

Invest Genius

LLM(LLaMA3, ChatGPT3.5) 기반 챗봇을 활용한 포트폴리오 추천 서비스

기업멘토(신한AI/전주형)

팀장: 이성도

팀원: 황경상, 기현수, 이승준, 박건우

Content

1. 제안배경

- 제안배경
- 필요성
- 방향성

2. 제품소개

- 개발목표
- 개발내용
- 주요기능

3. 설계 및 구현

- ER 다이어그램
- 사용언어 및 도구
- 시스템아키텍처

4. 트러블슈팅

- 트러블슈팅
- 시연영상

Content

5. 기대효과 및 활용방안

6. 참고문헌

01

제안배경

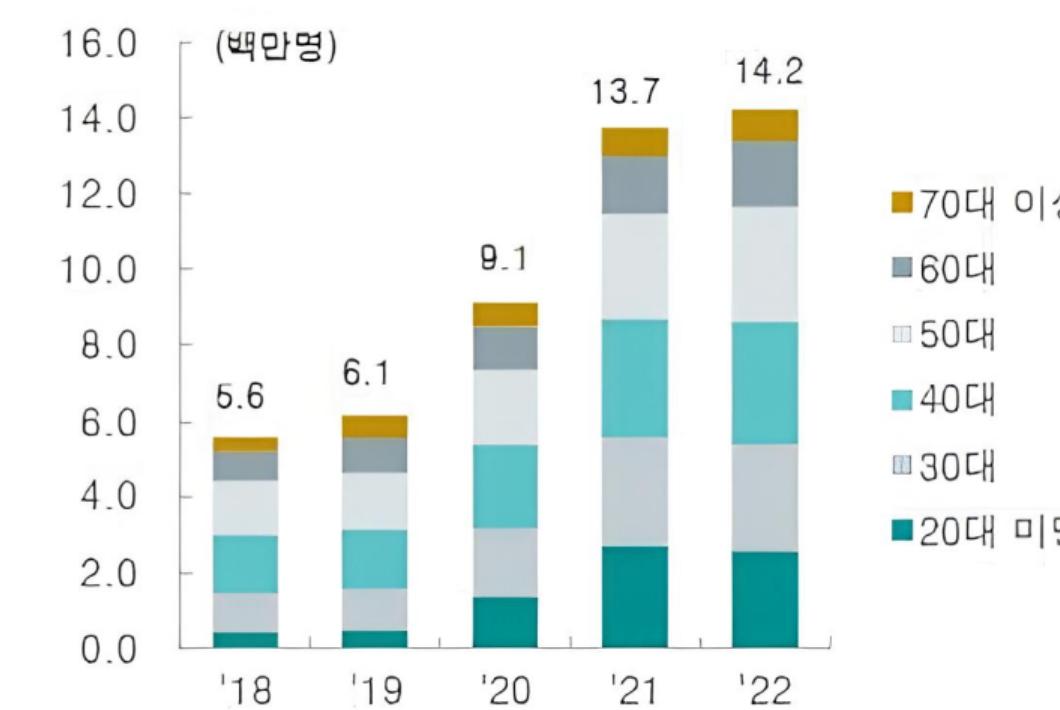
제안배경

필요성

방향성

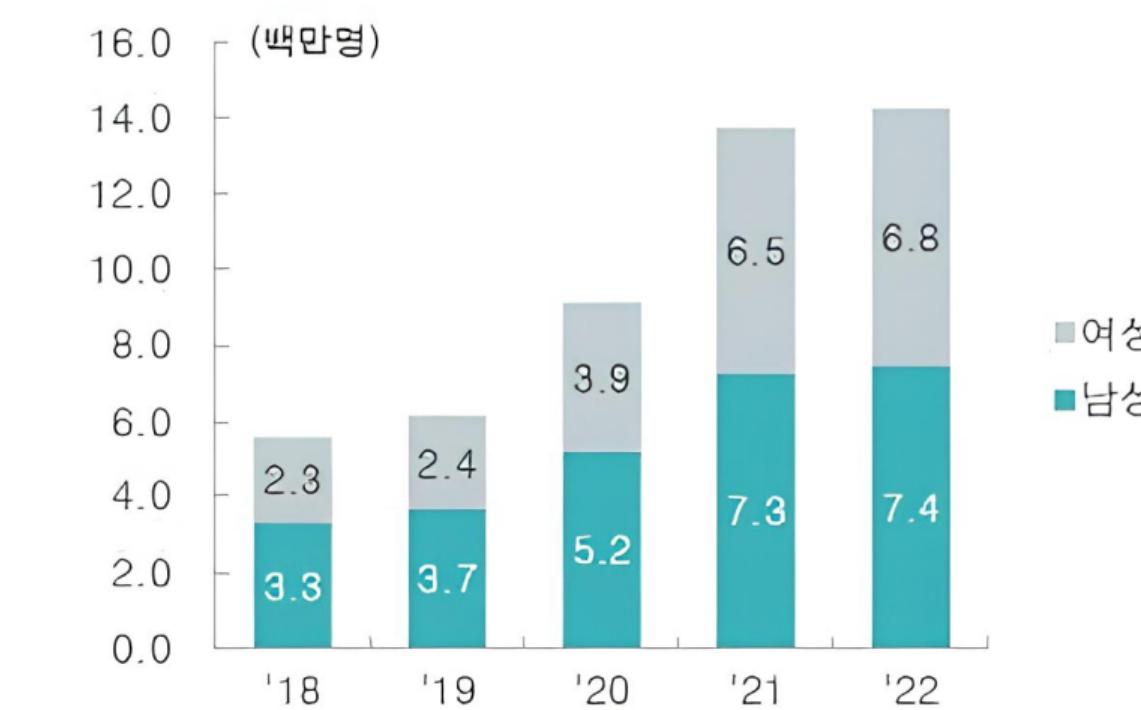
제안배경

그림4 | 연령별 상장법인 개인소유자(주주) 현황



주 : 각 연도말 시점의 소유자 내역 자료 기준
자료 : 한국예탁결제원

그림5 | 성별 상장법인 개인소유자(주주) 현황



주 : 각 연도말 시점의 소유자 내역 자료 기준
자료 : 한국예탁결제원

<연도별 30세 미만 투자자 동향>

“ 개인투자자 502만 명에서 1424만 명 약 3배 증가
이로인한, 온라인과 모바일에서의 주식거래량 급격한 증가 ”

“ 고도의 전문지식과 수준높은 이해를 요구하는 전문분야
이로인해 쉽게 피로감을 느끼는 초보 투자자들 ”

방향성

1. 대화형 AI챗봇을 활용한 개인 맞춤형 투자상담 서비스

대화형 AI과의 대화를 통해 개인투자성향 분석 ➡ '공격', '중립', '방어형' !

