

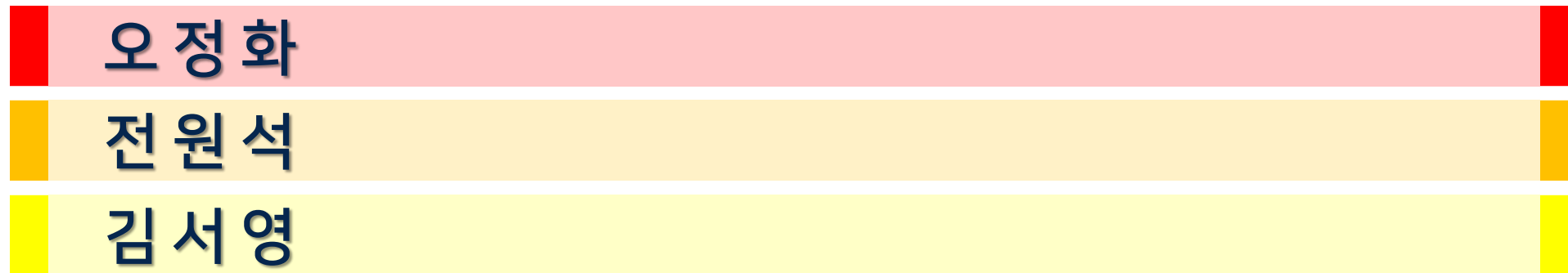


RAINBOW COMPANY

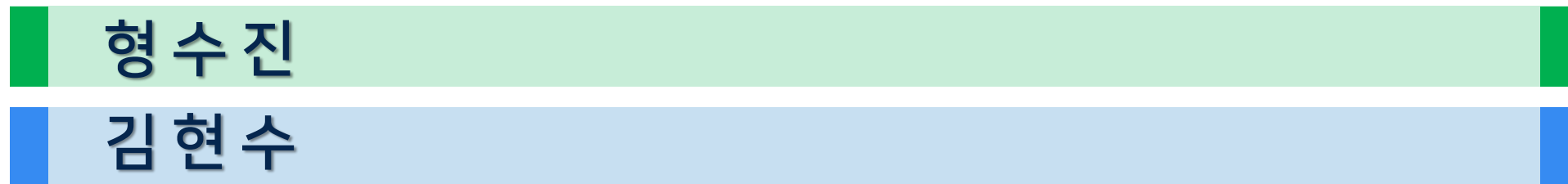
infinite of possibilities

Part 1 팀 원 소 개

데 이 터 전 처 리



강 화 학 습



웹

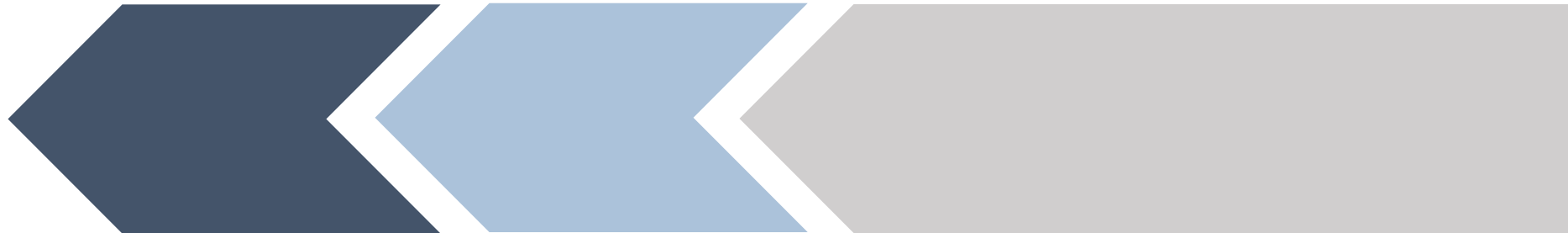


기 획 및 총 괄



Contents

- 1 개요 Introduction
- 2 수집 및 전처리 Data collection & Preprocessing
- 3 강화 학습 모델링 Reinforcement Learning Model
- 4 웹 구축 Web Development
- 5 개선 사항 및 결론 Room for improvement & Conclusion



