Reselling platform

누가 C2C 온라인 플랫폼에서 중고 명품을 구매할까?
(C2C: Consumer To Consumer)

Contents

1. Exploration

어떤 사용자가 구매를 할까?

Vestiaire Collective

AUTHENTICATED PRE-OWNED LUXURY FASHION

Vestiaire Collective 는 온라인 빈티지 몰

C2C 플랫폼이 성공하기 위해서
거래액을 늘려야 함

이번 분석에서는, 이 플랫폼에 가입한
사용자 중 구매자의 특성을 살펴봄

Vestiaire Collective의 사용자 데이터

identifierHash
type
country
language
socialNbFollowers
socialNbFollows
socialProductsLiked
productsListed
productsSold
productsPassRate
productsWished
productsBought
gender
civilityGenderId
civilityTitle
hasAnyApp
hasAndroidApp
hasIosApp
hasProfilePicture
daysSinceLastLogin
seniority
seniorityAsMonths
seniorityAsYears
countryCode

98,913 users
24 features

identifierHash
type
country
language
socialNbFollowers
socialNbFollows
socialProductsLiked
productsListed
productsSold
productsPassRate
productsWished
productsBought
gender
civilityGenderId
civilityTitle
hasAnyApp
hasAndroidApp
hasIosApp
hasProfilePicture
daysSinceLastLogin
seniority
seniorityAsMonths
seniorityAsYears
countryCode

98,913 users
15 features

- 데이터셋은 캐글에서 수집[1)]
  - Vestiaier Collective의 사용자 정보
- 전처리 이전
  - 98,913 users, 24 features
  - No missing values
  - No duplicate data
- 9개 특성 삭제
  - 중복되는 의미: type, gender, civilityTitle, hasAnyApp, seniorityAsMonths, seniorityAsYears
  - 특성의 값이 너무 다양함: identifierHash, country, countryCode
- 전처리 후
  - **98,913 users, 15 features**

1) E-commerce - Users of a French C2C fashion store (contributed by JEFFREY MVUTU MABILAMA)

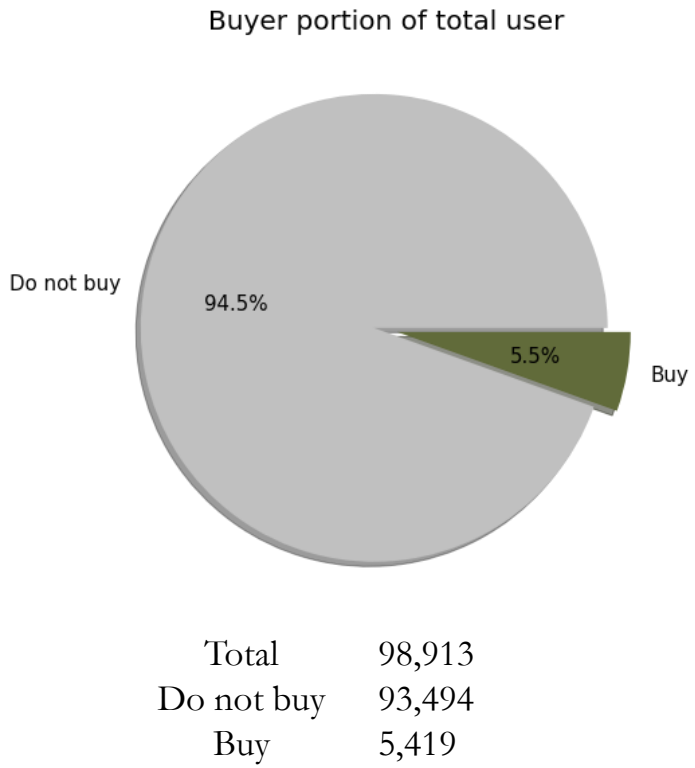
특성 기술

Variable	Description[1)]
language	The user's preferred language
socialNbFollowers	Number of users who follow this user's activity. New accounts are automatically followed by the store's official
socialNbFollows	Number of user account this user follows. New accounts are automatically assigned to follow the official partners
socialProductsLiked	Number of products this user liked
productsListed	Number of currently unsold products that this user has uploaded.
productsSold	Number of products this user has sold
productsPassRate	% of products meeting the product description. (Sold products are reviewed by the store's team before being shipped to the buyer)
productsWished	Number of products this user added to his/her wish list.
productsBought	**Number of products this user bought (Target of this analysis)**
civilityGenderId	1, 2, 3 (1 is Mr., 2 is Mrs, 3 is Miss)
hasAndroidApp	If user has ever used the official Android app
hasIosApp	If user has ever used the official iOS app
hasProfilePicture	If user has a custom profile picture
daysSinceLastLogin	Number of days since the last login
seniority	Number of days since the user registered

