o @ ®

MANUFACTURER'S COUPON

**$1.25 O**

Save $100

OFF

*WEB*

ac

2 WEEKS ONLY! MFG Coupon on any ONE (1) BUNDLE PACK® of

THREE (3) boxes or more of KLEENE Facial Tissue (not valid on trial size)

*A*NTAC 360°\* PRODUCT

**364**

***lantac***

ANY ONE (1) ICY HOT PRAI (Excludes 1.25oz cream, 1cti

trial and travel sizes)

IXIMUM STUNGIH

Klee

:

***leen*er**

**tac**

360

SHWA

**Predicting Coupon**

**Redemption**

**ww**

WWWWWW

MANUFACTURER'S COUPON - EXPIRES 01*/*

**w**

빅데이터 분석가 양성과정

**S*A*VE $6**

Team Project

7

*ON* ANY ONE *(1*) BIOTRUE" PURPOSE S*OLUTIO*N 2x10

SMARTV PANTS

CONSUMER: UMIT 1 C | PRODUCT SIZES AND

any other offers. reproduced or law. Consum Puerto Rico

same sh

\*

CONSUMER: Only one coupon *purc*hase on pro*du*ct Specifie*d* only. Any *oth*er use constit transferre*d, expir*e*d, pm* purchase *produc*ts

p*er household CONSUMI RETAILE*R: "

**RETA**

S SMART

S*IZES ANT 8t hande*

팀장 김채린 팀원 권용재

김지은 박가온 이자영

0851356004-130622

*offers. Voit*

the Baust

(CRPF1) inco, *from origir* use constitutes *sales tax. / expir*ed, tax*ed, G*s. *military ha to purchase products*

*per purchase of spediti*e *tran*s*actio*n upons Der househol*d* pe *or credit in* of *1c. Not jedeemable for moll consumer. Good only in U*S4. *Mail to: Baus submitted i Dept. 10119,1 f*aw*ati DrDen*

*is a trademark of Bausch & Lo*

*I affiliates. O 2022 Baust d/b/a Sanofi* its affiliates. PN *Box 880001 participates in the CIC*

*0041167*

***on any O*N*E (1) Sm*artyPants Kids, *T*e*en,* T*oddl*er or Baby Multivitamin**

*MANUFACTURER*S COUPON DO NOT DOUBLE I EXP. DATE: 03*/*06/2022 Con*sumer* This coupon is to be used on the purchase of any KIDS, TEEN *TOOD*L*ER OR B*ABY SMARTY*P*ANTS product. Void it altered reproduced, *transf*e*rred,* sol*d* of sucioned. Retailer Maximum Value $3.00. Smantypants wil *tem*buse tace value of coupon plus 8c shipping and handling, which vites you compliance with SmartyPants redempton policy, which is ble upon request Coupon rembursement not to be deducted from its product invoices Do not double or combine with any other operty redeemed coupons to SmartyPants, MPS PO.

6. CINNAMINSON NJ 08077

E02

**Content**

빅데이터 분석가 양성과정

Team Project

주제 선정 배경

데이터 설명

데이터 전처리

데이터 분석

- 주제 선정 배경 소개

- 사용 데이터 소개 - 피쳐 구성 소개

- 결측치 대체 - Label Encoding - Data Merge

■ 결측값 처리

- 학습 데이터 분리

■ 하이퍼파라미터 튜닝 적용 후

LightGBM - Feature Importance

E02

**주제 선청 배경**

● 판매 촉진을 위한 쿠폰 발행 - 기업에서는 판매 촉진을 위해 쿠폰 발행 이벤트를 진행

| 치열한 경쟁상황 속에서 온라인 쇼핑몰은 고객 유치를 위해 쿠폰 발행, 행사 할인 등 다양한 프로모션에 대해 많은 돈을 투입

● 마켓컬리 재구매율 77.3%, 고객 충성도 지속 상승 이유는?