Part 1

개요

Introduction



Question

Targeting Stock Investment Trading



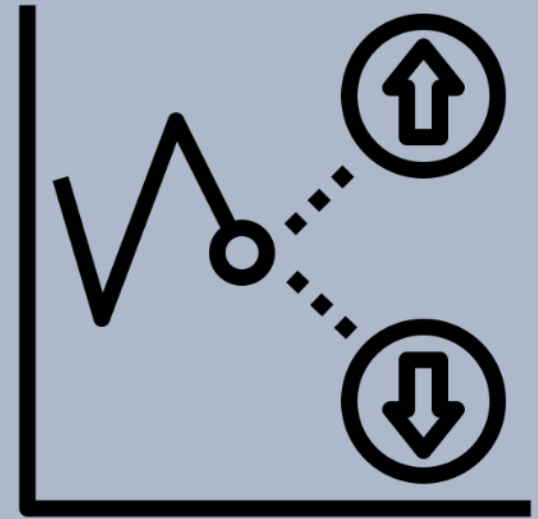
Part 1 개요 - 목적 및 방향성



고객 원하는 종목/투자금액



단기적 탐색



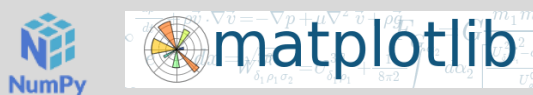
매수/매도 예측

Part1 개요 - 환경

수집



데이터분석



연동 시스템

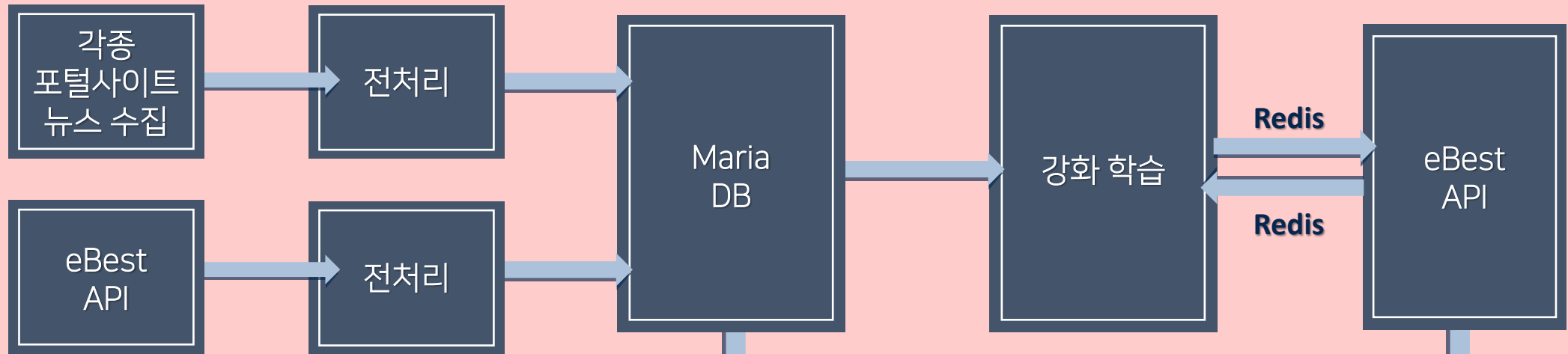


Web 구축



Part 1 개요 - 프로세스

<Linux Google VM Computer>



<Web>



Part 1 개요 - 개발 일정

* 2023년 07월																													
분류	항목			월		수		금			월		수		금			월		수		금							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
데이터수집	과거뉴스 (키워드 포함)					시작									완료														
	종목별 과거(2년치)	시작													완료														
	코피스 (2년치)					시작									완료														
	환율 (USD,CNY)						시작								완료														
	증권사 리포트					시작						완료																	
전처리	증권데이터 전처리(학습데이터로)	시작													완료														
	종목 데이터(분봉으로 병합)	시작													완료														
	종목 데이터(지표추가 및 기타 병합)								시작						완료														
	환율 데이터(USD,CNY)							시작							완료														
	코스피 데이터							시작							완료														
	chart(실시간)+sentiment(감성분석 결과)										시작				완료														
	수집 데이터 DB에 연결													시작			완료												
텍스트처리	텍스트모델링					시작								완료															
	데이터 전처리					시작								완료															
	모델 평가											시작		완료															
	DB와 연결										시작				완료														
증권사 API	매수																	시작			완료								
	매매																	시작			완료								
강화 학습	모델링														시작							1차			2차				
	실시간 매수매도 연결																	시작			완료								
웹구축	환경설정							시작		완료																			
	DB연동							시작																	완료				
	디자인								시작																완료				

Part 2

수집 및 전처리

Data collection & Preprocessing



Part 2 수집 및 전처리 - 과거 데이터 수집

- 수집 기간 : 2021년 6월 28일 ~ 2023년 7월 20일
- 수집 단위 : 틱 (Tick), 분 (minute)

데이터	수집 출처	수집 방법	특성
종목	대신증권 eBest증권	대신증권 CYBOS eBest XING	- 신세계, 신세계 푸드, 이마트, 신세계건설, 신세계 I&C, 현대건설, CJ제일제당
코스피	대신증권	대신증권 CYBOS	
환율	신한은행	Selenium	- 달러 (USD/KRW), 위안 (CNY/KRW)
종목 뉴스	'네이버 증권' 뉴스 검색	BeautifulSoup	- 종목 키워드 검색 (종목 명, 종목 관련 키워드, 트렌드 키워드)

Part 2 수집 및 전처리 - 과거 데이터 전처리

- Datetime 형식 : '%Y-%m-%d %H:%M:%S'
- TimeLine : 매일 09:01 ~ 15:20
- 종가 데이터 1분 씩 위로 당기기 (과거 분봉 종가는 1분씩 밀려서 제공됨)

구분		전처리 내용	Concat
종목	2021.06.28 ~ 2022.07.09	<ul style="list-style-type: none"> - 데이터 일별 제공 -> 모두 병합 - Tick data (주문 발생 건 데이터) -> 분봉 (minute) 데이터로 변환 - 신세계 I&C 매매거래 중지일 (22년 4월 6,7,8일) 전일 종가 데이터로 채움 	
	2021.07.10 ~ 2023.07.14	<ul style="list-style-type: none"> - 데이터 일별 제공 -> 모두 병합 	
코스피		<ul style="list-style-type: none"> - 데이터 일별 제공 -> 모두 병합 - 개장시간 1시간 지연(10시 개장) 일자 (수능일, 매년 첫 시장) TimeLine 맞춤 	
환율	달러 / 환율	<ul style="list-style-type: none"> - 데이터 일별 제공 -> 모두 병합 - Tick data (주문 발생 건 데이터) -> 분봉 (minute) 데이터로 변환 - 폐장일 (21년 12월31일) 데이터 Drop 	

Part 2 수집 및 전처리 - 과거 데이터 전처리

- 지표 컬럼 추가 : 모든 컬럼의 단위를 맞추고자 모두 비율로 변환

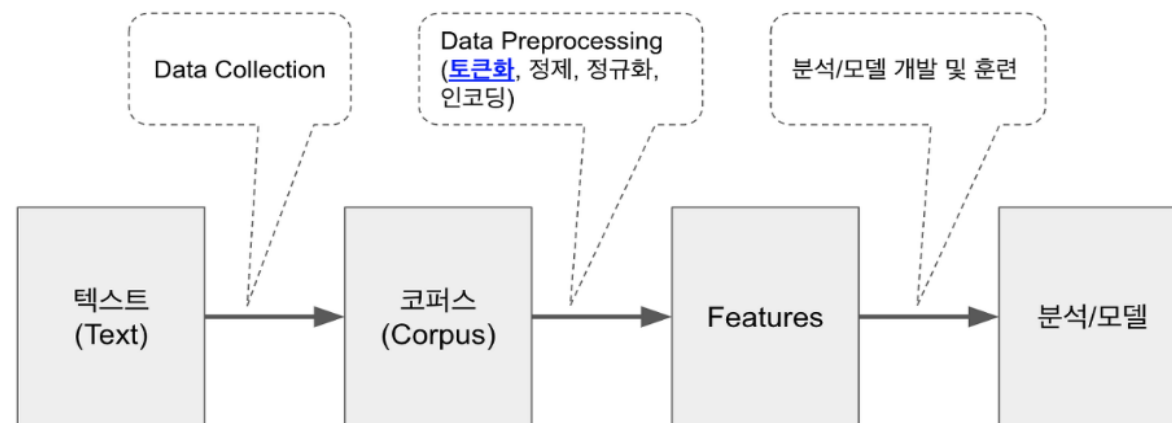
구분	추가 지표	지표 설명
종목	Ratio_close $(\text{close} - \text{last close}) / \text{last close}$	직전 종가 대비 현재 종가
	Ratio_high_close $(\text{high} - \text{last close}) / \text{last close}$	직전 종가 대비 현재 고가
	Ratio_low_close $(\text{low} - \text{last close}) / \text{last close}$	직전 종가 대비 현재 저가
	Ratio_open_close $(\text{open} - \text{last close}) / \text{last close}$	직전 종가 대비 현재 시가
	Ratio_volume $(\text{volume} - \text{last volume}) / \text{last volume}$	직전 거래량 대비 현재 거래량
	Ratio_MA20_close	직전 20분 동안의 가격을 산술평균한 값
	Ratio_MA20_volume	직전 20분 동안의 거래량을 산술평균한 값
	Sign(label)	직전가 대비 현재가 상승 (1) / 하락(-1) / 보합(0)
코스피 / 환율	Change $(\text{close} - \text{last close}) / \text{last close}$	직전가 대비 현재가
	Ratio_MA20_close	직전 20분 동안의 가격을 산술평균한 값

자연어 Model

KR-FinBERT 을 활용하여 감성분석을 진행하였습니다.

