

Data based Targeting collaboration Retailor X Manufacturer

- Insight based Category Look-alike
- Propensity Modeling

**Junyoung Lee
DaeKyung Lim**

P&G

Downy^{ULTRA}

Pampers®

BRAUN

Gillette®

head & shoulders

Oral-B

febreze

SK-II

Gillette®
Venus®

PANTENE



Junyoung Lee
P&G Analytic & Insights Manager



DaeKyung Lim
P&G Data Science Team Leader

Carlson Funds Enterprise
Fixed Income Analyst



Corporate & Marketing Strategy



Analytic & Insights



GBS Consultant



Product Reco. Team



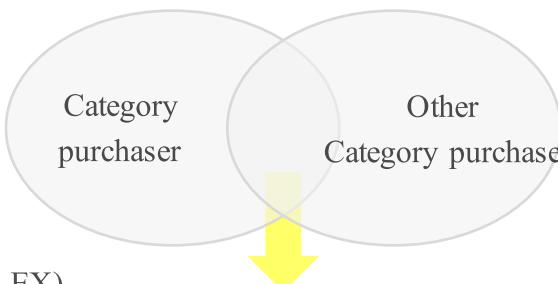
Data Science Team

Project overview

Data analysis & targeting projects

Drive Category growth & recruit new users

Phase 1) Insight based shopper clustering

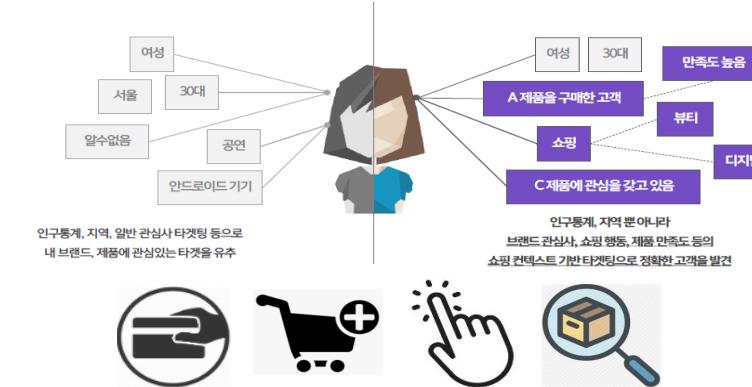


Understand the category
growth shoppers
characteristic

EX)
[Sports LOVER] [New hired] [Grooming]

Develop the promotion theme & target to wide rage of
shoppers

Phase 2) Propensity modeling



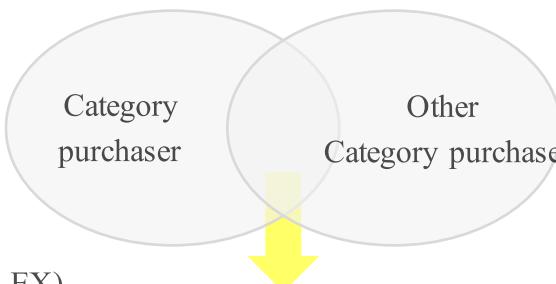
Sorting out shoppers who have high probability to become
new users

Project overview

Data analysis & targeting projects

Drive Category growth & recruit new users

Phase 1) Insight based shopper clustering



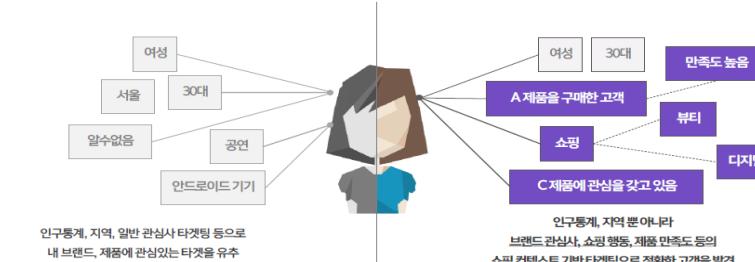
Understand the category
growth shoppers
characteristic

EX)

[Sports LOVER] [New hired] [Grooming]

