



오픈소스SW기초 6분반

분산 투자 전략 지원, SABU

#주식 #분산 투자

사고8조

정보통계학과	32200472	김동혁
컴퓨터공학과	32211228	김태형
소프트웨어학과	32232597	엄세훈

목차

1

문서화

2

클러스터링

3

백테스팅

4

Testing

5

프론트엔드

6

추가 기능 및 시연

1. 문서화

소제목



- 발표에서 필수적인 것만 소개
- 옆에 코멘트로 깃허브에 다 올려놔고 어디에 뭐가 있다 설명 달기

프로젝트 설계 문서 모음

이 레포지토리는 프로젝트의 기획, 설계, 구조, 아키텍처 등의 **비개발 산출물**을 정리하기 위한 공간입니다.
프론트엔드/백엔드 개발 이전의 **기획·디자인·요구사항 정의서** 등을 모두 포함합니다.

문서 목록

-  [기획 개요](#)
-  [디자인 문서](#)
-  [유스케이스 / UML](#)
-  [API 명세](#)
-  [WBS 및 일정표](#)
-  [주차별 회의록](#)
-  [발표자료 및 포스터](#)
-  [향후 개선 계획](#)
-  [의사결정 기록](#)
-  [회고 및 문제 해결](#)

폴더 구조

```

docs/
├── design/      # 디자인 설계 문서
├── uml/         # 유스케이스, 클래스 다이어그램 등
├── api-spec/    # API 명세
├── schedule/    # 일정표, WBS 등
├── presentation/ # 발표용 자료
├── meeting/     # 회의록
├── project-overview.md
├── future-work.md
└── README.md
    
```

Frontend (React)

이 디렉토리는 사용자 인터페이스(UI)를 담당하는 **프론트엔드 애플리케이션**입니다.
React 기반으로 개발되었으며, REST API를 통해 백엔드와 통신합니다.

주요 기능

- 사용자 인증 (로그인/회원가입)
- 프로젝트/서비스 대시보드
- 시각화 컴포넌트 (차트, 노드 기반 UI 등)
- 반응형 인터페이스

기술 스택

- React + Vite
- useContext 상태 관리
- TailwindCSS
- Axios

폴더 구조

```
frontend/
├── docs/                # 화면 설계 문서, 컨벤션 등
│   ├── Code-Convention.md
│   ├── Commit-Message-Convention.md
│   └── Issue-Convention.md
├── src/
│   ├── src/
│   │   ├── api/        # API 호출 관련 함수
│   │   ├── components/ # 재사용 가능한 컴포넌트
│   │   ├── contexts/   # 전역 상태 관리 컨텍스트
│   │   └── pages/      # 페이지 컴포넌트
```

컴포넌트 가이드

개요

이 문서는 프로젝트에 사용된 주요 **공통 컴포넌트**들의 역할과 props 인터페이스, 내부 구조를 설명합니다.

1. 공통 컴포넌트 목록

컴포넌트 이름	위치	설명
Button	src/components/Button.jsx	기본 버튼
Card	src/components/Card.jsx	대시보드 카드 UI
Modal	src/components/Modal.jsx	팝업 레이어

2. 컴포넌트 상세 설명

Button

```
<Button variant="primary" onClick={handleClick}>
  확인
</Button>
```

상태 흐름 구조 (Zustand)

개요

이 문서는 Zustand 기반 상태 관리 흐름과 store 정의, 주요 전역 상태 변수, 사용 위치 등을 설명합니다.

1. 상태 관리 라이브러리

- 사용: [Zustand](#)
- 도입 이유:
 - Redux 대비 가볍고 보일러플레이트 적음
 - 전역 상태와 로컬 상태 분리가 쉬움

2. 전역 상태 구조

Store 파일: `src/store/projectStore.js`

```
import { create } from 'zustand'

export const useProjectStore = create((set) => ({
  projects: [],
  setProjects: (projects) => set({ projects }),
}))
```


UI 디자인 문서

개요

이 문서는 전체 화면의 흐름과 레이아웃 구조를 설명하기 위해 작성되었습니다.
와이어프레임 이미지, 페이지 전환 흐름, 사용자 인터랙션 등을 포함합니다.

1. 전체 와이어프레임

`./wireframes/main-ui.png`

 [Main Wireframe](#)

2. 페이지별 설명

2-1. 로그인 페이지

- 구성 요소: 이메일 입력, 비밀번호 입력, 로그인 버튼
- UX 고려 사항: 잘못된 입력 시 실시간 에러 메시지 표시

2-2. 대시보드

- 주요 기능: 프로젝트 리스트, 생성 버튼, 최근 알림 표시
- 사용자 흐름: 로그인 → 대시보드 → 프로젝트 상세

3. 사용자 플로우



Login

2. 클러스터링

소제목

2. 클러스터링

- ✓ 구현 과정 상세히 설명
- ✓ 고민 내용, 실패 경험, 처리 방법 등

고민 내용

- ✓ 기업분석을 하는 것은 정형화를 하기 어렵기 때문에 차트분석으로 진행하는 것이 좋아 보임
- ✓ - 지표 기반 유사도 계산 (코사인 유사도: feature 벡터 간의 방향 유사성) / 피어슨 상관계수: 시계열 간의 선형 상관관계
 - 상관계수는 본래 두 변수 간의 선형관계를 측정 -> feature "벡터"로 묶어 비교하면 새로운 유사도(or 거리) 매트릭스가 필요
 - 피어슨 상관계수 = 두 벡터를 평균 0으로 중앙정렬한 뒤 코사인 유사도 계산
 - Feature 벡터 생성: 종목별 이동평균, 변동성, RSI, MACD 등 여러 지표 계산해 하나의 벡터에 담음
 - StandardScaler 등 이용해 스케일링
 - 그 사이에 PCA 등을 이용하여 주성분을 선택해 noise를 줄이는 방법도 있음
 - K-means, 계층적 클러스터링, DBSCAN 등 다양한 클러스터링을 진행해보는 방법
 - > 코사인 유사도 행렬 or 피어슨 상관계수 행렬을 거리로 변환하여 클러스터에 적용