KR-FinBERT 는 BERT 모듈 중 한국 금융 분야에 특화된 모듈로 텍스트에 대하여 토큰화 와 불용어 제거 과정을 수행할 수 있습니다.

전처리된 텍스트가 입력되면 긍정, 중립, 부정의 부문에서 각자 어느 정도의 비율을 내포하는지 환산해줍니다.



[자연어 처리 과정 토큰화의 위치]

```
# 토큰라이저 지정
tokenizer = AutoTokenizer.from_pretrained("snunlp/KR-FinBERT" return_tensors='pt')

Downloading (...)okenizer_config.json: 100% 336/336 [00:00<00:00, 26.8kB/s]
Downloading (...)solve/main/vocab.txt: 100% 143k/143k [00:00<00:00, 414kB/s]
Downloading (...)main/tokenizer.json: 100% 294k/294k [00:00<00:00, 574kB/s]
Downloading (...)cial_tokens_map.json: 100% 112/112 [00:00<00:00, 8.33kB/s]
```

[KR-finBERT 활용]

Part2 수집 및 전처리 - 뉴스 데이터셋 전처리

Pipeline 를 통해 "토큰화 -> KR-FinBERT -> 결과 전송" 과정을 연결

(긍정,부정 비율 환산처리를 단계별로 동시에 수행하여 예측 값 추출)

```
models = AutoModelForSequenceClassification.from_pretrained("snunlp/KR-FinBERT")
classifier= pipeline("sentiment-analysis",tokenizer=tokenizer,model=models,device=0)

results=[]
label_=[0]

for text in data_kr:
    result = classifier(text)
    results.append(result)

# label_에 긍,부정과 그에 대한 scores 확률이 담김
for i, result in enumerate(results):
    for r in result:
        label_.append(r['label']) # label로 긍정과 부정만 추출
data['predict']=label_ # predict 컬럼에 담기

data['predict']=data['predict'].map({'LABEL_0':1,'LABEL_1':2})
```

[KR-finBERT 활용 감성분석]

ABC NewsDate	ABC NewsTitle
2022-12-08 06:00:00	¶"2030 눈길 쏠리네" 신세계百貨, 신진 디자이너 모시기 열전¶
2022-08-10 14:17:00	¶"2분기·상반기 사상 최대 실적" 신세계, 온·오프 고른성장 빛¶
2022-08-10 14:17:00	¶"2분기·상반기 사상 최대 실적" 신세계, 온·오프 고른성장 빛¶
2022-11-10 14:46:00	¶"2천만원대 카라반 쏜다"...신세계, 랜더스 우승 기념 전사 이¶
2023-05-01 09:42:00	¶"5월엔 대전신세계 아트앤사이언스서 신나게 놀아볼까?"¶
2022-08-11 06:00:00	¶"60만병 와인, 최대 85% 할인" 신세계百貨, 와인 할인행사¶
2023-01-12 13:43:00	¶"6병 한정" 신세계백화점, 발베니 42년산 선보여¶
2023-04-27 09:10:00	¶"75종의 도어 옵션"...신세계까사, 불박이장 신제품 출시¶
2023-02-06 08:53:00	¶"7일 자정부터" 신세계라이브쇼핑, 갤럭시 S23 예약판매¶
2022-08-03 09:13:00	¶"80% 할인부터 환율 보상까지" 신세계免, 대규모 행사¶

[DB 뉴스데이터셋]

Part 2 수집 및 전처리 - 뉴스데이터셋 전처리

자연어 Model 감성분석결과

```
data['predict']=data['predict'].map({'LABEL_0':1,'LABEL_1':2}) # LABEL_0은 긍정, LABEL_1은 부정이므로 1,2로 긍정,부정을 나눔
```

kor_sentence	predict
국제 전자산업 회사인 엘코텍은 탈린 공장에서 수십 명의 직원을 해고했으며, 이전의 ...	2
새로운 생산공장으로 인해 회사는 예상되는 수요 증가를 충족시킬 수 있는 능력을 증가...	1
2009-2012년 회사의 업데이트된 전략에 따르면, Basware는 20% - 4...	2
ASPOCOMP의 성장기에 대한 자금 조달은 기술적으로 더 까다로운 HDI 인쇄 회...	2
2010년 4분기 Componenta의 순매출은 전년 동기의 7600만 유로에서 2...	1
...	...
헬싱키 톰슨 파이낸셜 - 카고텍의 주가는 화물 취급 그룹이 4월부터 6월까지 3개월...	1
런던 마켓워치 -- 은행주의 반등이 FTSE 100지수의 약세를 상쇄하지 못하면서 ...	1
영업이익은 2007년 68.8 mn에서 35.4 mn으로 떨어졌으며, 선박 판매 이...	2
페이퍼 부문 순매출은 2008년 2분기 241.1 mn에서 2009년 2분기 221...	1
핀란드에서의 판매는 1월에 10.5% 감소한 반면, 국외에서의 판매는 17% 감소했다.	2

Part 2 수집 및 전처리 - 뉴스데이터셋 전처리

자연어 Model 검증

[금융뉴스 기사 제목 데이터셋 활용 검증]

* 종목과 관계없는 데이터를 통한 검증진행

```
20 TP = 0
21 TN = 0
22 FP = 0
23 FN = 0
24
25 dl = data['labels']
26 dp = data['predict']
27
28 for idx, da in enumerate(dl):
29     if (dp[idx] == 1) and (dl[idx] == 1):
30         TP += 1
31     elif (dp[idx] == 2) and (dl[idx] == 2):
32         TN += 1
33     elif (dp[idx] == 1) and (dl[idx] == 2):
34         FP += 1
35     elif (dp[idx] == 2) and (dl[idx] == 1):
36         FN += 1
37
38 precision = TP / (TP + FP)
39 recall = TP / (TP + FN)
40 accuracy = (TP + TN) / (TP + TN + FP + FN)
41 f1_score = 2 * (precision + recall) / (precision + recall)
42
43 print('정밀도(Precision): ', precision)
44 print('재현율(Recall): ', recall)
45 print('F1 score: ', f1_score)
46 print('정확도(Accuracy): ', accuracy)
47
```

