

2021년(제2회) NH투자증권 빅데이터 경진대회 본선

NAMUH 키우기



휴먼스럽게

 NH투자증권

박채희, 김진성, 한현수

목차

- 과제소개
- 데이터 탐색 결과
- 알고리즘 설명
- 서비스 아이디어 제안
- 아이디어 소개
- 서비스 전략 및 기대효과
- 한계점 및 발전가능성

- 과제소개
- 데이터 탐색 결과
- 알고리즘 설명
- 서비스 아이디어 제안
- 아이디어 소개
- 서비스 전략 및 기대효과
- 한계점 및 발전가능성

과제소개

"데이터, 문화가 되다"

- 2021년(제2회) NH투자증권 빅데이터 경진대회

데이터 결합을 통한 주식 보유기간 예측 및 서비스 아이디어 제안

- NH투자증권 고객의 주식 보유기간을 예측하고

이를 활용할 수 있는 서비스 아이디어 제안



- 과제소개
- 데이터 탐색 결과
- 알고리즘 설명
- 서비스 아이디어 제안
- 아이디어 소개
- 서비스 전략 및 기대효과
- 한계점 및 발전가능성

데이터 탐색 결과

기업의 분기별 실적 발표에 따라, 주식 거래량이 증가
 -> 분기시점에 매도, 매수가 많아진다는 점을 활용해서 주식 보유기간을 예측해보자!

≡ 매일경제 뉴스 | 경제 기업 사회 국제 부동산 증권 정치 IT과학 문화

로그인 Q

[원포인트 투자 레슨] 분기별 실적 발표 보면 주식 성공의 길 열린다

직전 분기 성장률보다는 계절 호황 특수성 배제한 전년 동기대비 실적 중요
 실적 규모도 중요하지만 시장 전망치와의 격차가 주가 등락의 주요 원인

입력 : 2018.02.02 04:01:05 0



Q. 최근 수년 만에 코스피와 코스닥이 활황장이고, 주식투자에 대한 관심이 어느 때보다 높아지고 있습니다. 이런 상황에서 4분기 실적발표 시즌이 시작됐는데, 주식 투자자 입장에서 분기실적 발표를 어떻게 해석하고 활용해야 투자에 도움이 될까요?

이 질문에 대한 답변은 이미 모든 투자자가 알고 있다. 기본적으로 실적이 좋은 회사의 주가가 역시 좋다는 것이다. 조금 더 구체적으로 분기 실적을 활용할 수 있는 방법을 고민해 보자.

먼저, 분기실적에 어떤 지표를 어떻게 비교하며 봐야 할 것인가를



주요뉴스

캘리셰이드급 '아이오닉7...현대차 '세븐' 공-

해운업계 비명 "수년간 번돈 다 토해내란 얘기.

번호판 1억6000만원 낙찰...이 숫자 4개면 1억-

재난지원금 강행하는 이재명, 돈뿌리기인가 소-

여의도 시범아파트, 오늘 신속통합기획접수

인기뉴스

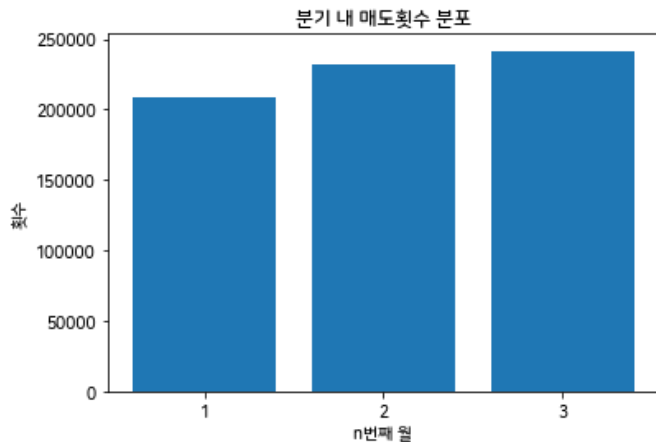
종합

연예

스포츠

데이터 탐색 결과

실제로 학습 데이터의 한 분기 내 매도 횟수 분포를 확인해 보았고,
한 분기내에서의 마지막달이 가장 매도매수 횟수가 많은것을 알아냄 (아래 그래프 참고)