2. 실시간 주식시장 변화를 반영한 투자전략 제시

개인투자자의 성향에 알맞는 실시간 주식시장 정보를 반영한 투자전략을 제시

02

제품소개

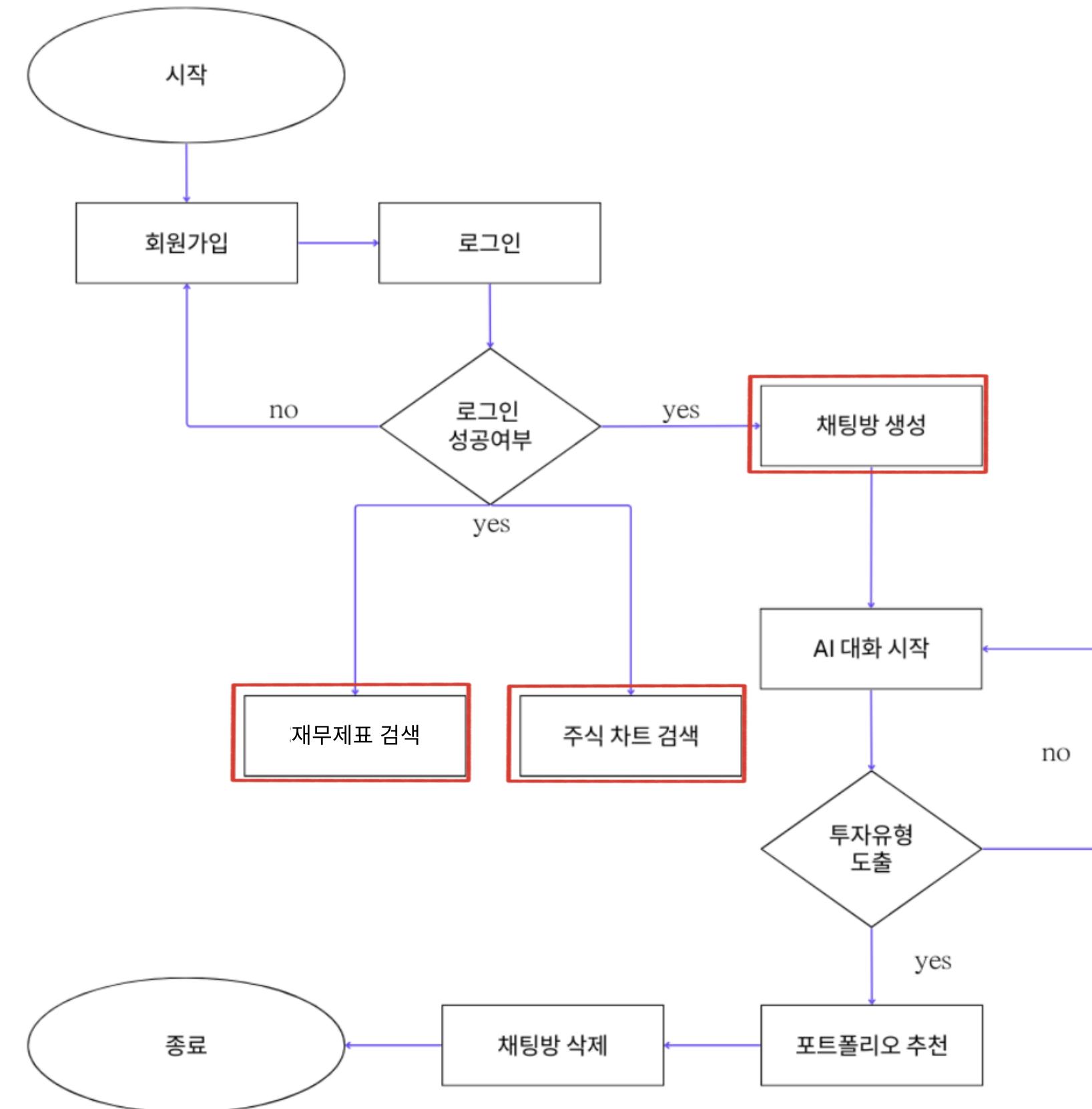
개발목표

개발내용

주요기능

개발목표

1. LLM(거대언어모델) 기반 투자자의 투자성향 분석
2. Yahoo Finance와 Finance Data Reader(FDR)API 활용한 실시간 주식데이터 확보
3. 투자자의 투자성향 맞춤형 포트폴리오(투자전략)제시



안녕하세요? 당신의 투자를 도와드리는 InvestGenius입니다. 어떻게 도와드릴까요?

투자유형을 분석받고 싶으시군요.
주로 어떤 자산에 투자하고 계시나요?

알겠습니다. 투자를 통해 안정적인 수익을 기대하시나요,
아니면 높은 수익을 기대하시나요?

알겠습니다.
당신은 중립형 투자자에 가깝습니다.

✓ 중립형 투자를 선호하시는 경우에는 HD현대일
렉트릭 10%, 한미반도체10%,
아모레퍼시픽 10&을 매수하는 것이며,
연평균 기대수익은 42.64% 입니다.

메시지를 입력하세요...

안녕?

내 투자유형을 분석받고 싶은데

잘 모르겠어. 그래서 그냥 한국 ETF에만 투자하고 있고,
너한테 추천을 받고싶은 상태야.

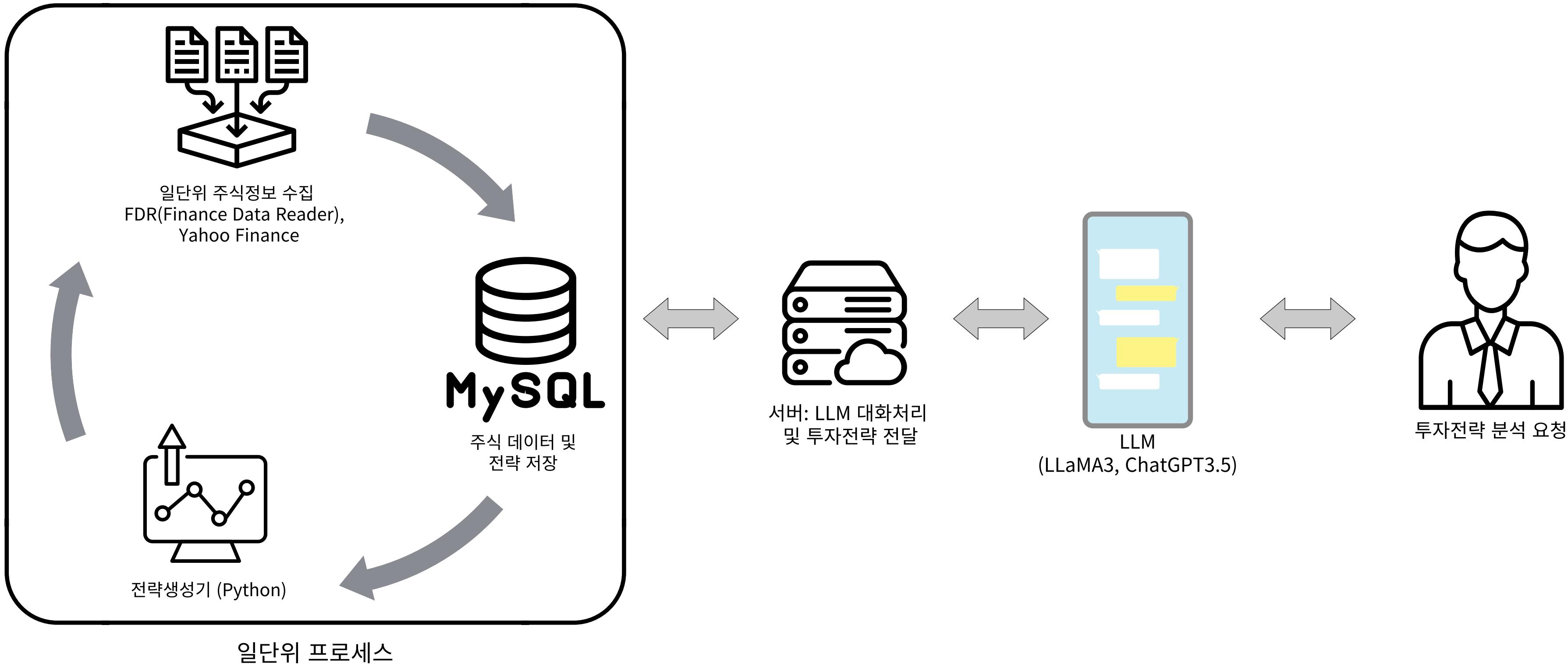
나는 적절한 수익을 추구해. 1년에 대략 7-10% 정도

그럼 포트폴리오 추천해줘

전송

투자전략 생성 흐름

개발내용

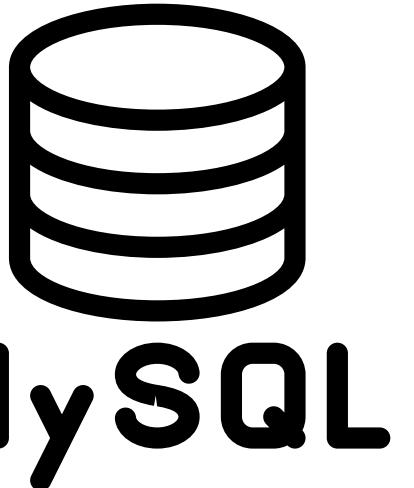


개발내용

2) 실시간 주식 데이터 수집 : 총 145만건 (07.30기준)