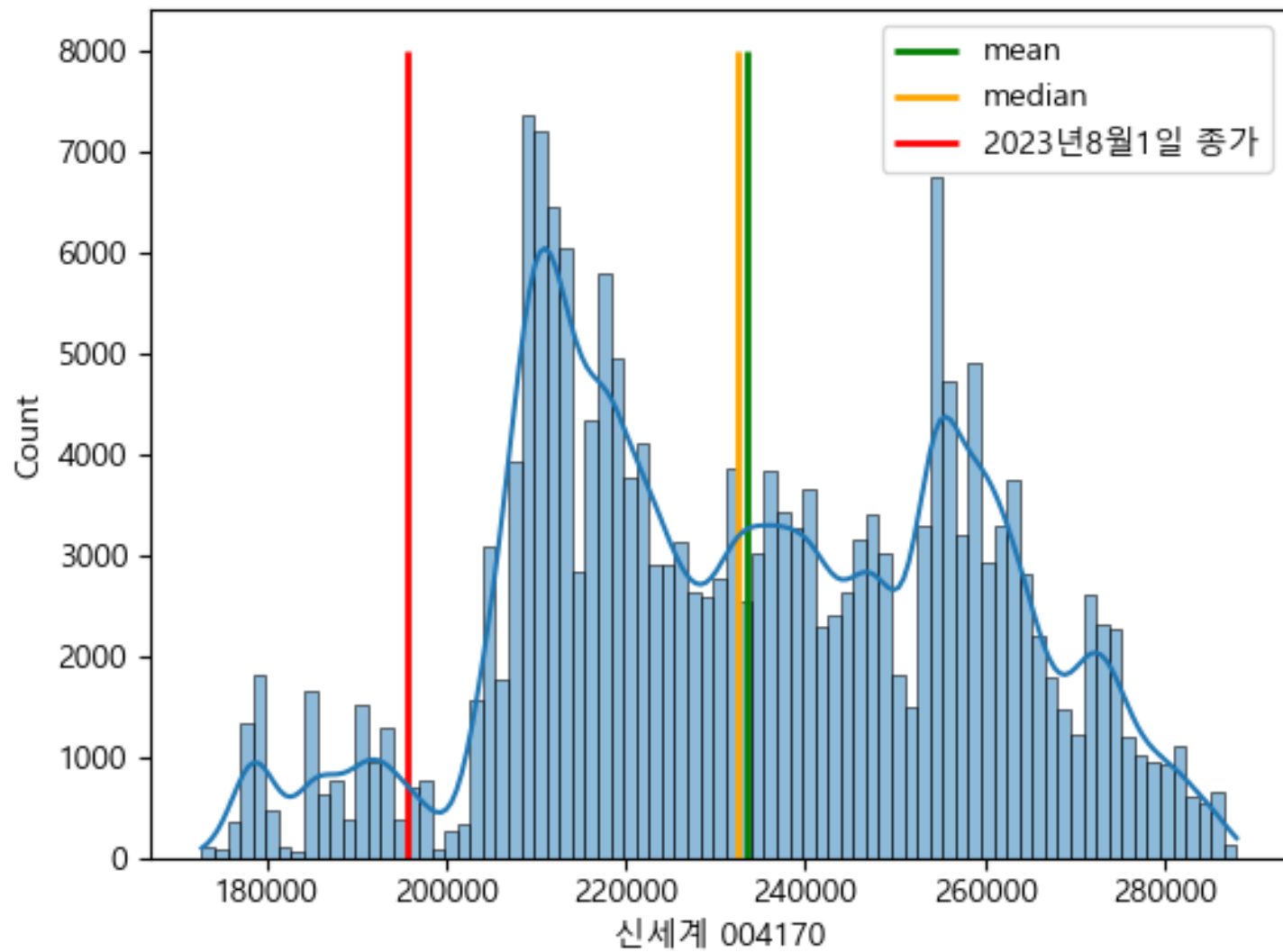
정밀도(Precision): 0.7010036979341257
재현율(Recall): 0.9743024963289281
F1 score: 0.8153609831029186
정확도(Accuracy): 0.7243031536113937

[종목 뉴스기사 제목 데이터셋 활용 검증]