빅데이터 분석가 양성과정

Team Project

치열한 경쟁상황 속에서 살아남기 위해 고객 유치를 위한 쿠폰, 인지도를 높이기 위한 광고, 고객을 유인할 수 있는 프로모션 등에 대해 많은

돈을 투입 - 그 결과, 2019년에는 이 수치가 61.2%였는데 2020년에는 65.2%, 올해는 71.3%로 계속 성장 - 그만큼 고객들의 충성도도 점점 높아짐

● 기업의 쿠폰을 받은 고객이 물품을 구매할 때 쿠폰을 사용하는지에 대해 알아보고자 프로젝트를 진행

o@

**데이터 설명**

= kaggle

Search

+

Create

- 25

VASUDEVA - UPDATED 3 YEARS AGO

25

**New Notebook**

New Notebook

**Download (12 MB)**

**Download (12 MB)**

0

Home

Competitions

**Predicting Coupon Redemption**

M

Datasets

<> Code

E

Discussions

빅데이터 분석가 양성과정

9

Learn

Data

Code (4)

Discussion (1)

Team Project

More

**About Dataset**

**Usab**ility o

4.71

É

Your Work

RECENTLY VIEWED

Problem Statement

**License** Unknown

BOOKMARKS

Predicting Coupon Redemption

빅분기 이용자 가이드

**Expected update frequency** Not specified

XYZ Credit Card company regularly helps its merchants understand their data better and take key business decisions accurately by providing machine learning and analytics consulting. ABC is an established Brick & Mortar retailer that frequently conducts marketing campaigns for its

diverse product range. As a merchant of XYZ, they have sought XYZ to assist them in their discount marketing process using the power of machine learning.

Discount marketing and coupon usage are very widely used promotional techniques to attract new customers and to retain & reinforce loyalty of existing customers. The measurement of a consumer's propensity towards coupon usage and the prediction of the redemption behaviour are crucial parameters in assessing the effectiveness of a marketing campaign.

ABC promotions are shared across various channels including email, notifications, etc. A number of these campaigns include coupon discounts that are offered for a specific product/range of products. The retailer would like the ability to predict whether customers redeem the coupons received across channels, which will enable the retailer's marketing team to accurately design coupon construct, and develop more precise and targeted marketing strategies.

6

View Active Events

**Data Explorer**

**train.csv**

**78,000JH)**

*V*ariable

id

**Name**

Definition 138 L 19 ID

7

campaign\_data.csv(1KB)

campaign\_id

0119 D

coupon\_id

201419 D

coupon\_item\_mapping.csv(810KB)

customer\_id redemption\_status(TARGET)

11. ID

0:018*/*1:18

customer\_demographics.csv(17KB)

5

customer\_transaction.csv(50,358KB)

**campaign\_data.csv**

**287H)**

빅데이터 분석가 양성과정

Definition

Team Project

item\_data.csv(2,387KB)

01

10 D

test.csv(859KB)

*V*ariable campaign\_id campaign\_type

start\_date end\_date

할인 행사 유형

3

할인 시작일

train.csv(1,456KB)

할인 종료일

**couponitemmapping.csv(&#93,000JH)**

Definition

Variable

coupon\_id

item\_id

01210 D

42

**customer\_demographics.csv(약 760개)**

*V*ariable

Definition customer\_id

고객 고유 ID

age\_range

고객 연령대

**| customer\_transaction.csv(약 1,324,000개)**

Variable

Definition

date

거래일

customer\_id

고객 고유 ID

martial\_status

기혼 / 미혼

item\_id

상품 고유 ID

rented

10 : 임대하지 않음 / 1 : 임대함

구매한 아이템 수량

family\_size

가족 구성원

quantity selling price other\_discount

15 16 |

판매 가격

| noofchildren

| 제조업체 쿠폰/포인트 카드와 같은 다른 출처에서 할인

|

가족 자녀의 수 레이블 인코딩된 소득(소득이 높을수록 숫자가 큼) |

income\_bracket

소매점 쿠폰에서 사용 가능한 할인

빅데이터 분석가 양성과정

coupon\_discount

Team Project

**item\_data.csv(약 74,000개)**

**test.cs**v(약 50,000개)

Variable

Definition

Variable

Definition 쿠폰 고객 노출 고유 ID

item\_id

상품 고유 ID

id

brand

상품 브랜드 고유 ID

12 |

할인 행사 고유 ID)