소제목

고민 내용

 feat: 추출 데이터 논의 #3

Closed

Feature

DKUOpenSource-SABU/backend

Public

 hyuk-dd opened last month · edited by hyuk-dd

Edits ...

설명

데이터 분석 및 클러스터링 진행 시 필요한 데이터, 추출 가능한 데이터에 대해 논의

작업 항목

☒ 가격 지표 추출

...

☒ 이동 평균선, 수익률, 변동성, RSI, MDD, 거래량 등의 지표에서 추출 가능한 지표 논의

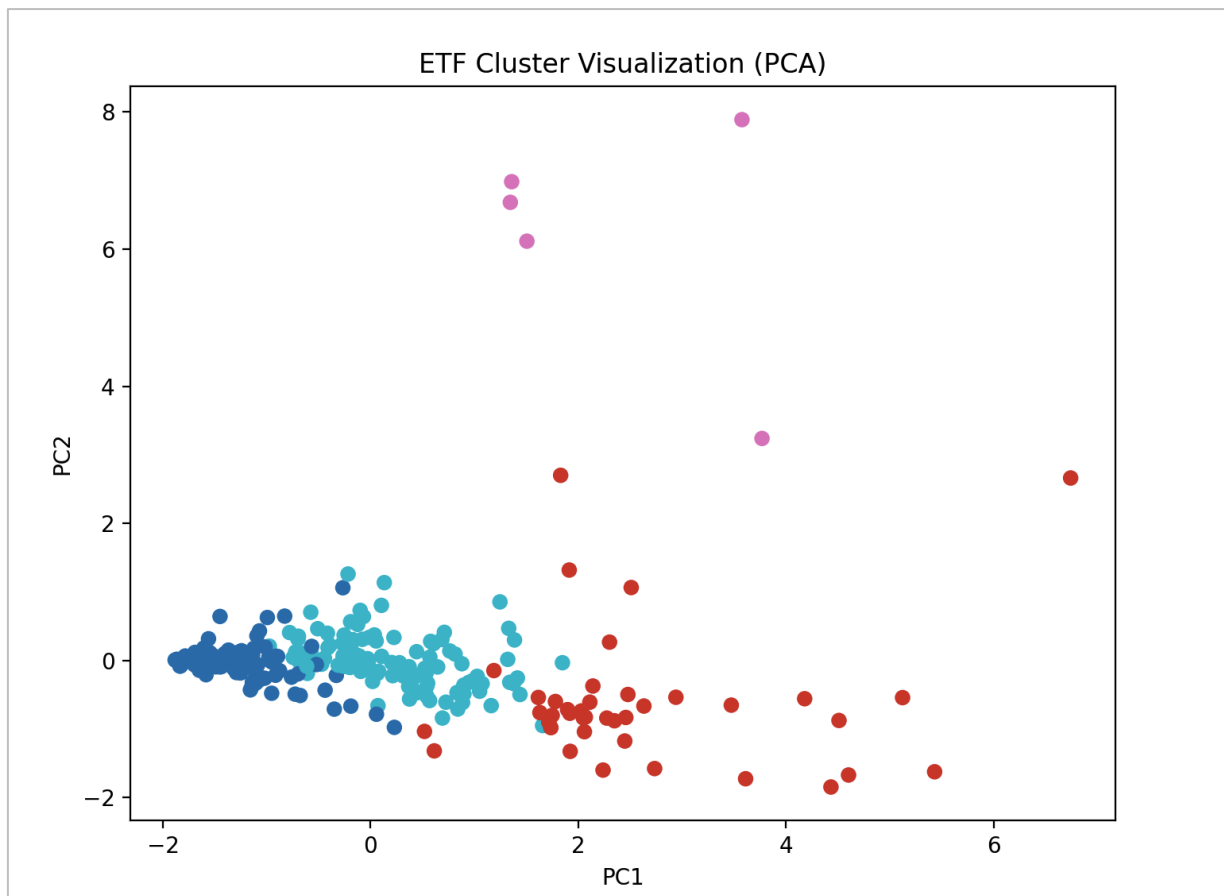
...