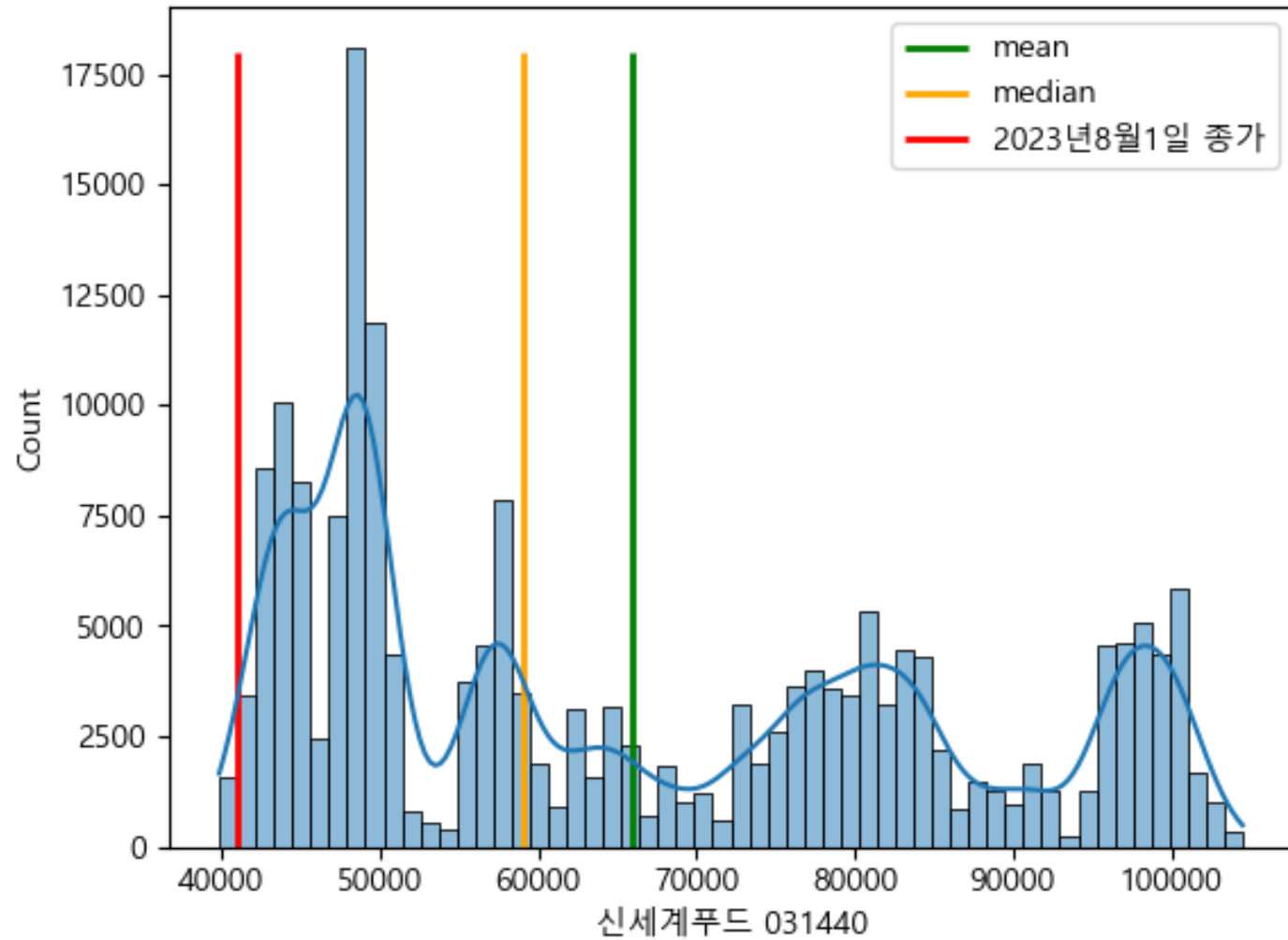
```
1 TP = 0
2 TN = 0
3 FP = 0
4 FN = 0
5
6 sent=shin['sentiment']
7 pred=shin['predict']
8 for idx, da in enumerate(sent):
9     if (pred[idx] == 1) and (sent[idx] == 1):
10         TP += 1
11     elif (pred[idx] == 2) and (sent[idx] == 2):
12         TN += 1
13     elif (pred[idx] == 1) and (sent[idx] == 2):
14         FP += 1
15     elif (pred[idx] == 2) and (sent[idx] == 1):
16         FN += 1
17
18 precision = TP / (TP + FP)
19 recall = TP / (TP + FN)
20 accuracy = (TP + TN) / (TP + TN + FP + FN)
21 f1_score = 2 * (precision + recall) / (precision + recall)
22
23 print('정밀도(Precision): ', precision)
24 print('재현율(Recall): ', recall)
25 print('F1 score: ', f1_score)
26 print('정확도(Accuracy): ', accuracy)
27
```

정밀도(Precision): 0.8189245887159103
재현율(Recall): 0.8779592803030303
F1 score: 0.8474150242787775
정확도(Accuracy): 0.7958610846812559

