

미래에셋대우/미래에셋생명 주관

2020 금융 빅데이터 페스티벌

Building on principles



팀 소개

팀 명	데이터 좀 아는 개미들
학교	연세대학교
학과	디지털 애널리틱스 융합 협동과정
학년	석사 2학기
연락처	변영목 (2020312084@yonsei.ac.kr/010-8834-3474)
	양명한 (suneverset@yonsei.ac.kr/010-4809-8701)
	이경환 (khrock1@yonsei.ac.kr/010-2051-1091)

Table of Contents

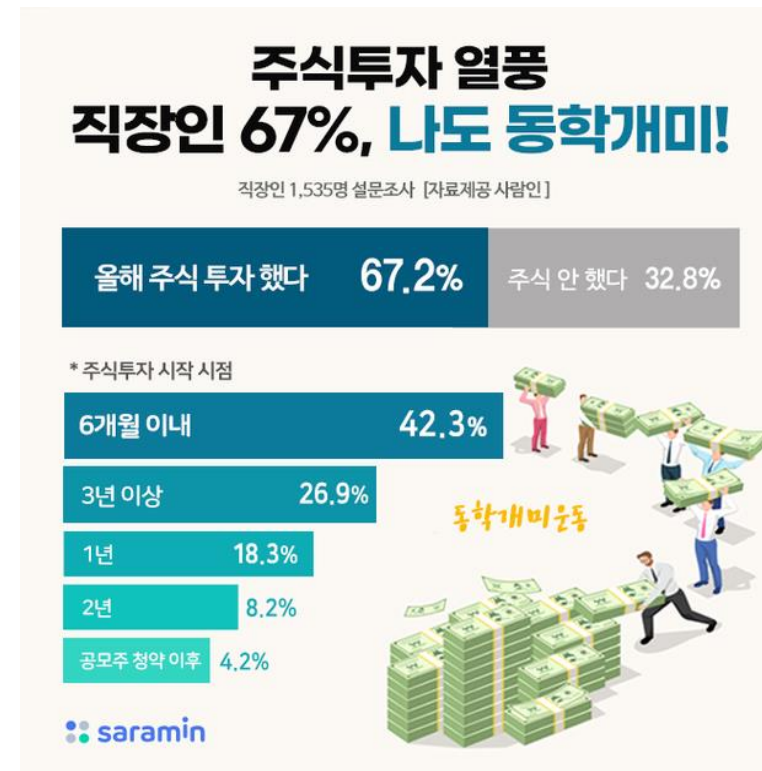
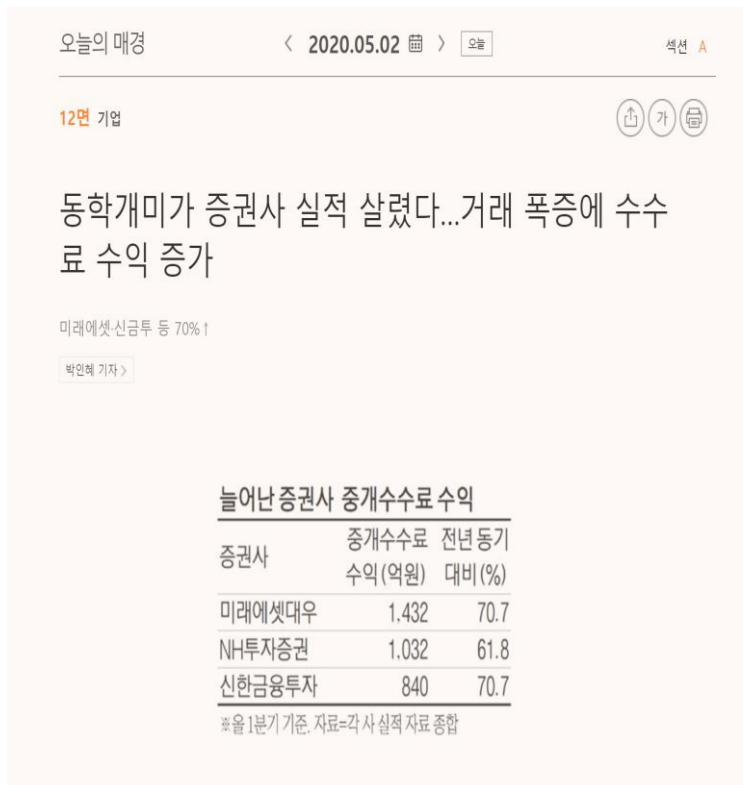
- I. 과제 배경 및 문제 정의
- II. 분석 개요
- III. EDA / 데이터 전처리
- IV. 모델링 및 학습
- V. 결론 – 모델평가 및 개선방안
- VI. 비즈니스 활용 방안



I. 과제 배경 및 문제 정의



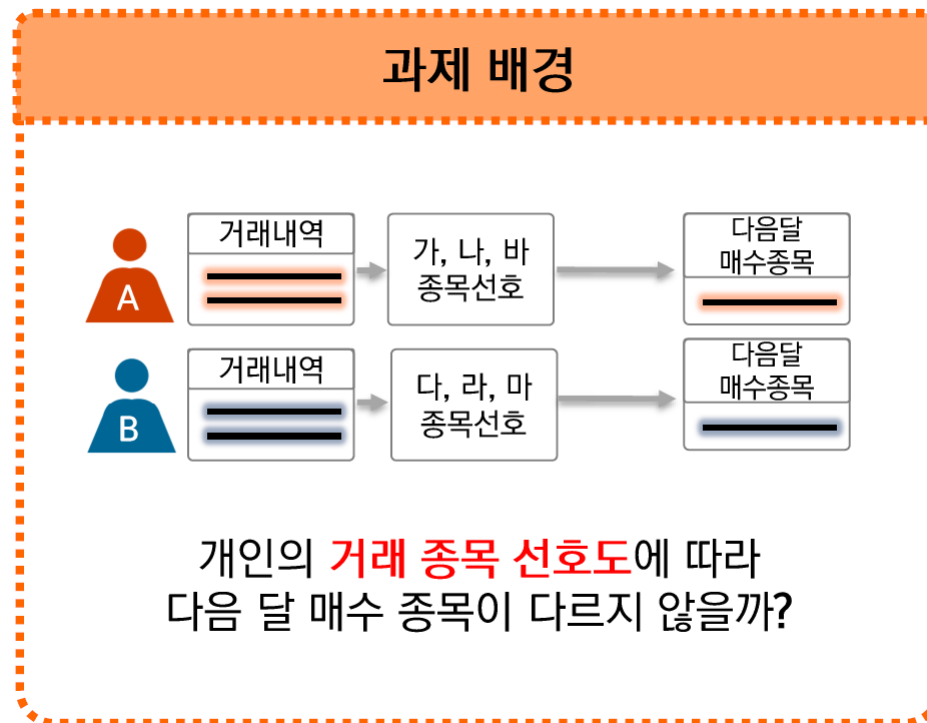
1. 과제 배경



“개미”들은 더 이상 무시할 수 없는 존재 :

과거에는 기관투자와 외국인 투자 규모가 압도적이었지만,
주식투자 열풍으로 개인투자자의 규모가 폭발적으로 증가

1. 과제 배경

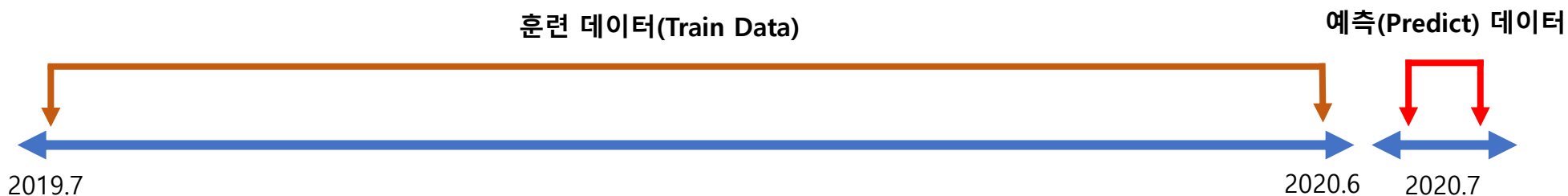


이번 과제 주제 또한 “**개인의 거래 종목 선호도**”가 핵심

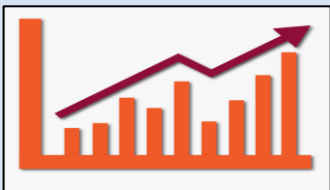
개인 투자자들의 **선호도** 및 그들을 위한 **서비스**를 염두에 두고 분석 진행