Create sub-issue

😊

소제목

고민 내용

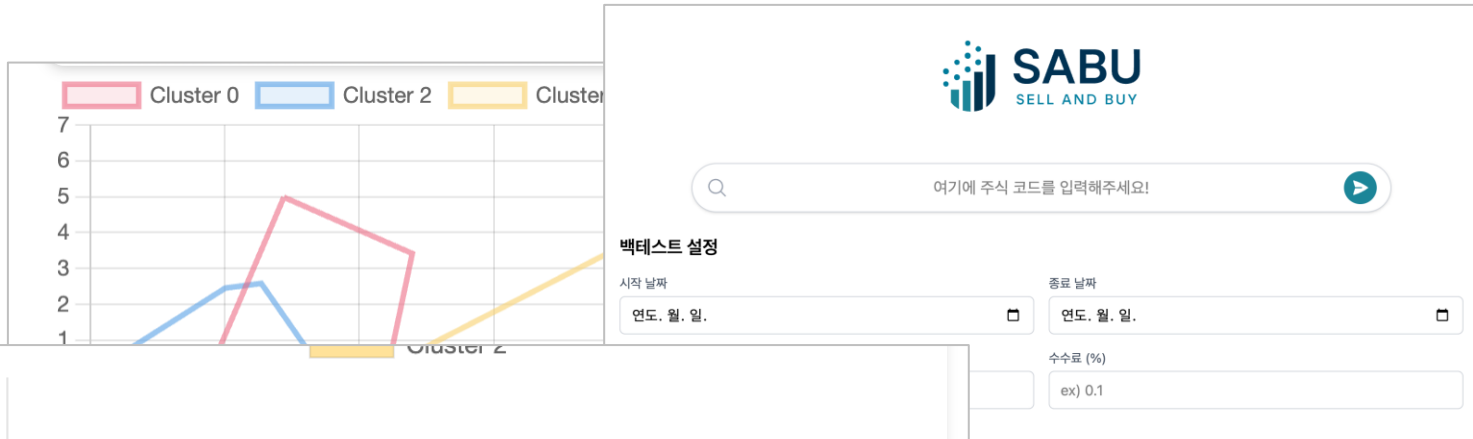


초기 수집된 소수의 데이터로 진행한 클러스터
- 종가, 거래량, 현금배당, 주식분할 조정계수

소제목

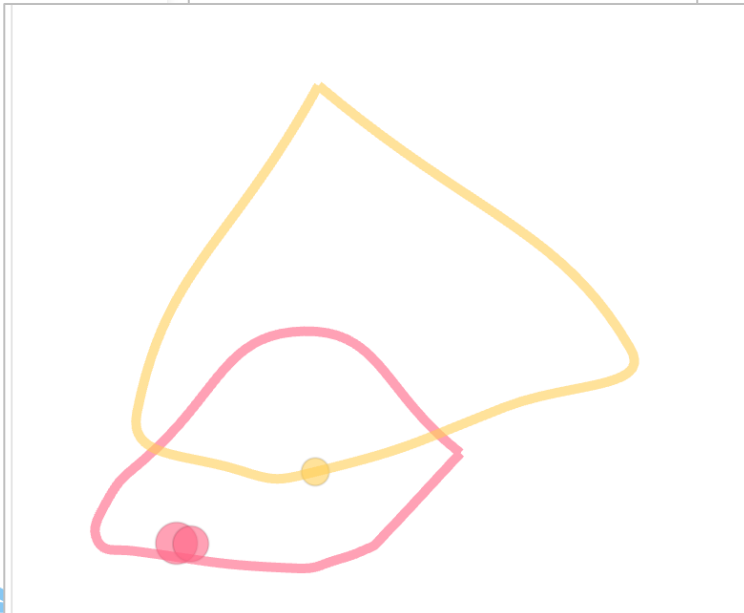
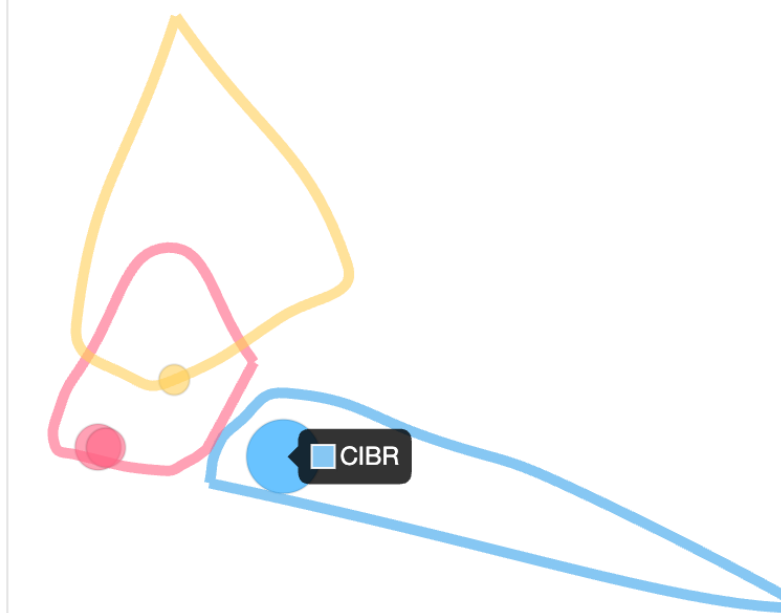
2.클러스터링

