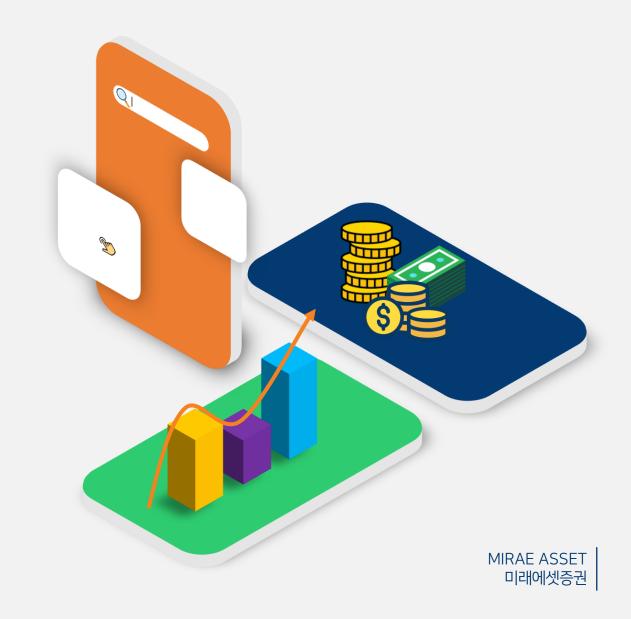
유사도 기반 양방향 종목 추천시스템



Contents

유사도 기반 양방향 종목 추천시스템

채지피티

윤지영 채소연 채희지

01 서론

아이디어 개요

문제 배경

분석 목표

서비스 상세 소개

분석 흐름

02 본론

데이터 소개

모델 소개

국내 데이터

미국 데이터

03 서비스시연

04 결론

종목 추천 결과 예시

서비스 상세 소개

Appendix

유사도 기반 양방향 종목 추천시스템 채지피티

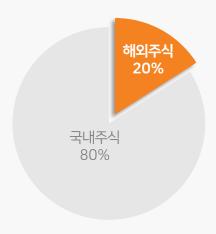
아이디어 개요

아이디어 수립을 위한 문제 파악

1 국내 주식 시장의 특징

주식 용어를 익히고 변동하는 내수 시장에 대한 이해가 <mark>필수적임</mark>

미국 경제에도 크게 영향을 받아 두 시장에 대한 종합적인 고려 필요



2030 주식 투자자 중 20%만 해외 주식 매매 경험 有 (출처: 신한금융투자)

아이디어 수립을 위한 문제 파악

1 국내 주식 시장의 특징

주식 용어를 익히고 변동하는 내수 시장에 대한 이해가 필수적임

미국 경제에도 크게 영향을 받아 두 시장에 대한 종합적인 고려 필요

이국 vs. 국내 주식 시장 간의 차이

주식 시장의 규모

주가 변동의 **상·하한** 존재 여부

위험과 수익에 대한 관점에 따라 **하나의 주식** 시장으로 투자 편중

아이디어 수립을 위한 문제 파악

M-STOCK 경쟁사 분석 서비스의 한계

경쟁사 선정 기준으로 <mark>시가총액</mark>만 고려된다는 것을 확인함

경쟁사로 묶인 기업들에 대해 제공되는 <mark>정보의 양</mark> 이 충분하다고 보기 어려움

회사규모 경쟁사 자산총계 데이터 기준일: 마지막 금요일 아이온큐 ⓒ 시가총액 : 발행주식수에서 자기주식수를 차감 한 주식수로 계산된 시가총액 닫기 자본총계 시세 및 주가배수 단위: USD, 백만, %, 배 (?) 부채총계 케이월드컴 아이온큐 현재가 17.93 28.20 0.00 시가총액 3,122 15 **Valuation** 목표주가 11.50 **PER** 상승여력 -25.95 아이온큐 케이… 투자의견 O_Weight 디지… 95.74 PER(LTM) **PBR** 5.63 2.42 아이온큐 배당수익률 0.00 0.00 0.00 디지…

