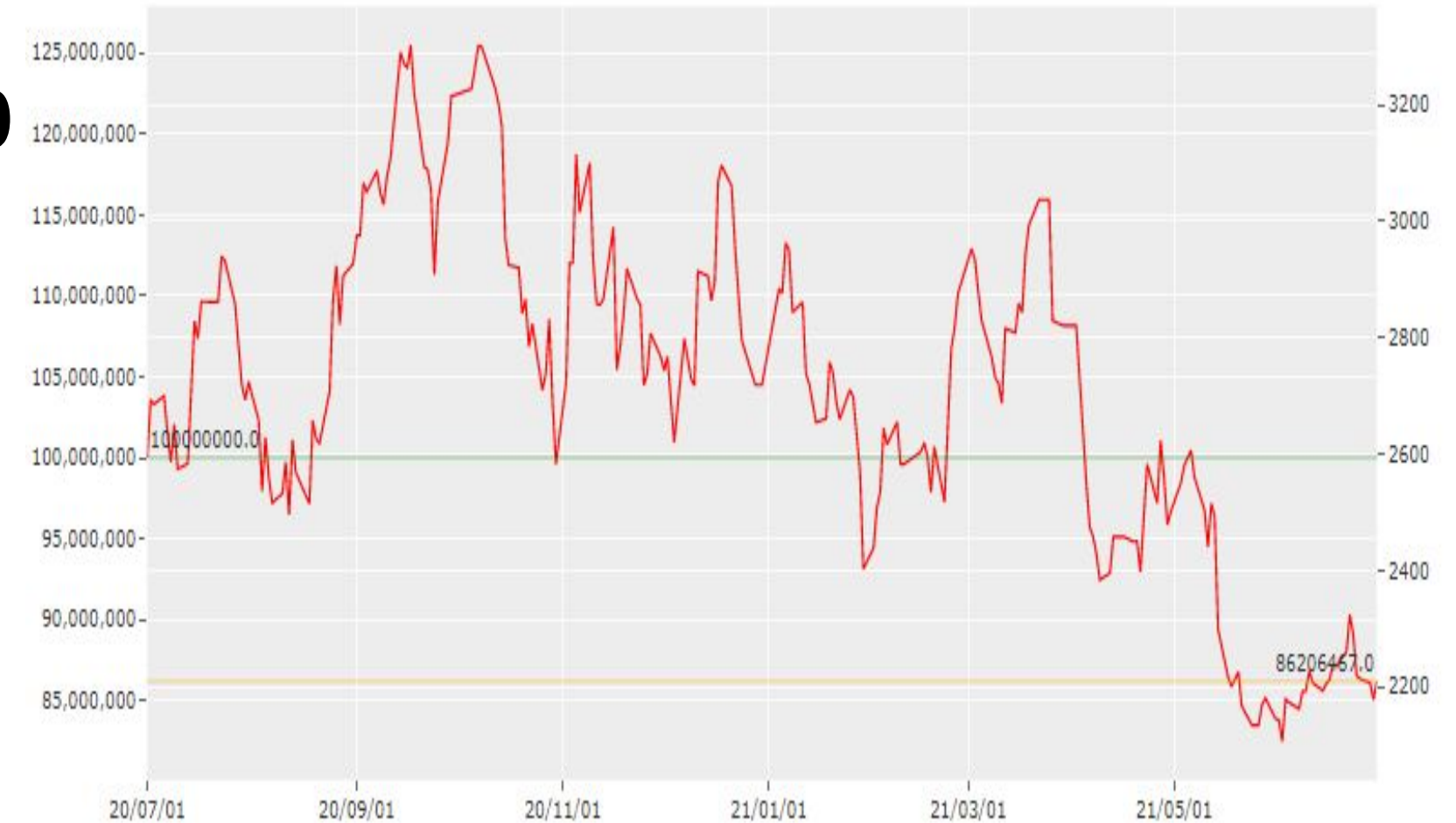
# 평가

MODEL	MSE	RMSE	수익률
Linear Regression	61.27	7.83	-13.79 %
Lasso Regression	46.46	6.81	24.48 %
Ridge Regression	46.25	6.80	33.39 %

## 수익률 산출 방식

기간 : 2020/07/01 ~ 2021/06/30  
 타겟 : 모든 상장 주식  
 현금 : 1억원  
 최소보유기간 : 10일  
 최대보유종목수 : 10개  
 익절 : 7%  
 손절 : 3%

