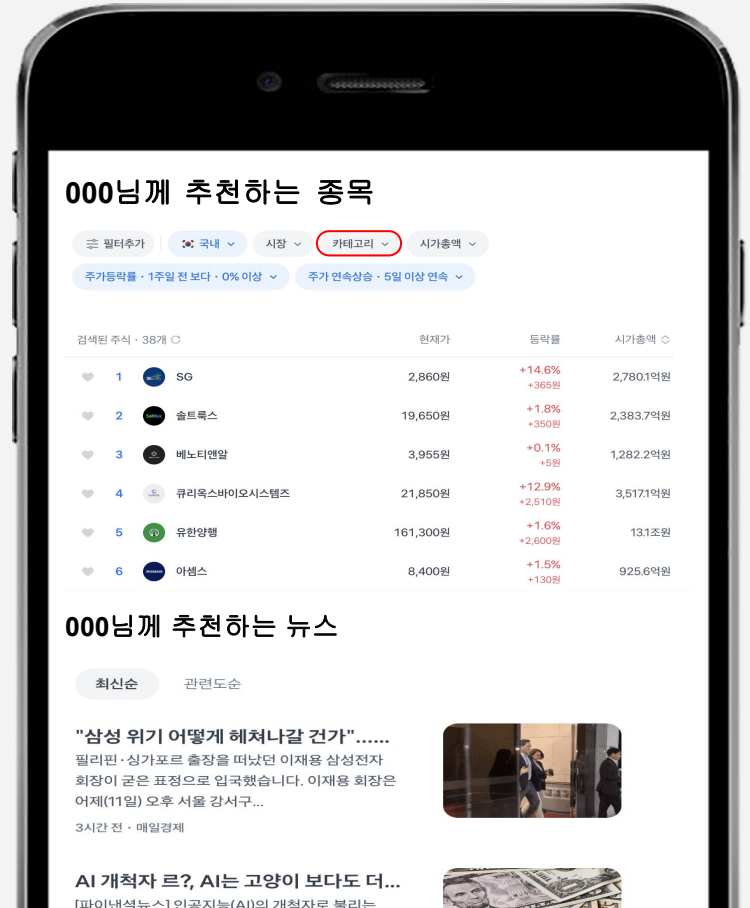


개인 투자자의 건강한 금융 습관 형성 프로젝트

투자자 종목 & 뉴스 추천 서비스 제안



Contents 목차

01 | 제안 배경

02 | 서비스 개요

03 | 데이터 및 분석 과정

04 | 기대효과 및 한계점

서비스 선정 배경 - 시장 현황 분석

초보 투자자도 쉽게 투자생활을 이어갈 수 있도록 '투자자 눈높이 서비스' 제공

- **신한투자증권** - '쉬운 공시'
 - ex) 취득 많은 - 자기주식을 취득할 예정이예요, '발행 주식 수 대비 취득할 주식의 비율'로 비교했어요, 회사는 공시한 수량을 3개월 안에 반드시 취득해야 돼요 등의 **쉬운 용어로 명시**
- **KB증권** - '오늘의 콕'
 - 주식부터 부동산, 세무, 법률까지 재테크에 필요한 모든 정보를 고객이 이해하기 쉽도록 **일상적인 언어로 객관적인 정보를 제공**하는 콘텐츠 서비스
- **한국투자증권** - 'MZIP'
 - **초보 투자자도 이해하기 쉬운 금융 관련 콘텐츠를 제공**하고, 투자에 실질적인 도움을 주는데 초점
- **나무증권** - '나무팜 게임'
 - '나무팜'은 주식을 마치 농작물을 키우듯이 관리하며 성장시키는 투자 농장 개념으로, **매일의 투자 루틴을 형성하고자 하는 투자자를 위한 플랫폼**

그럼에도 불구하고, 몰타기식 투자 피해는 증가

몰타기 투자 피해 기사

이윤우 한국경제교육학회 홍보이사는
“투자에 대한 개념이 불확실한 MZ세대가
**일부 투자 성공사례만 믿고 뛰어들다 보니
빚더미에 오르는 상황이 발생한 것**”이라며

“교육기관에서 금융과 관련된 정보를 접하기도
어렵고, 특히 온라인으로 투자 종목을 찍어주는
'리딩방' 등 **비전문가의 투자 의견을 맹신**하다
하락기에 큰 충격을 받게 됐다”고 설명했다.