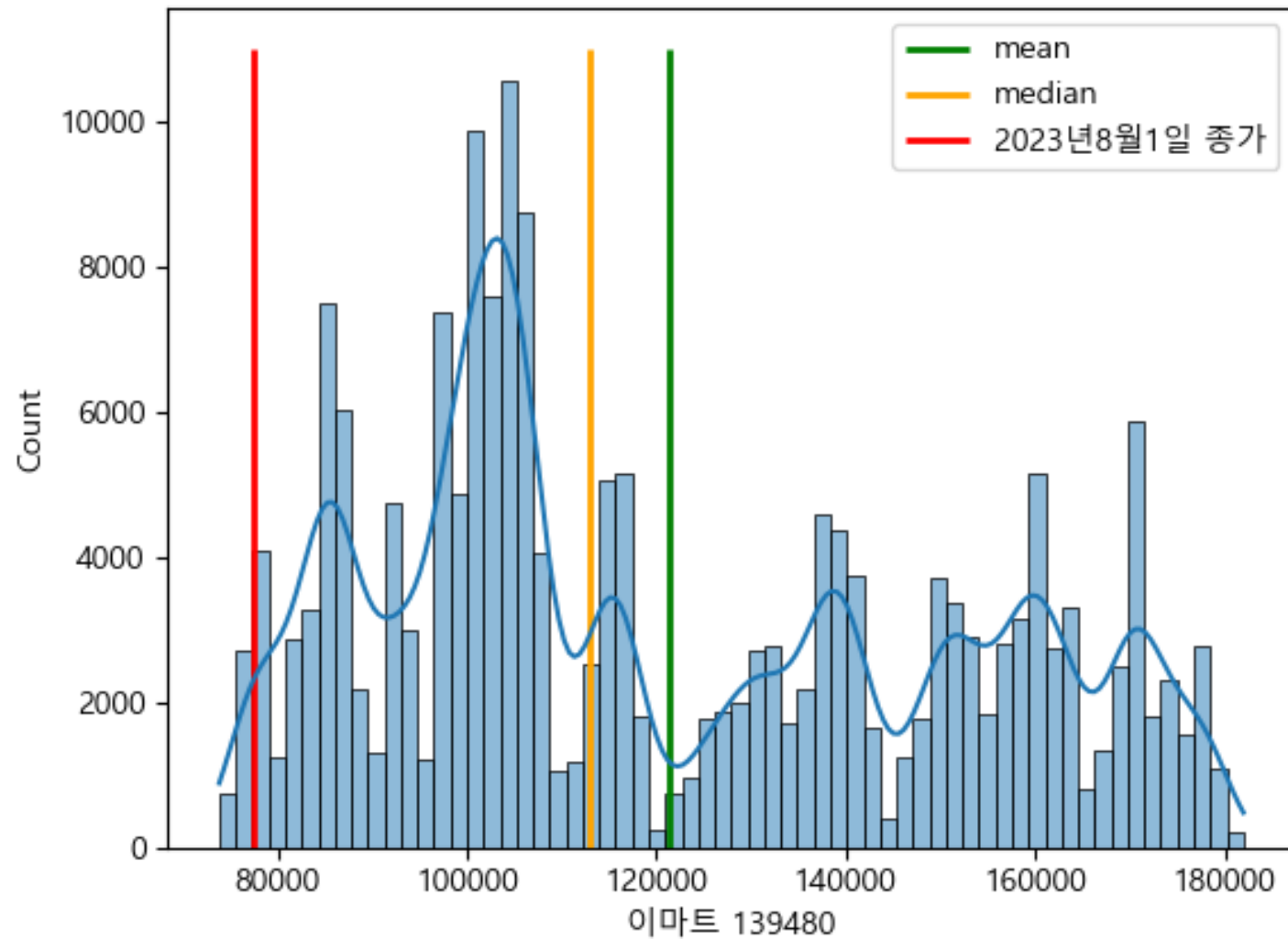
Part 2 수집 및 전처리 - 분석 시각화



Part 2 수집 및 전처리 - 분석 시각화



Part 2 수집 및 전처리 - 분석 시각화



Part 2 수집 및 전처리 - 분석 시각화



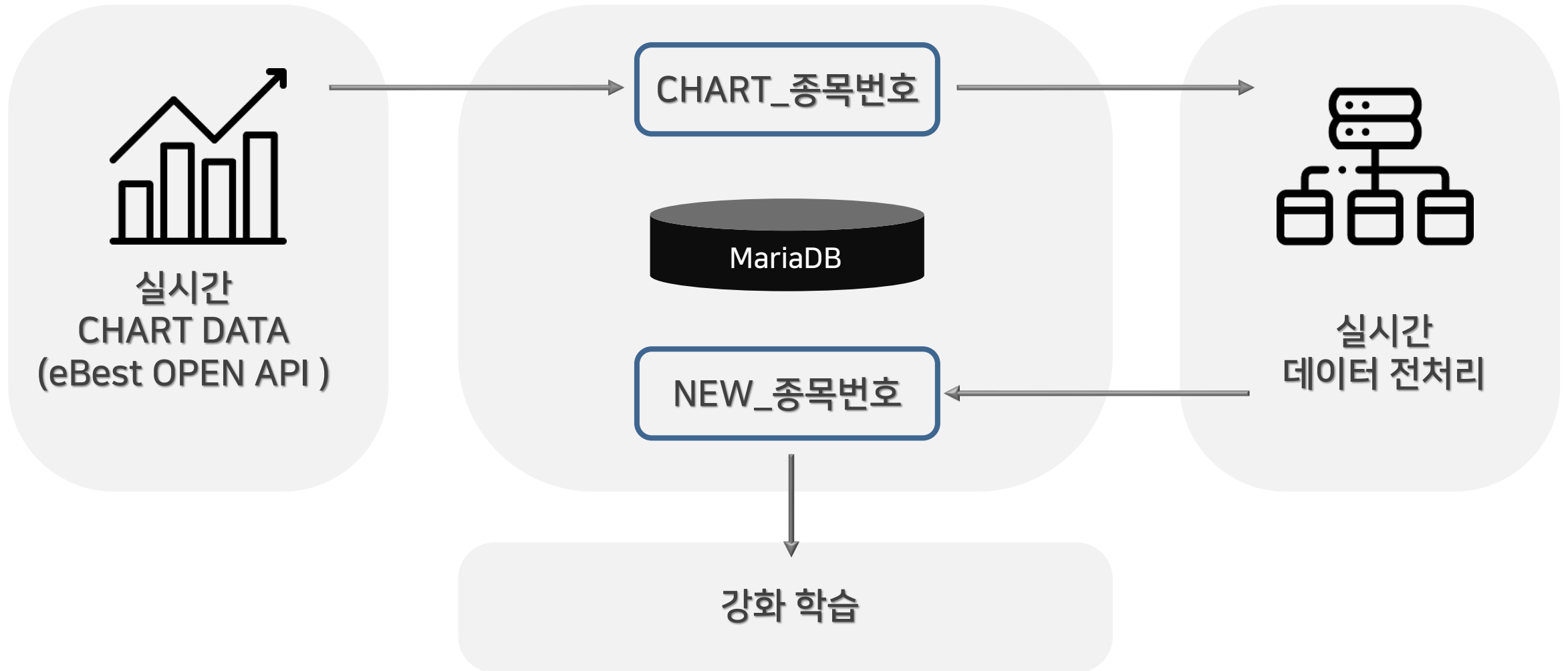
Part 2 수집 및 전처리 - 실시간 데이터 수집

- 수집 기간 : 2023년 7월 21일 ~

- 수집 단위 : 분 (minute)

데이터	수집 출처	수집 방법	저장 방법
종목	이베스트증권	eBest OPEN API	- Chart (Open High Low Close Volme) => DB : CHART_종목번호, Redis : 시간, 현재가(Close)
코스피	네이버 증권	BeautifulSoup	- DB : kospi_crawling, Redis : 시간, 현재가
환율	네이버 증권	BeautifulSoup	- DB : exchange_crawling, Redis : 시간, 현재가
종목 뉴스	'네이버 증권' 뉴스 검색	BeautifulSoup	- DB : shin/shinFood/eMart