Yahoo Finance -> 미국 주식 50개									
AAPL	2024-07-16 00:00:00.000000	234.82	235.00	236.27	232.33	43234300	NASDAQ		
AAPL	2024-07-17 00:00:00.000000	228.88	229.45	231.46	226.64	57345900	NASDAQ		
AAPL	2024-07-18 00:00:00.000000	224.18	230.28	230.44	222.27	66034600	NASDAQ		
AAPL	2024-07-19 00:00:00.000000	224.24	224.22	226.20	222.20	49151500	NASDAQ		
AAPL	2024-07-22 00:00:00.000000	225.01	224.37	226.94	222.00	48201800	NASDAQ		
AAPL	2024-07-23 00:00:00.000000	225.01	224.37	226.94	222.00	39855700	NASDAQ		
FDR(Finance Data Reader) -> 한국 주식 50개									
11982	005930	2024-06-27 00:00:00.000000	81500.00	81900.00	81900.00	80800.00	9455929		
11983	005930	2024-06-30 00:00:00.000000	81800.00	81500.00	82100.00	81300.00	11317202		
11984	005930	2024-07-01 00:00:00.000000	81800.00	82500.00	82600.00	81500.00	14471904		
11985	005930	2024-07-02 00:00:00.000000	81800.00	82300.00	82300.00	81000.00	11440328		
11986	005930	2024-07-03 00:00:00.000000	81500.00	82000.00	81700.00	82400.00	43857229		
11987	005930	2024-07-04 00:00:00.000000	81500.00	82000.00	81700.00	82400.00	45791193		
11988	005930	2024-07-05 00:00:00.000000	81400.00	87900.00	88600.00	86900.00	24035809		

데이터 pipeline



주식 데이터 및
전략 저장

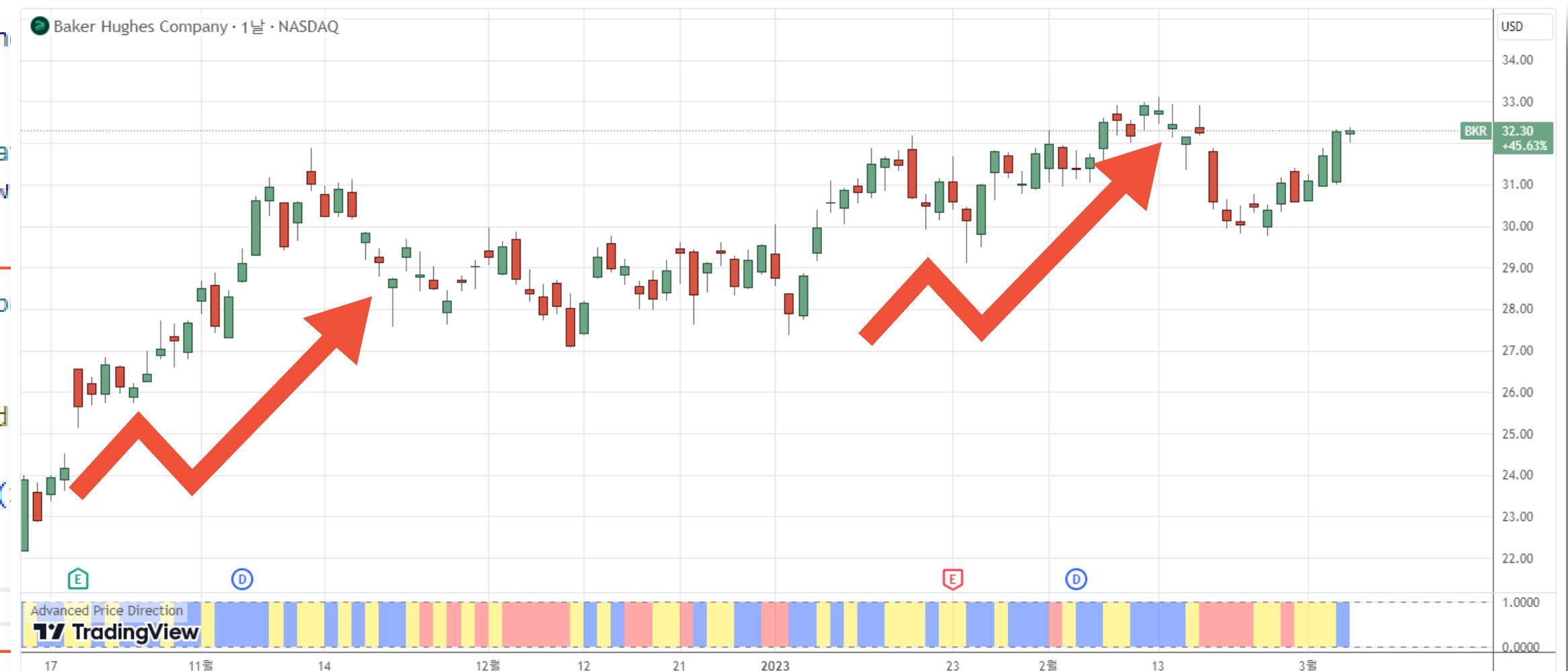
개발내용

2) 전략 : 모멘텀 전략

```
class MomentumAlgo(ValueBasedAlgo):
    def __init__(self, window: int = 20, **kwargs):
        super().__init__(indicator_type="momentum", window=window)
        self.window = window

    def prepare_data(self, data: pd.DataFrame) -> pd.DataFrame:
        smoothed_data = data.rolling(window=self.window).mean()
        return smoothed_data.dropna()

    def calculate_values(self, data: pd.DataFrame) -> pd.DataFrame:
        def momentum(ts: np.ndarray) -> float:
            ts_normalized = (ts - np.mean(ts)) / np.std(ts)
            x = np.arange(len(ts_normalized))
            slope, _, r_value, _, _ = stats.linregress(x, ts_normalized)
            annualized_slope = slope * np.sqrt(252)
            score = annualized_slope * (r_value**2)
            return score
        momentum_scores = data.apply(momentum, raw=True)
        momentum_scores = np.clip(
            momentum_scores,
            momentum_scores.mean() - 3 * momentum_scores.std(),
            momentum_scores.mean() + 3 * momentum_scores.std(),
        )
        return momentum_scores
```