1. 문제 정의

분석 목표: 2019.7~2020.6의 데이터로 각 그룹에서 2020년 7월에 가장 많은 고객이 매수할 상위 3개 종목 예측



제공 데이터:



<Stocks Data>

주식종목들의 시가, 거래량, 산업구분 등의 정보



<Train Data>

고객 그룹별 매수/매도 고객수, 평균매수수량 등의 정보

예측:

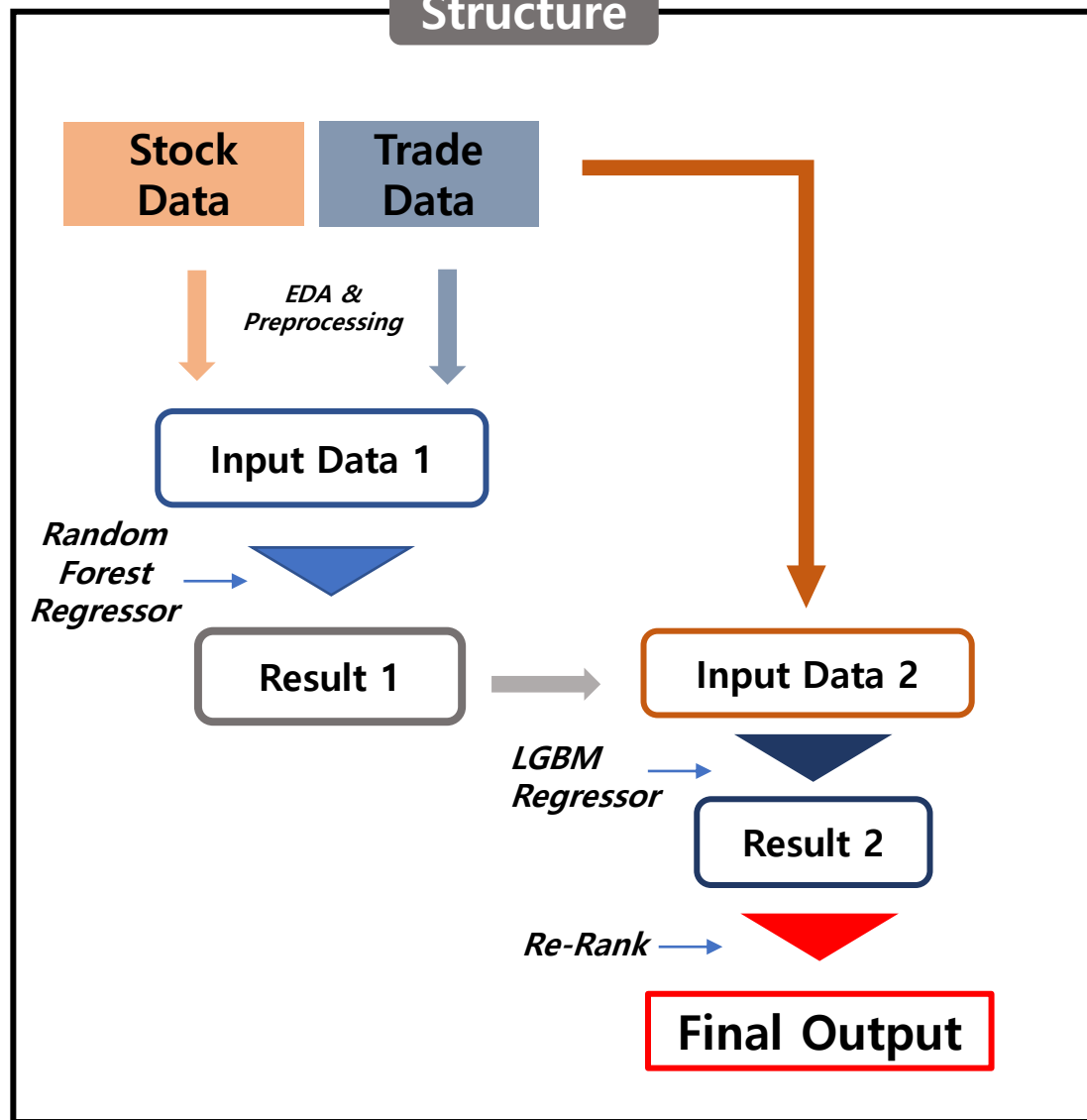
**2020년 7월 그룹별
고객들이 가장 많이
매수한 상위 종목 3개**