[출처 : 중앙일보 (2022)]

기업의 가장 중요한 요소이자 객관적인 가치를 나타내는 지표임에도 불구하고 강조되지 않는 **‘재무 정보’**

세계 최고의 투자가 워렌 버핏

**‘10년 이상 볼 것이 아니라면
10분도 갖고 있지 말라’**



성공적인 투자를 위해서는
기업의 가치를 보고
‘장기적으로 투자’해야 함을
강조

재무 정보 관련 서비스



재무제표에 대한 어려움을 호소하는 투자자들



@SosoTechMom 5년 전

저 거의 재무제표는 포기하고 있었는데..이제부터라도 해볼까요...



@user-y17rs2gm6d 4년 전

안녕하세요 ~ 재무제표 좀 공부하려고 했었는데 천만 단위, 억단위, 백만단위..도 헷갈리지만 원단위는 대체 어떻게 관어서 읽어야 하는지**가장 중요한 금액을 볼 줄 모르니 힘이 빠지네요....**



@gogojerry6518 4년 전

재무제표 보는거 너무 어려운 것 같아요ㅠㅠ 투자만 3년째인데 아직도 모르겠네요 ㅋㅋㅋㅋ



@advil-1225 2년 전

홍진경님 재무제표는 어디서 봐야하냐는데서 땀 터짐 ㅋㅋㅋㅋㅋ 저도 몰라서 공감됩니다.....^^

[출처 : 유튜브]

서비스 특징

000님의 보유 종목을 바탕으로 선호하시는 재무 유형 및 추천할 종목 및 뉴스를 찾는 중입니다

삼성전자 59,300원
+400 (+0.6%)



000님의 보유 종목은
고안정 추구 유형(AABCBI)
으로
이러한 유형은 ...

재무안정성과 수익성을 중요하게 생각하는 유형입니다.
해당 기준을 중요하게 여긴다면,
유동비율, 부채비율, 자본 유보율
관련 재무 정보를 파악하는 것이
투자하는데 도움이 될 수 있으며...

000님께 추천하는 종목

종 필터추가 X 국내 시장 카테고리 시가총액
주가동향률 · 1주일 전 보다 · 0% 이상 추가 연속상승 · 5일 이상 연속

검색된 주식 · 38개	한재가	등락률	시가총액
1 SG	2,860원	+14.6% +365원	2,780.1억원
2 솔트룩스	19,650원	+1.8% +350원	2,383.7억원
3 베노티엔탈	3,955원	+0.1% +5원	1,282.2억원
4 큐릭룩스바이오시스템즈	21,850원	+12.9% +2,610원	3,517.1억원
5 유한양행	161,300원	+1.6% +2,600원	131.8억원
6 아셀스	8,400원	+1.5% +130원	925.6억원

000님께 추천하는 뉴스

최신순 관련도순

"삼성 위기 어떻게 헤쳐나갈 건가".....
필리핀·싱가포르 출장을 떠났던 이재용 삼성전자
회장기 굳은 표정으로 입국했습니다. 이재용 회장은
이제(11일) 오후 서울 강서구...

3시간 전 · 매일경제



AI 개척자 르?, AI는 고양이 보다도 더...
[파이낸셜뉴스] 인공지능(AI)의 개척자로 불리는
프랑스의 전문가가 AI가 고양이를 비롯한
반려동물의 지능에 비해 크게 뒤떨어져있다고 AI...