한 분기내 1, 2, 3번째 달의 매도 횟수 합산 그래프

-> 분기에 속하는 매도 데이터의 경우 분기별로 나눠서 주식 보유기간 예측에 활용

데이터 탐색 결과

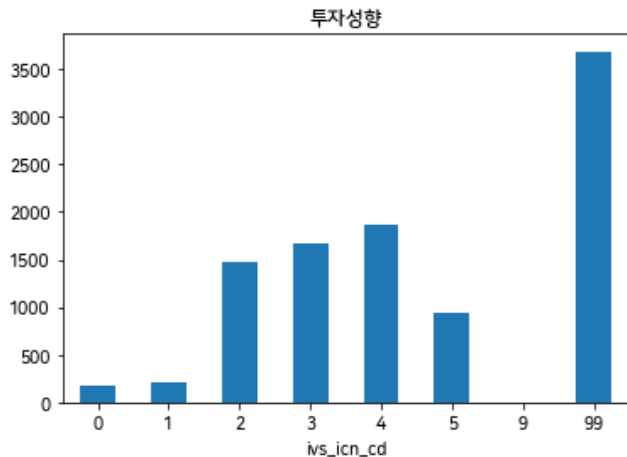
또한, 투자성향에 따라 주식 보유기간이 차이가 날 것 이다라고 판단
 -> 분기에 속하지 않은 매도 데이터의 경우 투자성향으로 주식 보유기간 예측에 활용
 (예시 - 구매일은 2018.07.02, 판매일은 2018.07.09 일 때 분기에 속하지 않는다고 판단)

고객 및 주거래계좌 정보(CUS_INFO.CSV)			
원천시스템명	한글시스템명	예시	비고
ACT_ID	계좌ID	27666E123D	사내 계좌번호를 토대로 임의 생성한 값으로 사용
SEX_DIT_CD	성별	01	1:남성 / 2:여성
CUS_AGE_STN_CD	연령구간	30	사내 CUS_AGE를 토대로 연령대 구분 01: 20세~25세미만 02: 25세~30세미만 03: 30세~35세미만 04: 35세~40세미만 05: 40세~45세미만 06: 45세~50세미만 07: 50세~55세미만 08: 55세~60세미만 09: 60세~65세미만
IVS_ICN_CD	투자성향	04	계좌 가입시 설문조사로 결정되는 투자성향 01: 안정형 02: 안정추구형 03: 위험중립형 04: 적극투자형 05: 공격투자형 09: 전문투자자형 00: 정보제공미동의 99: 미정의
CUS_AET_STN_CD	고객자산구간	01	고객총자산 금액을 구간별로 구분한 코드 01: 0원이상-1천만원미만 02: 1천만원이상-3천만원미만 03: 3천만원이상-5천만원미만 04: 5천만원이상-1억원미만 05: 1억원이상-3억원미만 06: 3억원이상

제공받은 데이터 스키마 중 일부

데이터 탐색 결과

하지만, 투자성향 데이터의 경우 약 37%가 99(미정의) 데이터로 구성됨 (아래 그래프 참고)



cus_info.csv의 ivs_icn_cd 분포 그래프

-> 데이터 탐색을 통해, 고객의 매도, 매수 횟수를 이용하여 고객의 투자성향을 재정의 해주기로 함

- 과제소개
- 데이터 탐색 결과
- **알고리즘 설명**
- 서비스 아이디어 제안
- 아이디어 소개
- 서비스 전략 및 기대효과
- 한계점 및 발전가능성