1) EDA: Online C2C fashion store - user behaviour (Kaggle, JEFFREY MVUTU MABILAMA)

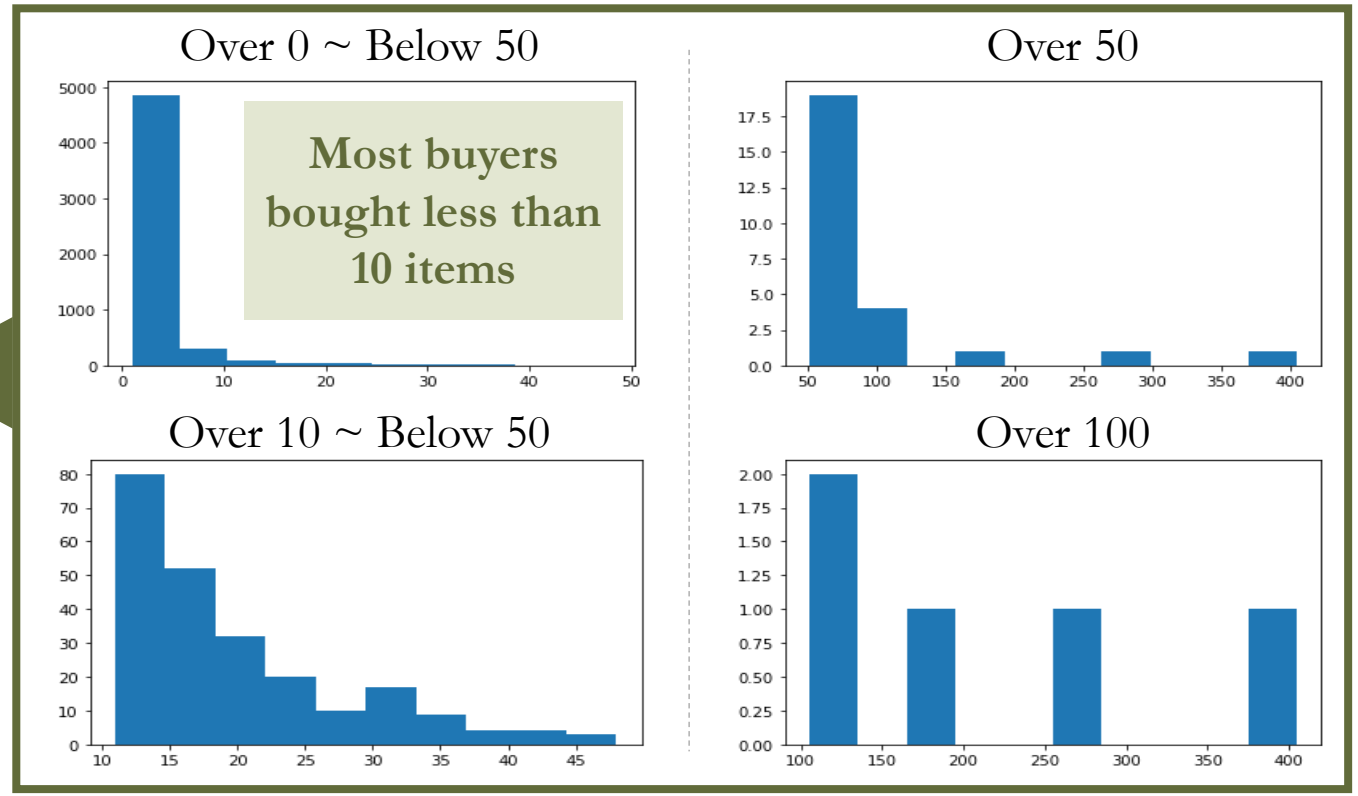
타겟이 불균형하고, 비대칭도가 큼

5.5% of total users ever bought an item



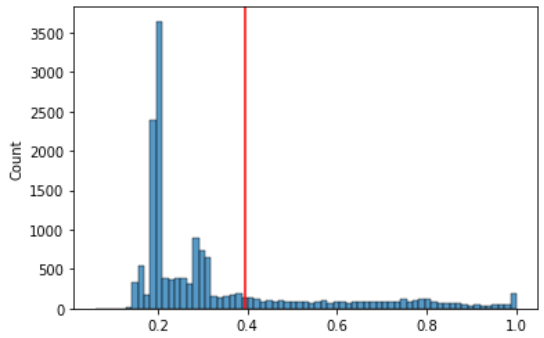
Total	98,913
Do not buy	93,494
Buy	5,419