II. 분석 개요



2. 분석 개요

Structure



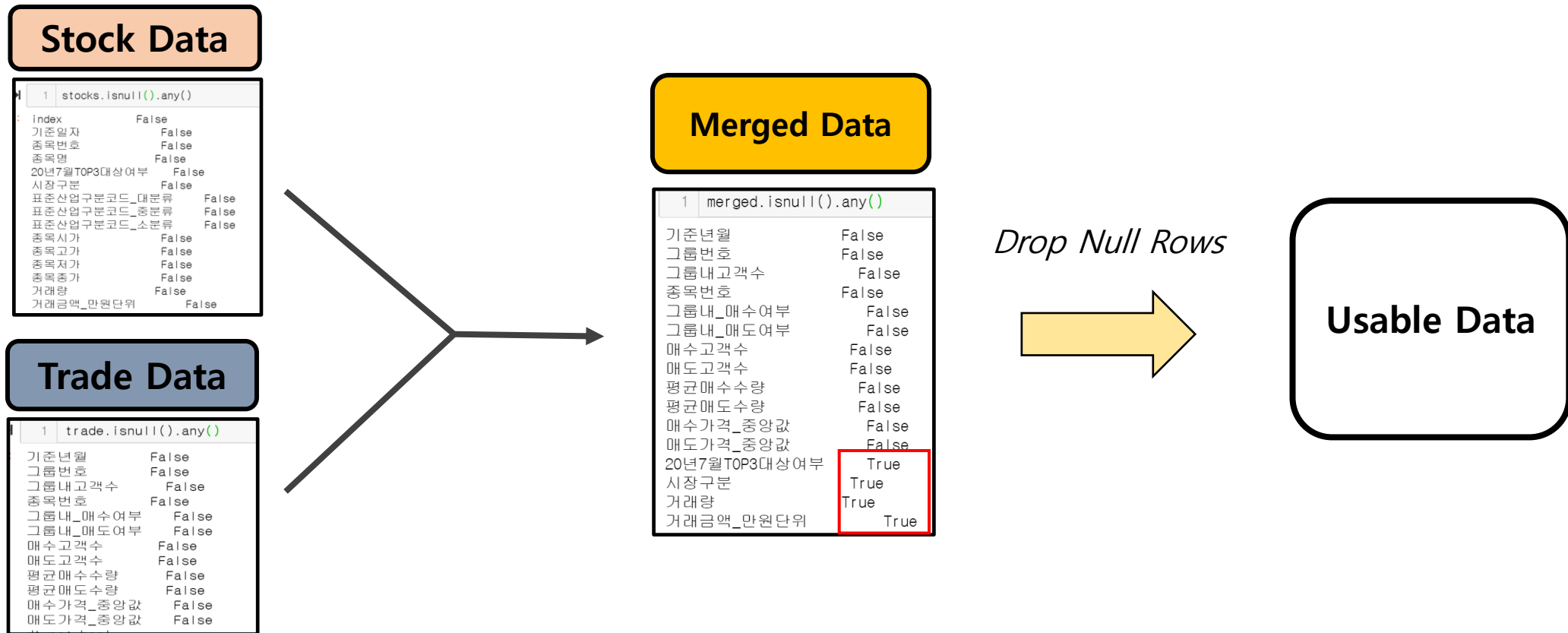
Analysis Flow



Ⅲ. EDA / 데이터 전처리



3. EDA 및 데이터 전처리- 결측치 처리



→ 주어진 Stocks Data와 Trade Data에는 결측치 없음

→ Stocks Data 에 포함된 종목들은 1087개, Trade Data에 포함된 종목들은 1066개로 두 데이터를 합쳐주면서 결측치 발생

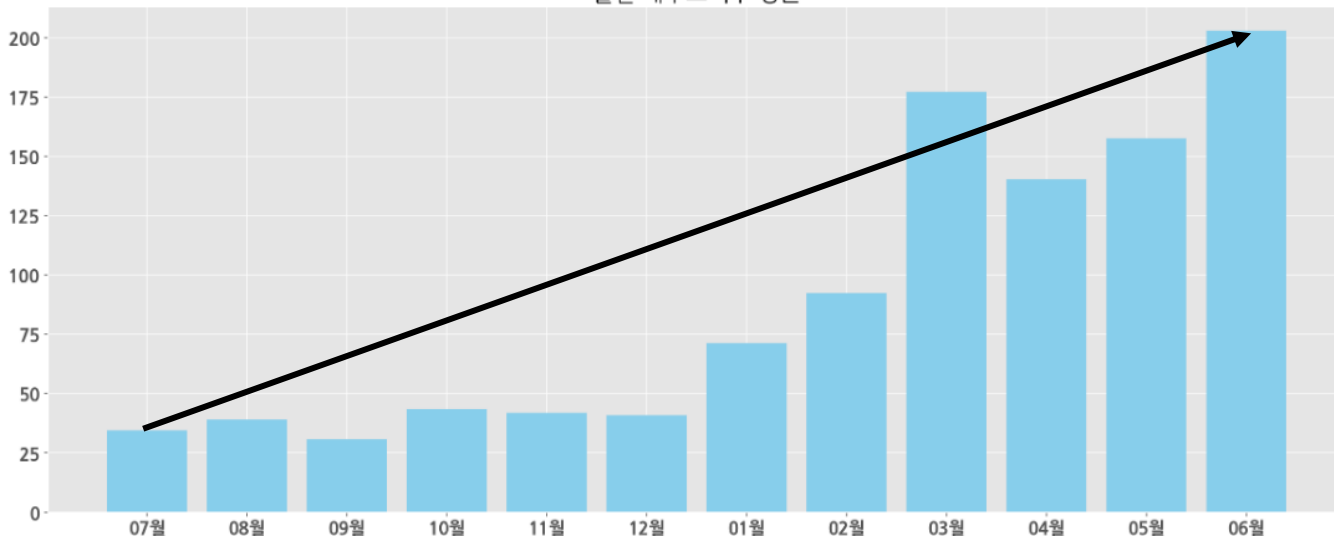
→ Trade Data에 존재하지 않는 종목들은 학습에 활용할 수 없으므로, 제거해주고 시작

3. EDA 및 데이터 전처리 - 분석 활용 데이터

“ 7월의 매수 고객 수 예측을 위해 몇 개월치의 데이터를 활용해야 할까? ”

그룹번호	201907	201908	201909	201910	201911	201912	202001	202002	202003	202004	202005	202006
MAD01	42.0	55.0	37.0	48.0	53.0	53.0	85.0	111.0	171.0	157.0	164.0	212.0
MAD02	7.0	11.0	0.0	11.0	6.0	3.0	15.0	31.0	49.0	48.0	38.0	53.0
MAD03	26.0	32.0	26.0	60.0	28.0	24.0	38.0	46.0	90.0	90.0	78.0	153.0
MAD04	70.0	82.0	61.0	94.0	91.0	91.0	143.0	169.0	309.0	247.0	243.0	323.0
MAD05	54.0	76.0	50.0	67.0	53.0	63.0	99.0	152.0	267.0	197.0	281.0	370.0
MAD06	145.0	158.0	124.0	176.0	188.0	188.0	348.0	403.0	738.0	569.0	721.0	972.0
MAD07	4.0	3.0	3.0	3.0	13.0	9.0	13.0	17.0	45.0	37.0	43.0	57.0
MAD08	37.0	37.0	30.0	78.0	45.0	24.0	44.0	73.0	141.0	107.0	94.0	126.0
MAD09	42.0	43.0	32.0	47.0	56.0	52.0	90.0	114.0	271.0	167.0	199.0	239.0
MAD10	3.0	10.0	0.0	3.0	0.0	0.0	13.0	21.0	41.0	32.0	64.0	93.0

월별 매수고객수 평균



[그룹별 - 월별 총 매수고객수]
“매수 고객수가 아예 없는 달이 존재함”

+

[월별 전체그룹의 매수고객수 평균]
“매수고객수가 지속적으로 늘어나는 경향”

=

“과거의 매수 성향과 현재의 매수 성향이
다를 가능성 ↑”

∴ 최근 3개월의 데이터를 활용하자.
(4, 5, 6월)

3. EDA 및 데이터 전처리 - 우량주 선호도

“ 그룹별로 우량주를 어느 정도 선호할까? ”

(* 우량주 기준 : 3개월 동안 매수 고객수가 가장 많았던 상위 5개 종목
→ 삼성전자, SK하이닉스, 카카오, 삼성전자우, 현대차)

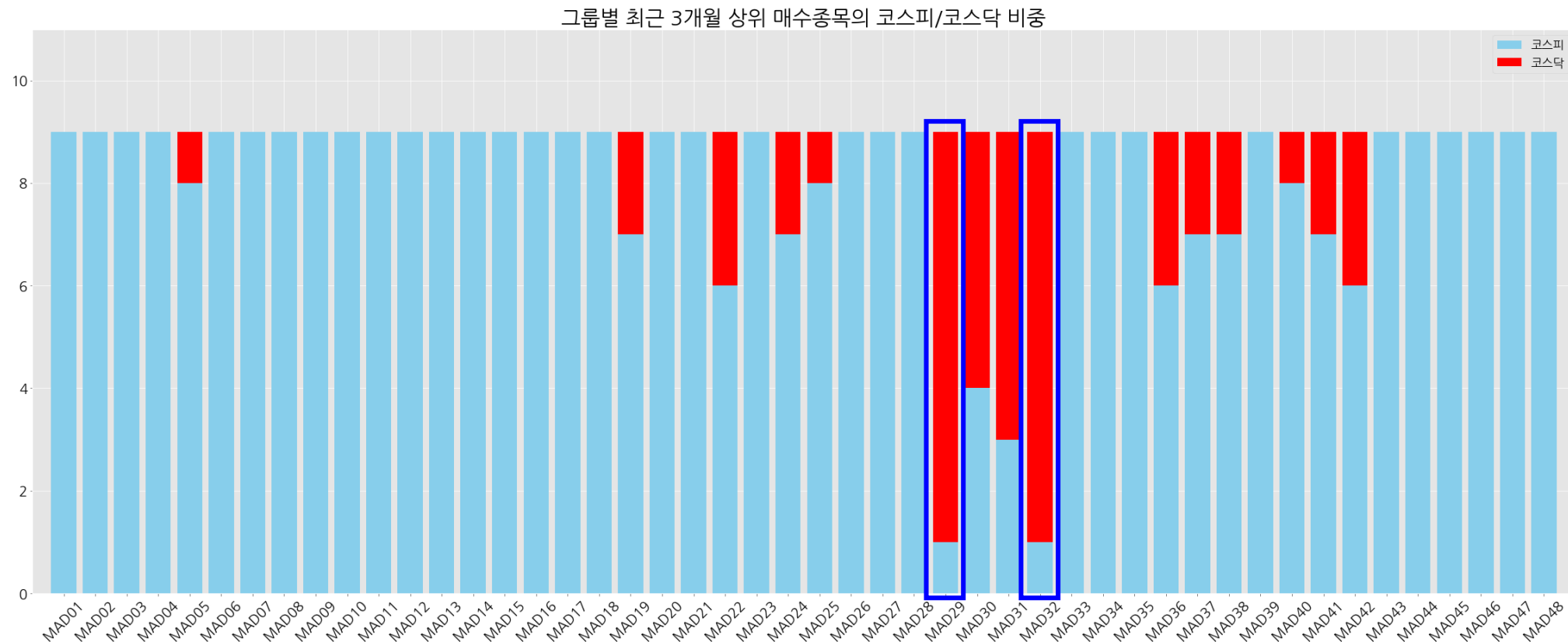
우량주 선호도
:
4~6월 매수 상위 종목
총 9개 중 우량주에
해당하는 종목의 비중

그룹번호	우량주 선호도	
0	MAD01 1.000000	→ 9개의 상위 매수 종목 모두 우량주에 해당함.
1	MAD02 0.666667	
2	MAD03 0.111111	→ 9개의 상위 매수 종목 중 한 종목만 우량주에 해당함.
3	MAD04 0.666667	
4	MAD05 0.444444	
5	MAD06 0.666667	
6	MAD07 0.444444	
7	MAD08 0.000000	→ 9개의 상위 매수 종목 중 한 종목도 해당 없음.
8	MAD09 0.555556	

→ 우량주 선호그룹, 중간 그룹 및 비선호 그룹의 특성을 구분할 수 있다.

3. EDA 및 데이터 전처리 - 코스피 선호도

"각 그룹별 코스피 종목 매수 비중은 어떻게 되는가?"
(코스피 종목 비중이 작을수록 위험선호도(Risk Appetite)가 크다고 판단)



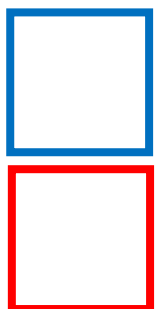
→ **MAD29, MAD32** 그룹의 코스피 종목 비중이 두드러지게 **낮음**을 확인할 수 있음.
(위험을 선호하는 그룹과, 그렇지 않은 그룹의 특성을 구분할 수 있다.)