campaign\_id

coupon\_id

brand\_type

브랜드 유형(Local/Established)

할인 쿠폰 고유 ID

category

|

상품 목록

customer\_id

고객 고유 ID

o@

campaign\_id

Campaign Data

campaign\_id campaign\_type start\_date end\_date

**•**

Item Data

item id brand brand\_type category

Train **id** campaign\_id coupon\_id customer\_id redemption\_status

coupon\_id

•

item\_id

빅데이터 분석가 양성과정

Coupon Item Mapping

coupon\_id item\_id

Team Project

customer\_id

customer\_id

item\_id

customer\_id

Customer Demographics

customer\_id age\_range marital\_status rented family\_size no\_of\_children income\_bracket

Customer Transaction Data

date customer\_id item\_id quantity selling\_price other\_discount coupon\_discount

o@

**EDA - >EX80190|| [[EEE 1**

0.0025

5000

0.0020

4000

0.0015

3000

**brand**

Density

campaign\_duration

Density

2000

0.0010

1000

0.0005

0.0000

0

1000

2000 3000

**brand**

4000

5000

6000

ALLA 30 35 40 45 50 55

campaign\_duration

60

65

redemption\_status

redemption status

빅데이터 분석가 양성과정

Team Project

**age\_range**

Density

marital status

Density

-1

0

i

2 2 3 age range

-0.25

0.00

0.25 0.50 0*7*5

marital status

100

125

redemption status

redemption\_status

**EDA - 쿠폰사용여부에 따른 분포 확인**

0.08

0011

0.06 1

income\_bracket

Density

Density

0.03

001

001

000)

|

10

20

0

| 02 46 8 10 12

income bracket

redemption\_status

redemption\_status

빅데이터 분석가 양성과정

Team Project

2001

1751

150

Density

**month**

Density

075

0.50

015

2

4

6

|

10

12

14

redemption status

month

**redemption\_status**

o

**결측값 대체** no\_of\_children

• Nane 009 CHF

marital status

· family size - no\_of\_children = 1, 'Single'

family size - no\_of\_children >= 2, 'Married'

빅데이터 분석가 양성과정

**marital status family size no\_of\_children**

**Married**

**NaN** 175

1 63 2 31

3+ 48

**Single**

**NaN 65**

Team Project

**marital\_status family\_size no\_of\_children**

**NaN**

**NaN 183 2 NaN 89**

18 31 26

**2 4** 2 12

3+ 1 **5+**

3+ 6

**NaN**

**1**

**5**

3+

3

o@

**Label Encoding**

- Marital Status 0

· age\_range . campaign\_type

• brand\_type

· category

from sklearn.preprocessing import Label Encoder encoder = Label Encoder)

*#Label Encoding Marital Status* --*- Ois Single and 1 is Married* cust\_demo["marital\_status"] = encoder.fit\_transform(cust\_demo["marital\_status"])

*# Label Encoding age\_range. 18-25 is 0,26-35 is 1, 36-45 is 2, 46-55 is 3, 56-70 is 4 and 704 is 5* cust\_demo["age\_range" ] = encoder.fit\_transform(cust\_demo["age\_range"])

빅데이터 분석가 양성과정

Team Project

*#Label Encoding Campaign type* campaign["campaign\_type" ] = encoder.fit\_transform(campaign.campaign\_type)