Part 2 수집 및 전처리 - 실시간 데이터 전처리



Part 3

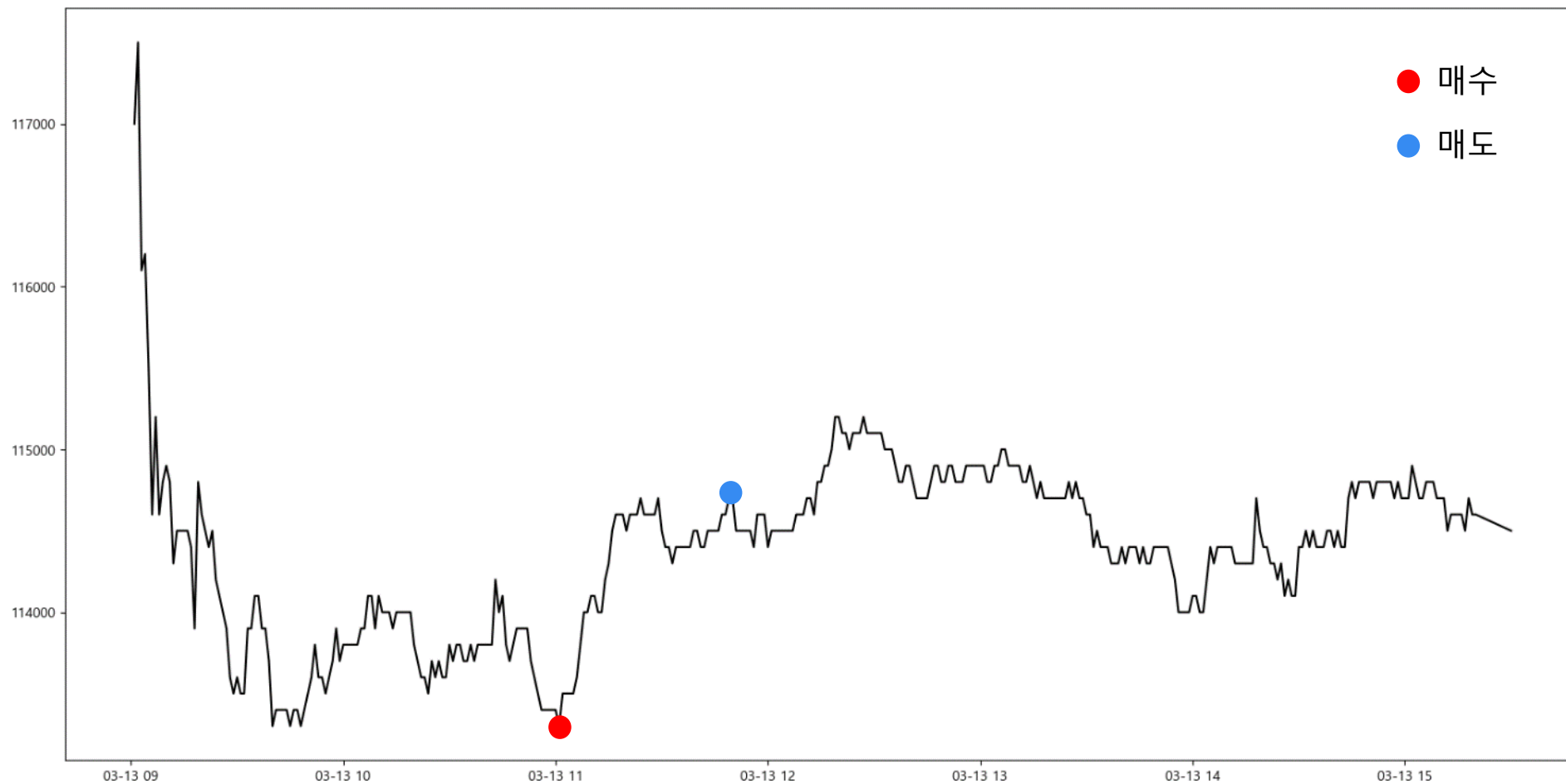
강화 학습 모델링

Reinforcement Learning Modeling



Part 3 강화 학습 모델링 - 강화 학습이란?

Part 3 강화 학습 모델링 - 매매 내역



Part 4

결론 (웹 구축)

Wep Development



Part4 웹 구축 - 회원가입

회원 가입

프로필 이미지

파일 선택 선택된 파일 없음

이름

아이디

비밀번호

비밀번호 표시: ☐

가입하기

이미 계정이 있으십니까? [로그인하기](#)

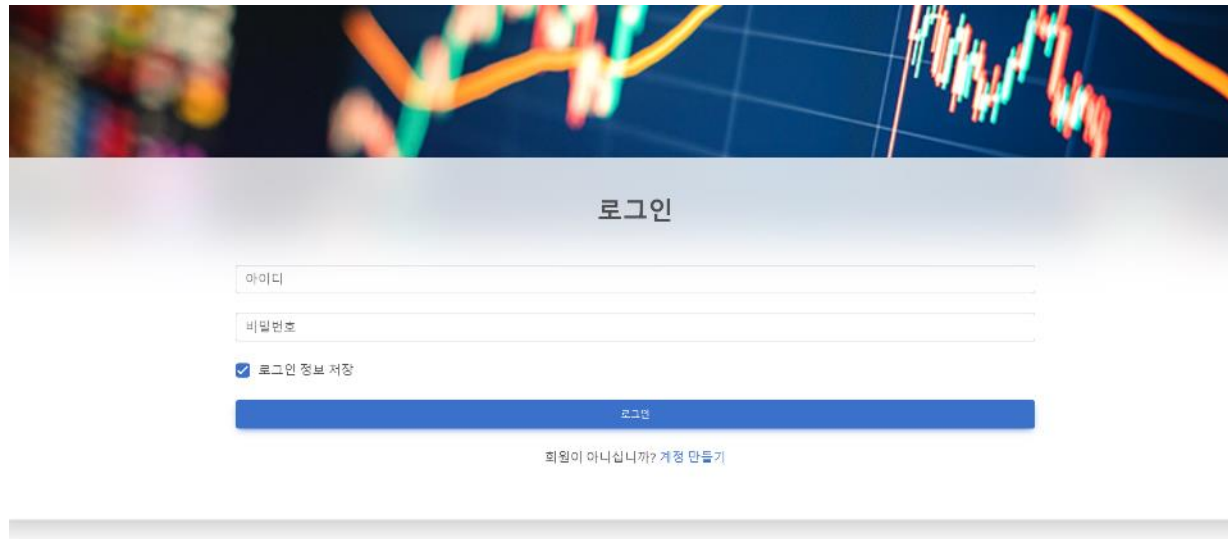
[회원 가입 화면]

컬럼명	#	Data Type	Not Null	Auto Increment	Key	디폴트
username	1	varchar(100)	[v]	[]		
imgnm	2	varchar(2000)	[]	[]		NULL
userid	3	varchar(500)	[v]	[]	PRI	
userpasswo...	4	varchar(500)	[v]	[]		

username	imgnm	userid	userpassword
유저	/images/사람.png	user	jeonchan

[DB에 저장된 정보 화면]

Part4 웹 구축 - 로그인



로그인

아이디

비밀번호

☒ 로그인 정보 저장

로그인

회원이 아니십니까? 계정 만들기

[로그인 화면]

localhost:8080 내용:
로그인에 성공하셨습니다.

확인

localhost:8080 내용:
로그인에 실패하셨습니다. 다시 로그인해주세요

확인

[로그인 성공 및 실패 확인 메시지 화면]

Part4 웹 구축 - 로그인

```
public void doFilter(ServletRequest request, ServletResponse response, FilterChain chain) throws IOException, ServletException {  
    HttpServletRequest req = (HttpServletRequest) request;  
    HttpServletResponse res = (HttpServletResponse) response;  
    HttpSession session = req.getSession();  
    String logincheck = (String) session.getAttribute("userid");  
    if(logincheck == null) {  
        res.sendRedirect("/rainbowcompany/main");  
    }else {  
        }  
    chain.doFilter(request, response);  
}
```

[Filter를 통해 로그인 회원만 페이지 이동 가능 코드 화면]

```
if (check == 1) {  
    HttpSession session = request.getSession();  
    session.setAttribute("userid", userid);  
    session.setAttribute("userpassword", userpassword);  
    red.addFlashAttribute("msg", "<script>alert('로그인에 성공하셨습니다.');} else {  
    red.addFlashAttribute("msg", "<script>alert('로그인에 실패하셨습니다. 다시 로그인해주세요');    return "redirect:/rainbowcompany/signin";  
}
```

[Session에 로그인 정보 저장 코드 화면]