3. EDA 및 데이터 전처리 - 종목 순위 변동성

"4, 5, 6월 그룹별 상위 3개 종목 추출 → 9개 종목 중 중복되는 종목이 몇 개일까?"
(중복 종목이 없다 → 매수 종목이 자주 변한다 → 순위변동성이 심하여 예측이 어려운 그룹이다.)

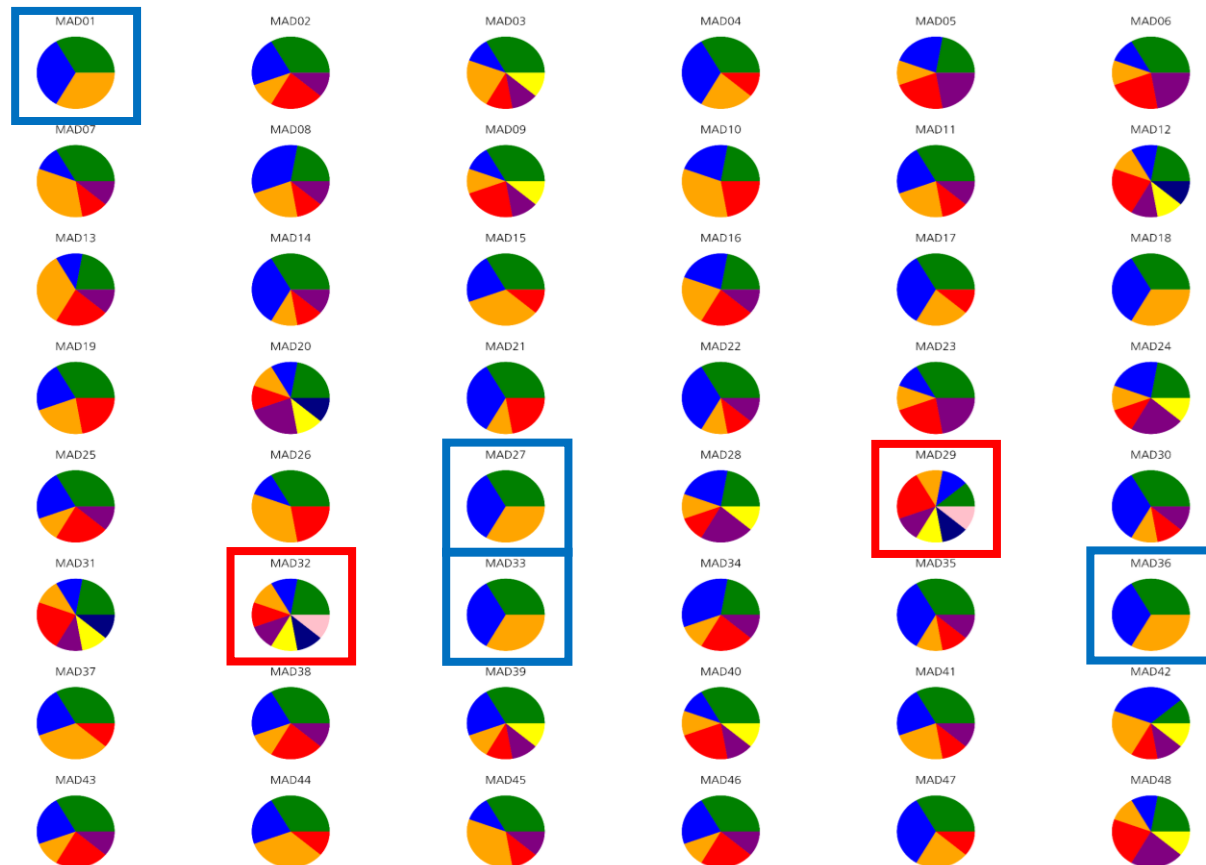
MAD01 [('삼성전자', 3), ('SK하이닉스', 3), ('삼성전자우', 3)]
MAD02 [('삼성전자', 3), ('SK하이닉스', 2), ('현대차', 1), ('KB금융', 2), ('신한지주', 1)]
MAD03 [('신용제약', 3), ('진원생명과학', 1), ('파미셀', 2), ('삼성전자', 1), ('부광약품', 1), ('일양약품', 1)]
MAD04 [('삼성전자', 3), ('SK하이닉스', 3), ('삼성SDI', 2), ('부광약품', 1)]
MAD05 [('삼성전자', 2), ('셀트리온', 1), ('카카오', 2), ('NAVER', 2)]
MAD06 [('삼성전자', 3), ('SK하이닉스', 1), ('셀트리온', 1), ('카카오', 2), ('NAVER', 2)]
MAD07 [('삼성전자', 3), ('셀트리온', 1), ('SK이노베이션', 3), ('SK하이닉스', 1), ('SOIL', 1)]
MAD08 [('남신알미늄', 2), ('신용제약', 3), ('파미셀', 2), ('부광약품', 1), ('일양약품', 1)]
MAD09 [('삼성전자', 3), ('셀트리온', 1), ('삼성바이오로직스', 1), ('카카오', 2), ('NAVER', 1), ('LG화학', 1)]
MAD10 [('삼성전자', 2), ('삼성전자우', 2), ('한국전력', 3), ('KT', 2)]
MAD11 [('삼성전자', 3), ('SK하이닉스', 2), ('미래에셋대우', 2), ('SK', 1), ('삼성바이오로직스', 1)]
MAD12 [('삼성전자', 2), ('현대차', 1), ('셀트리온', 1), ('카카오', 2), ('미래에셋대우', 1), ('NAVER', 1), ('SK', 1)]
MAD13 [('삼성전자', 2), ('삼성전자우', 1), ('한국전력', 3), ('KT', 2), ('카카오', 1)]
MAD14 [('삼성전자', 3), ('한국전력', 3), ('현대차', 1), ('SK하이닉스', 1), ('삼성전자우', 1)]
MAD15 [('삼성전자', 3), ('SK하이닉스', 2), ('SOIL', 3), ('SK이노베이션', 1)]
MAD16 [('삼성전자', 2), ('현대차', 2), ('SK이노베이션', 2), ('SOIL', 2), ('카카오', 1)]
MAD17 [('삼성전자', 3), ('현대차', 3), ('SK하이닉스', 2), ('SK이노베이션', 1)]
MAD18 [('삼성전자', 3), ('SK하이닉스', 3), ('LG화학', 3)]
MAD19 [('삼성전자', 3), ('삼성전자우', 2), ('셀트리온헬스케어', 2), ('SK하이닉스', 2)]
MAD20 [('삼성전자', 2), ('현대차', 1), ('미래에셋대우', 1), ('카카오', 1), ('NAVER', 2), ('SK', 1), ('신한지주', 1)]

→ **Pie Chart**로 시각화하여 확인



: 3개월동안 같은 종목만 상위매수
(MAD01, MAD27, MAD33, MAD36)

: 3개월동안 상위 매수 종목 변동성이 심함
(MAD29, MAD32)