M-STOCK 앱 > 메뉴 > 주식 > 투자 정보 > 미국리서치 > 종목분석 > 경쟁사분석

단위 : USD, 백만

587.92

23.85

552.24 17.26

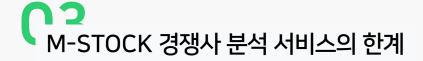
35.69

단위 : USD, 배

95.74

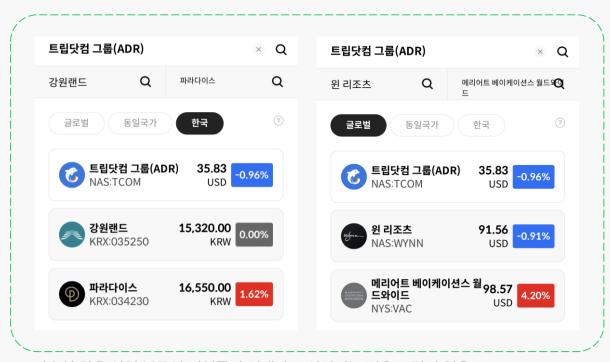
5.63 2.42

아이디어 수립을 위한 문제 파악



경쟁사 분석 시 고려되는 선정 기준이 모호함

주식에 영향을 미치는 <mark>다양한 정보</mark>들이 적절히 반영되고 있는지 **알** 수 **없음**



단순히 같은 산업 분류인 기업들이 경쟁사로 선정되는 경우도 많이 있음

분석 목표

분석의 방향성 정의



미국 기반 국내 종목 추천

현재 시행중인 M-Stock의 경쟁사 분석 서비스 개선



사용자가 직접 기준을 필터링하여 기준에 따라 1:1 비교 결과 확인



뉴스 키워드, 감정 점수 등 다양한 방면의 데이터 종합 비교



국내 기반 미국 종목 추천

초보 주식 투자자들을 타겟으로 한 **초기 투자 유입** 정보 제공 서비스



국내 종목의 정보를 바탕으로 쉽게 미국 종목의 정보 습득 가능



객관적 정보 제공으로 <mark>투자 가능한 주식 시장의 범위 확대</mark>

분석 목표

문제 개선 및 새로운 서비스 구축을 위한 분석 목표 설정



유사도 기반 양방향 주식 종목 추천 서비스 구축

미국 기반 국내 종목 추천 서비스의 종합적 개선 + 국내 기반 미국 종목 추천 서비스 구축



언어의 장벽 & 정보 접근성의 한계점 해결



각 시장의 특성을 고려해 세부 종목들에 대해 구체적으로 정보 제공



시장 간 경계를 허물어 다양한 종목 비교 가능

서비스 상세 소개

각 서비스별 타겟층 소개



미국 기반 국내 종목 추천 서비스



윤미국 (미국 주식 전문 투자자)

미국 주식으로 돈을 좀 벌었는데, 비슷한 국내 주식 종목에도 투자를 해볼까?



미국 주식 투자 경험이 있는 고객들이 유사한 국내 주식 종목 투자 또한 고려해볼 수 있도록 하는 서비스



서비스 상세 소개

각 서비스별 타겟층 소개



국내 기반 미국 종목 추천 서비스 🏻 😂 → 👛



채국내 (미국 주린이)

미국 주식에 투자를 해보고 싶은데, 미국 기업에 대해서는 아는 게 별로 없어. 국내 기업들처럼 쉽게 정보를 접할 수 있다면 좋을 텐데.

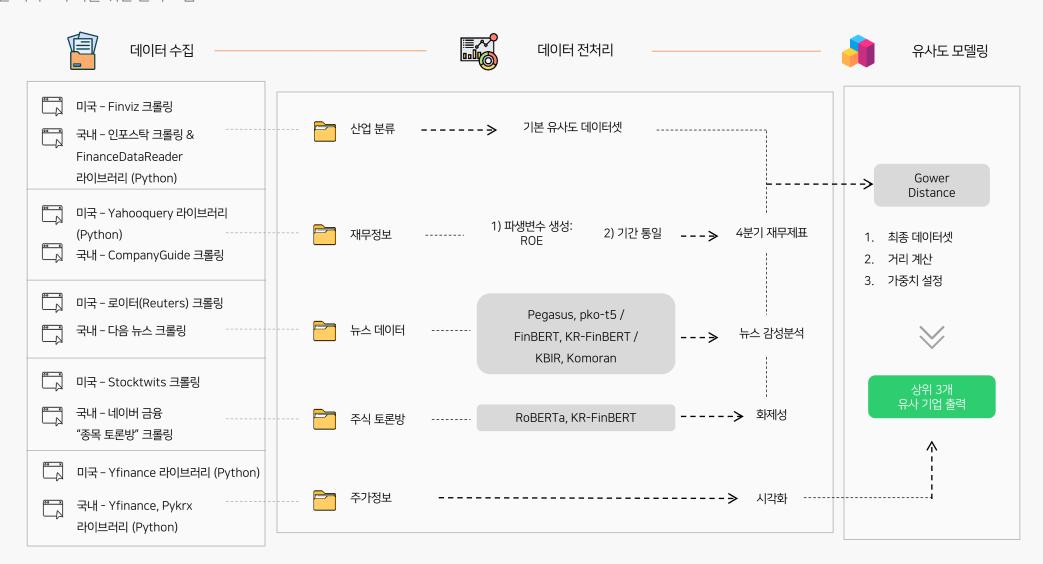


SERVICE

비교적 친숙한 국내 기업과의 비교를 통해 미국 기업에 대한 심적인 거리감을 줄이고 정보에 대한 접근성을 높이는 서비스

분석 흐름

전체적인 서비스 구축을 위한 분석 흐름



유사도 기반 양방향 종목 추천시스템 채지피티

데이터 소개



데이터 수집 -----

최종 데이터셋 구축

산업 분류

미국 - Finviz 크롤링 미국_산업 분류(최종).csv [4882 X 4]