Develop the promotion theme & target to wide rage of
shoppers

Phase 2) Propensity modeling



인구통계, 지역, 일반 관심사 타겟팅 등으로
내 브랜드, 제품에 관심있는 타겟을 유추



Sorting out shoppers who have high probability to become
new users

WHAT THEY PURCHASE – Category adjacency

How to analyze

*Top selling category = Mean (Log(Purchase)) > Mean (Log(Purchase)) + (Std (Log(Purchase)) * 2)

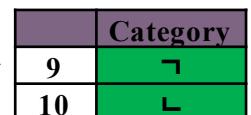
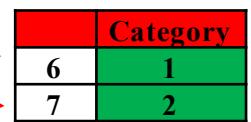
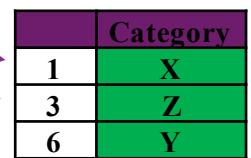
1 Identifying Top selling category* among men

2 Identifying unique characteristics of BnR category users.

**Exclude out categories that show high rank in
TTL Men shoppers purchase category from
BnR Shoppers Top selling category**

Shave Shopper's Purchase category

Grouping the categories that shows similarity & commonality



Target ‘Look-alike’ shoppers with customize communication

Target users =
Category X/Z/Y buyers – BnR
purchaser

**Target users =
Category 1/2 buyers – BnR
purchaser**

**Target users =
Category 1/2 buyers – BnR
purchaser**

Potential Actions

- 1) Expand to external media & develop commercial ideas for boarder use**
- 2) Develop Customized promotion page with targeted app-push**

III. 조사 결과 상세

1. 남성의 외모 관리 및 콘텐츠 이용 경험 | 4) 구독 중인 외모 관리 관련 인플루언서

구독하는 외모 관리 인플루언서: 디렉터 짱구대디 > 디렉터 파이 > 깡스타일리스트·씬님

- 전체 응답자 10명 중 1명(15.8%)은 남성 외모 관리 관련 인플루언서를 구독하고 있음.
- 구독 중인 남성 외모 관리 관련 인플루언서는 디렉터 짱구대디(10.1%), 디렉터 파이(6.3%), 깡스타일리스트·씬님(5.1%) 순임.
- 구독하는 인플루언서는 대부분 남성으로, 주로 같은 성별의 인플루언서에게 외모 관리 관련 정보를 얻는 편임.

구독 중인 외모 관리 관련 인플루언서

[Base: 미응답자 제외, n=79, 복수, 주관식]

*없음, 모름 응답 제외

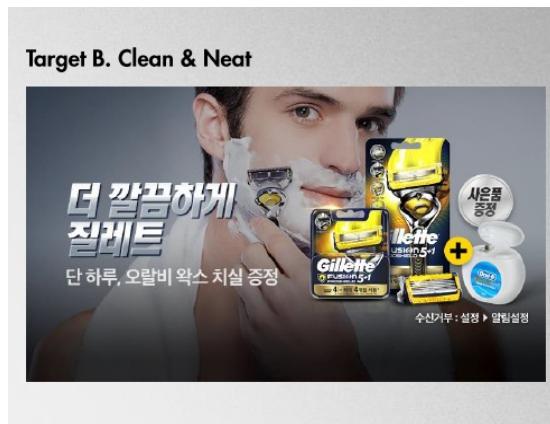
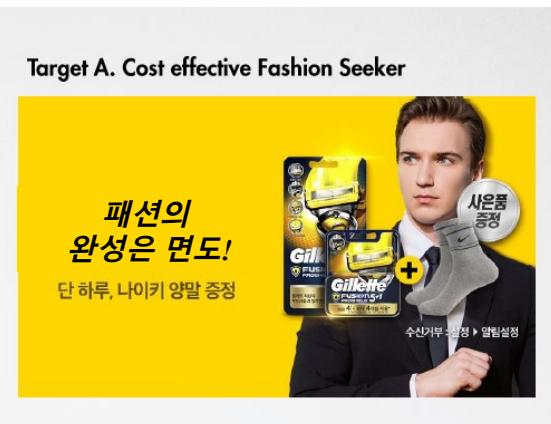
1위	2위	3위		5위		
디렉터 짱구대디	디렉터 파이	깡스타일리스트	씬님	금강연화	신쿡	스타일가이드 최겨울
패션	뷰티	패션	뷰티	헤어	뷰티	패션
10.1%	6.3%	5.1%	5.1%	3.8%	3.8%	3.8%