모멘텀 전략

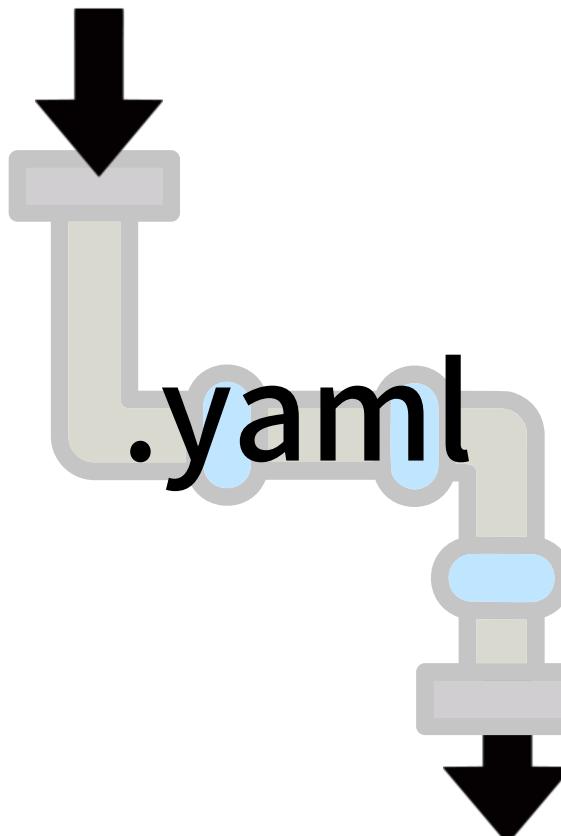
개발내용

2) 다양한 투자 전략 도출

```
1  strategy:
2    name: MomentumStrategy
3    module: "modules.strategy.strategy"
4    params:
5      train_period: 1200
6      valid_period: 20 # Holding periods
7      selection_method: "top_n" # "threshold", "relative", "top_n"
8      selection_param: 20 # top n
9      min_stocks: 5 # select stocks at least
10     max_stocks: 20 # select stocks maximum
11
12
13  algorithm:
14    name: MomentumAlgo
15    module: "modules.algo.momentum"
16    params:
17      window: 20
18
19  data_pipelines:
20    name: FinanceDataReader
21    module: "modules.data.fdr_korea"
22    base_path: "data/KOR"
23    interval: "1d"
24    start_date: "1970-01-01"
25    end_date: "TODAY"
26    stocks:
27      - symbol: "005930"
28        full_name: 삼성전자
29        exchange: KRX
30
31      - symbol: "000660"
32        full_name: SK하이닉스
33        exchange: KRX
34
35      - symbol: "373220"
36        full_name: LG에너지솔루션
37        exchange: KRX
38
39      - symbol: "005380"
40        full_name: 현대차
41        exchange: KRX
```

```
1  strategy:
2    name: MomentumStrategy
3    module: "modules.strategy.strategy"
4    params:
5      train_period: 30
6      valid_period: 5 # Holding periods
7      selection_method: "relative" # "threshold", "relative", "top_n"
8      min_stocks: 5 # select stocks at least
9      max_stocks: 20 # select stocks maximum
10
11  algorithm:
12    name: MomentumAlgo
13    module: "modules.algo.momentum"
14    params:
15      window: 20
16
17  data_pipelines:
18    name: YahooFinance
19    module: "modules.data.yahoo_finance"
20    base_path: "data/USA"
21    interval: "1d"
22    period: "max"
23    start_date: "1970-01-01"
24    end_date: "TODAY"
25    stocks:
26      - symbol: AAPL
27        full_name: Apple Inc.
28        exchange: NASDAQ
29
30      - symbol: MSFT
31        full_name: Microsoft Corporation
32        exchange: NASDAQ
33
34      - symbol: AMZN
35        full_name: Amazon.com Inc.
36        exchange: NASDAQ
37
38      - symbol: GOOGL
39        full_name: Alphabet Inc.
40        exchange: NASDAQ
```

미국주식 50개, 한국주식 50개
총 100개의 주식데이터 수집.관리



다양한 투자전략 조합 기능 구현

개발내용

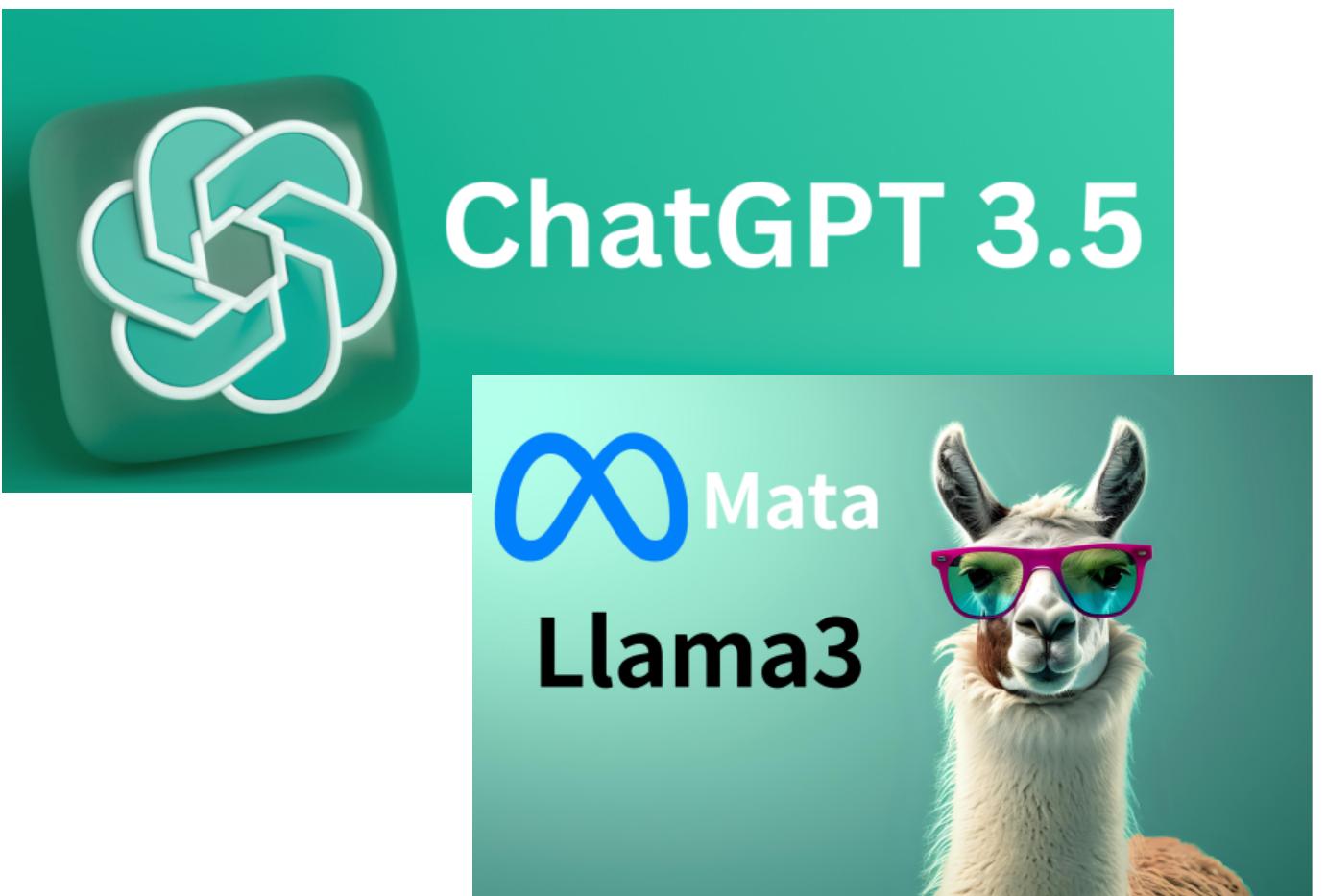
2) 투자성향 분석 : 공격형, 중립형, 방어형 투자자

2024-07-28의 추천 전략을 말씀 드리겠습니다.
오늘의 추천 전략은 전체 4 개의 투자 전략을 검토하였고,
그 결과 고객님의 투자 유형인 중립투자형에 적합한 투자 포트폴리오 입니다.
크래프톤 5.00%, SK하이닉스 5.00%, SK스퀘어 5.00%, 신한지주 5.00%, 기아 5.00%, 한화에
어로스페이스 5.00%, 하나금융지주 5.00%, 우리금융지주 5.00%, 메리츠금융지주 5.00%, 기업
은행 5.00%, HD현대일렉트릭 5.00%, 한미반도체 5.00%, KB금융 5.00%, 삼성화재 5.00%, 삼
성전자우 5.00%, 삼성전자 5.00%, 현대차 5.00%, 삼성물산 5.00%, SK텔레콤 5.00%, 셀트리온
5.00% 을 매수하는 것 이며,
해당 포트폴리오로 투자 하였을 때 예상할 수 있는 연평균 기대수익은 9.03% 이며,
최대 낙폭은 -4.32% 입니다.

포트폴리오 추천해줘

추천 주식있어?