알고리즘 설명

처음 과정으로, 데이터 항목 정의서를 만들어 항목별로 데이터 탐색 결과를 작성하고, 데이터 전처리 과정을 진행

항목명 / 단위	한글명	데이터 타입	변경사항	비고
ACT_ID	계좌ID			
SEX_DIT_CD	성별			
CUS_AGE_STN_CD	연령구간			
IVS_ICN_CD	투자성향			
CUS_AET_STN_CD	고객자산구간			
MRZ_PDT_TP_SGM_CD	주거래상품군		삭제	단순 feature만 삭제 가 아닌, 국내주식 (label : 2) 외 데이터 삭제 처리 후, 해당 feature 삭제
LSG_SGM_CD	LIFESTAGE		삭제	삭제이유 코드 참고
TCQ_CUS_GRD_CD	고객등급		삭제	삭제이유 코드 참고
TOT_IVS_TE_SGM_CD	총투자기간			
MRZ_BTP_DIT_CD	주거래업종구분		삭제	x년도 12월기준 한 달 데이터만 파악하 기 때문에, 부족한 데이터라고 판단. 대신 실제 고객의 거 래 내역을 보고 주거 래업종을 파악해 보 기로 함
BUY_SELL_CNT	매도 매수 횟수	int	생성 / 삭제	inv_type, buy_sell_quant 생성 후 삭제
INV_TYPE	투자유형	1 : 방어형, 2 : 중립형, 3 : 공격형, 4 : 초단타	생성	생성이유 코드 주석 참고

작성한 데이터 항목 정의서 중 일부

알고리즘 설명

그 다음, 테스트 데이터에서 주식 보유기간을 예측하는 column인 "hold_d"의 값의 예측을 위해 "hist_d"의 값이 많은 영향을 준다고 판단 (baseline 코드 중 hint 참고)

-> 학습 데이터에는 생성되어 있지 않은 "hist_d"를 생성하는 것을 1차 타겟으로 하고,
 학습 데이터로 학습시킨 모델을 사용하여 테스트 데이터의 "hold_d"를 예측하는 것을 2차 타겟으로 정함

	act_id	iem_cd	byn_dt	hold_d		act_id	iem_cd	byn_dt	hist_d	submit_id	hold_d
0	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A006360	20180726	11	0	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A032640	20200522	153	DX00001	0
1	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A005930	20180131	80	1	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A160600	20190823	335	DX00002	0
2	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A005070	20180517	5	2	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A234340	20200617	139	DX00003	0
3	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A003520	20201112	22	3	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A131760	20200120	236	DX00004	0
4	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A002310	20180905	324	4	0ad104dbed99be0cd858aa772765ddedade554601a981b...	A293490	20201217	9	DX00005	0

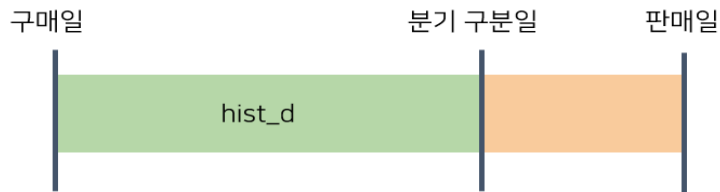
학습 데이터

테스트 데이터

1차 타겟인 hist_d 생성 방법

분기 구분일: 판매일 기준으로 직전의 분기를 구분

-> 구매일 부터 분기 구분일 까지의 주식 개장일을 계산하면 hist_d *주식 개장일은 외부 데이터를 사용



예시1) 구매일은 2018.06.02, 판매일은 2018.07.09 일 때
hist_d = 구매일~ 직전 분기인 2018.07.01 까지의 주식 개장일 수

예시2) 구매일은 2017.12.23, 판매일은 2018.07.09 일 때
hist_d = 구매일~ 직전 2분기인 2018.04.01까지의 주식 개장일 수

예시3) 구매일은 2018.07.02, 판매일은 2018.07.09 일 때
-> 분기에 속하지 않은 데이터이므로 고객의 투자성향을 활용하여 처리
hist_d = 고객 매도 매수 총 횟수 / 주식 운용 총 기간