국내 - 인포스탁 크롤링 & FinanceDataReader 라이브러리 (Python) 국내_산업 분류(최종).csv [3976 X 4]

▶ 147개의 공통된 산업 분류 생성

재무제표

미국 - Yahooquery 라이브러리 (Python) 미국 재무제표.csv [08.31 기준]
국내 - CompanyGuide 크롤링 국내 재무제표.csv [08.31 기준]

2022.06 - 2023.03 (분기별) 데이터 수집

뉴스 데이터

[소형주 제거 알고리즘]

크롤링 과정에서 규모가 작은 기업들은 재무정보나 주가정보, 뉴스 등에서 NA값이 많았고 시가 총액이 낮거나 거래가 활발하지 않으면 투자 리스크 측면에서 너무 위험이 커질 것 같아 제거

종목 토론방

" 미국 - Stocktwits 크롤링 Stocktwits 크롤링(.pickle) [08.31 기준]
" 국내 - 네이버 금융 '종목 토론방' 크롤링 네이버 종토방 크롤링(.xlsx) [08.31 기준]

주가 데이터

미국 - Yfinance 라이브러리 (Python) 미국 주가 데이터(.csv) [08.31 기준]

국내 - Yfinance, Pykrx 라이브러리 (Python) 국내 주가 데이터 (.csv) [08.31 기준]

▶ 2023.08.31 기준 약 <mark>2년치</mark> 데이터 수집

유사도 기반 양방향 종목 추천시스템 채지피티

모델 소개



국내 데이터

국내 뉴스 데이터 & 국내 종목 토론방

요약 모델



Pko-t5

OOV(Out-of-Vocabulary)를 해결한 BBPE(Byte level Byte Pair Encoding)을 사용한 모델



감성분석 모델



(R-FinBert

BERT 모델 중 한국 금융 분야에 특화된 모듈

키워드 추출 모델



Komoran

Java 기반의 대표적인 오픈소스 형태소 분석기

요약 진행 후 사용

뉴스 기사 감성분석의 경우, 전문을 모두 사용할 시 불필요한 내용들이 섞여 정확한 감성분석이 어려울 것이라 판단

Example.

검색어: 금호건설

제목: 회사채 발행 속속 성공하는 건설사들…"미래 리스크 대비" [투자360]

기사내용(전처리): 연합 헤럴드경제신동윤 기자 원자재 가격 상승과 부동상 프로젝트파이 낸싱PF 부실 우려 속에서 향후 발생할 수 있는 리스크 대비에 나섰던 건설사들이 속속회사채 발행에 성공하고 있다30일 금융투자업계에 따르면 최근 …

요약 column



'건설 경기가 바닥을 치고 반등할 것이란 기대 심리 덕분에 건설채에 투자하려는 투자자들이 늘면서 회사채 시장도 활기를 띨 것이란 기대감이 높아지는 상황이다.'

국내 데이터

국내 뉴스 데이터 & 국내 종목 토론방

요약 모델

JEWS

Pko-t5

OOV(Out-of-Vocabulary)를 해결한 BBPE(Byte level Byte Pair Encoding)을 사용한 모델

감성분석 모델





KR-FinBert



키워드 추출 모델



Komorar

Java 기반의 대표적인 오픈소스 형태소 분석기

│ 감성분석 결과 Sentiment Analysis

label: positive / neutral / negative

score: 0 ~ 1 (정도값)



기사 개수에 따라 가중치 적용

[{'label': 'positive',

'score': 0.9872146844...}]