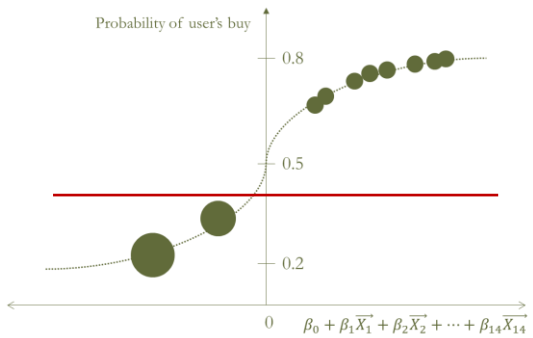
Distribution of buyers by the number of purchased item

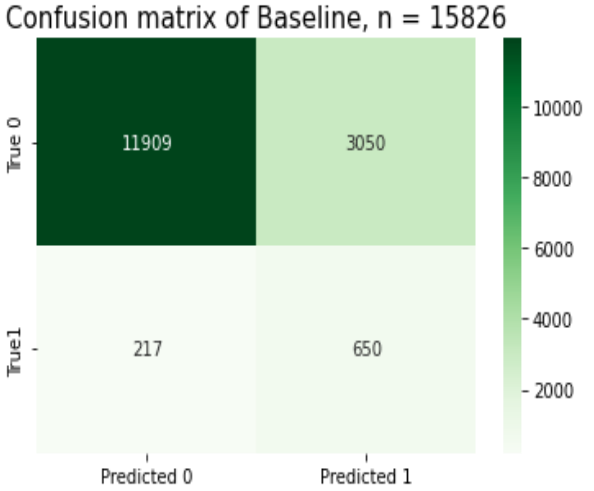


2. Model

Logistic regression, recall is 0.75







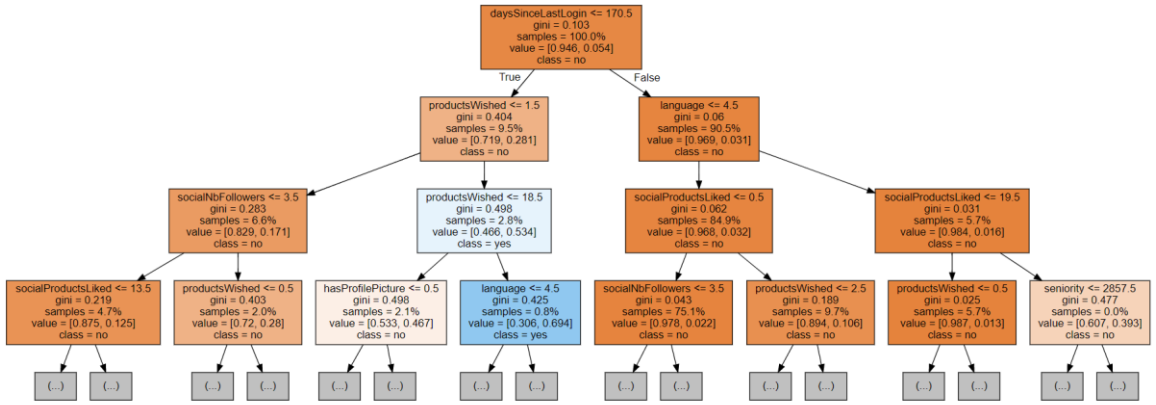
- Logistic regression은 특성과 타겟의 선형식에 sigmoid라는 함수를 씌워 확률을 정하는 모델임
- Threshold에 따라서 이중 분류함
- 최적[1]의 threshold는 0.39로 계산되어, 구매확률이 39%이상인 사용자를 구매자로 분류함