*#Label Encoding the brand\_type and category columns* # 브랜드 타임과 카테고리 라인 코딩 items.brand\_type = encoder.fit\_transform( items["brand\_type"]) items.category = encoder.fit\_transform(items["category"])

o@

**Data Merge & Aggregate**

Step 1 쿠폰상품매핑 데이터와 상품데이터 item\_id로 merge

campaign\_id

Campaign Data

• campaign\_id

campaign\_type start\_date end\_date

•

Item Data item\_id brand brand\_type category

In [48] : coupons\_items = pd. merge(coupons, items, on="item\_id", how="left")

Train id campaign\_id coupon\_id customer\_id redemption\_status

coupon\_id

item\_id

빅데이터 분석가 양성과정

Coupon Item Mapping

coupon\_id item\_id

In [49] : coupons\_items. head() Out [49]

**coupon\_id item\_id brand brand\_type category**

Team Project

customer\_id

customer\_id

105

37

56

item\_id

Customer Demographics

customer\_id age\_range marital\_status rented family\_size no\_of\_children income\_bracket

1 2 3

Customer Transaction Data

**date** customer\_id item id quantity selling price other\_discount coupon\_discount

customer\_id

107 *75* 561 494 76 209 06 522*77 278*

518

77

278

CO2

**Data Merge & Aggregate**

Step 2 item\_id 별로 고객 구매정보 통계내기

상품별 구매정보에 대한 평균 aggreggte

상품별 쿠폰 할인 총액, 상품 구매자 수, 다른할인총액, 팔린 수량, 상품별 총 수익 aggregate

**item\_id**

**coupon\_discount coupon\_used no\_of\_customers other\_discount quantity**

00

| 0

2

0 0 10 0.0

0 1

10

**se**lling price

124.31 35.26

100

00

**item\_id cd\_sum t\_counts od\_sum** | 1 0 2 0 12 0 1 0 13 0 0 1 0 14 0 0 1 0 0

5 0 1 0

0 0

**qu\_sum price\_sum**

2 248.62 1 35.26 1 56.64 1 54.85 1 81.57

00

56.64

10

00

00

10

빅데이터 분석가 양성과정

54.85

00

00

10

31.57

Team Project

상품별 통계정보 취합 후, 파생변수 생성

\* 상품별 통계정보 취한 후 | 상품별 총 할인 평군, 상품별 총 할인의 한 생성

transactions1 ['total\_discount\_mean'] = transactions1 ['coupon\_discount'] + transactions['other\_discount'],

secull "tot**al\_discount\_mean total\_discount\_sum** transactions['total\_discount\_sum'] = transactions ['od\_sum'] + transactions1 ['cd\_sum'] transactions1.head()

00

0 0

o@

**Data Merge & Aggregate** Step 3 1872 item\_id& merge

campaign\_id

Campaign Data

• campaign\_id

campaign\_type start\_date end\_date

item\_coupon\_trans = pd. merges coupons\_items, transactions, on="item\_id', how="left')

item\_coupon\_trans.head()

•

Item Data item\_id brand brand\_type category

Train id campaign\_id coupon\_id customer\_id redemption\_status

coupon\_id

item\_id

**0**

빅데이터 분석가 양성과정

Coupon Item Mapping

coupon\_id item\_id

0.0

coupon\_id item\_id brand brand\_type category coupon\_discount coupon\_used no\_of\_cu**stomers**

105 37 56

0.0

0.0

2.0

107 *75* 56

4.0 494 76 2090 6

0.0

0.0

1.0 522 *77 2*78 0 6

0.0 0.0

2.0 518 *77* 278 0 6

0.0

2.0

Team Project

customer\_id

customer\_id

item

id

2 3 4

customer\_id

Customer Demographics

customer\_id age\_range marital\_status rented family\_size no\_of\_children income\_bracket