감정분포

해당 종목의 전체 기사들에 대해 각 감정 label (positive / neutral / negative)의 분포 계산

감정지수

감정 label별 score 값의 평균을 계산

→ 두 수치를 사용하여 점수화

국내 데이터

국내 뉴스 데이터 & 국내 종목 토론방

요약 모델

NEWS

Pko-t5

OOV(Out-of-Vocabulary)를 해결한 BBPE(Byte level Byte Pair Encoding)을 사용한 모델

감성분석 모델







KR-FinBert

BERT 모델 중 한국 금융 분야에 특화된 모듈

키워드 추출 모델





차트분석 시각화

주가 정보와 해당 기간의 관련 키워드 한 눈에 보기

Komoran

Java 기반의 대표적인 오픈소스 형태소 분석기





종목 토론방

워드클라우드 시각화

주식 토론방의 상위 50개의 키워드 한 눈에 보기

미국데이터

미국 뉴스 데이터 & 미국 종목 토론방

요약 모델



PEGASUS

중요 문장 단위로 마스킹(Masking)을 진행하여 학습을 수행하는 요약 특화 모델

감성분석 모델



FinBert

국내 뉴스와 동일한 이유로 FinBERT 모델 사용

RoBERTa



감성분석 모델

키워드 추출 모델



KBIR

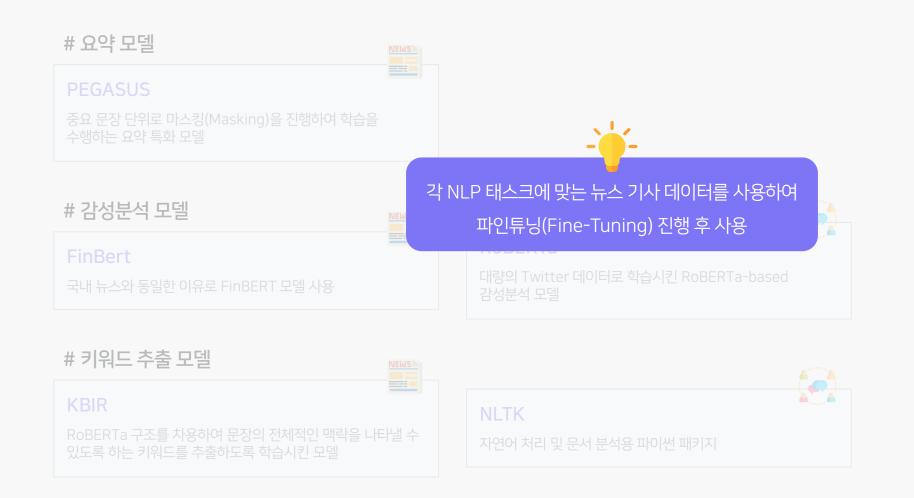
RoBERTa 구조를 차용하여 문장의 전체적인 맥락을 나타낼 수 있도록 하는 키워드를 추출하도록 학습시킨 모델

NLTK

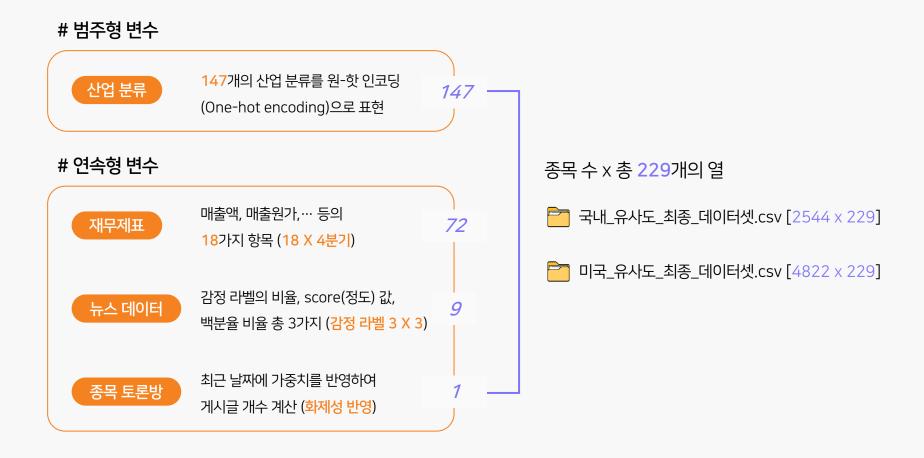
자연어 처리 및 문서 분석용 파이썬 패키지

미국데이터

미국 뉴스 데이터 & 미국 종목 토론방



미국 데이터와 국내 데이터 간의 유사도 계산 진행



미국 데이터와 국내 데이터 간의 유사도 계산 진행

범주형 변수

산업 분류

147개의 산업 분류를 원-핫 인코딩 (One-hot encoding)으로 표현

연속형 변수

재무제표

매출액, 매출원가,… 등의 18가지 항목 (18 X 4분기)

뉴스 데이터

감정 라벨의 비율, score(정도) 값, 백분율 비율 총 3가지 (<mark>감정 라벨 3 X 3</mark>)

종목 토론방

최근 날짜에 가중치를 반영하여 게시글 개수 계산 (<mark>화제성 반영</mark>)