3시간 전 · 파이낸셜뉴스



카테고리

카테고리

Q 직접 검색해볼 수 있어요

과도안개설

고위험

중위험

저위험

수익성

고위험

중위험

저위험

변동성

고위험

중위험

저위험

성장성

고위험

중위험

저위험

과도안개설

고위험

중위험

저위험

*종목 & 뉴스 추천 기준 커스텀 가능

재무 정보 기반의 분류 필터 설정

통상적으로 사용되는 기준을 바탕으로
재무 안정성, 수익성, 변동성, 성장성,
배당안정성 등 총 5가지 분류 기준 제공

투자자의 투자 종목 유형 분석

투자자 개인의
관심/보유 종목의 재무 상태 분석 &
선호 투자 유형 추출 및 관련 정보 제공

투자 종목 & 뉴스 추천 기능

분석한 유형을 기반으로
유사한 종목과 관련 뉴스 추천 제공
원하는 기준의 필터를 선택하여 해당 기준에
따른 투자 종목 및 뉴스 추천도 제공

상세 데이터 분석 파이프라인

데이터 수집 및 전처리

[재무제표 데이터]



재무 정보 기반 분류 카테고리 (5가지)
: 재무안정성, 수익성, 변동성, 성장성,
배당안정성

[소비자 데이터]

자체 생성

[뉴스데이터]



클러스터링

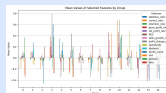
[카테고리 별 최적의 군집 수 결정]

(K-Means
DBSCAN
Hierarchical Clustering → **K-Means**
채택!

클러스터링 결과 기반, 분류기준 수정

[클러스터 그룹화]

Hamming Distance 기반
총 **15그룹** 생성



소비자 매칭

[투자자 데이터 기반 종목 매칭]

step 1) 투자자 보유주식과 같은
군집에 속한 종목 선택

step 2) 상위 20개의 종목 추천

[추천종목 관련 뉴스 추천]

추천 종목과 매칭된 뉴스 추천



데이터셋 설명

출처	자료원명	연도	산출지표	활용부분
<u>전자공시시스템</u> <u>DART API</u>	사업보고서 (연간)	2023	재무상태표(부채총계, 자본총계, 유동자산, 유동부채, 이익잉여금, 자본금, 자본잉여금), 손익계산서(매출총이익, 매출액, 당기순이익), 연결재무제표(배당수익률, 배당성향, 배당연속지급)	재무제표 데이터
<u>KRX</u> <u>정보데이터</u> <u>시스템</u>	주식 - 종목시세	2024.03.01 ~ 2024.07.20	종목명, 시가, 종가, 변동폭, 거래량, 거래대금, 등락률	주가 데이터
<u>네이버 증권</u>	국내증시 - 주요시세정보	2023	52주 베타값	주가 데이터
<u>빅카인즈</u>	뉴스 기사 (10만개)	2024.08.01 ~ 2024.10.14	뉴스 제목, 본문, 언론사, 통합분류, 사건/사고	소비자 데이터 기반 뉴스 매칭
<u>임의 생성</u>	유저 데이터		유저아이디, 보유 종목코드 및 종목명, 투자성향	소비자 데이터 기반 뉴스 매칭