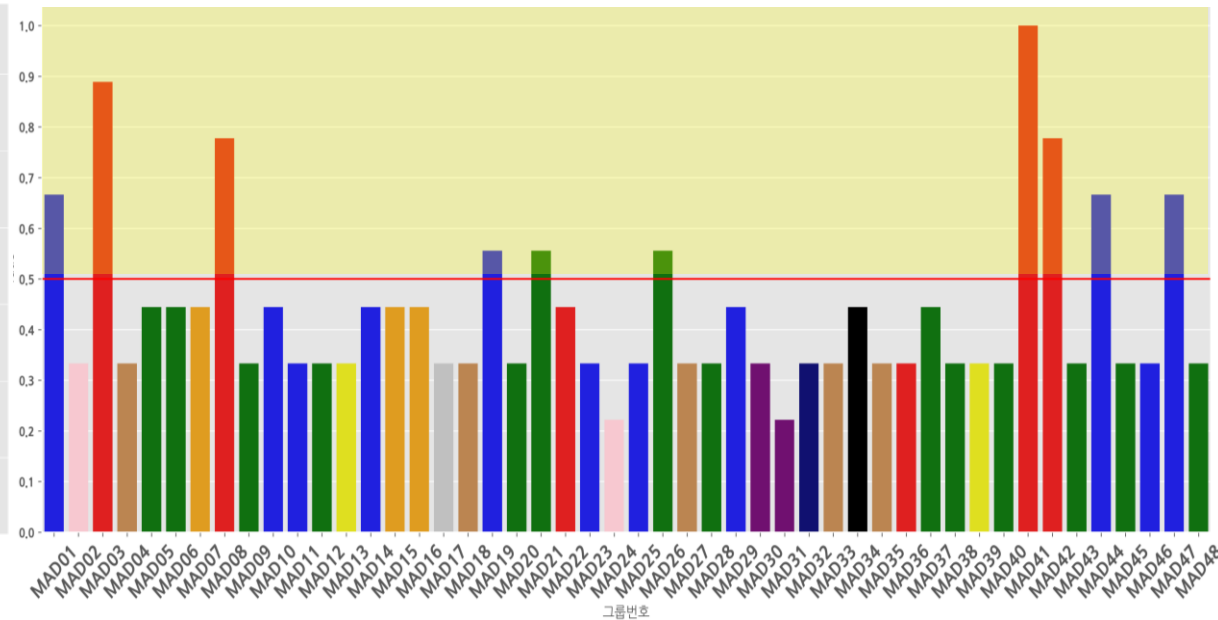
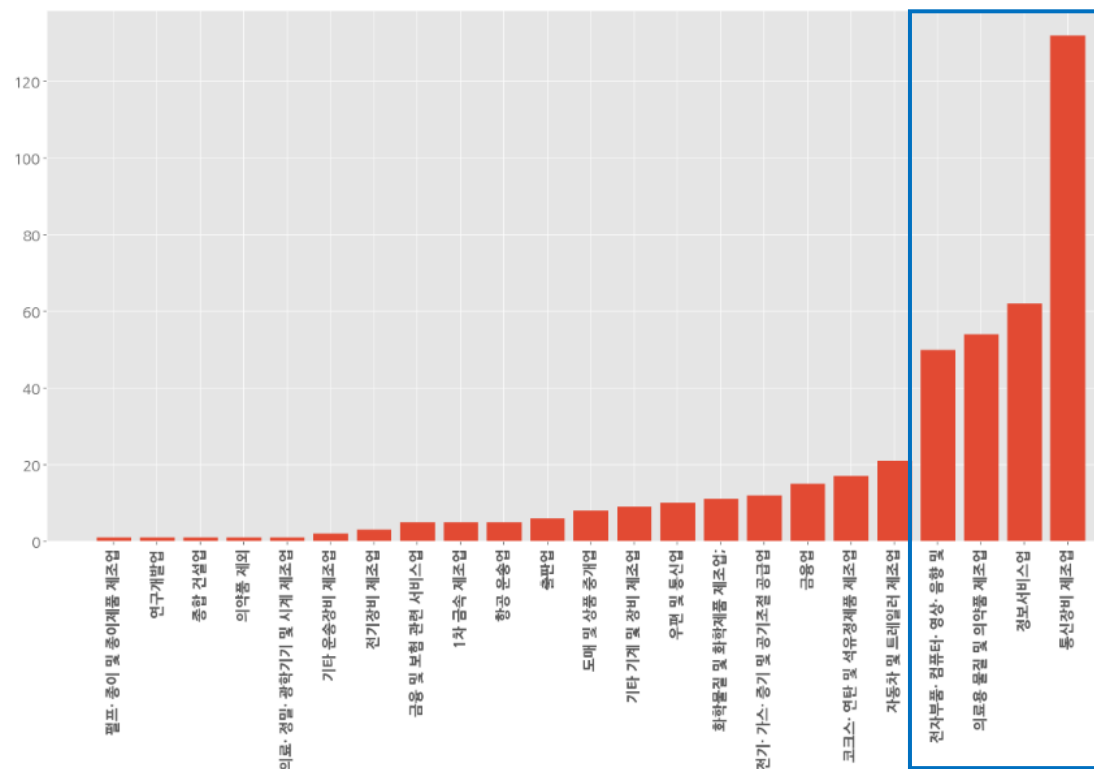
→ 3개월 동안의 상위 매수종목이 같은 그룹도 있고 다양한 그룹도 있다.

3. EDA 및 데이터 전처리 – 산업 선호도

“그룹들은 주로 어떤 산업에 선호도를 가질까?”



“특정 산업이 절반 이상을 차지한다면,
이를 해당 그룹이 선호하는 산업으로 정의”



- 통신장비 제조업
- 의료용 물질 및 의약품 제조업
- 정보서비스업

→ 선호 여부가 나타나는 산업군은 **통신장비 제조업, 의료용 물질 및 의약품 제조업, 정보서비스업.**
(선호하는 특정산업이 있는 그룹과, 그렇지 않은 그룹의 특성을 구분할 수 있다.)

3. EDA 및 데이터 전처리 - 파생변수 생성

“우량주선호도, 코스피선호도, 순위변동성, 특정산업 선호도를 통해 그룹의 특성을 도출”

	그룹번호	우량주선호도	코스피선호도	순위변동성	통신장비제조업_선호도	의약품제조업_선호도	정보서비스업_선호도
0	MAD01	1.000000	1.000000	0.0	1	0	0
1	MAD02	0.666667	1.000000	0.4	0	0	0
2	MAD03	0.111111	1.000000	0.6	0	1	0
3	MAD04	0.666667	1.000000	0.2	0	0	0
4	MAD05	0.444444	0.888889	0.4	0	0	0
5	MAD06	0.666667	1.000000	0.4	0	0	0
6	MAD07	0.444444	1.000000	0.4	0	0	0
7	MAD08	0.000000	1.000000	0.4	0	1	0
8	MAD09	0.555556	1.000000	0.6	0	0	0
9	MAD10	0.444444	1.000000	0.2	0	0	0
10	MAD11	0.555556	1.000000	0.4	0	0	0
11	MAD12	0.555556	1.000000	0.8	0	0	0

→ “우량주 및 코스피 종목, 그 중 통신장비제조업을 선호하는 순위 변동성이 낮은 그룹”

→ “비우량주 및 코스피 종목, 그 중 의약품제조업을 선호하며 순위 변동성이 중간인 그룹”

→ “순위 변동성이 심하며, 특정 산업 선호도가 없는 그룹”

IV. 모델링 및 학습



4. 모델링 및 학습 - 1차 학습

1) 1차 학습의 목적

: 2차 학습을 위한 그룹별 후보 종목 선정 → 예측 매수고객수 기준 상위 (최대) 30종목 선정

2) Features

: 우량주 선호도, 코스피 선호도, 순위변동성, 특정산업 선호도 및 관련 Guide Features

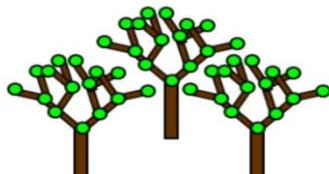
3) Guide Features?

: 그룹의 특성 변수들에 대응하는 개별 종목들의 특징

그룹의 특성 변수	Guide Features
우량주 선호도	(종목의) 우량주 여부
코스피 선호도	(종목의) 코스피 여부
순위변동성	(종목의) 이번 달 상위종목 해당 여부
특정산업 선호도	(종목의) 산업코드

4) 사용 알고리즘

: *Random Forest Regressor*



4. 모델링 및 학습 - 1차 학습

그룹의 특성

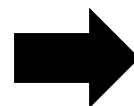
기준년 월	그룹번호	종목번호	20년7월 TOP3대상 여부	우량주선 호도	코스 피션 호도	순위 변 동성	통신장비 제조업_선 호도	의약품제 조업_선 호도	정보서비 스업_선 호도	시장구 분	우량 주 여 부	이번달_ 순위권_ 여부	의약품 제조업 여부	통신장 비제조 업여부	정보서 비스업 여부
202004	MAD01	A000210	0	1.000000	1.0	0.0	1	0	0	1	0	0	0	0	0
202004	MAD01	A000660	1	1.000000	1.0	0.0	1	0	0	1	1	1	0	0	0
202004	MAD01	A000990	1	1.000000	1.0	0.0	1	0	0	1	0	0	0	0	0
202004	MAD01	A005380	1	1.000000	1.0	0.0	1	0	0	1	1	0	0	0	0
202004	MAD01	A005930	1	1.000000	1.0	0.0	1	0	0	1	1	1	0	1	0
...
202005	MAD48	A091990	1	0.666667	1.0	0.6	0	0	0	0	0	0	0	0	0
202005	MAD48	A096530	1	0.666667	1.0	0.6	0	0	0	0	0	0	1	0	0
202005	MAD48	A096770	1	0.666667	1.0	0.6	0	0	0	1	0	0	0	0	0
202005	MAD48	A105560	1	0.666667	1.0	0.6	0	0	0	1	0	0	0	0	0
202005	MAD48	A139480	0	0.666667	1.0	0.6	0	0	0	1	0	0	0	0	0

종목의 특성(Guide Features)

→ "그룹의 특성과 종목의 특성을 함께 고려하여 매수 후보 종목 선정 "