Gower's Distance 가워거리

가워 차이 계수(Gower's dissimilarity coefficient)라고도 하며, 연속형 변수와 범주형 변수를 모두 갖는 혼합 데이터(Mixed Data)일 때 두 데이터셋 간의 거리 차이를 측정하는 방법

$$D_{x_i x_j} = \frac{\sum_{r=1}^{p} W_{x_i x_j z_r} D_{x_i x_j z_r}}{\sum_{r=1}^{p} W_{x_i x_j z_r}} + \frac{\sum_{r=1}^{q-p} W_{x_i x_j c_r} D_{x_i x_j c_r}}{\sum_{r=1}^{q-p} W_{x_i x_j c_r}}$$

W: 가중치값 p: 범주형 변수 개수 q: 연속형 변수 개수

▶ 범주형 변수
$$D_{x_ix_jz_r} = \begin{cases} 0, & z_r^i = z_r^j \\ 1, & otherwise \end{cases}$$

• 연속형 변수
$$D_{x_i x_j c_r} = \frac{\left|c_r^i - c_r^j\right|}{\max(c_r) - \min(c_r)}$$

미국 데이터와 국내 데이터 간의 유사도 계산 진행

01.

가중치 필터 선택

산업 분류, 감정지수, … 등 사용자가 가중치를 주고 싶은 필터를 자유롭게 선택

미국 데이터와 국내 데이터 간의 유사도 계산 진행

01.

가중치 필터 선택

산업 분류, 감정지수, … 등 사용자가 가중치를 주고 싶은 필터를 자유롭게 선택 02.

가워 거리 계산

필터 값에 가중치를 곱한 후 기준이 되는 Target 기업과 다른 기업들간의 가워 거리 계산

미국 데이터와 국내 데이터 간의 유사도 계산 진행

01.

가중치 필터 선택

산업 분류, 감정지수, … 등 사용자가 가중치를 주고 싶은 필터를 자유롭게 선택

02.

가워 거리 계산

필터 값에 가중치를 곱한 후 기준이 되는 Target 기업과 다른 기업들간의 가워 거리 계산 03.

종목 추천

Target 기업을 기준으로 유사도가 높은(거리 결과 값이 작은) 상위 3개 기업 추천

유사도 기반 양방향 종목 추천시스템 채지피티

서비스 시연



유사도 기반 양방향 종목 추천시스템 채지피티

종목 추천 결과 예시



○ 셀트리온

론 **4** 결 론



산업 분류를 고려하여 동일 산업 내에서 기업별로 어떠한 공통점과 차이점을 갖는지 확인 가능

피F-	리	항목
_	10	\circ

유사 기업

산업 분류, 재무제표 종합

SGEN (Seagen Inc)
INCY (Incyte Corp)
REGN (Regeneron Pharmaceuticals Inc)

산업 분류, 매출액, 영업이익, PBR, PER EXEL (Exelixis Inc)

UTHR (United Therapeutics Corp)

BMRN (Biomarin Pharmaceutical Inc)

산업 분류, 매출액, 영업이익, PBR, PER, 당기순이익, EV/EBITDA, 뉴스 감정점수, 화제성 RPRX (Royalty Pharma plx)
REGN (Regeneron Pharmaceuticals Inc)
NVAX (Novavax, Inc.)

종합 (<mark>산업 분류</mark>, 재무제표 종합, 뉴스 감정점수, 화제성) SGEN (Seagen Inc)
RPRX (Royalty Pharma plx)
INCY (Incyte Corp)



기업별 특성을 바탕으로 유사 기업 선정 기준을 선택할 수 있도록 하여 타 산업 분류에 속한 기업과의 비교 가능

필터링 항목	유사 기업	
재무제표 종합	MPWR (Monolithic Power Systems Inc.)	
	VNO (Vornado Realty Trust)	
	TYL (Tyler Technologies, Inc.)	
매출액, 영업이익,	NGVT (Ingevity Corp)	
	CACC (Credit Acceptance Corp.)	
PBR, PER	EXEL (Exelixis Inc)	
매출액, 영업이익, PBR, PER, 당기순이익, EV/EBITDA, 뉴스 감정점수, 화제성	CABO (Cable One Inc)	
	RRR (Red Rock Resorts Inc)	
	EXAS (Exact Sciences Corp)	
	ENOV (Enovis Corp)	
재무제표 종합, 뉴스 감정점수, 화제성	HP (Helmerich and Payne Inc)	
	MPWR (Monolithic Power Systems Inc.)	



