

MY GROUND

우리의 생각이 세상을 바꿉니다.
OUR THOUGHTS MAKE DIFFERENCE

: CONTENTS

온라인 가맹점 방문 고객 예측 알고리즘 개발

—
쉽고 빠른 알고리즘 개발

— SAMSUNG CARD DATA COMPETITION

MY GROUND

이 구역 정복자는 나야!

WE ARE

'MY GROUND(마이그라운드)'는 마치 게임에서 미션을 달성해 나가듯 우리의 생각으로 주변을 바꾸어 나가며 결국에는 모두 우리 세상으로 만들어 보자는 의미를 담고 있습니다.

SKILLS

분석력	<div></div>	100%
통찰력	<div></div>	100%
창의력	<div></div>	100%
도전정신	<div></div>	600%

MEMBERS

김태언 : 국민대학교 빅데이터 경영통계학과

김예은 : 숙명여자대학교 미디어학과

이현아 : 성신여자대학교 수학과

데이터 탐색

본격적인 데이터 분석에 앞서 데이터의 구조를 파악하고,
모델링을 위한 기본적인 분석의 틀을 마련하는 과정

— 데이터 탐색

samp_cst_feat.csv	온라인 가맹점 이용 회원의 특징 정보 샘플
samp_train.csv	온라인 가맹점 이용 이력
mrc_info.csv	온라인 가맹점 업종 정보

1차로 제공받은 샘플 데이터를 종합적으로 분석하기 위해 MRC_ID_DI (온라인 가맹점 번호)를 기준으로 데이터를 병합하여 하나의 데이터프레임을 생성하였습니다

	cst_id_di	VAR002	VAR003	VAR004	VAR005	VAR006	VAR007		VAR222	VAR223	VAR224	VAR225	VAR226	VAR227	MRC_ID_DI	Category
0	90000000089	-0.06610	0.5280	-0.13607	0.10945	0.06557	0	...	0.27091	0.01931	0.02938	0.17105	0.12537	0.22197	0	미이용
1	900000000176	-0.09537	0.1347	-0.13541	0.17331	-0.19657	0	...	0.30207	0.06053	-0.01107	0.12413	0.29702	-0.31717	8	홈쇼핑
2	900000000210	-0.01048	0.8360	0.37797	-0.10970	0.52032	1	...	-0.56565	-0.17840	-0.06314	-0.17111	-0.32239	0.33962	0	미이용
3	900000000212	0.05194	0.7505	0.04611	-0.16512	0.07413	0	...	-0.16357	-0.05697	0.01587	-0.04022	0.31213	-0.00559	5	종합몰
4	900000000213	-0.08536	0.3767	-0.12288	0.10023	-0.43414	0	...	0.46971	0.07964	-0.04698	0.03581	0.22588	-0.34868	6	홈쇼핑

분석 결과, 고객수는 10124명으로 중복값은 없었으며, 한 고객당 하나의 가맹점만을 이용하였다는 사실을 파악하였습니다.

cst_id_di	
0	900000000089
2	900000000210
5	900000000263
6	900000000322
7	900000000354
...	...
10118	90000459958
10120	90000460117
10121	90000460233
10122	90000460310
10123	90000460313