재무 정보 기반 분류 카테고리 소개 - 재무 안정성, 수익성

1. 재무 안정성

: 부채 비율, 유동 비율, 자본 유보율 지표를 바탕으로

재무 안정성이 높은 기업은 저위험, 중간 정도인 기업은 중위험, 높은 기업은 고위험군으로 분류



*부채 비율 = (부채총계/자본총계) * 100

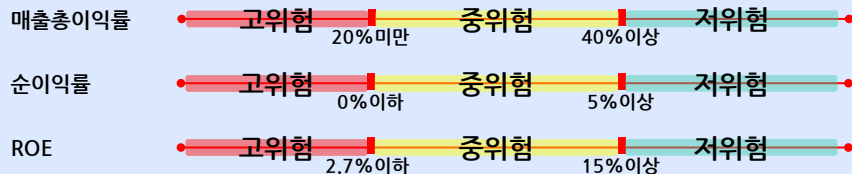
유동비율 = (유동자산/유동부채) * 100

유보율(단위 %) = ((이익잉여금 + 자본 잉여금) / 자본금) * 100

2. 수익성

: 매출총이익률, 순이익률, 3년 연속의 연평균 ROE(자기자본이익률) 지표를 바탕으로 수익성이 안정적으로

높은 기업을 저위험, 중간 정도의 수익률을 보이는 기업을 중위험, 수익률이 저조한 기업을 고위험군으로 분류



*매출총이익률 = 매출총이익 / 매출액

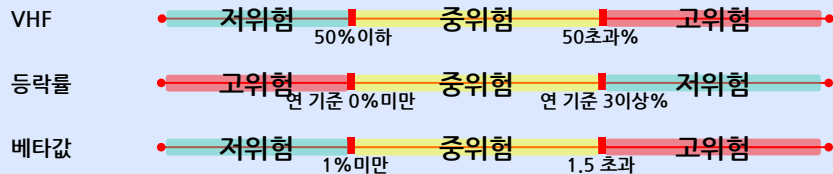
순이익률 = (당기순이익/매출액) * 100

ROE = (당기순이익/자본총계) * 100

재무 정보 기반 분류 카테고리 소개 - 변동성, 성장성, 배당안정성

3. 변동성

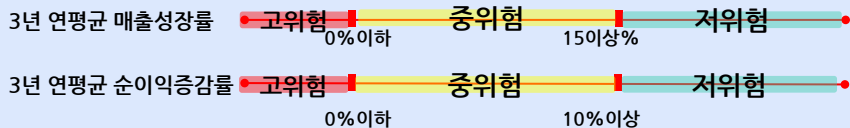
:VHF, 등락률, 베타값을 정보를 바탕으로 변동성이 적은 기업은 저위험, 보통인 기업은 중위험, 심한 기업은 고위험군으로 분류



* 변동성 지표 = $VHF \quad VHF = 100 * (\text{현재종가} - n\text{일 전 종가}) / ATR(n)$
 등락률, 베타값은 네이버 증권 크롤링값으로 활용

4. 성장성

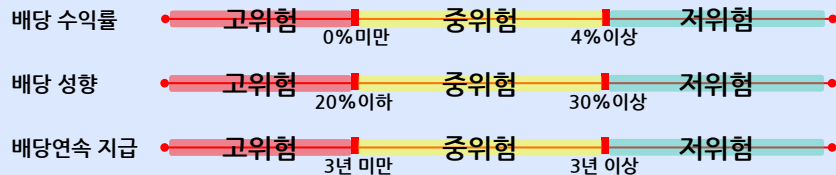
:3년간 연평균 매출 성장률, 연평균 순이익 증감률 정보 바탕으로
 성장성이 안정적으로 높은 경우 저위험, 성장률이 낮으면 고위험으로 분류



*연평균 매출 성장률(CAGR) = $(\text{최종 년도의 매출액} / \text{최초 년도의 매출액})^{(1/\text{기간})} - 1$
 연평균 순이익 증감률은 재무제표 자료 값을 활용

5. 배당안정성

:배당수익률, 배당성향, 배당 연속 지급 정보를 바탕으로
 배당 안정성이 높은 것은 저위험, 낮은 것은 고위험으로 분류



*배당수익률 = $\text{배당금} / \text{현재 주가} * 100$
 배당성향 = $(\text{주당배당금} / \text{주당순이익}) * 100$

데이터분석 - 클러스터링 (검증 용도)