지호정(48) 간호사 / 제약 관련 전문 투자자 " 미국의 제약 산업 내에서는 또 어떤 이슈가 있나요? "

Situation

이번에 셀트리온의 호재를 뉴스에서 접했어요 이 뉴스는 국내뿐만 아니라 미국에서도 주목하니 미국의 제약 관련 기업들에 대해서 정보를 얻고 싶어요



4 결 론



산업 분류

뉴스



지호정(48) 간호사 / 제약 관련 전문 투자자 " 미국의 제약 산업 내에서는 또 어떤 이슈가 있나요? "

Situation

이번에 셀트리온의 호재를 뉴스에서 접했어요 이 뉴스는 국내뿐만 아니라 미국에서도 주목하니 미국의 제약 관련 기업들에 대해서 정보를 얻고 싶어요

Seagen Inc.

SGEN | 생명공학 | NASDAQ

92.8486.4897.2384.27종합점수재무제표뉴스화제성

Incyte Corporation

INCY | 생명공학 | NYSE

92.2182.5996.8286.82종합점수재무제표뉴스화제성

Regeneron Pharmaceuticals, Inc.

REGN | 생명공학 | NYSE

91.6581.9895.7280.36종합점수재무제표뉴스화제성





지호정(48) 간호사 / 제약 관련 전문 투자자

" 미국의 제약 산업 내에서는 또 어떤 이슈가 있나요? "

Situation

이번에 셀트리온의 호재를 뉴스에서 접했어요 이 뉴스는 국내뿐만 아니라 미국에서도 주목하니 미국의 제약 관련 기업들에 대해서 정보를 얻고 싶어요





지호정(48) 간호사 / 제약 관련 전문 투자자 " 미국의 제약 산업 내에서는 또 어떤 이슈가 있나요? "

Situation

이번에 셀트리온의 호재를 뉴스에서 접했어요 이 뉴스는 국내뿐만 아니라 미국에서도 주목하니 미국의 제약 관련 기업들에 대해서 정보를 얻고 싶어요



상세보기 - 뉴스



4 결 론

Q 셀트리온



화제성

매출액

매출총이익

PER

PBR



황기혁(34) IT 종사자 / 쏠쏠한 주식 수익 보유 " 요즘 미국 종목 중에서 셀트리온과 규모가 비슷하면서 화제성 높은 종목이 뭐가 있어요? "

Situation

이번에 셀트리온이 호재로 화제가 되면서 수익이 50% 이상 났어요! 미국 주식 투자가 더 수익이 잘 난다고 하던데 화제성 높은 미국 주식에 투자해서 더 많은 수익을 내보고 싶어요!

Exact Sciences Corporation

EXAS | 헬스케어 서비스 | NASDAQ

92.8486.4897.2394.27종합점수재무제표뉴스화제성

SoFi Technologies, Inc.

SOFI | 금융 서비스 | NASDAQ

92.2182.5985.8296.82종합점수재무제표뉴스화제성

Teladoc Health, Inc.

TDOC | 생명공학 | NYSE

91.6581.9895.7291.36종합점수재무제표뉴스화제성



황기혁(34) IT 종사자 / 쏠쏠한 주식 수익 보유 " 요즘 미국 종목 중에서 셀트리온과 규모가 비슷하면서 화제성 높은 종목이 뭐가 있어요? "

Situation

이번에 셀트리온이 호재로 화제가 되면서 수익이 50% 이상 났어요! 미국 주식 투자가 더 수익이 잘 난다고 하던데 화제성 높은 미국 주식에 투자해서 더 많은 수익을 내보고 싶어요!



4 결 론

셀트리온 X

산업 분류

화제성

매출액

매출총이익

PER

PBR

Novavax, Inc.

NVAX | 생명공학 | NASDAQ

85.14 종합점수

재무제표

85.64 87.28 뉴스

98.27 화제성

SRPT | 생명공학 | NASDAQ

Sarepta Therapeutics, Inc.

84,36

86.84 79.72

94.11 뉴스

종합점수

재무제표

화제성

Seagen Inc.

SGEN | 생명공학 | NASDAQ

84,24 종합점수

87.06 재무제표

97.23

뉴스

84.27 화제성

4 결 론

뉴스 산업 분류

화제성

셀트리온

매출액

매출총이익

X

PER

PBR

Regeneron Pharmaceuticals, Inc.

REGN | 생명공학 | NYSE

91.6584.8895.7280.36종합점수재무제표뉴스화제성

Moderna, Inc.

MRNA | 생명공학 | NASDAQ

86.9784.6887.8886.82종합점수재무제표뉴스화제성

Gain Therapeutics, Inc.