5. 1차 학습 결과

	그룹번호	종목번호	20년7월TOP3대상여부	시장구분	평균매수수량	거래량	거래금액_만원단위	매수고객수
0	MAD01	A005930	1	1	219	517439261.0	2.756149e+09	7.597827
1	MAD01	A005935	1	1	149	59384055.0	2.758364e+08	7.597827
2	MAD01	A000660	1	1	121	91590220.0	7.928776e+08	7.253585
3	MAD01	A035420	1	1	7	26395703.0	6.582895e+08	5.217119
4	MAD01	A005380	1	1	32	35327068.0	3.684219e+08	4.983211
5	MAD01	A035720	1	1	15	27742274.0	7.331739e+08	4.885504
6	MAD01	A068270	1	1	18	44753636.0	1.251311e+09	3.250212
7	MAD01	A090430	1	1	31	10385251.0	1.744200e+08	3.172859
8	MAD01	A105560	1	1	27	64921629.0	2.335226e+08	3.172859
9	MAD01	A096770	1	1	636	21121685.0	2.658133e+08	3.172859
10	MAD01	A006800	1	1	1265	91305168.0	6.214274e+07	3.172859
11	MAD01	A051910	1	1	16	15173513.0	7.009731e+08	3.172859
12	MAD01	A010140	1	1	492	611713472.0	4.037859e+08	3.172859



그룹 당 매수고객수가 높은 순으로
최대 30개의 후보종목을 선정함.



해당 종목들로 2차 학습 수행

			그룹번호	종목번호	20년7월TOP3대상여부	시장구분	평균매수수량	거래량	거래금액_만원단위	매수고객수	
13	MAD01	A000990									
14	MAD01	A086520	0	MAD02	A005930	1	1	107	517439261.0	2.756149e+09	9.518243
15	MAD01	A091990	1	MAD02	A105560	1	1	217	64921629.0	2.335226e+08	7.162446
16	MAD01	A218410	2	MAD02	A005935	1	1	19	59384055.0	2.758364e+08	5.904016
			3	MAD02	A000660	1	1	34	91590220.0	7.928776e+08	5.852930
			4	MAD02	A035720	1	1	174	27742274.0	7.331739e+08	5.625996
			5	MAD02	A035420	1	1	11	26395703.0	6.582895e+08	5.004699
			6	MAD02	A003490	1	1	364	39258842.0	8.002379e+07	2.826071
			7	MAD02	A006800	1	1	1811	91305168.0	6.214274e+07	2.826071
			8	MAD02	A010140	1	1	11888	611713472.0	4.037859e+08	2.826071
			9	MAD02	A042670	1	1	4147	410297246.0	2.906399e+08	2.826071

6. 2차 학습

1) 2차 학습의 목적

: 1차 학습을 통해 선정된 후보종목 중 상위 3개의 매수종목을 선별함.

2) Features

: Trade data 및 Stock data 에서 유래한 특성

	기준년월	그룹번호	종목번호	20년7월TOP3대상여부	시장구분	평균매수수량	거래량	거래금액_만원단위	이번달_매수고객수
0	202004	MAD01	A000210	0	1	189	11527212.0	9.136904e+07	2.236068
1	202004	MAD01	A000660	1	1	107	91942271.0	7.587119e+08	7.211103
2	202004	MAD01	A000990	1	1	544	25807588.0	6.101197e+07	2.449490
3	202004	MAD01	A005380	1	1	73	41304577.0	3.862312e+08	3.464102
4	202004	MAD01	A005930	1	1	182	424101937.0	2.078519e+09	8.774964
...

3) Target

: 다음달 매수 고객 수

4) 사용 알고리즘

: *LightGBM Regressor*



6. 2차 학습 결과

“그룹별 매수 고객 수 상위 3개 종목을 선정”

[MAD01]

그룹번호	종목번호	종목명	20년7월TOP3대상여부	매수고객수
0	MAD01	A005930	삼성전자	1 8.771530
1	MAD01	A000660	SK하이닉스	1 5.965353
2	MAD01	A005935	삼성전자우	1 5.852187
3	MAD01	A005380	현대차	1 3.014045
4	MAD01	A035720	카카오	1 2.895483
5	MAD01	A010140	삼성중공업	1 2.801532
6	MAD01	A096770	SK이노베이션	1 2.743587
7	MAD01	A035420	NAVER	1 2.589025
8	MAD01	A068270	셀트리온	1 2.589025
9	MAD01	A000990	DB하이텍	1 2.589025
10	MAD01	A006800	미래에셋대우	1 2.589025
11	MAD01	A051910	LG화학	1 2.589025
12	MAD01	A090430	아모레퍼시픽	1 2.589025
13	MAD01	A105560	KB금융	1 2.589025
14	MAD01	A086520	에코프로	1 2.556215
15	MAD01	A091990	셀트리온헬스케어	1 2.556215
16	MAD01	A218410	RFHIC	1 2.556215

[MAD03]

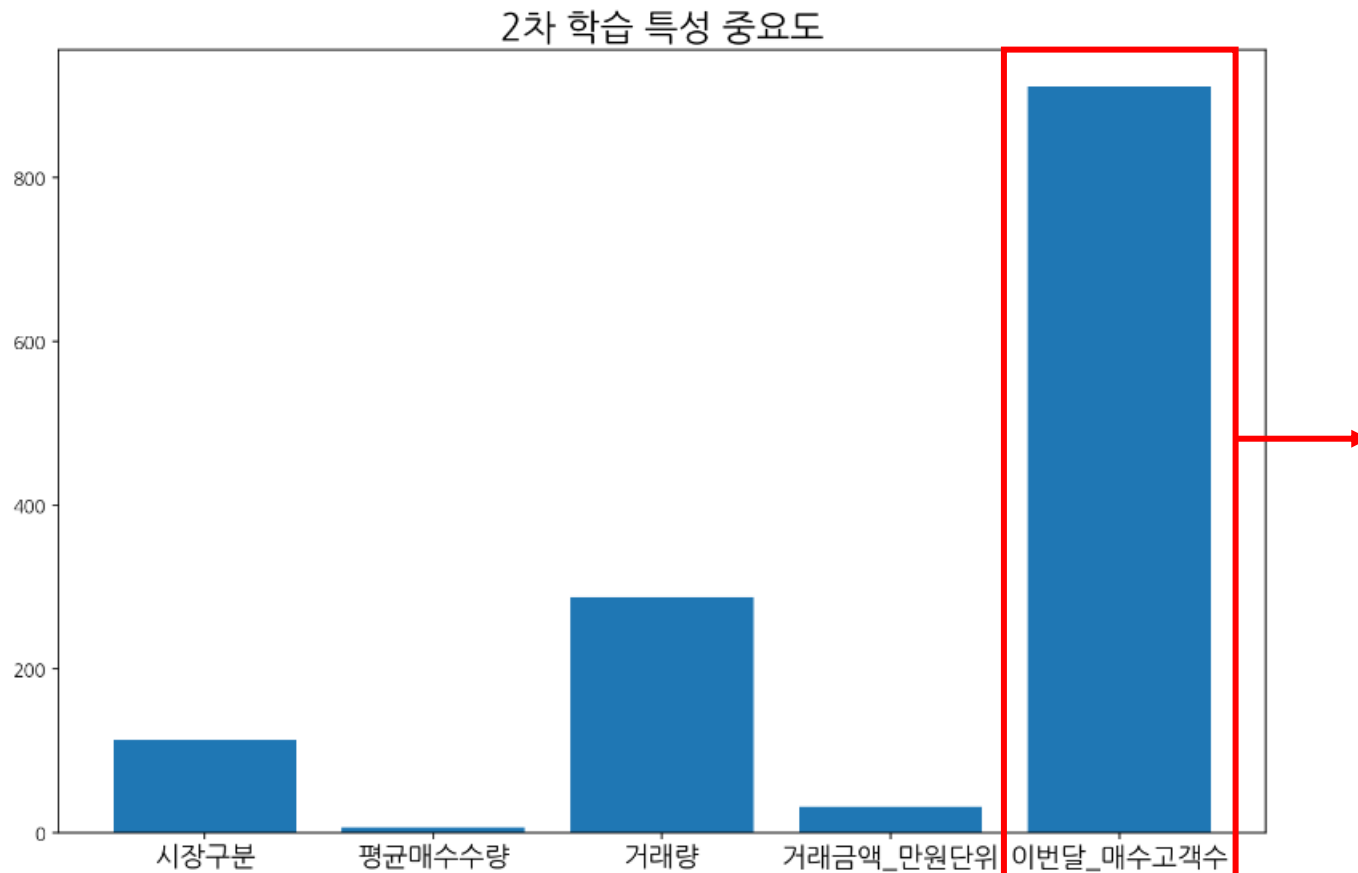
그룹번호	종목번호	종목명	20년7월TOP3대상여부	매수고객수
0	MAD03	A007570	일양약품	1 5.759158
1	MAD03	A003000	부광약품	1 5.525200
2	MAD03	A019170	신풍제약	1 5.525200
3	MAD03	A068270	셀트리온	1 3.894006
4	MAD03	A005930	삼성전자	1 3.503950
5	MAD03	A010145	삼성중공업	1 3.145922
6	MAD03	A010140	삼성중공업	1 2.937175
7	MAD03	A005380	현대차	1 2.895483
8	MAD03	A000100	유한양행	1 2.891514
9	MAD03	A096530	씨젠	1 2.824986
10	MAD03	A091990	셀트리온헬스케어	1 2.731036
11	MAD03	A035420	NAVER	1 2.589025
12	MAD03	A000660	SK하이닉스	1 2.589025
13	MAD03	A003010	혜인	1 2.589025
14	MAD03	A003490	대한항공	1 2.589025
15	MAD03	A005360	모나미	1 2.589025
16	MAD03	A006800	미래에셋대우	1 2.589025
17	MAD03	A047400	유니온머티리얼	1 2.589025
18	MAD03	A288620	에스퓨얼셀	1 2.556215