8200 rows × 229 columns

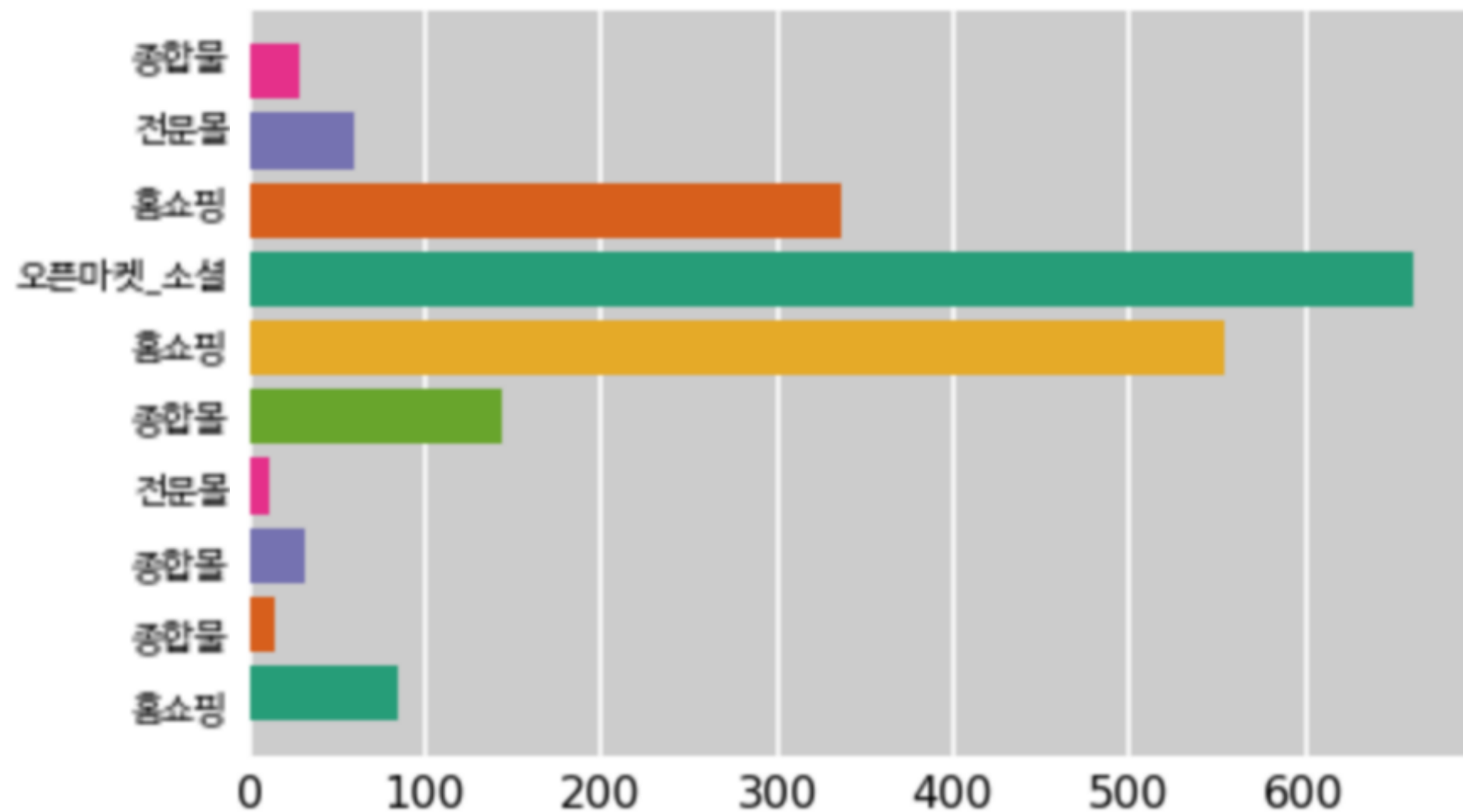
데이터 탐색

고객의 가맹점 이용을 예측하는 알고리즘을 개발하고 이를 활용한 마케팅 아이디어를 제시하는 것이 저희의 목표이기에 가장 먼저 가맹점 별 이용 고객 현황을 파악하고자 했습니다.

MRC_ID_DI (온라인 가맹점 번호) 별로 고객수를 분석한 결과, 미이용 고객 수가 8200명으로 전체 데이터의 약 80.9%를 차지하고 있습니다.

미이용과 이용 고객의 수가 상당한 차이를 보이는 것으로 보아 이를 고려하는 것이 알고리즘 개발에 효과적일 것이라 사료됩니다.

— 데이터 탐색



가맹점 별 이용고객의 수를 비교하기 위해
MRC_ID_DI (온라인 가맹점 번호) 별 고객수를 분석한 결과,

오픈마켓소셜의 이용 고객이 661명 으로 가장 많았으며
홈쇼핑이 그 뒤를 이은것으로 보입니다.

반면, 전문몰의 이용고객 수는 11명으로 가장 적었으며
종합몰의 이용고객 수 또한 상대적으로 낮았습니다.

모델링 및 마케팅 전략

앞서 분석한 데이터를 기반으로
전반적인 모델링 과정 및 마케팅 전략을 수립

— 모델링 프로세스

1. 데이터 전처리 및 파생변수 선정

가맹점별 혹은 미이용고객과 이용고객을 구별할 파생변수를
집중하여 선정하고 이를 위한 데이터 전처리 진행

2. 다양한 모델 구축

Regression, Boosting, Stacking,
SVM, Tree 등 다양한 모델 구축

3. 모델별 파라미터 튜닝

구축한 모델 별 최고 성능을 끌어내기 위한
하이퍼 파라미터 튜닝

4. 최종 앙상블 모델 개발

위의 모델들을 종합하여 앙상블을 진행하고
최종 모델 구축 및 검증

— 데이터로 바라본 마케팅 전략

빅데이터 시대가 도래한 만큼, 카드사가 보유한 데이터를 활용한
개인 맞춤형 마케팅 전략이 요구된다고 생각합니다.

이용고객 대비 미이용고객의 수치가 큰 차이를 보이는 변수들을 파악하여
미이용 고객의 카드사용을 촉진 및 기존 고객의 이탈을 방지할 수 있는 전략이 필요합니다.

예를 들어, 오픈소셜마켓을 이용한 고객의 연령대가 20-30대 로 집중되어 있을 경우
다른 연령대의 고객들을 대상으로 하는 카드사용혜택 이벤트를 개설하고,
20-30대에게는 지출의 부담을 덜어 줄 캐시백, 무이자할부와 같은 맞춤형 이벤트를 진행하는 것이 좋습니다.



MARKETING

2020.09.25

THANK YOU!

MY GROUND

SAMSUNG CARD

DATA ANALYSIS