K-means

```
#군집수별 평가 지표 산출 함수
def cluster_and_evaluate(data, k_range):
    silhouette_scores = []
    for k in k_range:
        kmeans = KMeans(n_clusters=k, random_state=42)
        labels = kmeans.fit_predict(data)
        silhouette_avg = silhouette_score(data, labels)
        silhouette_scores.append(silhouette_avg)
    return silhouette_scores
```

클러스터링 평가 지표 계산

- **Silhouette Score** : 각 데이터 포인트가 자신의 군집에 얼마나 잘 속해 있는지를 측정하는 값으로, 값이 높을수록 더 나은 군집화 결과를 의미

DBSCAN

```
f_cluster
0    1407
1         7
-1        2
Name: count, dtype: int64
p_cluster
0    1408
-1         6
1         2
Name: count, dtype: int64
v_cluster
0    1408
1         5
-1         3
Name: count, dtype: int64
g_cluster
0    1413
1         2
-1         1
Name: count, dtype: int64
d_cluster
0    1408
-1         4
1         2
2         2
Name: count, dtype: int64
```

클러스터 별 데이터 불균형 문제 발생

Hierarchical Clustering

```
f_cluster
0    1414
2         1
1         1
Name: count, dtype: int64
p_cluster
0    1415
1         1
Name: count, dtype: int64
v_cluster
0    1415
1         1
Name: count, dtype: int64
g_cluster
0    1413
1         3
Name: count, dtype: int64
d_cluster
0    1409
1         5
2         2
Name: count, dtype: int64
```

클러스터 별 데이터 불균형 문제 발생

데이터분석 - 클러스터링

K-means

클러스터링 평가 지표 계산

○ Silhouette Score

```
#군집수별 평가 지표 산출 함수
def cluster_and_evaluate(data, k_range):
    silhouette_scores = []
    for k in k_range:
        kmeans = KMeans(n_clusters=k, random_state=42)
        labels = kmeans.fit_predict(data)
        silhouette_avg = silhouette_score(data, labels)
        silhouette_scores.append(silhouette_avg)
    return silhouette_scores
```

최적의 군집 수

Silhouette index를 고려한 최적의 군집 수:

- 재무 안정성 기준 최적의 군집 수: 3
- 수익성 기준 최적의 군집 수: 2
- 변동성 기준 최적의 군집 수: 2
- 성장성 기준 최적의 군집 수: 3
- 배당 기준 최적의 군집 수: 4

주식 분류
기준 수정

각 기준별 클래스 개수

```
f 클러스터 레이블의 개수:
financial_stability
C    918
A    452
B     46
Name: count, dtype: int64
p 클러스터 레이블의 개수:
profitability
B    1332
A     84
Name: count, dtype: int64
v 클러스터 레이블의 개수:
volatility
B    1046
A     370
Name: count, dtype: int64
g 클러스터 레이블의 개수:
growth
C     900
B     438
A      78
Name: count, dtype: int64
d 클러스터 레이블의 개수:
dividend_stability
C    1224
B     138
A      54
Name: count, dtype: int64
```

-> 재무안정성

-> 수익성

-> 변동성

-> 성장성

-> 배당

분류 기준 수정 - 클러스터링 결과를 참고하여 최적의 군집 수로 변경

수익성

군집 수 변경 : 3개 -> **2개**

[저위험]

- 매출총이익률: **20% 이상**
- 순이익률 : 0% 이상
- ROE : 1년 연속 15% 이상

[고위험]

- 매출총이익률: **20% 미만**
- 순이익률 : 0% 미만
- ROE : 1년 연속 15% 미만

변동성

군집 수 변경 : 3개 -> **2개**

[저위험]

- VHF : 50 이하
- 등락률 : 연 기준 3% 이상
- 베타값 : 0~1 미만 (절대값)

[고위험]

- VHF : 50 초과
- 등락률 : 연 기준 3% 미만
- 베타값 : 절대값 1 이상

성장성

군집 수 변경 없음. **연평균 순이익 증감률 중/고위험 기준 변경**

[중위험]