*그 외: 김인호TV, 무신사TV, 아우라M, 이사배, 패션TV정대, 포맨트, 피부는민동성 등

*이미지 출처: Youtube 공식계정

Potential Actions

- 1) Expand to external media & develop commercial ideas for boarder use
- 2) Develop Customized promotion benefit with targeted app-push

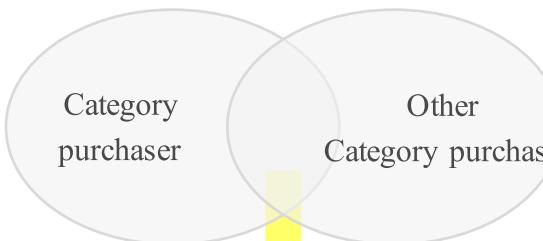


Project overview

Data analysis & targeting projects

Drive Category growth & recruit new users

Phase 1) Insight based shopper clustering

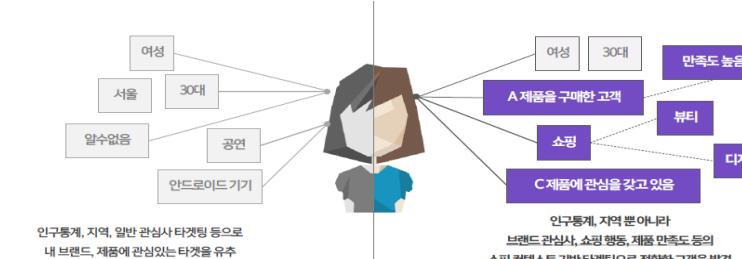


EX)
[Sports LOVER] [New hired] [Grooming]

Develop the promotion theme & target to wide rage of shoppers

Understand the category growth shoppers characteristic

Phase 2) Propensity modeling



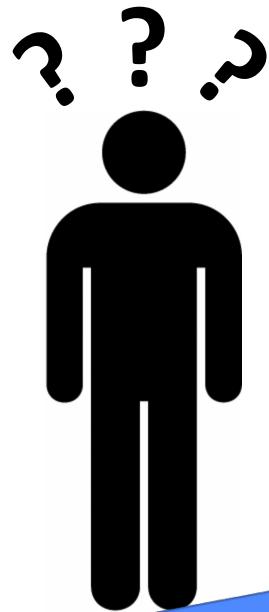
인구통계, 지역, 일반 관심사 타겟팅 등으로
내 브랜드, 제품에 관심있는 타겟을 유추



Sorting out shoppers who have high probability to become Gillette new users

Big Data – Propensity Modeling

How to identify “WHO”?



“WHAT” to sell?



...can answer the question!