Customer Transaction Data

**date** customer\_id item id quantity selling price other\_discount coupon\_discount

408

**Data Merge & Aggregate**

Step 4 상품별 통계 정보 3번 table을 coupon id 별로 통계내기

# 쿠폰별 통계내기 # 쿠폰별 사용가능한 상품 수, 최빈 브랜드, 최빈 브랜드 타입, 최빈 카테고리, # 쿠폰별 할인액 평균, 쿠폰 사용 고객 평균, 다른 할인액 평균, 쿠폰별 상품 구매 수량 평균 # 쿠폰별 상품 평균구매액 평균, 쿠폰 사용수, 사용자 수,

# 다른할인액 합계, 쿠폰할인액 합계, 총할인액 평균, 총할인액 합계

빅데이터 분석가 양성과정

Team Project

coupon. head)

**item\_counts no\_of\_customers**

**od\_sum**

**other\_discount**

**price\_sum**

**qu\_sum**

**quantity**

**total\_discount\_**

39

-20306.33 -1163.52

14.794872 -18780.020833 15,000000 -1163.521667

8.588235 -4055.343333 22.333333 -25895.740000 6,000000 -1228.880000

-16.620713 89796.868333 | -21.343885 8940.520000

-14.728021 27756.490000 -36.718597 142874.023333

| -27.265786 16636.570000

17 24

1034.

01.221644 103.0 1.137500 248.0 1.121525 702.0 1,020872

44.0 1,000000

**selling price t\_counts total\_discount\_mean**

101.183245 | 826.0

-17.942237 122.534500 | 181.0

-21.343885 131.655894 212.0

-17.475379 211.708369 | 676.0

-38.203749

403.970000 144.0

1-27.265786

-4634.15

-26777.33

| -1228.88)

o@

**Data Merge & Aggregate**

Step 5 고객구매정보 데이터 customer id 별로 통계내기

*# Aggregate tra&actions by customer\_id* transactions3 = pd.pivot\_table(cust\_tran, index = "customer\_id",

values=['item\_id', 'quantity', 'selling\_price', 'other\_discount', 'coupon\_discount','coupon\_used', 'day', 'dow', 'month'], aggfunc={'item\_id': lambda x: len(set(x)),

'quantity': np. mean, 'selling\_price': np. mean, 'other\_discount":np. mean, 'coupon\_discount': np. mean, 'coupon\_used': np. sum, 'day': lambda x: mode(x) [O] [O],

dow': lambda x: mode(x) [O] [O], 'month': lambda xi mode(x) [O] [O]}

빅데이터 분석가 양성과정

Team Project

transactions3.reset\_index(inplace=True) transactions3.rename( columns={'item\_id': 'no\_of\_items'}, inplace=True) transactions3. head()

**customer\_id coupon\_discount coupon\_used day dow no\_of\_items month other\_discount quantity selling\_price**

-2.019876

*78* 3 3

463

5 -12.837537 1.170802 97.470480

-0.595084

4 13 5 3526 -13.4321951.131265 107.805783 -3.091546

53 16 4

- 14.074853 11.578723 85.082452

-0.404773

1 14 5

-8.883656 1.27*272*7 138.256770

-0.114684

2 11 1

5 - 11.260696 117.869949 115.482842

406

125

490

o@

**Data Merge & Aggregate**

Step 6, 7,8581 240|| CICO]El campaign\_id */*

7w1 13H G E|O|El customer\_id = / 687 88 customer\_id merge

빅데이터 분석가 양성과정

def merge\_all(df):

df = pd. merge(df, coupon, on="coupon\_id", how="left") df = pd. merge(df, campaign, on="campaign\_id", how="left") df = pd. merge(df, cust\_demo, on="customer\_id", how="left") df = pd. merge(df, transactions, on='customer\_id', how="left') return df