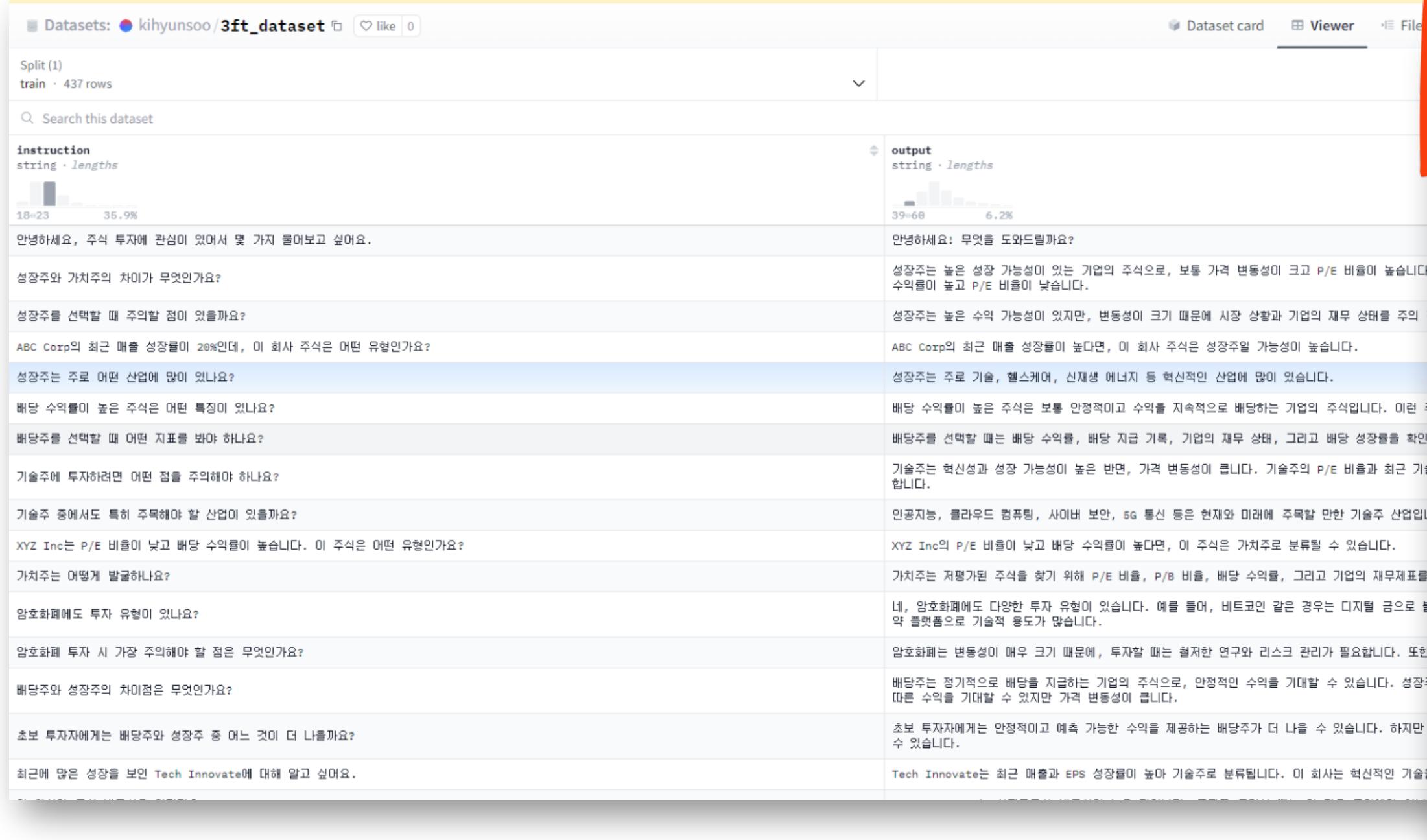
네, 다음 주식을 추천합니다: 크래프톤, SK하이닉스, SK스퀘어, 신한지주, 기아, 한화에어로스
페이스, 하나금융지주, 우리금융지주, 메리츠금융지주, 기업은행, HD현대일렉트릭, 한미반도
체, KB금융, 삼성화재, 삼성전자우, 삼성전자, 현대차, 삼성물산, SK텔레콤, 셀트리온



개발내용

3) LLM(LLaMA3, ChatGPT3.5) 기반 투자자의 투자유형 분석

중요 ★★★



```
from peft import LoraConfig
from trl import SFTTrainer

# PEFT 파라미터 (Parameter-Efficient Fine-Tuning)
peft_params = LoraConfig(
    lora_alpha=16,
    lora_dropout=0.1,
    r=64,
    bias="none",
    task_type="CAUSAL_LM",
)

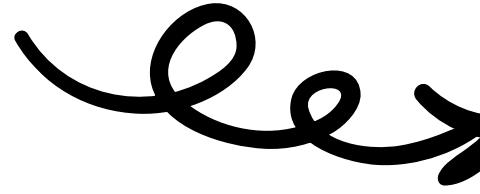
# 모델 훈련
trainer = SFTTrainer(
    model=model, # 학습할 모델 (사전 학습된 모델)
    train_dataset=ft_dataset3, # 학습 데이터셋
    peft_config=peft_params, # SFT 모델의 설정
    dataset_text_field="text",
    max_seq_length=None, # 최대 시퀀스 길이
    tokenizer=tokenizer, # 토크나이저
    args=training_params, # 학습 인자들
    packing=False,
)
```

[LLaMA3 파인튜닝]
투자.금융 관련 내용 테이터셋 수집



TRL library SFT(Supervised Fine-Tuning) Trainer ? PEFT(Parameter-Efficient Fine-Tuning) library?

모델이 커짐에 따라 Fine-tuning 비용이 증가하며,
모든 파라미터를 튜닝하는 것이 점차 어려워지는 문제가 발생



TRL - Transformer Reinforcement Learning

SFTTrainer :

- 지도 학습에서 파인튜닝에 최적화
- 상대적으로 더 적은 양의 데이터셋으로 짧은 시간 안에 빠르게 학습
- 메모리 또한 PEFT 를 사용하므로 적게 사용 할 수 있다는 장점

PEFT :

- 대부분의 모델 파라미터를 고정하고 필요한 부분만 튜닝
- 저장 공간 및 계산 비용을 크게 줄이며, catastrophic forgetting*과 같은 문제를 극복
- 적은 양의 데이터로 뛰어난 일반화 성능 보여줌

(*catastrophic forgetting: 인공 신경망이 새로운 정보를 학습할 때 이전에 학습한 정보를 갑자기 급격하게 잊어버리는 경향)

```
prompt = "성장주에 대해서 알려줘"
pipe = pipeline(task="text-generation", model=model, tokenizer=tokenizer, max_length=500)
result = pipe(f"<s>[INST] {prompt} [/INST]")

result[0]['generated_text']
```

'<s>[INST] 성장주에 대해서 알려줘 [/INST] 적<没钱><没钱> 성장주는 주식이 많으면 주식이 적으면 성장주이고, 주식이 적으면 주식이 적으면 주식주이고, 주식주는 주식이 아니라는<没钱> 성장주는 주식주의 대표적인 주식주이고, 주식주는 주식주의 대표적인 주식이에 투자하는 것이 주식주의 한 형태로도 있지만, 주식주는 주식에 특정 기술적 요소도 포함할 수 있다. </没钱> </INST> </没钱> </INST>는 주식에 특정 기술적 요소를 포함하는 것이<没钱> </没钱> </INST> </没钱> </INST> 성장주는 주식주의 대표적인 주식으로, 주식주는 주식에 특정 기술적 요소를 포함하는 주식으로 분류된다. </没钱> </INST> </没钱> </INST> 성장주는 주식주의 대표적인 주식으로, 주식주는 주식에 특정 기술적 요소를 포함하는 주식으로 분류된다. </没钱> </INST> </没钱> </INST> </没钱> </INST> 성장주는 주식주의 대표적인 주식으로, 주식주는 주식에 특정 기술적 요소를 포함하는 주식으로 분류된다. </没钱> </INST> </没钱> </INST> </没钱> </INST> 성장주는 주식주의 대표적인 주식으로, 주식주는 주식에 특정 기술적 요소를 포함하는 주식으로 분류된다. </没钱> </INST> </没钱> </INST> </没钱> </INST> 성장주는 주식주의 대표적인 주식으로, 주식주는 주식에 특정 기술적 요소를 포함하는 주식으로 분류된다. </没钱> </INST> </没钱> </INST> </没钱> </INST> 성장주는 주식주의 대표적인 주식으로, 주식주는 주식에 특정 기술적 요소를 포함하는 주식으로 분류된다. </没钱> </INST> </没钱> </INST> </没钱> </INST>'

TrainOutput(global_step=270, training_loss=0.570550556536074, metrics={'train_runtime': 778.6922, 'train_samples_per_second': 1.708, 'train_steps_per_second': 0.347, 'total_flos': 4770270016856064.0, 'train_loss': 0.570550556536074, 'epoch': 10.0})