P&G Knowledge Database



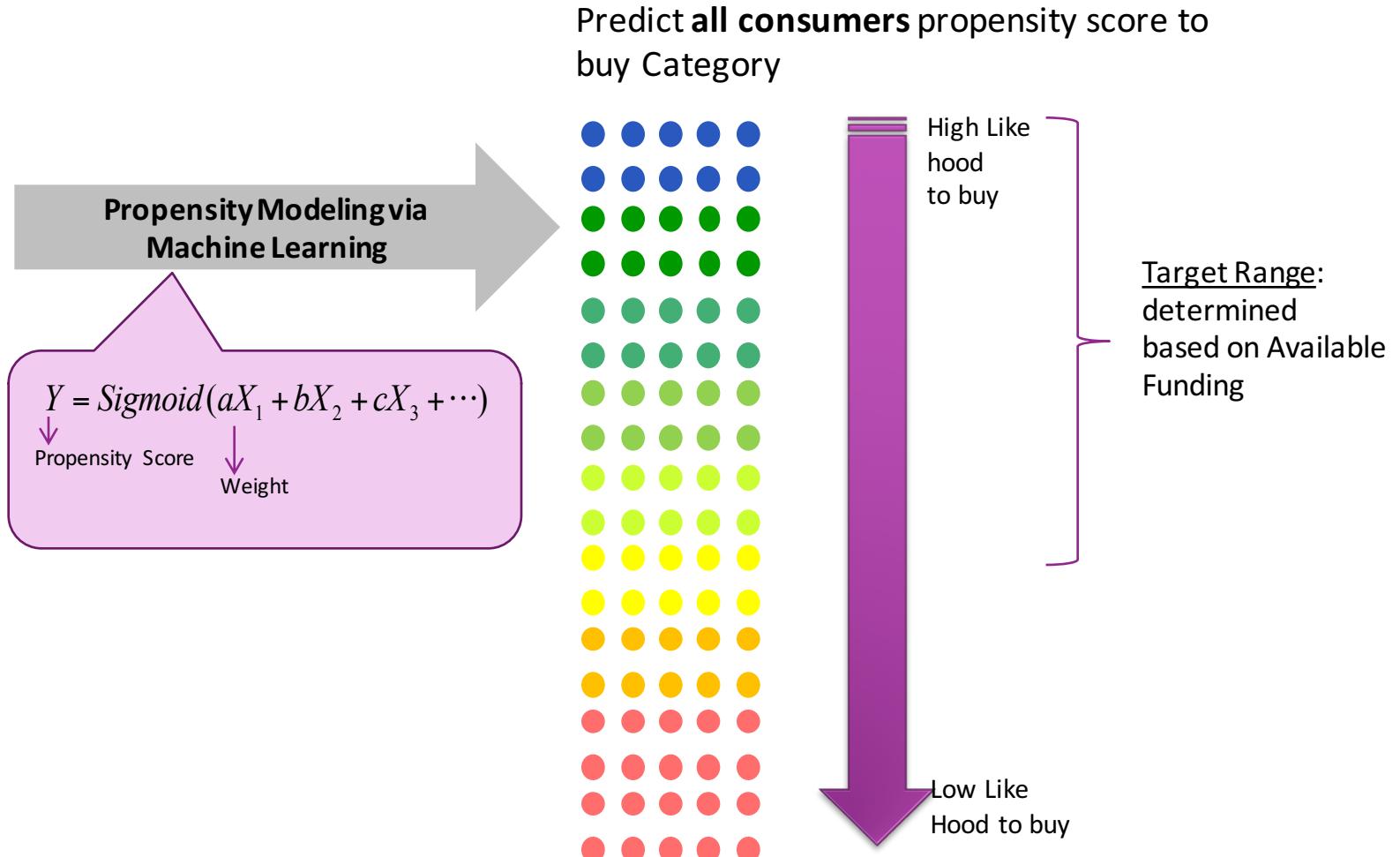