고민 내용

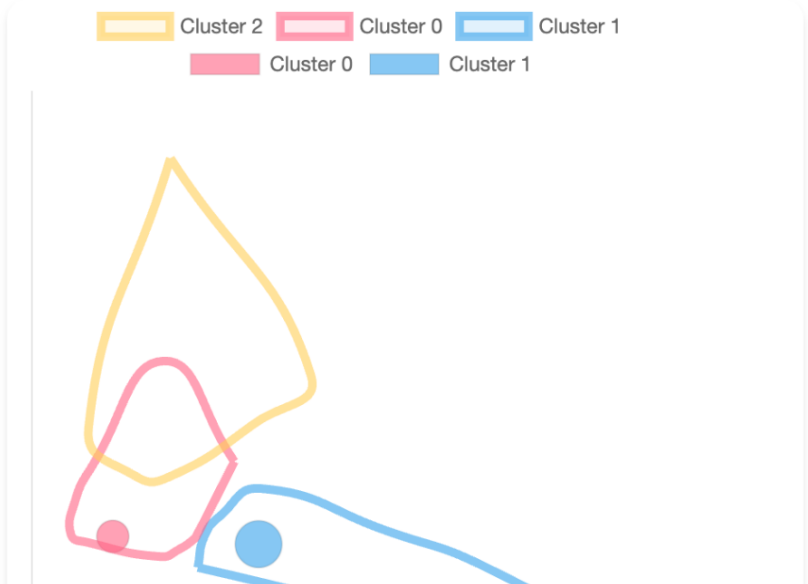


시각화에 대한 고민

- 산점도로 표현하기 보다 클러스터의 범위로 시각화
- 선택 종목의 비율에 따라 종목 크기 변경
- 점 위에 마우스를 올리면 해당 ticker 표시
- 범례 선택에 따라 표시되는 정보 필터링 가능



클러스터링



소제목

고민 내용



- ✓ 초기 구현한 클러스터링 모델에 약 10,000 개의 데이터를 넣었을 때 결과

소제목

고민 내용



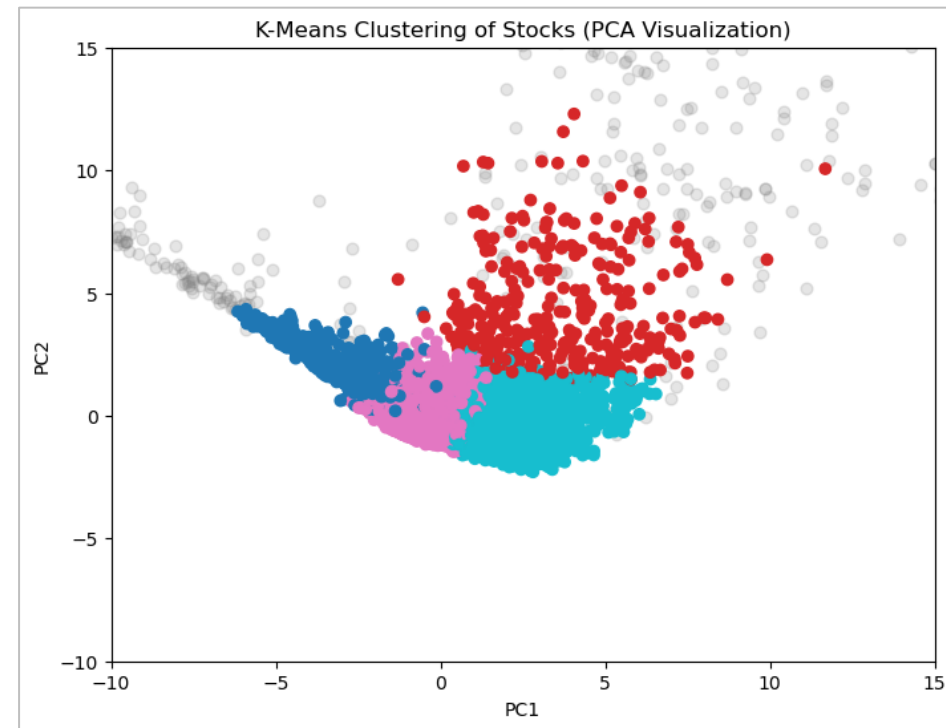
✓ 수정된 클러스터링

소제목

고민 내용



이상치 데이터 포함



소제목

2.클러스터링

고민 내용

```

np.min(X_pca), np.max(X_pca)
[89] ✓ 0.0s
... (-6.15544036102779, 12.29131458522084)

np.min(outlier_pca), np.max(outlier_pca)
[90] ✓ 0.0s
... (-125.78068040888974, 1267.0278680298686)
        
```