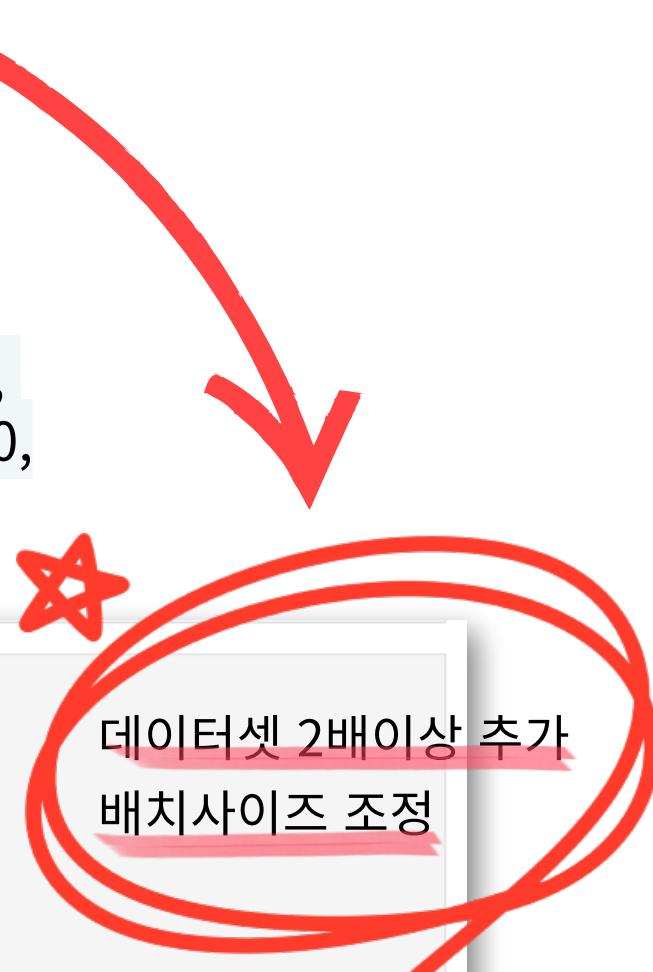
```
key="초보투자자에게 추천할 만한 투자전략이 있나요?"
prompt=f"""you are a assistant please answer in four korean sentences

### Instruction:
{key}

### Response:"""
generate_response(prompt, model)
```

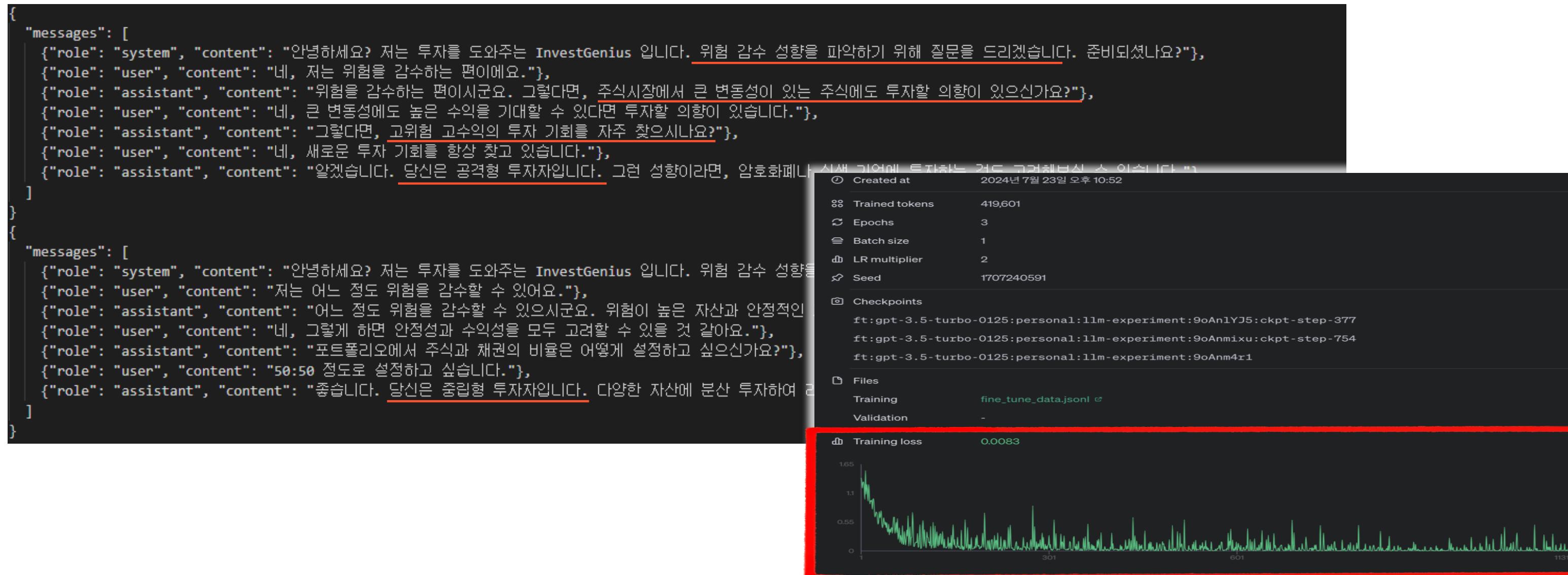
'<|begin_of_text|> \n네, 초보투자자에게는 장기적인 관점에서 꾸준히 투자하는 전략이 좋아요. 또한, 안정적인 배당주나 인덱스 펀드에 투자하는 것도 좋은 방법이에요. 주식 시장은 장기적으로 경제 성장과 함께 성장하니까요.<|end_of_text|>'

TrainOutput(global_step=1100, training_loss=0.4530192444541238, metrics={'train_runtime': 2985.5052, 'train_samples_per_second': 1.464, 'train_steps_per_second': 0.368, 'total_flos': 1.637098388729856e+16, 'train_loss': 0.4530192444541238, 'epoch': 10.0})



개발내용

3) LLM(LLaMA3, ChatGPT3.5) 기반 투자자의 투자유형 분석



[llama-3-Korean-Blossom-8B]

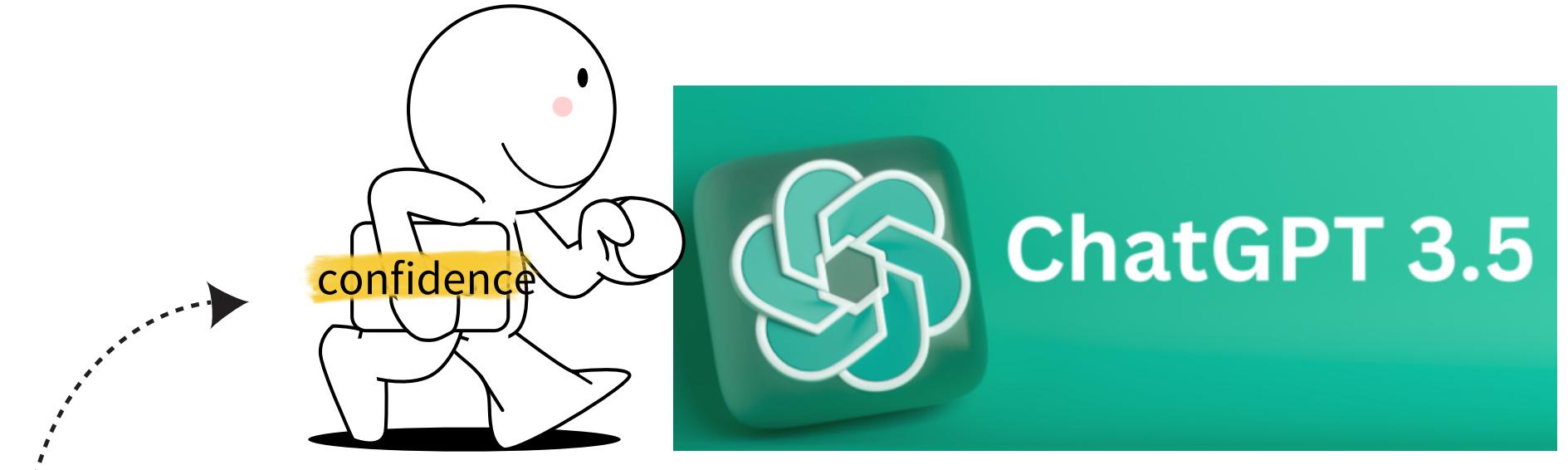


Update!

- [2024.06.18] 사전학습량을 250GB까지 늘린 Blossom ELO모델로 업데이트 되었습니다. 다만 단어확장은 하지 않았습니다. 기존 단어확장된 long-context 모델을 활용하고 싶으신분은 개인연락주세요!
- [2024.06.18] Blossom ELO 모델은 자체 개발한 ELO사전학습 기반으로 새로운 학습된 모델입니다. [LogicKor](#) 벤치마크 결과 현존하는 한국어 10B이하 모델중 SOTA점수를 받았습니다.