예측수익률 높은 순으로 매수 실시





# 재학습 - AutoML

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
rf	Random Forest Regressor	4.1221	31.7852	5.6373	0.2005	0.9262	2.7884
et	Extra Trees Regressor	4.1978	32.7869	5.7255	0.1755	0.9322	2.9113
lightgbm	Light Gradient Boosting Machine	4.3603	35.3214	5.9426	0.1118	0.9203	2.8408
gbr	Gradient Boosting Regressor	4.4359	36.9405	6.0773	0.0711	0.9203	2.8775
br	Bayesian Ridge	4.5237	39.2499	6.2642	0.0132	0.9914	2.6246
en	Elastic Net	4.5224	39.3650	6.2734	0.0103	0.9680	2.5977
omp	Orthogonal Matching Pursuit	4.5266	39.4230	6.2780	0.0088	0.9753	2.6058
lasso	Lasso Regression	4.5235	39.4379	6.2792	0.0084	0.9552	2.6117
ridge	Ridge Regression	4.5286	39.7394	6.3032	0.0009	0.9314	2.6792
lr	Linear Regression	4.5292	39.7471	6.3038	0.0007	0.9320	2.6730
llar	Lasso Least Angle Regression	4.5286	39.7837	6.3067	-0.0002	0.9280	2.7090
huber	Huber Regressor	4.7277	41.6535	6.4532	-0.0473	1.5488	1.3024
knn	K Neighbors Regressor	4.9981	46.2909	6.8032	-0.1642	0.9655	4.3829
ada	AdaBoost Regressor	5.1299	48.5030	6.9631	-0.2209	0.8228	6.0935
dt	Decision Tree Regressor	5.8947	66.6884	8.1656	-0.6776	0.9235	7.4199
par	Passive Aggressive Regressor	9.8769	923.5017	26.3360	-22.2969	1.2805	11.0136

Random Forest Regressor  
Extra Trees Regressor  
Light Gradient Boosting Machine