GANX | 생명공학 | NASDAQ

 86.61
 88.44
 85.29
 81.35

 종합점수
 재무제표
 뉴스
 화제성

-▼ 결 론

서비스 키워드

서비스 주요 키워드 한 눈에 확인해보기

상세한 정보 제공

- 단순히 유사하다는 단편적인 정보 제공에 그치지 않음
- 어떤 부분에서 유사한지 항목별 상세 비교 제공

사용자 편의성 증진

- 사용자가 중점적으로 고려하고 싶은 항목 자유롭게 선택
- 사용자가 얻을 수 있는 정보의 범위를직접 선택 가능하도록 함

주식 투자 범위 확대

- 미국 및 국내 시장의 신규 투자자 유입 확대에 도움
- 현재 시행되고 있는 서비스들과의 시너지 효과

유사도 기반 양방향 종목 추천시스템 채지피티

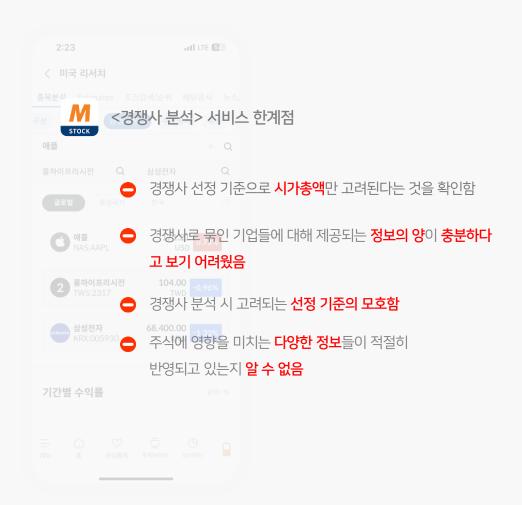
기대효과 및 의의



▼ ▼ 결 론

기존 서비스 한계 개선

서론에서 언급된 기존 서비스 한계점을 개선하여 채지피티만의 새로운 서비스 제안



제안 서비스

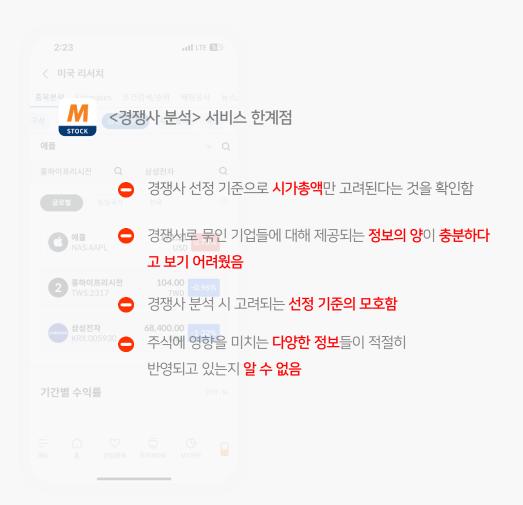


유사도 계산 시 충분한 정보 사용

4 결 론

기존 서비스 한계 개선

서론에서 언급된 기존 서비스 한계점을 개선하여 채지피티만의 새로운 서비스 제안



제안 서비스



유사도 계산 시 충분한 정보 사용



사용자가 보기에 모호하지 않은 직관적인 유사도 점수 결과 제공

4 결 론

기대효과

고객 유인 및 접근성 증대





현재 한 시장에만 투자하고 있는 고객층 을 대상으로 투자 범위를 넓힐 수 있게끔 유도하는 것이 필요!

◀ 결 론

기대효과

고객 유인 및 접근성 증대

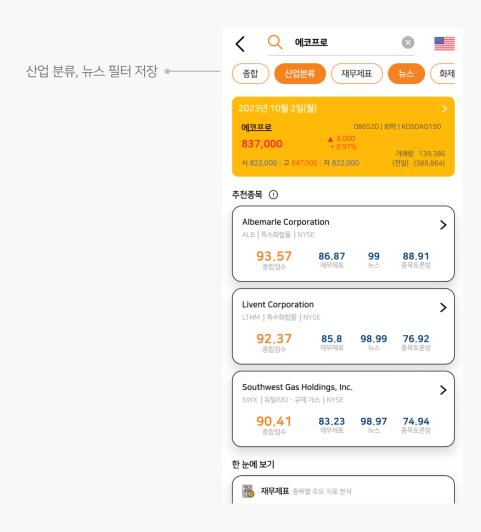


- ✓ 친숙한 대상과 연결지어 쉽고 자연스럽게 서비스를 접할 수 있음
- ✓ 사용자들이 국내 및 미국 주식에 쉽게 입문할 수 있도록 발판으로서의 역할 기대

-▼ 결 론

발전 가능성

기존 서비스와 접목시켜 더 나은 시너지 효과를 낼 수 있음



서비스 사용자의 데이터 수집 후, 사용자들이 주로 택한 필터링 항목을 저장 하여 자동으로 유사 기업 추천 가능

발전 가능성

기존 서비스와 접목시켜 더 나은 시너지 효과를 낼 수 있음



사용자의 투자성향을 파악하여 사용자 맞춤 서비스 제공 가능

네이버 클라우드 & 미래에셋이 준비하는 '투자 비서' 서비스와도 활용 가능!