선택한 종목

섹터	클러스터	티커	비율 (%)
Technology	Cluster -1	AAPL	100 %
N/A	Cluster -1	BHST	ex) 20 %

클러스터링 결과

클러스터링 점수 ★★★★★

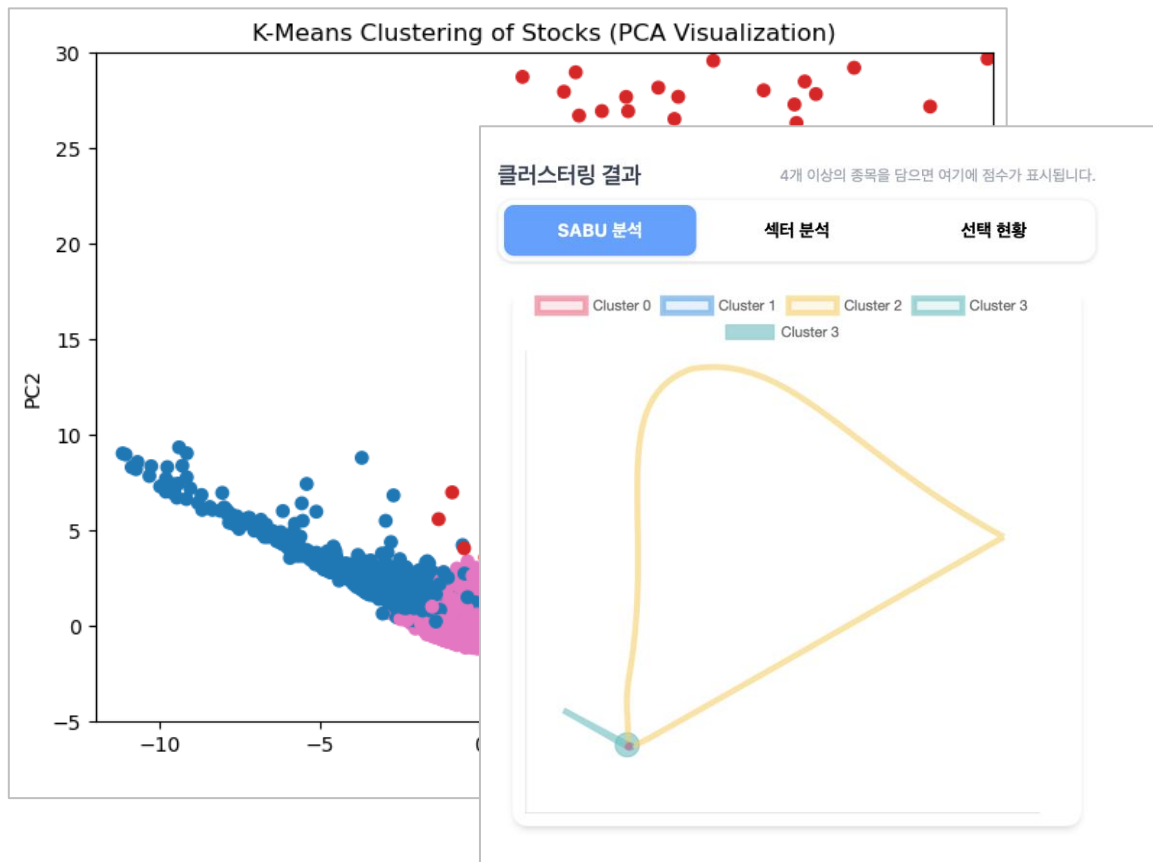
SABU 분석 섹터 분석 선택 현황

- ✓ 이상치 데이터의 2차원 축소값을 구해보니 정상치 데이터와 차이가 많이남



소제목

고민 내용



- ✓ 정상치 데이터로 사전 학습된 K-Means 클러스터링 모델에 이상치 데이터를 넣어 이상치 데이터도 분류
- ✓ 시각화가 잘 나오는 기준으로 좌표값을 지정하여 잘라야 하나?

고민 내용

SABU 시스템 WBS						
No.	업무 영역 (Level 1)	Level 2	Level 3	프로세스 설명	프로세스 ID	비고
1	클러스터링 기능			ETF 간 지표를 기반으로 클러스터링하여 시각화 제공 및 분산투자 포트폴리오 제안	CLU	
2	클러스터링 기능	ETF 종목 코드 검색		사용자가 조회할 ETF의 티커(Ticker) 입력	CLU-001	
3	클러스터링 기능	선택 ETF 종목 목록 조회		선택한 종목의 목록과 각 종목별 기간 내 수익률 조회	CLU-002	
4	클러스터링 기능	선택 ETF 종목 삭제		선택한 종목 중 조회하고 싶지 않은 종목 삭제	CLU-003	
5	클러스터링 기능	클러스터 필터		사용자가 직접 클러스터에 대한 조건 설정		
6	클러스터링 기능	클러스터 필터	군집 수(K) 선택	K-Means 알고리즘 기반 사용자가 조회하고 싶은 군집의 개수 설정	CLU-004	
7	클러스터링 기능	클러스터링 결과 조회		선택한 종목, 클러스터 필터에 따른 클러스터링 시각화 결과 조회	CLU-005	
8	클러스터링 기능	클러스터 추천		사용자에게 분산 투자 포트폴리오에 제안할 수 있는 특정 클러스터 추천	CLU-006	

소제목

고민 내용

- ✓ 또는, 차트분석에 필요한 지표를 고를 수 있게 하여 지표 선택 지표별 클러스터링 결과를 보여주는 방법도 있음
또는, 상관계수를 이용해 분산 투자를 잘 하고 있는지 확인 판단 가능
- ✓ 상위 100개 종목을 클러스터링 하여 back 단에 미리 구축
 - 기술적 지표 추출
 - 데이터 전처리 (표준화)
 - 차원 축소(PCA, 시각화 용도)
 - 엘보우 기법 이용하여 최적의 K값 탐색(SSE)
 - K-means 클러스터링 수행
 - 학습된 클러스터링 모델에 새로운 종목 넣으면 군집 레이블 예측 가능
- ✓ StandardScaler로 정규화하여 단위를 맞춰주기(평균 0, 분산 1)
- ✓ PCA(주성분 분석)를 이용해 먼저 변수를 줄이고 시작하는 방법도 고려

소제목

고민 내용

- ✓ 각 클러스터의 평균 Feature 값 비교해 클러스터를 고수익·고변동 클러스터 / 안정형·저변동 클러스터 등으로 분류?
-> 이런 성향을 띤 클러스터를 가지고 분산 투자를 추천
- ✓ - 일단 클러스터링 할 수 있게 아무거나 상관없이 하나 기준으로 만들기
- 그 후 모델 최적화
- ✓ [클러스터링 구성 전략]
- 기간 데이터 필터링
-> 기간을 맞춰서 구성해야 클러스터링 성능이 높아질 듯
-> 데이터 수집 개수가 최소 6개월 이상은 되어야 할 듯 (or 1년 이상)
-> 그 중에서 처음 날짜가 제일 최근인 것 기준으로 모두 기간을 동일하게 맞춰야 할 듯(전처리)
- ✓ - 이동평균, 수익률, 변동성, RSI(상대 강도 지수), MDD(최대 손실률, 최대 낙폭), 거래량 등 지표를 사용하기로 결정