(*LogicKor: 한국어언어모델 벤치마크)

```
LLAMA_PROMPT = """
입력받은 언어를 인식하여, 해당 언어로 대답 해주세요. 모든 질문에 대해 JSON 형식으로만 응답해주세요. Please respo
응답은 다음 구조를 따라야 합니다:
|
{
  "answer": "답변",
    "confidence": 0에서 1 사이의 숫자 (확신도),
  "user_invest_type": "귀하가 생각하는 사용자의 투자 유형",
  "strategy_request": "0에서 1 사이의 숫자 (주식추천을 원하는 질문을 할 때 대화의 흐름을 판단해서 비율로 반환)"
}
```



왜 LLaMA3와 GPT3.5를 함께 사용하였다?

- LLaMA3의 불안정함을 chatGPT가 보완하며 backend에서 필요한 기술구현으로 조금 더 안정적인 서비스 목적 달성

개발내용

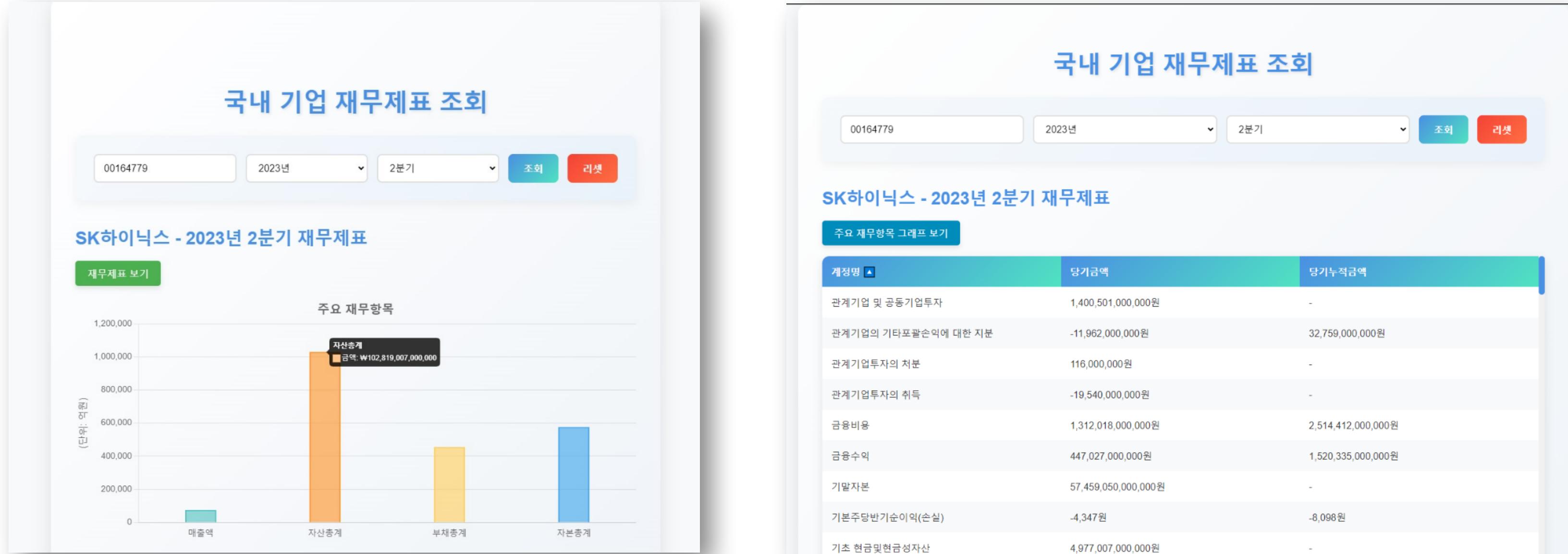
4) 실시간 국/내(미국)외 주식시장 정보 시각화: FDR, YahooFinance 라이브러리



데이터 수집: Ran Aroussi. (2018). yfinance. [Python Library]. <https://github.com/ranaroussi/yfinance>
Jo HyunJoon. (2018) Finance Data Reader. [Python Library]. <https://github.com/FinanceData/FinanceDataReader>

개발내용

5) 국내기업 재무제표 시각화: Open Dart Open API



데이터 수집: Financial Supervisory Service. (2024). [Open DART API] <https://opendart.fss.or.kr>