발전 가능성

기존 서비스와 접목시켜 더 나은 시너지 효과를 낼 수 있음

<종목의 발견> 테마

기존 <종목의 발견> 서비스에서 제공하는 종목에 대한 정보를 바탕으로 제안하는 서비스를 적용하여 사용자들에게 보다 다양한 정보를 한 눈에 접할 수 있도록 다채롭게 활용 가능



◀ 결 론

기대효과 정리

투자자 친화적인 주식 투자 환경 제공 이미지 확보

각 시장으로의 초기 투자 유입

⁶⁶ 유사도 기반 양방향 주식 종목 추천 서비스 ⁶⁰

타 서비스 보완 효과

투자 시장 범위 확대를 위한 징검다리 역할

개인 맞춤 정보제공

기대효과 정리

◀ 결 론

각 시장으로의 초기 투자 유입

투자자 친화적인 주식 투자 환경 제공 이미지 확보

♣ 유사도 기반 양방향 주식 종목 추천 서비스

타 서비스 보완 효과

투자 시장 범위 확대를 위한 징검다리 역할



기대효과 정리

◀ 결 론

각 시장으로의 초기 투자 유입

투자자 친화적인 주식 투자 환경 제공 이미지 확보

⁶⁶ 유사도 기반 양방향 주식 종목 추천 서비스 ⁶⁰

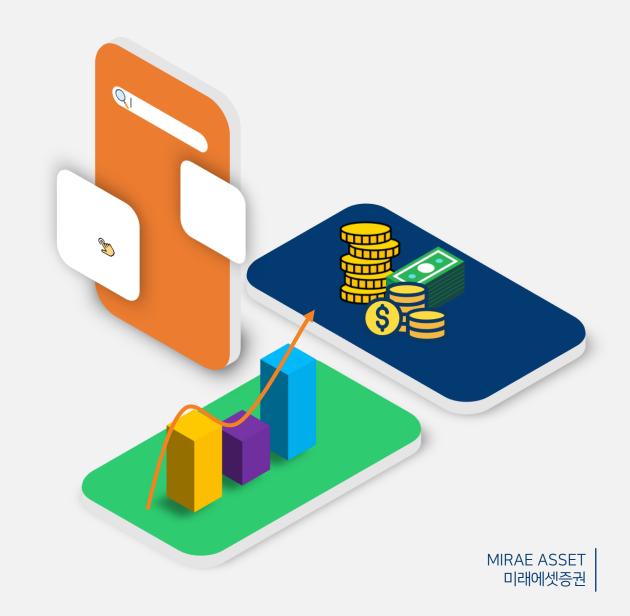


타 서비스 보완 효과

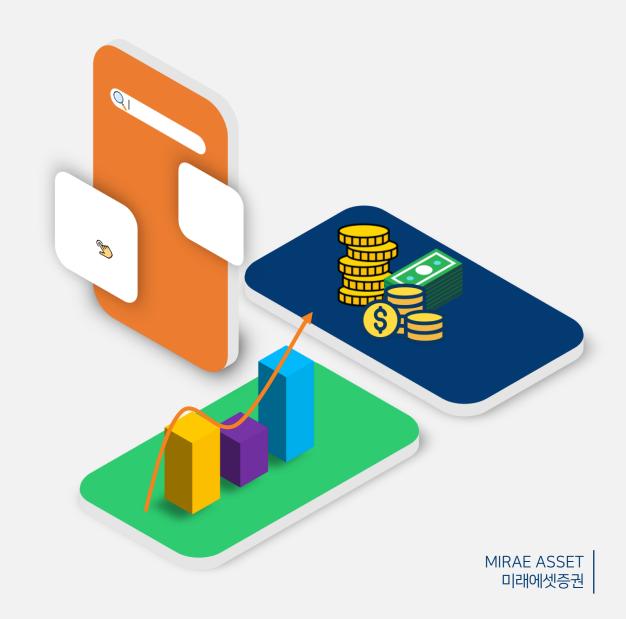
투자 시장 범위 확대를 위한 징검다리 역할

개인 맞춤 정보제공

감사합니다



QnA



유사도 기반 양방향 종목 추천시스템 채지피티

Appendix



서비스 정리



개인 맞춤 필터링



3 추천 종목 유사도 결과



뉴스 감성분석





재무제표

화제성

서비스 정리 - 상세보기