고민 내용

- ✓ 클러스터링 평가지표:
 - 실루엣 계수 (응집도 + 분리도 종합, 1에 가까울 수록 좋음)
 - 데이비스-볼딘 지수 (분산 / 거리 비율)
 - 칼린스키-하라바츠 지수 (분리도 대비 응집도 비율)
 - > 비지도 학습이므로 3가지 지표를 모두 종합해서 보아야 할 것 같음
- ✓ - 초기 클러스터를 구현해보니 클러스터가 쏠리는 현상이 발생해서 이상치 or 특이값 처리가 필요해 보인다
 - Isolation Forest의 이상치 판단 기준은 각 데이터 포인트가 여러 개의 랜덤 결정 트리에서 고립(isolated)되는 데 필요한 경로 길이(depth)를 기반
- ✓ - 가격 지표만 이용하는 것은 한계가 있을 수 있으므로 재무제표 같은 지표를 이용하는 것이 어떨지 검토
- ✓ 데이터가 들어있지 않은 파일과 상장 폐지된 종목 등에 대한 전처리도 필요

소제목

2. 클러스터링

고민 내용



1. 1년치 이상 / 6개월치 계산 지표(이동평균, 수익률, 변동성)

- 1년치 이상 데이터 보유 종목: 9277개

- $k = 4$

- 정상치: 8798 / 이상치: 464

- 클러스터링 평가 지표

- 실루엣: 0.35934526367893854

- 데이비스: 1.2018531854639207

- 칼린스키: 2420.7878505083745

2. 3년치 이상 / 1년치 계산 지표(이동평균, 수익률, 변동성)

- 3년치 이상 데이터 보유 종목: 7797개

- $k = 4$

- 정상치: 7403 / 이상치: 390

- 클러스터링 평가 지표

- 실루엣: 0.30196390869338263

- 데이비스: 1.292719271867659

- 칼린스키: 1970.553476542335

이외에 3년치 이상 / 6개월치 계산 지표 등 진행해보았지만 1년치 이상 / 6개월치 계산 지표를 이용하는 것이 제일 좋았음.
최신성을 반영할 수 있어서 인것으로 판단

소제목

고민 내용

- ✓ 이상치 데이터를 따로 표현하지 말고
학습 시킨 클러스터에 넣어서 예측 분류를 시키자
그렇게 해서 일단 시각화
PCA로 시각화 했을 때 y축 너무 많이 벗어나는 종목들은 제외시켜야 할 듯

3. 백테스팅

소제목



- bt는 살짝 설명
- 어떻게 돌아가는지랑 전략에 대해 알려줘야 할 듯
 - RSI + monthly / RSI + quarterly / RSI + none
 - buy & hold + monthly / buy & hold + quarterly / buy & hold + none
 - sma cross + monthly / sma cross + quarterly / sma cross + none
- 구현 내용은 최대한 줄이고
- 코드 설명할거면 다이어그램으로

Open Source - bt



내용을 입력해주세요.

작동 플로우



내용을 입력해주세요.

백테스팅 전략 소개



- RSI + monthly / RSI + quarterly / RSI + none
- buy & hold + monthly / buy & hold + quarterly / buy & hold + none
- sma cross + monthly / sma cross + quarterly / sma cross + none

백테스팅 전략 소개

투자 전략 선택 ①

rsi + monthly 🏆

rsi + quarterly

buy-and-hold + quarterly

rsi + none

buy-and-hold + monthly

sma cross + none

sma cr

초기 자산

\$10,000

최종 자산

\$22,377.762

총 수익률 ①

123.77761535234211%

CAGR ①

0%

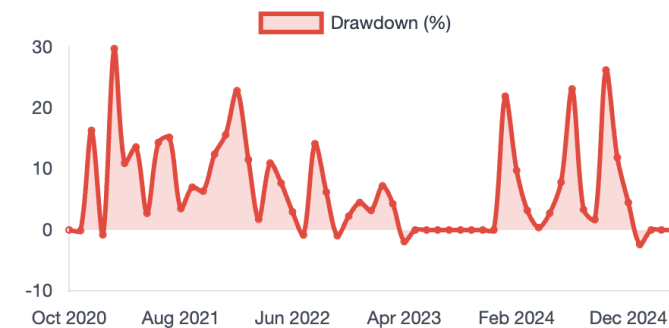
최대 낙폭 ①

76.43%

포트폴리오 성장



최대 낙폭 (Drawdown)