- 3년간 연평균 매출 성장률: 0 ~ 15 %
- 연평균 순이익 증감률: **0 미만**

[고위험]

- 3년간 연평균 매출 성장률: 0 % 이하
- 연평균 순이익 증감률: **0미만**

데이터분석 - Hamming Distance 기반의 클러스터 그룹화

클러스터의 벡터화

• 각 종목이 속한 클러스터를 벡터화

```
from sklearn.metrics.pairwise import cosine_similarity

# 각 클러스터의 조합을 벡터로 저장
df = pd.DataFrame({
    'f_cluster': new_df['f_cluster'],
    'p_cluster': new_df['p_cluster'],
    'v_cluster': new_df['v_cluster'],
    'g_cluster': new_df['g_cluster'],
    'd_cluster': new_df['d_cluster']
})

# 각 row의 클러스터 값을 하나의 벡터로 변환
cluster_vectors = df.values
```

• 해밍 거리(Hamming distance) 측정

- 각 종목 간의 클러스터 조합의 유사도를 측정
- 특정 임계값(0.6)을 설정하여, 유사도가 높은 클러스터를 동일한 그룹으로 그룹화

해밍 유사도 기반 그룹화 ⇒ 총 15 그룹 생성

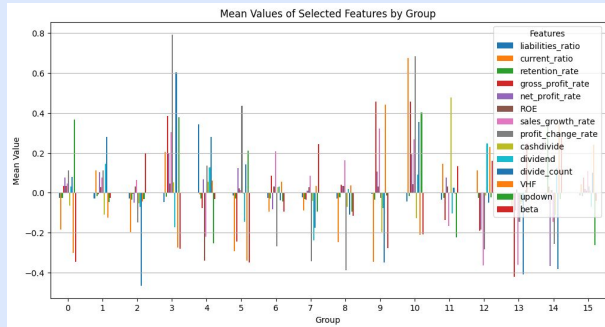
큰 그룹은
나누고,
작은 그룹은 병합

group	
4	600
1	397
3	165
0	145
2	109

group	
0	100
1	100
2	100
3	100
4	100
6	100
11	100
12	100
7	100
13	100
14	100
15	100
8	97
10	65
5	45
9	9

→ 제거

<1,2차 데이터 Grouping>



<15개의 최종 그룹별 데이터 분포>

기대 효과

개인 투자자



- 보유 주식에 관한 재무 정보를 이해하기 쉬운 형태로 제공함으로써 투자자의 금융 지식 함양 및 재무 정보 기반의 투자 습관 형성 기대
- 고객의 보유 자산 특징뿐만 아니라 개별 투자자의 성향을 고려한 ‘맞춤형 종목 및 뉴스 추천’ 기능을 통해 추천 서비스 만족도 향상

기대 효과

금융/증권 기관



- 고객의 금융 행동과 선호도 기반 맞춤형 서비스 개발에 중요한 기초 자료로 활용 가능
- 재무정보 기반 종목 & 뉴스 추천 기능을 새롭게 도입함으로써 서비스 활성화 및 신규 투자자 유입에 기여
- 15개의 유사한 특징을 지닌 종목 Group별 특징을 소개하는 재무 정보 기반 콘텐츠 기획을 통해 브랜드 이미지 구축에 기여

기대효과

투자자



- 보유 주식에 관한 재무 등급 및 정보를 제공함으로써 투자자의 금융 지식 함양
- 고객의 투자 성향 및 위험 선호도에 맞춘 종목과 기사 추천으로 고객 만족도 향상
- 뉴스 필터링 서비스를 통해 정보 과부하 해소 및 관심 정보를 빠르게 습득

금융 & 증권사



- 고객의 금융 행동과 선호도 기반 맞춤형 서비스 개발에 중요한 기초 자료로 활용 가능
- 재무정보 기반 종목 & 뉴스 추천 기능을 통해 서비스 활성화 및 신규 투자자 유입에 기여