Hyperparameter Tuning 후 CV fold 5 Ensemble

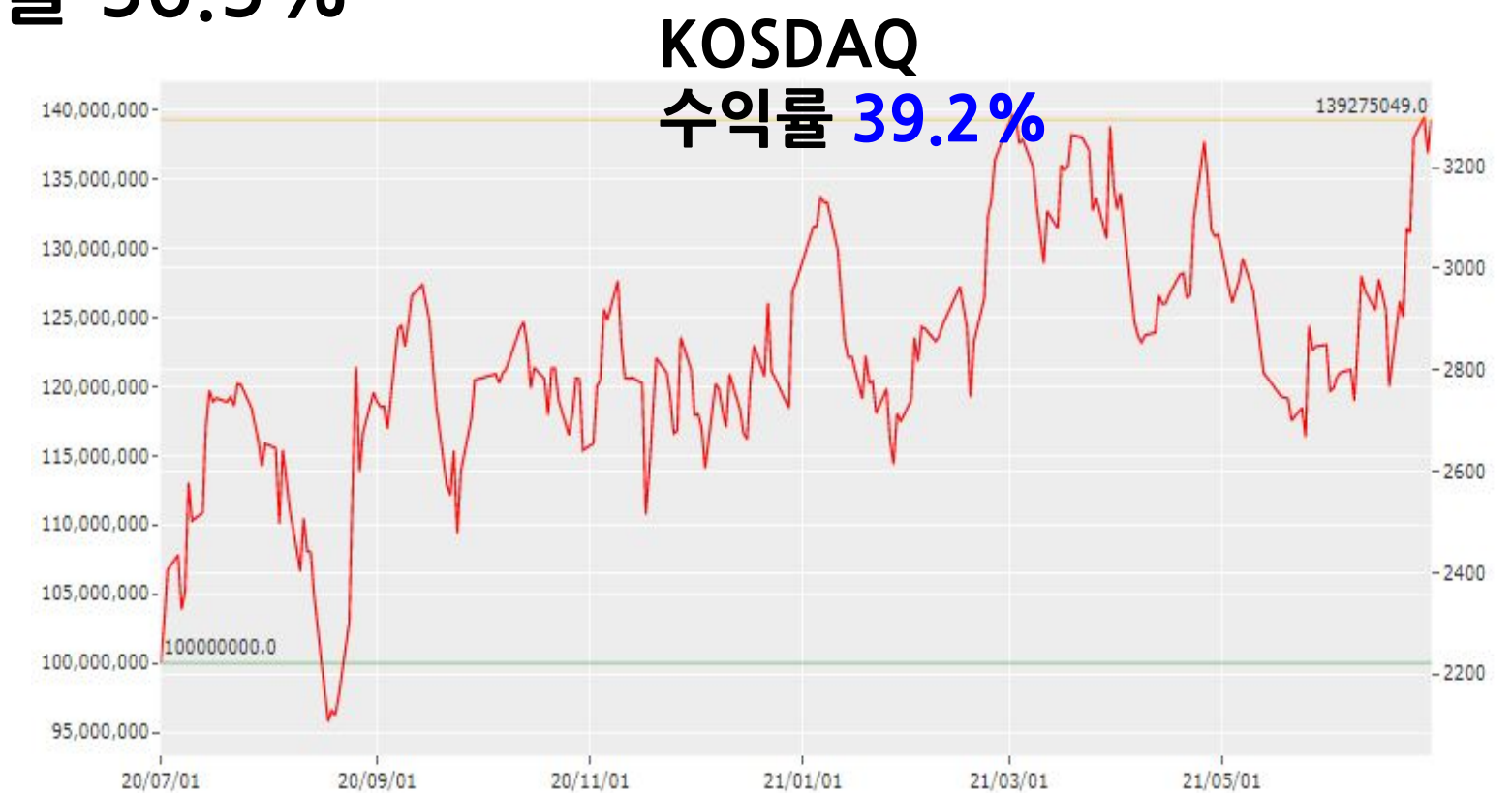
	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	4.1893	32.5101	5.7018	0.1788	0.9228	2.8829
1	4.1650	32.3088	5.6841	0.1690	0.9107	2.6893
2	4.1776	32.1302	5.6684	0.1590	0.9146	3.1445
3	4.2293	33.4137	5.7805	0.1810	0.9220	2.7639
4	4.2336	33.6368	5.7997	0.1826	0.9297	2.4705
Mean	4.1990	32.7999	5.7269	0.1741	0.9199	2.7902
SD	0.0276	0.6084	0.0530	0.0089	0.0067	0.2223