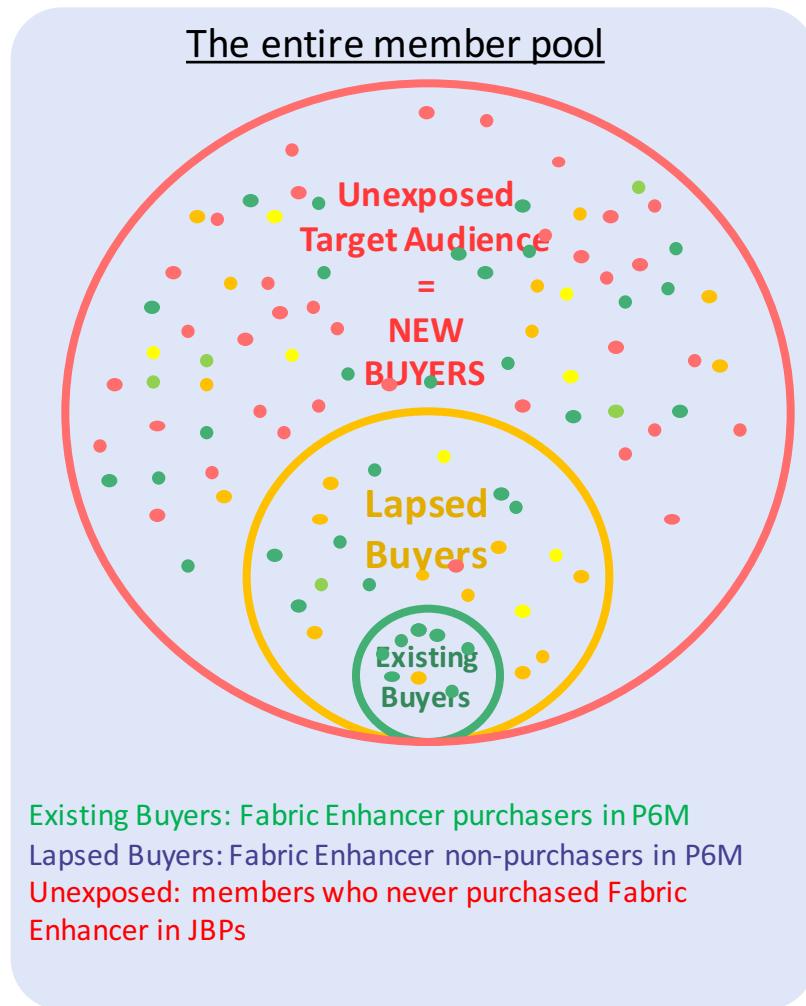
P&G Analytics Capability