Team Project

train = merge\_all(train) test = merge\_all(test)

train.shape, test . shape *((*78369, 46), (50226, 45))

o@

**결측값 처리**

함수를 이용하여 결측값 처리

def deal\_na(df):

for col in cust\_demo.columns.tolist([1:]:

df [col].fillna(model df [col]).mode[O], inplace=True) return df

1. age\_range 2. marital\_status 3. rented 4. family\_size 5. no\_of\_children 6. income\_bracket

train = deal\_na(train)

빅데이터 분석가 양성과정

test = deal\_na(test)

Team Project

age\_range marital\_status rented family\_size no\_of\_children income\_bracket

OOOOOO

o@

**학습 데이터 분리**

필요없는 변수 제거 후 학습과 테스트 데이터 분리

test\_id = test['id'] target = train['redemption\_status'] train.drop(['id','campaign\_id', 'start\_date', 'end\_date', 'redemption\_status'], axis=1, inplace=True) test .drop( ['id','campaign\_id', 'start\_date', 'end\_date'], axis=1, inplace=True)

train.columns

빅데이터 분석가 양성과정

Team Project

**+**

Index(['coupon\_id', 'customer\_id', 'brand', 'brand\_type', 'category', 'cd\_sum',

'coupon\_discount\_x', 'coupon\_used\_x', 'item\_counts', 'no\_of\_customers', 'od\_sum', 'other\_discount\_x', 'price\_sum', 'qu\_sum', quantity\_x', 'selling\_price\_x', 't\_counts', 'total\_discount\_mean', 'total\_discount\_sum', 'campaign\_type', 'campaign\_duration', 'age\_range', 'marital status', 'rented', 'family\_size', 'no\_of\_children', 'income\_bracket', 'coupon\_discount\_y', 'coupon\_used\_y', 'day', 'dow', 'no\_of\_items', 'month', 'other\_discount\_y', 'quantity\_y',

selling price\_y', 'cdd\_sum', 'customer\_id\_count', 'odd\_sum', 'qa\_sum', 'pprice\_sum'], dtype='object)

**I**

x\_train,x\_test,y\_train,y\_test = train\_test\_split(x,y,test\_size=0.3, random\_state=2439)

408

모델 비교

여러 분석기의 모델 정확도 비교 후, 불균형 데이터를 보완하기 위해 SMOTE 기법 적용

**Model Accuracy**

0

Logistic Regression

0.502287

Naive Bayes 0.638832 Decision Tree 0.509512

Random Forest 10,500000

빅데이터 분석가 양성과정

Logistic Regression ------ Weighted Accuracy: 0.9907484208511452 RandomForest Classifier ------ Weighted Accuracy: 0.99074842085114521 Gradient Boosting Classifier ------ Weighted Accuracy: 0.9893447329802846 LGBM Classifier ------ Weighted Accuracy: 0.990365596886365

Team Project

0.509661

Bagged-LR

Bagged-NB

0.623031

Bagged-DT

0.500762

Boosted-LR

0.500000

일부 모델에 대해 SMOTE기법 적용 결과, 너무 높은 정확도로 인해

과적합으로 판단하여 SMOTE 기법 적용하지 않기로 판단

Boosted-NB 0.566727

Boosted-DT

0.500000

LGBM

0.794657

o@

**LightGBM**

LightGBM + hyper parameter tuning

bayesian\_params = {

'max\_depth': (6, 16), "num\_leaves': (24, 64), 'min\_data\_in\_leaf': (10, 200), *#min\_child\_san /98* 'min\_child\_weight':(1, 50), 'bagging\_fraction':(0.5, 1.0), *# subsano le* 'feature\_fraction': (0.5, 1.0), *# co/ samo 1e\_bytr99* 'max\_bin':(10, 500). 'Tambda\_12':(0.001, 10), *#reg\_lambda* 'Tambda\_l1': (0.01, 50) *#reg\_alpha*

import light gbm as Igb

빅데이터 분석가 양성과정

def lgb\_f1\_score[y\_hat, data):

y\_true = data.get\_label().astypel int) y\_hat = np.round(y\_hat). astype(int) *soikits f1 doesn't like probabilities* return 'fl', f1\_score[y\_true, y\_hat, average="weighted'), True