[MAD30]

그룹번호	종목번호	종목명	20년7월TOP3대상여부	매수고객수
0	MAD30	A005930	삼성전자	1 3.426536
1	MAD30	A222080	씨아이에스	1 3.215097
2	MAD30	A131390	피앤이솔루션	1 2.821017
3	MAD30	A036540	SFA반도체	1 2.814306
4	MAD30	A003000	부광약품	1 2.699543
5	MAD30	A035720	카카오	1 2.589025
6	MAD30	A068270	셀트리온	1 2.589025
7	MAD30	A000660	SK하이닉스	1 2.589025
8	MAD30	A005380	현대차	1 2.589025
9	MAD30	A000990	DB하이텍	1 2.589025
10	MAD30	A018880	한온시스템	1 2.589025
11	MAD30	A082270	젬백스	1 2.556215
12	MAD30	A234100	세원	1 2.556215

6. 2차 학습 - 해석 및 보완

“매수 고객 수 예측에 각 특성은 어떠한 중요도를 가지는가?”



Feature Importance 측정 결과:

“다음달 매수고객수 예측에 있어
압도적으로 중요한 특성은
이번 달 매수 고객 수”



“1차 학습에 사용했던
**그룹의 특성을 반영하지
못할 위험성**”

7. 3차 Re-rank

1) Re-rank 의 목적

: 2차 학습을 통해 선정된 종목을 최대한 보존하되,
그룹의 특성이 반영되지 않은 경우 순위를 재정렬(Re-rank)해줌.

2) 방법

: EDA 수행 당시 독특한 특성을 보였던 **MAD29, MAD32** 그룹 (코스피 선호도 ↓, 위험선호도 ↑)
을 대상으로 특성 확인 및 Re-rank 수행.

[MAD29]

그룹번호	종목번호	종목명	시장구분	20년7월TOP3대상여부	매수고객수
0	MAD29 A222080	씨아이에스	0	1	3.244345
1	MAD29 A005930	삼성전자	1	1	2.699543
2	MAD29 A003000	부광약품	1	1	2.589025



코스피 선호도가 제일 낮은 그룹이지만
예상 매수 상위 종목에 코스피 종목이 2개.
→ 종목 변경 필요

[MAD32]

그룹번호	종목번호	종목명	시장구분	20년7월TOP3대상여부	매수고객수
0	MAD32 A222080	씨아이에스	0	1	3.215097
1	MAD32 A235980	메드팩토	0	1	3.209772
2	MAD32 A032850	비트컴퓨터	0	1	2.909851



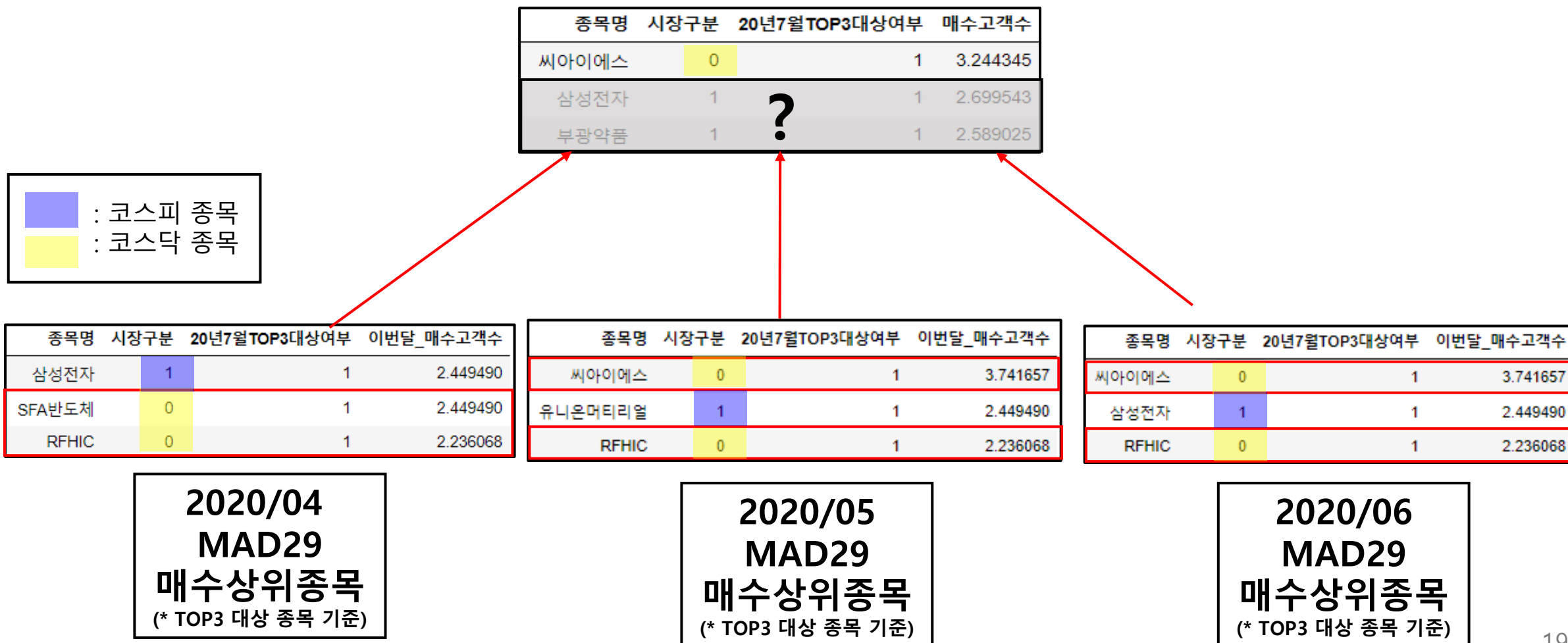
MAD29 그룹과 동일하게 코스피 선호도가 가장 낮음.
하지만 이미 모든 종목들이 코스닥으로 채워져 있음.
→ 종목 변경 불필요

7. 3차 Re-rank

"① 2차 학습의 결과를 보존하는 선에서 ② 그룹의 특성을 반영할 필요 有"

→ ① 기존 매수고객수가 많았던 종목(씨아이에스)은 유지하고

→ ② 코스피 종목들은 과거 매수고객수가 많았던 코스닥 종목 (RFHIC, SFA반도체)으로 변경



V. 결론 – 모델 평가 및 개선방안



5. 결론 - 모델 평가

1.

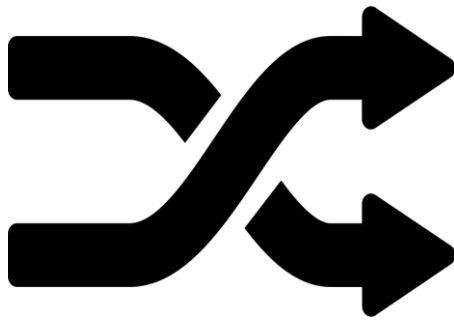
그룹별 특성 & 종목별 특성을
고려한 변수 생성



여러 형태의 파생변수를 바탕으로
그룹의 특징을 표현하고, 각 종목
과 매칭하여 효과적인 분석 진행

2.