eCommerce Data

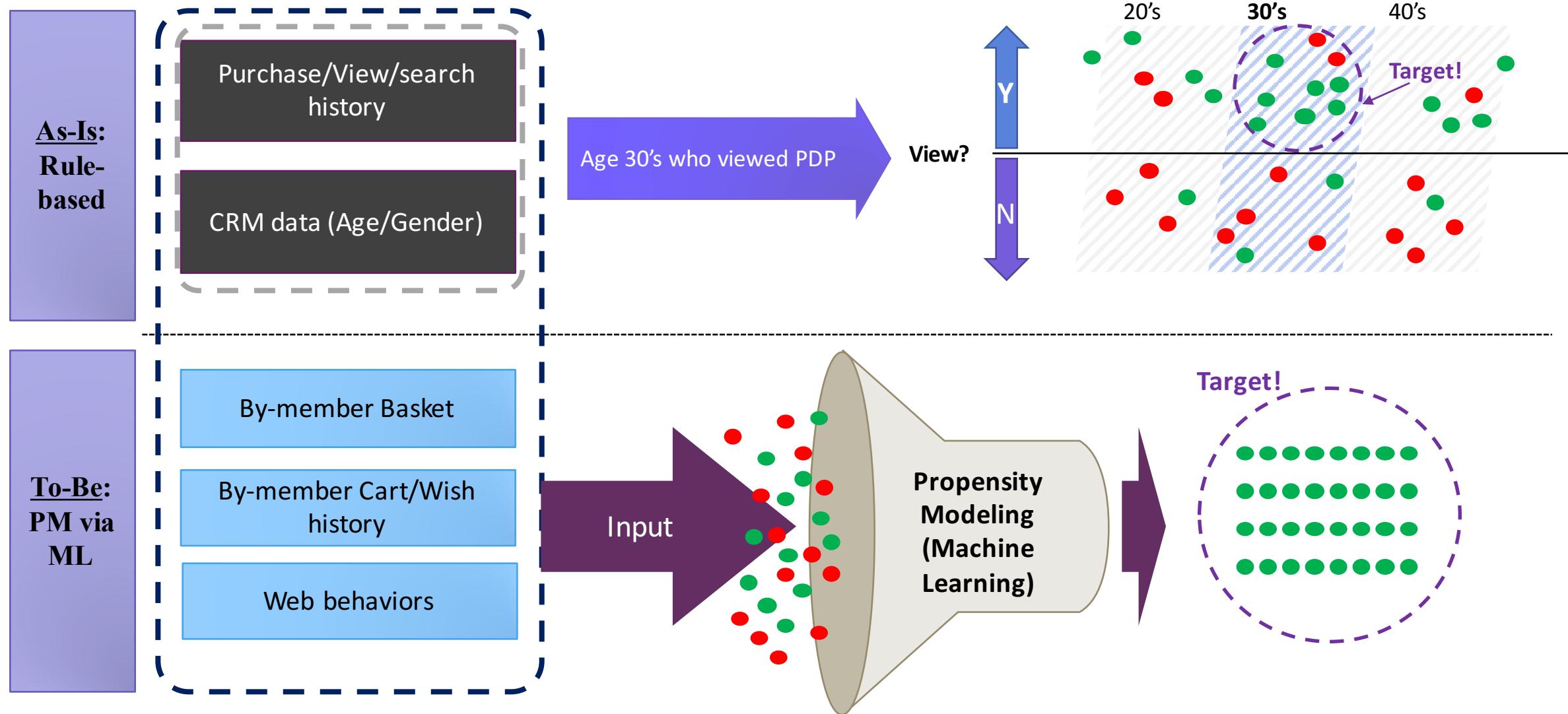
Mock-up Example – Targeting

Propensity modeling starts with the all members in the specific eCommerce, the model results in “likelihood to” by each member level.



Mock-up Example – Targeting

Currently, the way to target is rule-based method but this method cannot capture all the “high-propensity” to buy FE category (green-colored) members. However, propensity modeling via Machine Learning can capture “green-colored” members.



Mock-up Example – Inputs for Propensity Modeling

Data should be wrangled distinct-by-member level; Model trains on positive & negative samples and learns patterns for users.

Features(Tags) including purchasing history/demographic/
cart/wishlist/viewership

ID	Downy User?	Age Group	Gender	Purchased Fashion	Purchased Electronics	Viewed Kids	Viewed Beauty
A	1	1	0	13	0	5	20
B	1	2	1	59	5	0	7
C	1	3	0	0	0	55	0
D	1	1	0	2	0	3	0
E	0	2	1	0	9	21	8
F	0	3	0	1	1	4	22
G	0	4	0	0	0	2	36
H	0	6	1	0	10	0	6