RSI 전략

3. 백테스팅

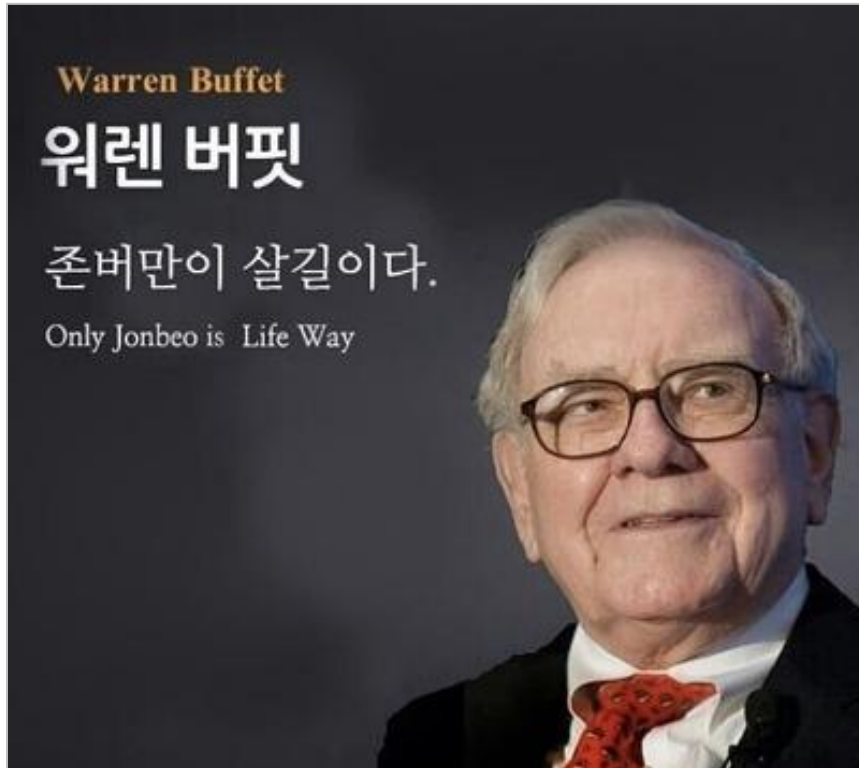
- ✓ **RSI(상대강도지수):** 일정 기간 동안의 상승 변화량과 하락 변화량의 비율
- ✓ 과매수 또는 과매도 상태를 파악하고 매매 시점을 결정

$$RS = \frac{\text{평균 상승}}{\text{평균 하락}}$$

$$RSI = 100 \frac{100}{1+RS}$$



Buy and hold 전략



- ✓ **buy and hold**(매수 및 보유): 주식을 매수한 후 장기 보유하는 전략
- ✓ 주가는 장기적으로 볼 때 지속적으로 상승하는 경향이 있으므로 장기 투자 시 투자 성공 확률이 높다는 것을 이용하는 전략

SMA 전략

- ✓ **SMA(단순이동평균):** 특정 기간 동안 주식의 평균 증가
- ✓ 단기 SMA가 장기 SMA를 상향 돌파할 때 매수 신호, 하향 돌파할 때 매도 신호로 판단
- ✓ 주가의 추세 변화를 예측하고 수익을 얻기 위해 활용



4. Testing

소제목

4. Testing



- gunicorn (서버)
- k6.js (테스팅)



내용을 입력해주세요.



내용을 입력해주세요.