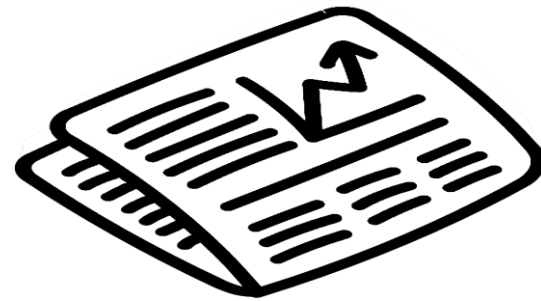
결과 예측 값에 대해
Re-rank 처리



2차 학습 결과에 대해, 그룹별
특성을 반영하여 Re-rank 진행

3.

높은 발전 가능성



외부 데이터 및 관련 논문을
활용하여, 모델 성능 향상 가능

5. 결론 – 개선방안 ① : Feature Augmentation



1. 뉴스 관련 변수

국내 대표 검색엔진 사이트를 관찰해보면, 오전 인기 검색어 순위에 특정 기업 이름이 있는 경우가 많다.

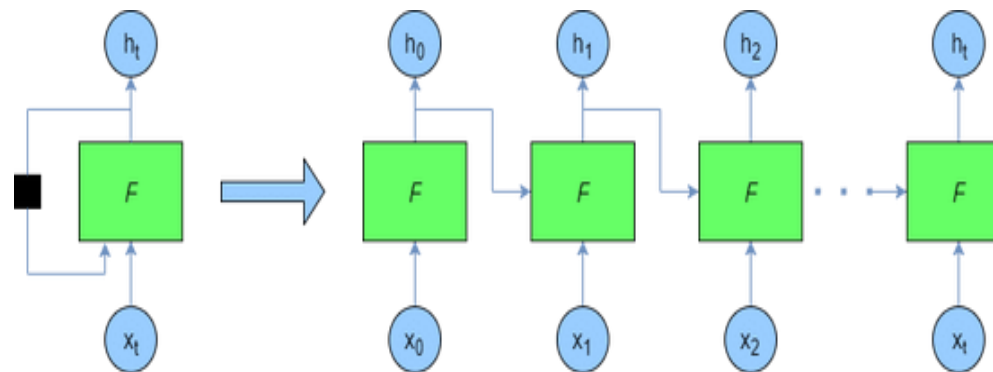


2. 기관 및 외국인 투자 관련 변수

기관 및 외국인의 매수/매도 동향 역시 개인 고객 매수에 큰 영향을 미친다.

→ 매수 고객 수에 영향을 미치는 외부 변수들을 추가하여 성능 개선

5. 결론 – 개선방안 ② : Model Concatenation



주식 데이터의 시계열(time-series) 특성

→ 주식 가격 데이터는 관측값들이 상호 종속적 관계에 있음. 따라서 추세성, 비선형성, 계절성 등의 특징을 고려한 모델 선정이 필요함.

RNN(순환신경망) 계열의 딥러닝 모델 접목

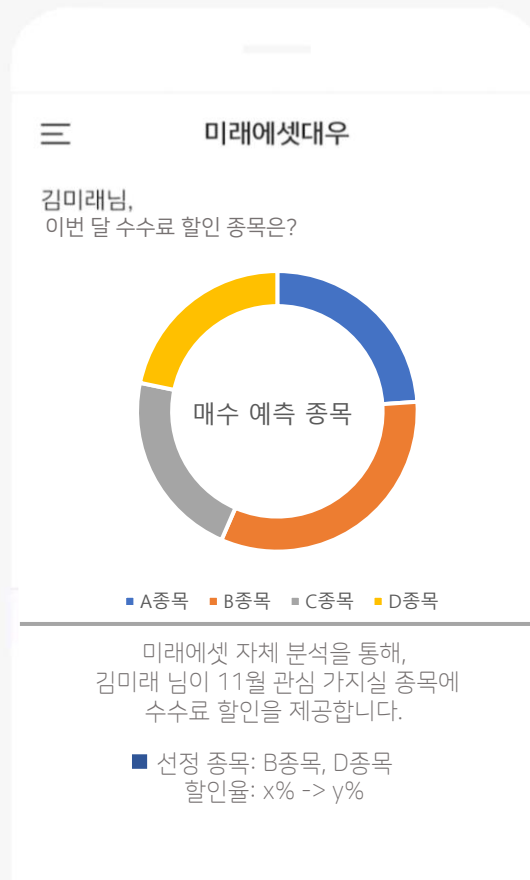
→ GRU, LSTM 등 RNN(Recurrent Neural Network) 계열의 우수한 모델을 선정하여 기존 모델에 접목시켜 성능 개선

VI. 비즈니스 활용방안



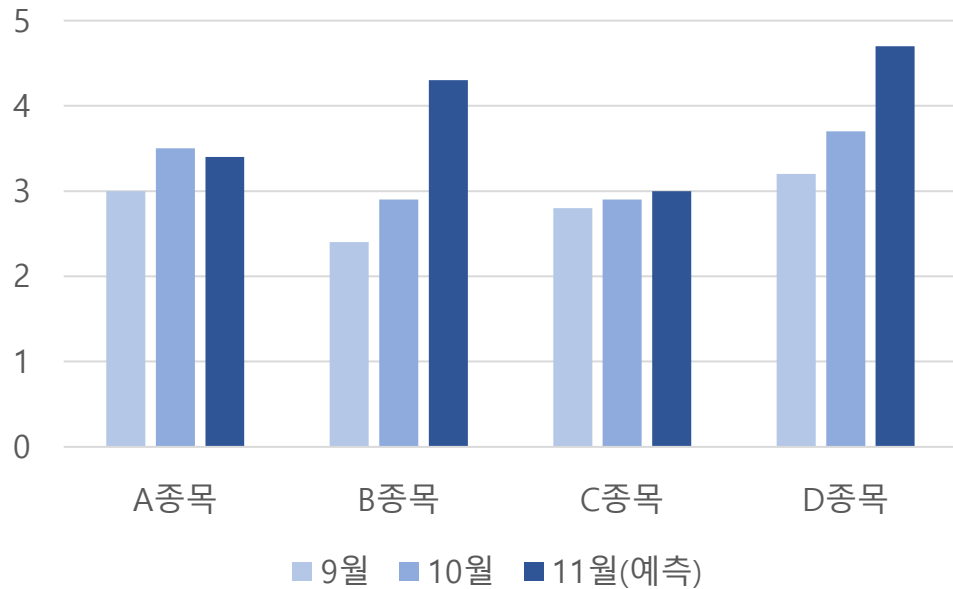
5. 결론 – 비즈니스 활용방안

나의 선호 종목 맞춤 수수료 할인

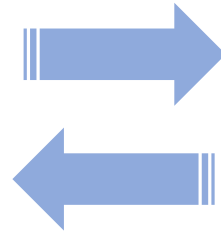


5. 결론 – 비즈니스 활용 방안

Ex) 김미래 고객의 11월 매수 종목 예측 지수



자체 데이터 기반 알고리즘 적용
타 플랫폼 대비 경쟁력 확보



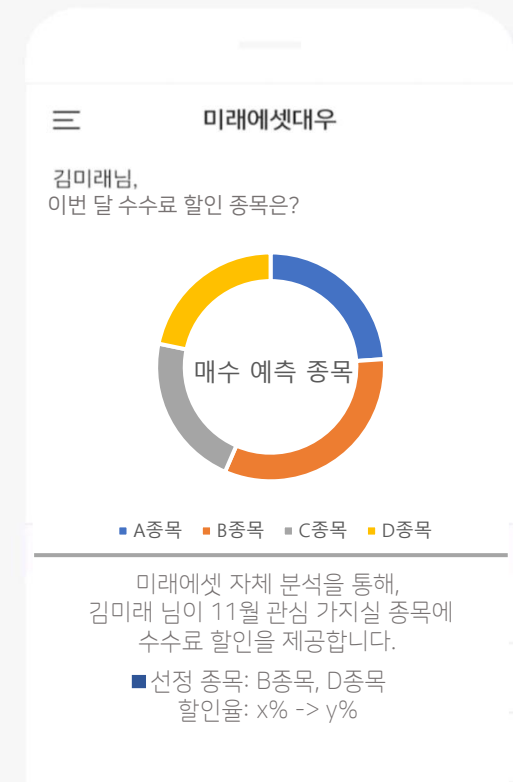
사용자 거래 내역을 통해
알고리즘 적합성 검증 및
추가 데이터 확보

자체 알고리즘 분석에 따라, 사용자 맞춤 수수료 인하 종목 선정

Ex) 위 경우, 11월 높은 매수가 예측되는 **B, D 종목의 수수료** 할인

+ **신규 고객**의 경우,
서비스 가입 신청서 및 과거 거래 내역서를 통해 데이터 확보

나의 선호 종목
맞춤 수수료 할인



Thank you !