Positive samples;
members
purchased
category in P6M

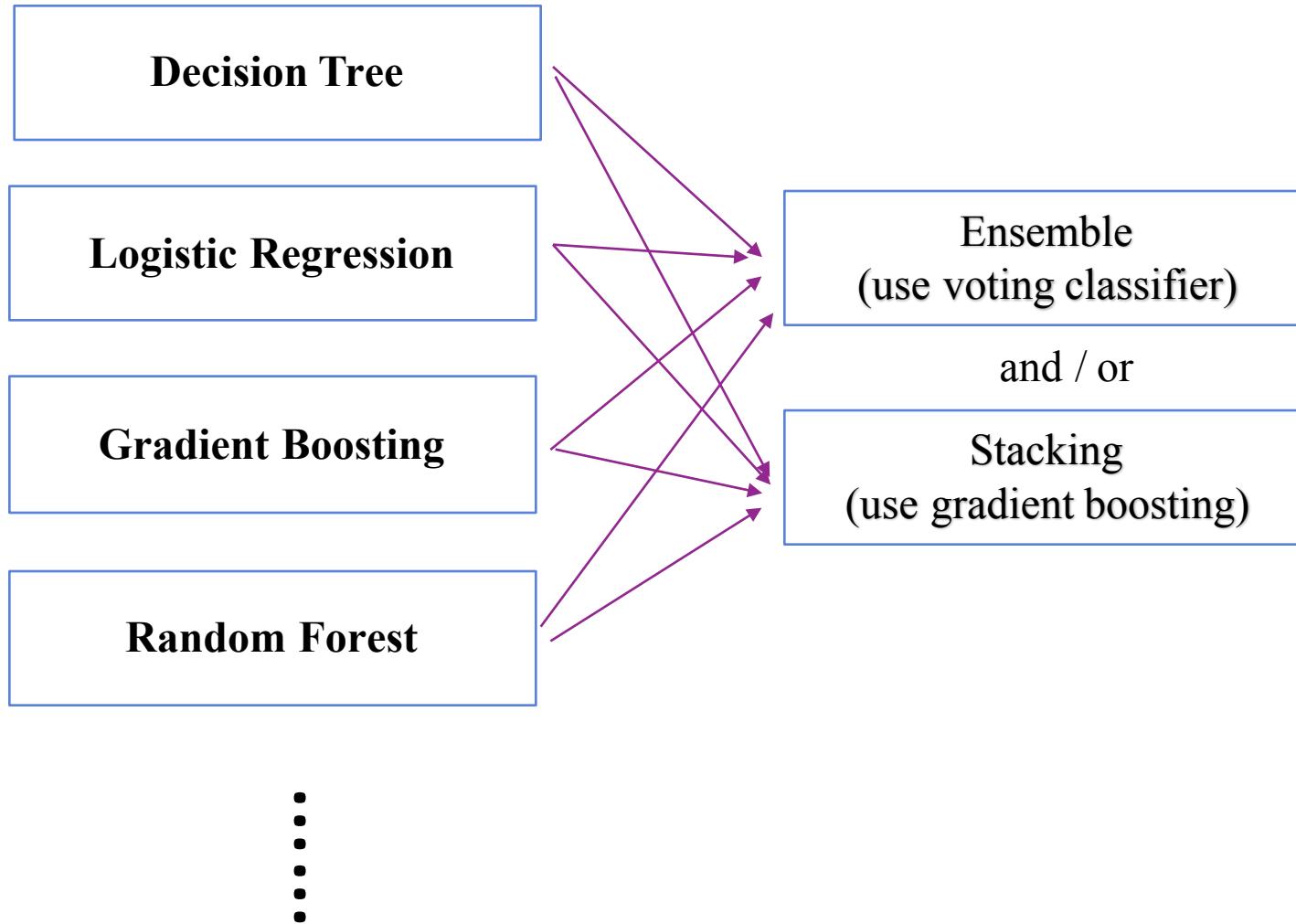
Negative samples;
members
not purchased
category in P6M

.....

•
•
•
•

Mock-up Example – Models for Propensity Modeling

There are so many algorithms to classify users. You can use only one algorithm, you can use ensemble or stacking with many algorithms



Use random forest algorithms as a Baseline model.

Use ensemble and stacking to beat baseline model.

Another choice can be user-based collaborative filtering.

Mock-up Example – Result for Propensity Modeling

How to evaluate Propensity Model

Confusion matrix

	P(predicted)	N(predicted)
P(Actual)	TP	FN
N(Actual)	FP	TN
Our customers	Potential new customers They didn't buy our product in this mall. But model said they have very similar shopping interest to our existing customers → Persuadables	

Watch out

No Interest

Mock-up Example – Result for Propensity Modeling

How to evaluate Propensity Model

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

Precision = $\frac{TP}{TP+FP}$

Recall = $\frac{TP}{TP+FN}$

Accuracy is not meaningful, most of case our data set is imbalanced.(Non Downy user >>> Downy user)

Focus on the recall score. High Precision is not useless, our goal is find new users(Persuadables – FP) that has similar behavior and interest to our existing users

Make to higher recall, you have to adjust threshold.

Like

`y_test_pred_prob[:,1] > 0.5`

`y_test_pred_prob[:,1] > 0.45`

`y_test_pred_prob[:,1] > 0.40 ...`

A blurred background image of a classroom or lecture hall. Numerous students are seated at wooden desks arranged in rows, facing towards the front of the room where a teacher or lecturer would typically stand. The students are dressed in casual attire, and the overall atmosphere appears to be a typical classroom setting.

Thank you