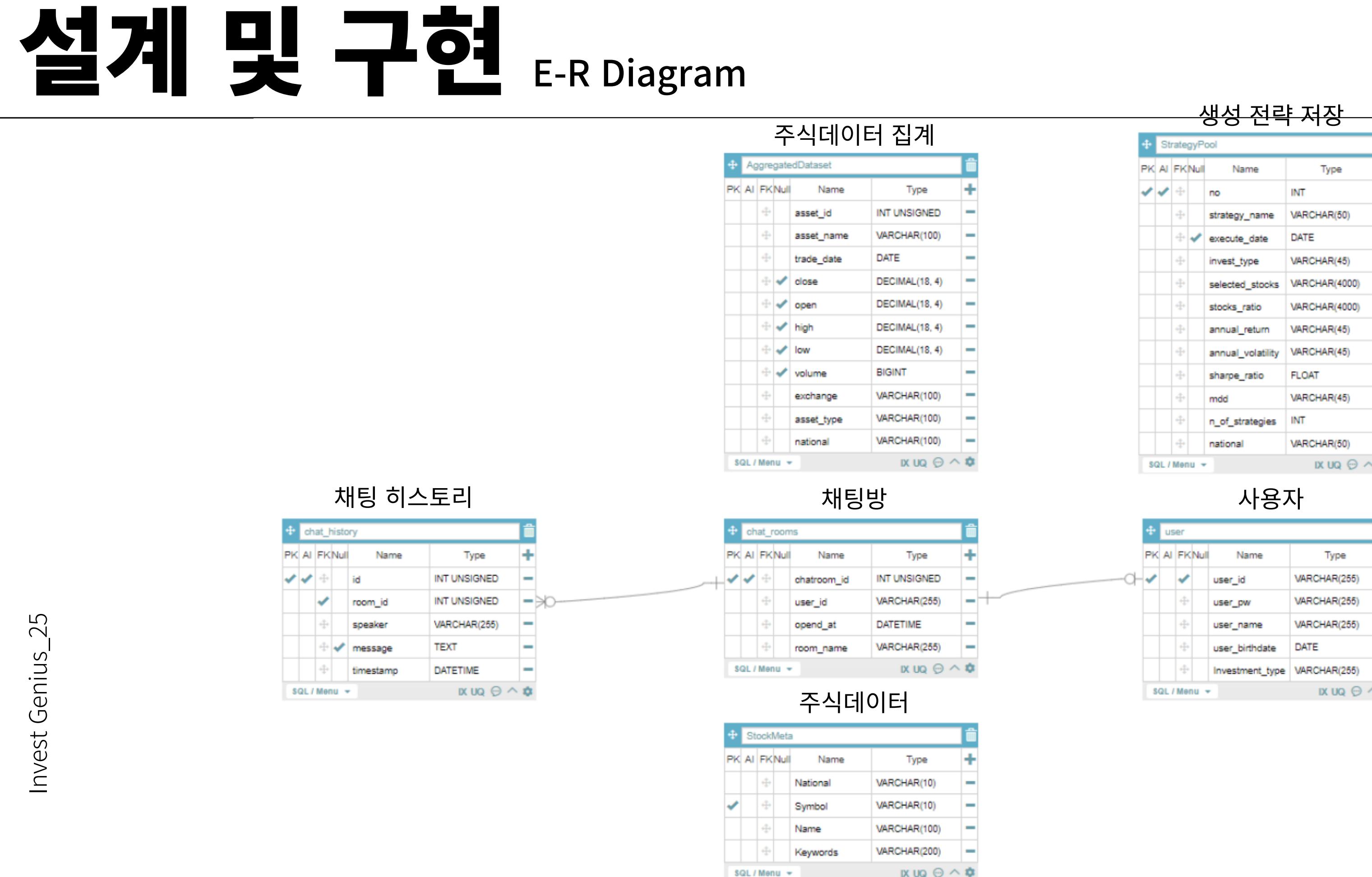
03

설계 및 구현

ER다이어그램
사용언어 및 도구
시스템아키텍처

설계 및 구현

E-R Diagram



설계 및 구현

사용언어 및 도구

Backend



Frontend



Server

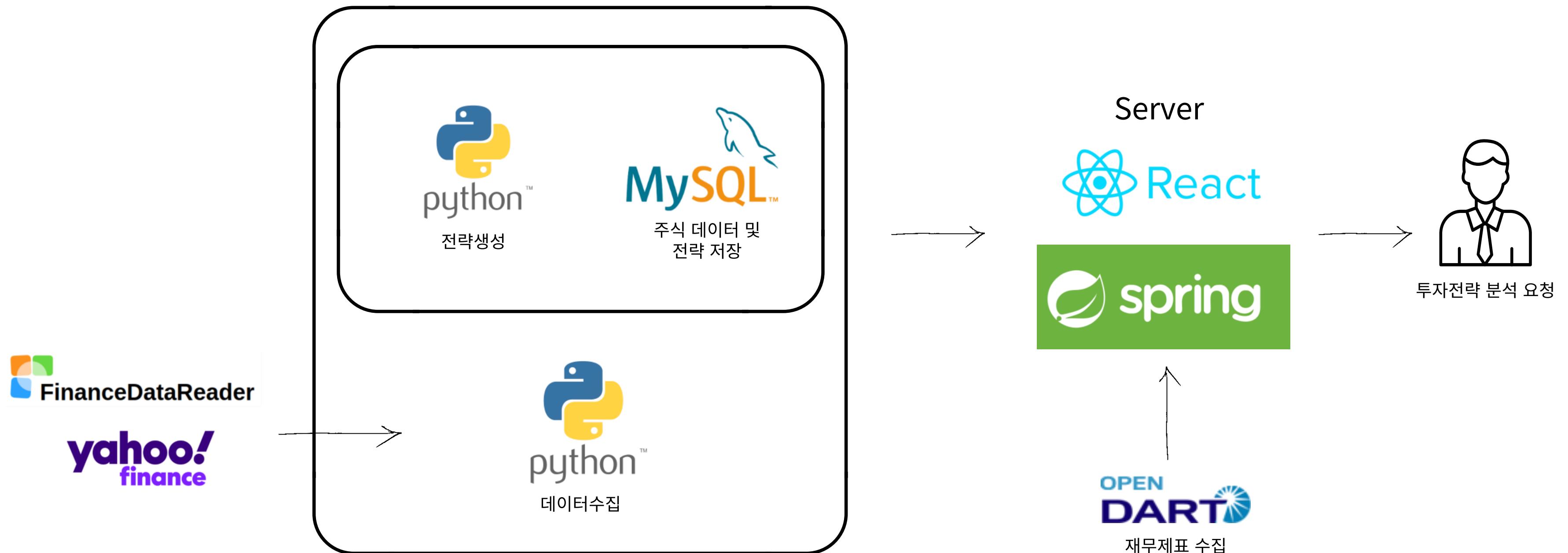


DataBase



설계 및 구현

시스템아키텍처



04 트러블슈팅

트러블슈팅
시연영상

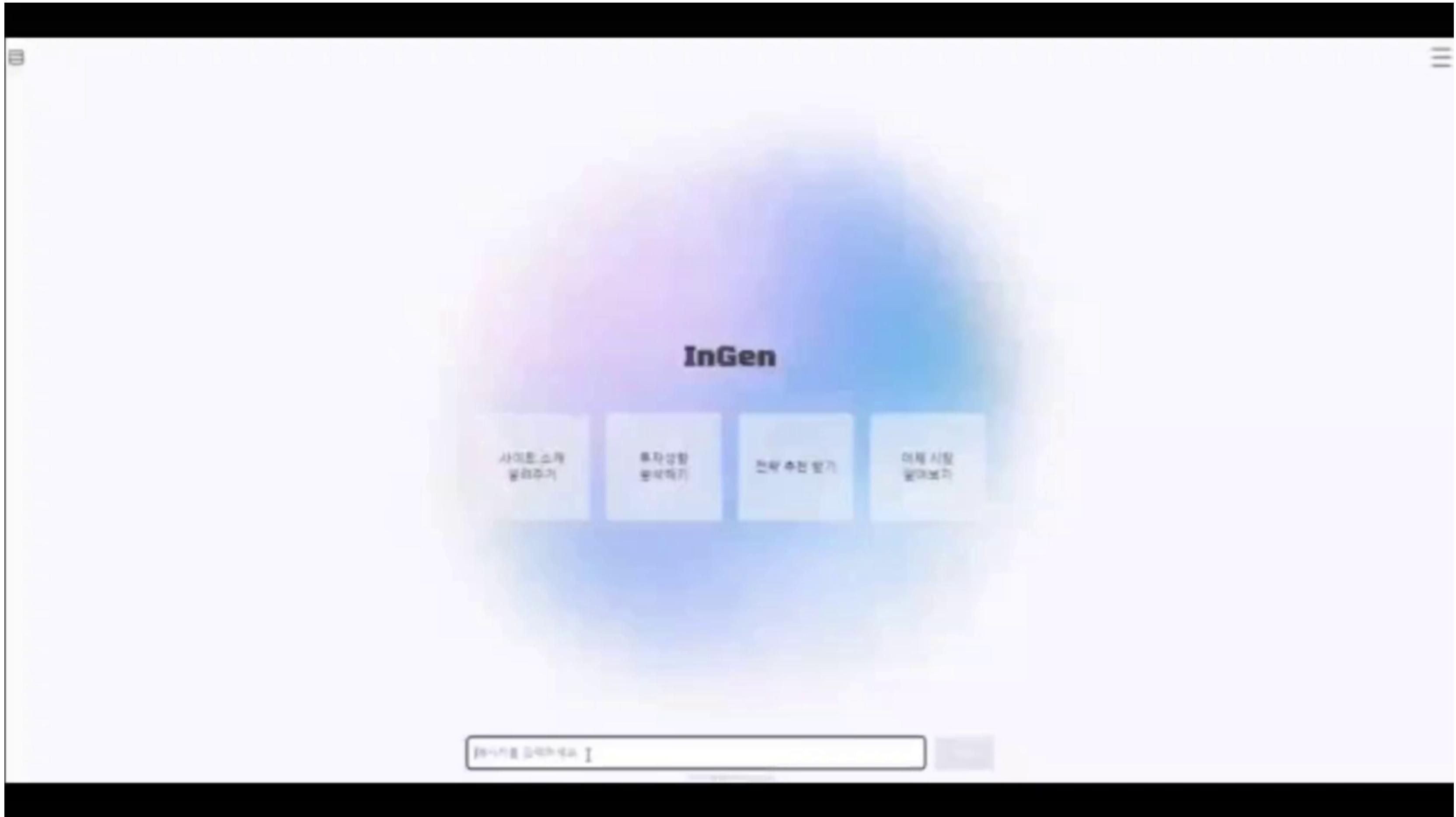
트러블슈팅

1. 채팅방 대화내용이 JSON으로 받아지지 않는 오류

프롬프트에 입력한 json 내용을 함수로 구현하여, json으로 반환

2. 기존 채팅방 입장 시, 기존 대화 히스토리까지 매번 타이핑 애니메이션이 들어가는 오류

기존 채팅방 입장 시, 기존에 있던 대화는 False 값으로 받아오고, 새롭게 입력되는 대화를 True 값으로 받아와서 True/False 값으로 구분하여 애니메이션 구현



05

기대효과 및 활용방안

기대효과 및 활용방안

1. 향상된 투자 성과 기대

LLM기반의 실시간 투자성향 분석과 투자전략 도출

2. 투자자들의 시간과 비용절감효과 기대

대화형 AI 기반의 자동화된 시스템

3. 투자자의 높은 만족도와 심리적 안정감 향상 기대

사용자 친화적인 투자상담 경험 및 안정적인 자산 형성이 심리적 안정감을 향상시킬 것으로 기대

06

참고문헌

Invest
Genius

참고문헌

1. LLM 모델

OpenAI. (2023). ChatGPT (3.5버전) [Large language model]. <https://chat.openai.com>

Meta. (2024). Llama (3버전) [Large language model]. <https://chat.openai.com>

2. 국/내외 실시간 주식데이터

Ran Aroussi. (2018). yfinance. [Python Library]. <https://github.com/ranaroussi/yfinance>

Jo HyunJoon. (2018) Finance Data Reader. [Python Library]. <https://github.com/FinanceData/FinanceDataReader>

3. 국내기업 재무제표

Financial Supervisory Service. (2024). [Open DART API] <https://opendart.fss.or.kr>

4. 데이터 수집

OpenAI. (2023). ChatGPT (3.5버전) [Large language model]. <https://chat.openai.com>

Thank You Genius

Invest