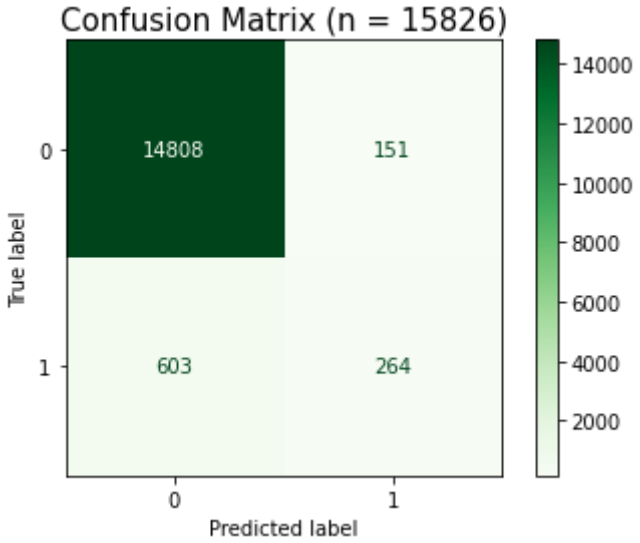
(Recall is 0.75)
This model retrieved 75% of buyers.
(Fail in retrieving 25% of buyers.)

1) 최적의 threshold 기준을 Area Under Curve(AUC) 사용

Random Forest, recall is 0.3

One of many trees

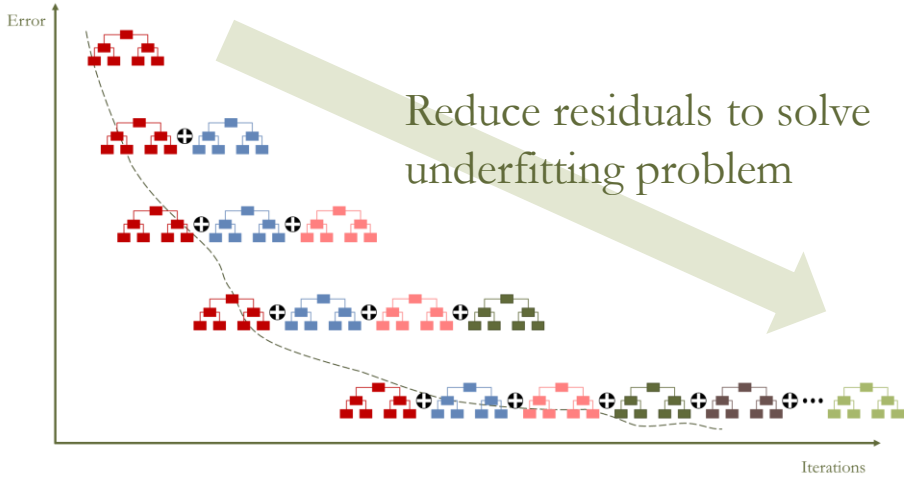


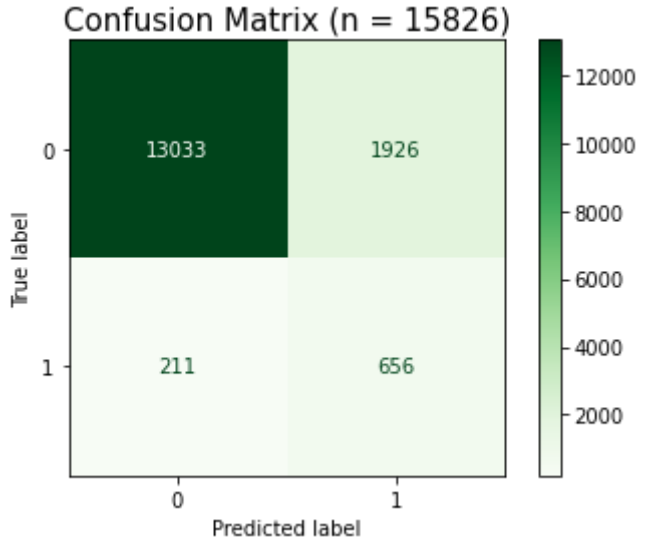


- Random Forest는 Tree를 여러 개 만들어보고, 다수결로 사용자를 구매자/비구매자로 분류함
- 복원 추출로 다양한 data set을 이용하고, 특성도 무작위로 추출한 후 Tree를 만들기 때문에 특정 data set에만 맞는 Tree의 과적합문제를 완화함

(Recall is 0.3)
This model retrieved **30%** of buyers.
(Fail in retrieving 70% of buyers.)