알고리즘 설명

2차 타겟인 hold_d 예측 방법

생성한 hist_d와 여러 고객 정보를 바탕으로
머신러닝 알고리즘인 lgbm(Light Gradient boosting machine)을 사용하여
학습 데이터를 학습시킨 모델로 테스트 데이터의 hold_d(고객 주식 보유 일자)를 예측

	hist_d	yield	sex_dit_cd	cus_age_stn_cd	ivs_icn_cd	cus_aet_stn_cd	mrz_pdt_tp_sgm_cd	tot_ivs_te_sgm_cd	mrz_btp_dit_cd	buy_sell_quant
0	1.0	1.016279	1	9	3	2	2	5	8	0.408
1	8.0	0.991984	1	9	3	2	2	5	8	0.408
2	0.0	1.053773	1	9	3	2	2	5	8	0.408
3	2.0	1.055844	1	9	3	2	2	5	8	0.408
5	10.0	0.810225	1	9	3	2	2	5	8	0.408



LightGBM

모델
생성

데이터 전처리 후 hist_d 값 생성한 학습 데이터 중 일부

	hist_d	yield	sex_dit_cd	cus_age_stn_cd	ivs_icn_cd	cus_aet_stn_cd	mrz_pdt_tp_sgm_cd	tot_ivs_te_sgm_cd	mrz_btp_dit_cd	buy_sell_quant	..
4	9	0.953419	1	9	3	2	2	5	8	0.408	..
13	6	1.009829	1	6	4	4	2	6	1	0.388	..
14	11	1.009829	1	6	4	4	2	6	1	0.388	..
15	3	1.009829	1	6	4	4	2	6	1	0.388	..
19	8	1.009829	1	6	4	4	2	6	1	0.388	..

생성한
모델

submit_id hold_d

0	IDX00001	245.0
1	IDX00002	431.0
2	IDX00003	225.0
3	IDX00004	324.0
4	IDX00005	84.0

데이터 전처리 후의 테스트 데이터 중 일부

테스트 데이터에서 예측한 hold_d 값 중 일부

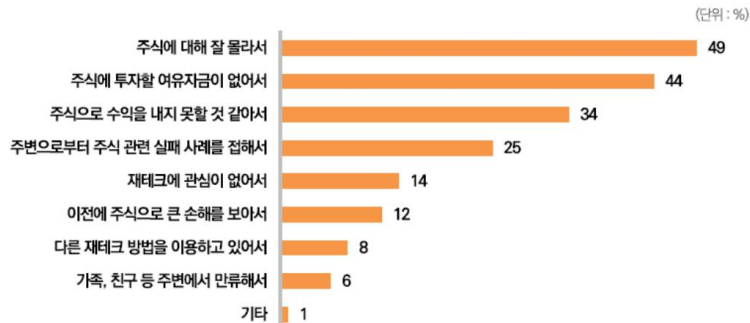
- 과제소개
- 데이터 탐색 결과
- 알고리즘 설명
- **서비스 아이디어 제안**
- 아이디어 소개
- 서비스 전략 및 기대효과
- 한계점 및 발전가능성

서비스 아이디어 제안

배경 : 주식을 시작하기 어려운 이유

주식을 하지 않는 사람들이 주식이란 단어를 들었을 때,
아래 이유와 같이 “주식을 어렵고 전문적이고 위험한 분야다” 라고 느끼기 때문에
주식을 선불리 시작하지 못한다는 문제점이 존재

주식을 하지 않는 이유, 주식에 대해 잘 몰라서 49%



질문 : 귀하께서 주식을 하지 않는 이유는 무엇입니까? (복수응답)

표본 수 : 주식 혹은 펀드를 하지 않는 사람 653명

조사기간 : 2020.11.27~11.30

한국리서치 정기조사 여론 속의 여론(hrcopinon.co.kr)