5. 프론트엔드

기능 수정이 가능하다면 작성
안 된다면 제외

6. 추가 기능 및 시연

소제목

6.추가기능및시연

- ✓ - 리더보드
- 감정분석
- ✓ 종목 추천, 섹터 집단 시각화, 분산투자 점수화 등을 여기에 넣을지,
클러스터링에 넣을지

SABU 종목 추천



사용자가 선택한 종목들을 기준으로 분산 투자에 도움이 되는 종목
5가지 추천

전체

Cluster 0

Cluster 1

Cluster 2

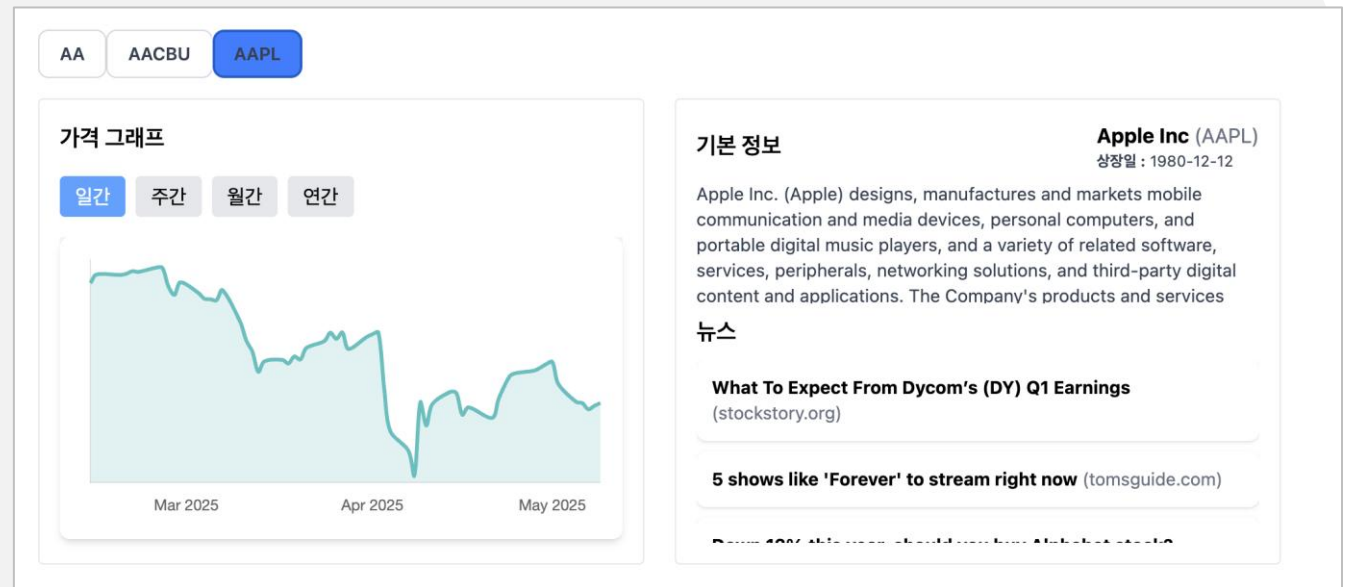
Cluster 3

섹터	클러스터	티커	이름	현재가	등락률
🔥 SABU 추천 종목					
Industrials	Cluster 2	AZ	A2Z Cust2Mate Solutions Corp. Common Shares	\$9.31	-0.21%
Consumer Disc...	Cluster 2	BQ	Boqii Holding Limited American Depositary Shares (...	\$2.35	8.29%
Health Care	Cluster 2	CCM	Concord Medical Services Holdings Limited Ameri...	\$8.45	0.48%
Finance	Cluster 2	DISTW	Distoken Acquisition Corporation Warrant	\$0.09	-13.46%
Consumer Sta...	Cluster 2	FAMI	Farmmi Inc. Ordinary Shares	\$1.77	-0.56%

주가 그래프 및 뉴스 정보 제공

6.추가기능및시연

✓ 내용을 입력하세요.



뉴스 감정 분석

✓ 내용을 입력하세요.

뉴스

Now or wait? (finance.yahoo.com)

☐ 중립적인 의견이에요.

South32 gains funding for Worsley Alumina
(australianmining.com.au)

☒ 긍정적인 뉴스예요!

✓ 내용을 입력하세요.

클러스터링 결과

클러스터링 점수 3.3848770939907746 ★★☆☆☆

SABU 분석

섹터 분석

선택 현황

Cluster 0 Cluster 1 Cluster 2 Cluster 3
Cluster 1 Cluster 2 Cluster 3

섹터별 거리 시각화

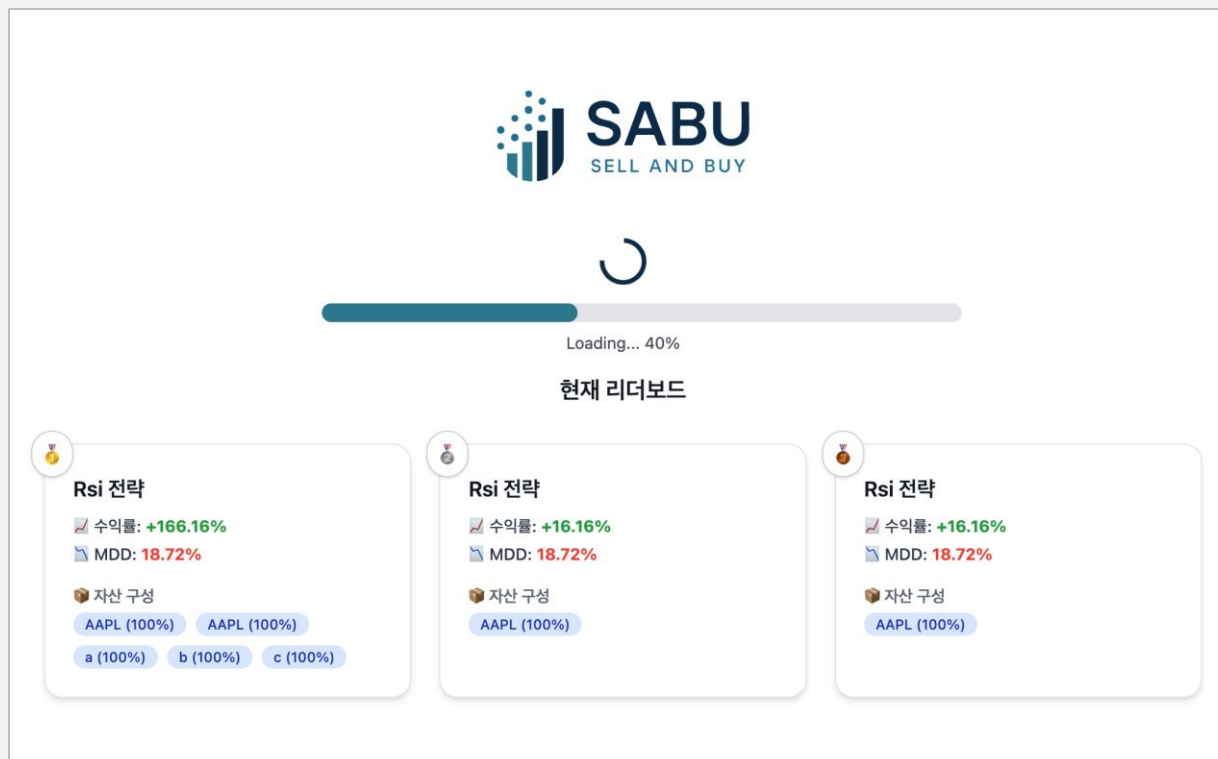
6.추가기능및시연

- ✓ Nasdaq에서 수집된, 섹터 정보가 있는 주식만 가지고 진행
- ✓ 섹터별로 IQR을 이용하여 이상치 처리
- ✓ 섹터를 그룹화 하여 평균 지점의 위치를 2차원으로 시각화



백테스팅 실시간 리더보드

✓ 내용을 입력하세요.



시연

6.추가기능및시연

Q & A