# 수익률 평가

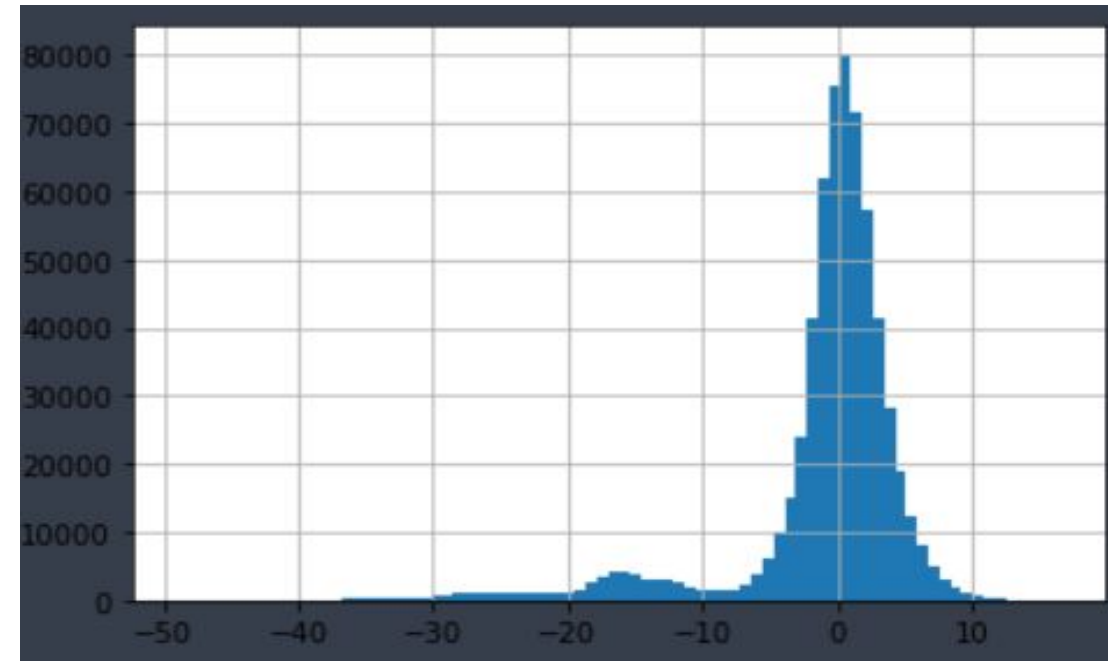


**코스피 지수 상승률 56.5%**



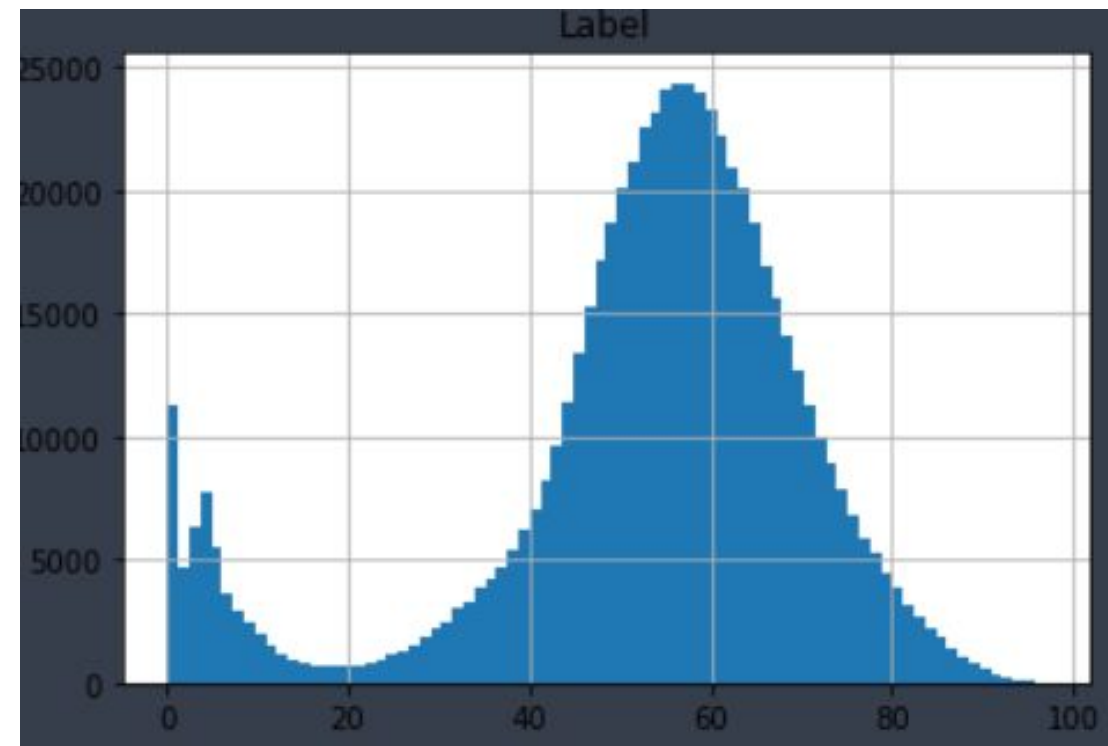
# 종합 점수화

예측 수익률(%)



count	624974
mean	-0.8572
std	6.0441
min	-48.9814
25%	-1.4905
50%	0.3401
75%	2.0860
max	16.5428

종합 점수



count	624974
mean	52.8569
std	18.6003
min	0.0066
25%	46.8373
50%	55.9581
75%	64.3055
max	97.0113

# 03

## Discussion

---

---

# Discussion

## 결론

주식데이터와 파생된 보조지표 및 타겟변수로부터  
미래 수익률을 예측하였으며 이를  
바탕으로 백테스트 진행시  
기대할 수 있는 수익성을 확인

## 기대효과

개인투자자들에게 종목 추천의  
기준이 될 수 있는 종합점수를  
제공하고 이를 자동매매 알고리즘에  
활용하여 높은 수익성을 기대

## 보완점

모멘텀, 볼린저 밴드 등의 다양한  
보조지표를 추가하여 학습에 사용  
  
딥러닝 기반 학습으로 모델 강화  
  
train, test 기간의 확대

# 04

## 시연

---

---