서비스 아이디어 제안

또한, 최근 MZ 세대가 특히 주식에 많은 관심을 갖고 투자하는 추세
하지만, MZ세대는 “투자”를 공부하고 하는 것이 아닌, 게임으로써 생각함



투자에 대한 MZ세대의 시각

- 1 그나마 가장 공정한 게임
사회적 배경 관계없이 동일한
투자 조건과 리스크
- 2 남이 하니 나도 해야 하는 것
디지털 기술 활용한 행태의 신속한
확대·재생산
- 3 부동산 구입을 위한 디딤돌
MZ세대 61% “재테크 이유는
주택 구입 재원 마련”

*자료=금융감독원

-> 이를 이용해서 주식에 다가가기 어려운 사람들을 위해,
주식을 쉽게 바라볼 수 있게 SNS, 게임 같이 하나의 문화로 만들자!

서비스 아이디어 제안

우리의 아이디어 <NAMUH 키우기>

- 퀴즈를 풀면서 자연스럽게 주식 공부를!
- 내가 매수 매도한 주식들이 나무로!
- 나무들을 모아모아 나만의 숲을!
- 나무가 클 때마다 보상도!



-> 주식에 관한 퀴즈를 풀면서 지식과 정보를 쌓음과 동시에
자신의 주식 매도 매수내역으로 나만의 홈페이지를 꾸며보며, 주식에 더욱 친근하게 접근 가능!

- 과제소개
- 데이터 탐색 결과
- 알고리즘 설명
- 서비스 아이디어 제안
- **아이디어 소개**
- 서비스 전략 및 기대효과
- 한계점 및 발전가능성

아이디어 소개

1. 주식을 사면 씨앗을 제공하고, 기업에 대한 업종을 맞추면 발아
-> 내가 매수한 기업에 대한 최소한의 지식 습득 제공



아이디어 소개

2. 나무를 키우는 방법

- 종목에 대한 정보로 퀴즈, 뉴스, 영상을 보면 경험치 제공
- 수익률에 따른 경험치 제공

(주식 매도시, 매도할때 수익률 기준 경험치 제공 / 주식 보유시, 매주 월 ~ 금요일 동안 얻은 수익률 기준 경험치 제공)



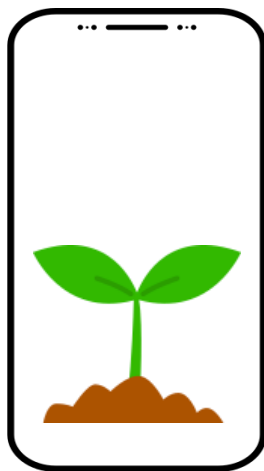
아이디어 소개

경험치에 따른 나무 6단계로 구성



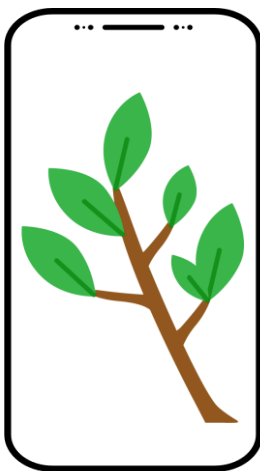
씨앗

->

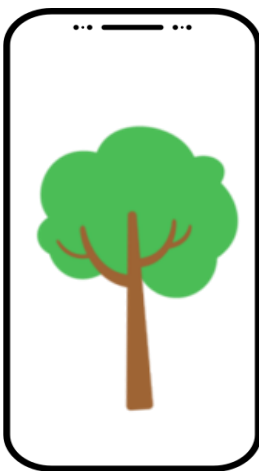


새싹

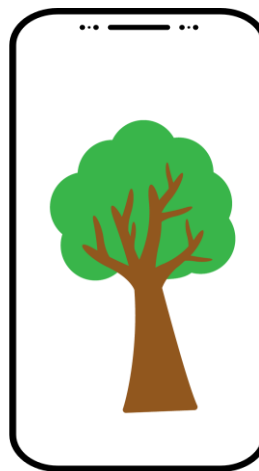
-> 나무 1단계



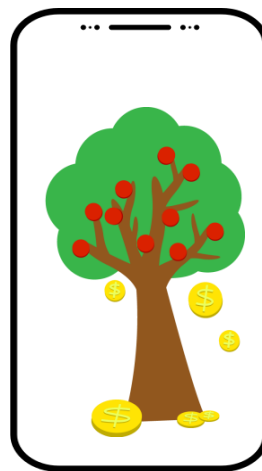
-> 나무 2단계



-> 나무 3단계



-> 열매열림



아이디어 소개

6단계 <열매> 단계가 되어 열매가 열리면?