Team Project

train\_data = |gb. Dat aset (data=x\_train, label=y\_train, free\_raw\_data=False)

*90*3 *93*

def lgb\_roc\_eval\_cv(max\_depth, num\_leaves, min\_data\_in\_leaf, bin\_child\_veight, bagging\_fraction,

feature\_fraction, max\_bin, lambda\_12, lambda\_11): params = {

"num\_iterations":500,"Tearning\_rate":0.02.

early stopping\_rounds':100, 'metric': 'auc", 'max\_depth": int(round(max\_depth)), #JË A +*2 o HDD*E +\* *50/1 Z+ZXL' ELE A* "num\_leaves': int(round(nul\_leaves)). 'min\_data\_in\_leaf': int(round(min\_data\_in\_leaf)). "min\_child\_weight': int(round(min\_child\_weight)). "bagging\_fraction': max(min(bagging\_fraction, 1), 0), "feature\_fraction': max(min(feature\_fraction, 1), 0), 'max\_bin': max(int(round(max\_bin)),10), 'Tambda\_12': max(lambda\_12,0), "lambda\_l1': max(lambda\_11, 0)

*# 740/*

*/ightgome or H*

*S*G 18.

cv\_result = 1gb.cv(params, train\_data, nfold-3, seed=0, verbose\_eval =100, early stopping\_rounds 50,

metrics=['auc'], feval=lgb\_f1\_score) return max(cv\_result['auc-mean'])

o@

**LightGBM**

LightGBM + hyper parameter tuning

IgbBO = BayesianOptimization(Igb\_roc\_eval\_cv, bayesian\_params, random\_state=0) lgbB0. maximizes init\_points = 5, n\_iter=25)

빅데이터 분석가 양성과정

L

iter

target

baggin... | featur... | lambda\_11 | lambda\_12 | max\_bin

| max\_depth / min\_ch... , min\_da... | num\_le...

Team Project

L

데이터그대로, 베이지안튜닝 | 30 | 0.9307 | 1.0 10.5 COE/2CH2, HOA10141+f1-score

| 30 | 0.8869 0.7962 0.8055

|

| 0.01 | 5.456

| 0.001 | 9.706

| 42.11 / 358.3

| 11.22 | 10.07

/ 31.87 | 6.75

| 200.0 | 32.92

64.0 | 36.68

o@

**LightGBM**

최적 파라미터 값 추출하여 선택

*#dictionaryol E target29* target\_list = [] for result in TgbB0.res:

target = result['target']

target\_list.append(target) print(target\_list)

*# 7 3 target 39 = (index)* print('maximum target index:', np.argmax(np.array(target\_list)))

빅데이터 분석가 양성과정

[0.8864175166355279, 0.887712830679772, 0.8647927053294614, 0.879825075223725, 0.8635599014391508, 0.8631 289938805627, 0.8921 796095474054, 0.8859578924833814, 0.8747843346804219, 0.888325609997563, 0.8781835755046505, 0.8878241537542477, 0.891 20023091 73741, 0.8700784624067529, 0.8380731049375193, 0.8759697006620364, 0.8855782834436727, 0.8647989986529114, 0.8840072795242543, 0.93064900347603, 0.8785446218844347, 0.8339259634390697, 0.8341285132610373, 0.8342656971352369, 0.8700970250666775, 0.8525469027117*77*2, 0.8872982646081259, 0.875378*7*09993501 4, 0.8787539938805627, 0.8868725926715676] maximum target index: 19

Team Project

max\_dict = |gbB0.res(np.argmax(np.array(target\_list))] print(max\_dict)