Part4 웹 구축 - Websocket

```
@Configuration
@EnableWebSocketMessageBroker
public class WebSocketConfig implements WebSocketMessageBrokerConfigurer {

    @Override
    public void configureMessageBroker(MessageBrokerRegistry config) {
        config.enableSimpleBroker("/stock");
        config.setApplicationDestinationPrefixes("/app");
    }

    @Override
    public void registerStompEndpoints(StompEndpointRegistry registry) {
        registry.addEndpoint("/websocket").withSockJS();
    }
}
```

[Websocket 연결 코드 화면]

```
Opening Web Socket... stomp.min.js:8
Web Socket Opened... stomp.min.js:8
>>> CONNECT stomp.min.js:8
accept-version:1.1,1.0
heart-beat:10000,10000

<<< stomp.min.js:8
CONNECTED
version:1.1
heart-beat:0,0
```

[Web에서 개발자 도구를 통한 연결 화면]

Part4 웹 구축 - Redis

```
@Component
public class RedisMessageListener implements MessageListener {

    private final SimpMessagingTemplate messagingTemplate;

    @Autowired
    public RedisMessageListener(SimpMessagingTemplate messagingTemplate) {
        this.messagingTemplate = messagingTemplate;
    }

    @Override
    public void onMessage(Message message, byte[] pattern) {
        String topic = new String(message.getChannel());
        String value = new String(message.getBody());
        messagingTemplate.convertAndSend("/stock/" + topic, value);
    }
}
```

[Redis 실시간 연결 코드]

```
CONNECTING...

2667.07

<<< MESSAGE stomp.min.js:8
destination:/stock/exchangeu
sd
content-
type:text/plain;charset=UTF-
8
subscription:sub-2
message-id:pndve4pi-1142
content-length:6

1290.0

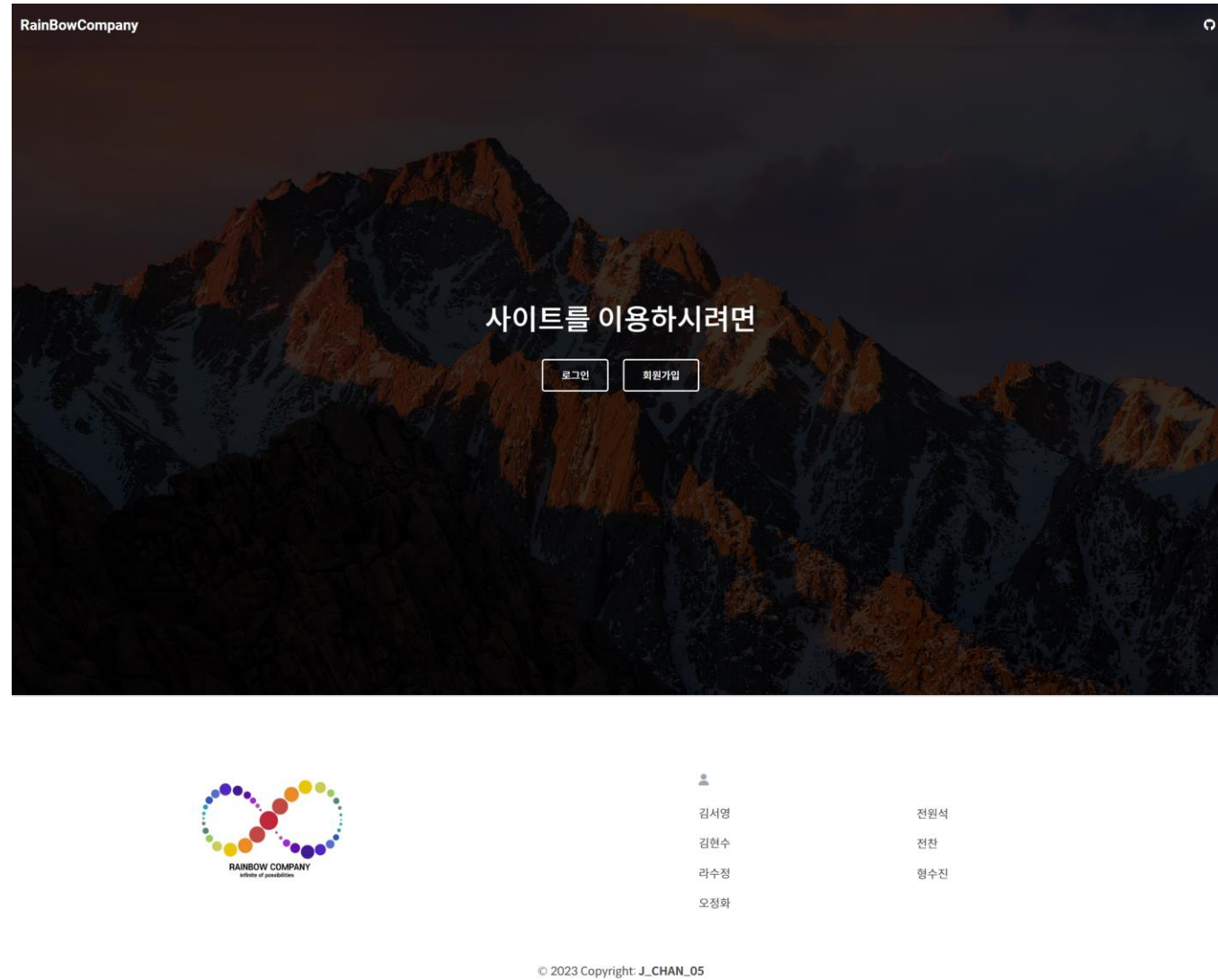
<<< MESSAGE stomp.min.js:8
destination:/stock/exchangec
ny
content-
type:text/plain;charset=UTF-
8
subscription:sub-3
message-id:pndve4pi-1143
content-length:5

179.7

>
```

[Web에서 개발자 도구를 통한 Redis 연결 화면]

Part4 웹 구축 - 홈페이지 시연



Part 5

개선사항 및 결론

Room for improvement & Conclusion



Part5 개선 사항 및 결론

001 >> 시간 부족

프로젝트에 주제 선정에서 많은 시간을 소모하여 수집 및 개발 시간이 부족.
증시 거래시간 제약 조건으로 인한 코드 검증 시간 부족.

002 >> 개념 학습

강화학습 개념에 대한 이해시간이 부족하여 개발하는데 어려움 겪음.
이러한 경험을 토대로 공부와 개발을 병행하여 완성도 있는 프로그램 개발을 위한 노력을 하고자 함.

003 >> 부가 기능 구현

고객의 편의성을 고려한 부가적인 기능(* 수익률 실시간 알림 기능)을 시간 부족으로 구현하지 못함.
향후엔 모바일 애플리케이션을 기획하고 개발하여 고객의 편의성을 개선하고자 함.