-> 수확 (수수료면제권, 우대쿠폰, 주식 1주, ETF등등)



아이디어 소개

3. 매도한 새싹~나무들이 고객 각각의 나만의 숲으로 올려짐

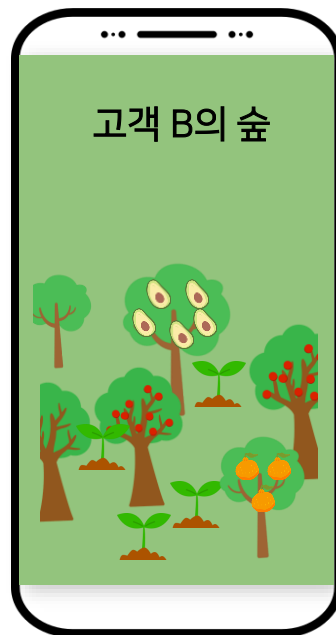
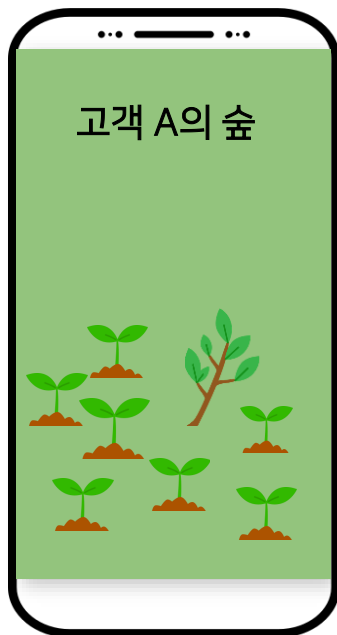


4. 고객의 투자성향 파악 가능

새싹이 많다 -> 단타 고객

vs

나무가 많다 -> 장투 고객



아이디어 소개

5. **주식 보유기간 예측**을 활용하여 월마다 핫한 테마주 + 해당 월에 개화하는 나무 매칭을 통한 매수 유도!