{'target': 0.93064900347603, 'params': {'bagging\_fraction': 0.6988204802600908, 'feature\_fraction': 0.9660757567475873, 'Tambda\_ll': 1.128 3877191876799, lambda\_12': 6.911608197236618, 'max\_bin': 68.10699765361503, 'max\_depth': 11.576110031361186, 'min\_child\_weight': 34.71860 8412164315, 'min\_data\_in\_leaf': 13.97742609628643, 'num\_leaves': 54.444802638132984}}

params={'bagging fraction': 0.6988204802600908,

'feature\_fraction': 0.9660757567475873,

lambda\_11': 1.1283877191 876799, lambda\_12': 6.911608197236618, 'max\_bin': 68, 'max\_depth': 11, 'min\_child\_weight': 34.718608412164315, 'min\_data\_in\_leaf': 14, 'num\_l eaves': 54}

o

**LightGBM**

최적 파라미터 값 적용

def train\_apps\_all(x,y, params):

train\_x, test\_x,train\_y, test\_y = train\_test\_split(x,y,test\_size=0.3, random\_state=2439)

cl f = LGBMClassifier

\*\*params

clf.fit(train\_x, train\_y, eval\_set=[(train\_x, train\_y), (test\_x, test\_y)],

eval\_metric= 'auc', verbose=100, early stopping\_rounds= 100)

빅데이터 분석가 양성과정

Team Project

return clf

clf = train\_apps\_all(x,y, params)

[Light GBM] [Warning] feature\_fraction is set=0.966075*7*567475873, col sample\_byt ree=1.0 will be ignored. Current value: feature\_fraction=0.9 660757567475873 [Light GBM] [Warning] min\_data\_in\_leaf is set=14, min\_child\_samples=20 will be ignored. Current value: min\_data\_in\_leaf=14 [Light GBM] [Warning] lambda\_11 is set=1.1283877191876799, reg\_alpha=0.0 will be ignored. Current value: lambda\_11=1.1283877191876799 [Light GBM] [Warning] bagging\_fraction is set=0.6988204802600908, subsample=1.0 will be ignored. Current value: bagging\_fraction=0.69882048 02600908 [Light GBM] [Warning] lambda\_12 is set=6.911608197236618, reg\_lambda=0.0 will be ignored. Current value: lambda\_12=6.911608197236618 [100] training's auc: 0.966346 training's binary\_logloss: 0.0297907 valid\_i's auc: 0.958094 valid\_i's binary\_logloss: 0.031240

o@

**Feature Importance**

coupon\_used\_y

customer\_id

coupon\_used\_x

coupon\_id

coupon\_discount\_x

selling\_price\_y

quantityy

campaign\_duration

**month**

빅데이터 분석가 양성과정

coupon\_discount\_y

Team Project

**day**

cd\_sum

other\_discount\_y

**qa\_sum**

odd sum

no\_of\_customers

cdd\_sum

no\_of\_items

customer\_id\_count

selling\_price\_x

Feature

item\_counts

quantity\_x

o @

**출처**

- CoE: Kaggle(https*://www*.kaggle.com*/*datasets*/v*asudeva009/predicting-coupon-redemption

- 0|0||1: flaticon(https*://ww*w.flaticon.com/kr*/*free-icon*/*presentation\_3534081?related\_id=3534081)

- O|0|1|2 : flaticon(http*s://www.*flaticon.com*/*kr*/*free-icon*/*coupons\_3706131?term=%EC%BF%A0%ED%8F%B0&pa ge=1&position=17&page=1&position=17&related\_id=3706131&origin=search)

빅데이터 분석가 양성과정

- 0|0|713 : flaticon(https*://www*.flaticon.com*/*kr*/*free-icon/big-data\_2980512?related\_id=2980512&origin=search)

Team Project

NWV

- 0|0|||4: flaticon(https*://ww*w.flaticon.com/free-icon*/*process\_4149678?related\_id=4149678)

빅데이터 분석가 양성과정

감사합니다:)

29

Team Project