-> 매수한 주식과 예상 매도 날짜에 따라 수익률이 높은 테마주를 다양한 나무와 매칭해서 추천하며 활발한 매수 유도

예시) 고객의 주식 평균 보유일이 2개월 일때, 8월에는 여행시즌으로 항공주 주가 올라감

-> 8월에 열매가 맺히는 복숭아로 매칭하여 6월에 항공주를 추천

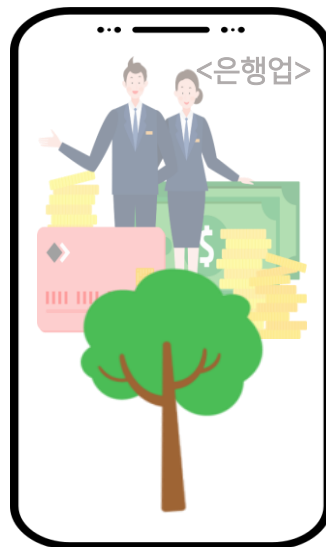


아이디어 소개

6. 제일 많이 구매한 업종의 테마로 배경을 꾸며줌

-> 고객이 가장 많이 매수한 업종의 테마를 바탕으로 고객의 개인 숲의 배경을 꾸며주어서,
직관적으로 자신의 매수 스타일 파악 가능

자신이 관심있는 업종의 배경을 가진 다른 고객의 숲을 보면서
해당 업종은 어떻게 매수, 매도 해야 하는지 배울 수도 있음

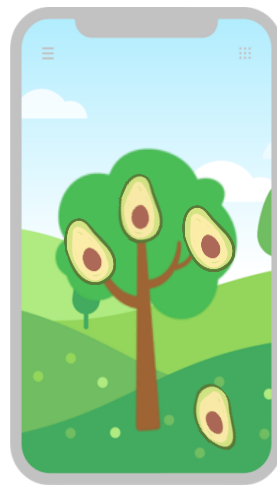
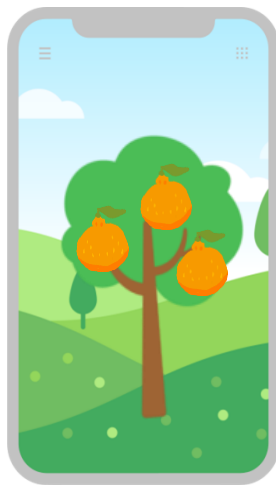


아이디어 소개

7. 국내주식은 한라봉, 해외주식 아보카도

-> 국내주식은 국내에서 유명한 과일로, 해외주식은 해외에서 유명한 과일(미국, 중국, 일본 등 상이하게)로 나타내서
고객의 거래 스타일을 파악

국내 혹은 해외 주식 한 곳만 하는 경우,
국내 또는 해외 주식을 사야지만 재배할 수 있는 과일들을 재배해서,
자신의 숲을 꾸미기 위해 다양한 시장에서 거래 하도록 유도 할 수 있음



- 과제소개
- 데이터 탐색 결과
- 알고리즘 설명
- 서비스 아이디어 제안
- 아이디어 소개
- **서비스 전략 및 기대효과**
- 한계점 및 발전가능성

서비스 전략

NH투자증권에서 NH수목원 만들어서 고객이 주식 거래를 통해,
나무를 10개를 잘 키워내면 실제 1개의 나무를 투자자 이름 또는 닉네임으로 심어주기

NH투자증권의 “친환경” 이미지 메이킹 가능

+ 많은 나무로 꾸며진 수목원에서 고객들이 사진찍고 sns에 올리며 홍보효과도 톡톡!

+ 팝업스토어로도 이목 집중 가능!



기대효과

- 투자에 대한 공부 + 본인의 투자 히스토리 + 본인에게 맞는 종목 추천 등을 하나의 공간에서 볼 수 있어 편리함
- 계절별, 월별 나무/꽃 도감을 통해, 고객 개인의 주식보유기간 예측을 이용한 종목 추천 으로 활발한 매수 유도 가능
- 고객들끼리 숲을 서로 비교하며, 투자에 관한 정보도 얻어 갈 수 있는 하나의 커뮤니티 공간 조성
- 수목원 조성을 통한 다양한 오프라인 이벤트를 개최하여 온라인+오프라인 환경에서 동시에 소통가능

- 과제소개
- 데이터 탐색 결과
- 알고리즘 설명
- 서비스 아이디어 제안
- 아이디어 소개
- 서비스 전략 및 기대효과
- **한계점 및 발전가능성**

한계점 및 발전가능성

- 프로그램의 시각화 요소도 다양하며, 이 프로그램은 투자의 과정에서 하나의 옵션으로 사용해야 해서 사용자가 이를 사용하고 싶게 UI/UX 디자인 모두 신경 써야 하기에 이목을 끄는 시각화 방법 필요
- 주식을 매도/매수하는 창과 이 시스템을 완전히 분리할 경우, 사용자는 이 시스템의 존재를 모르거나, 처음 시작이 귀찮아서 시도해보지 않을 수도 있음
- 사용자들의 나무를 공정하게 성장할 수 있도록 하기 위해, 문제를 계속해서 업데이트 해야함

-> 그러나, 투자에 대한 진입장벽을 낮추어, 모든 연령대가 쉽게 투자에 흥미를 갖고 즐길 수 있음

그 결과, 다양한 연령대가 투자에 참여하고 소통하여, 투자가 하나의 놀이와 문화로 발전할 수 있는 발판 마련