Gradient boosting decision tree, recall is 0.76



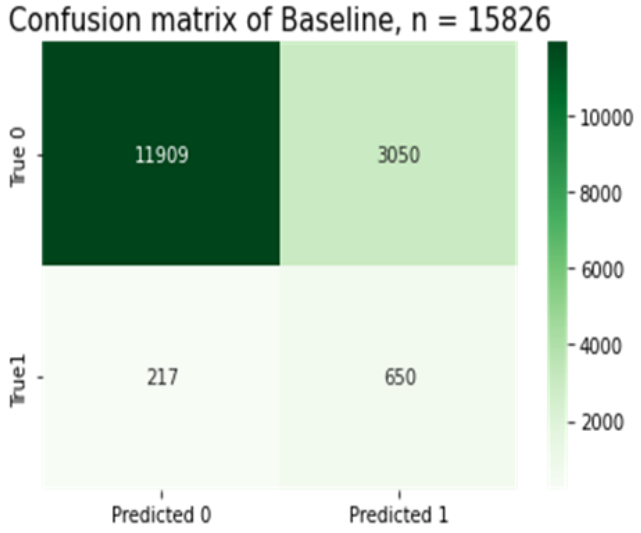


- Gradient boosting decision tree도 여러 Tree를 생성하여 타겟을 예측하는 앙상블 모델임
- Tree의 Leaf 수를 제한하여 과적합문제를 완화하고, 먼저 생성한 Tree의 잔차를 줄여 나가는 Tree를 계속 만들어서 과소적합문제도 완화함

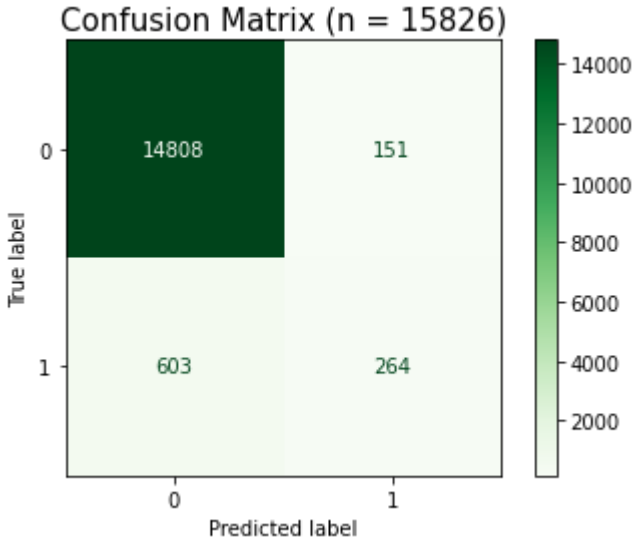
(Recall is 0.76)
This model retrieved **76%** of buyers.
(Fail in retrieving 24% of buyers.)

2-2. Improved model

Gradient boosting decision tree model 의 성능이 근소하게 향상됨

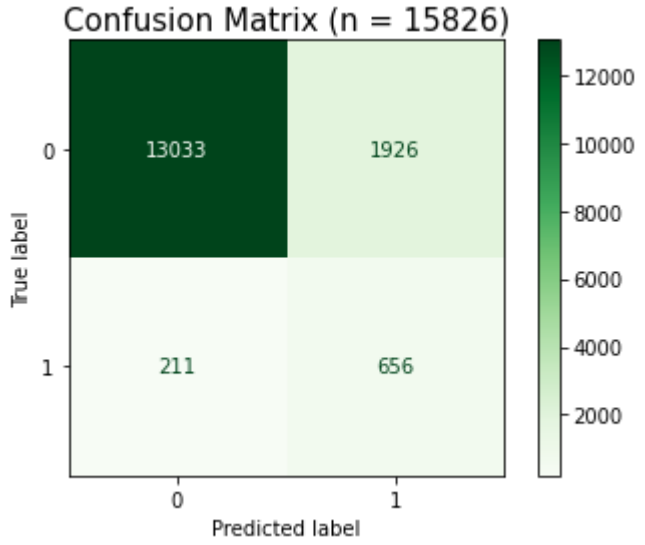


Logistic regression
(baseline) recall **0.75**



Random forest
recall **0.3**

Final model

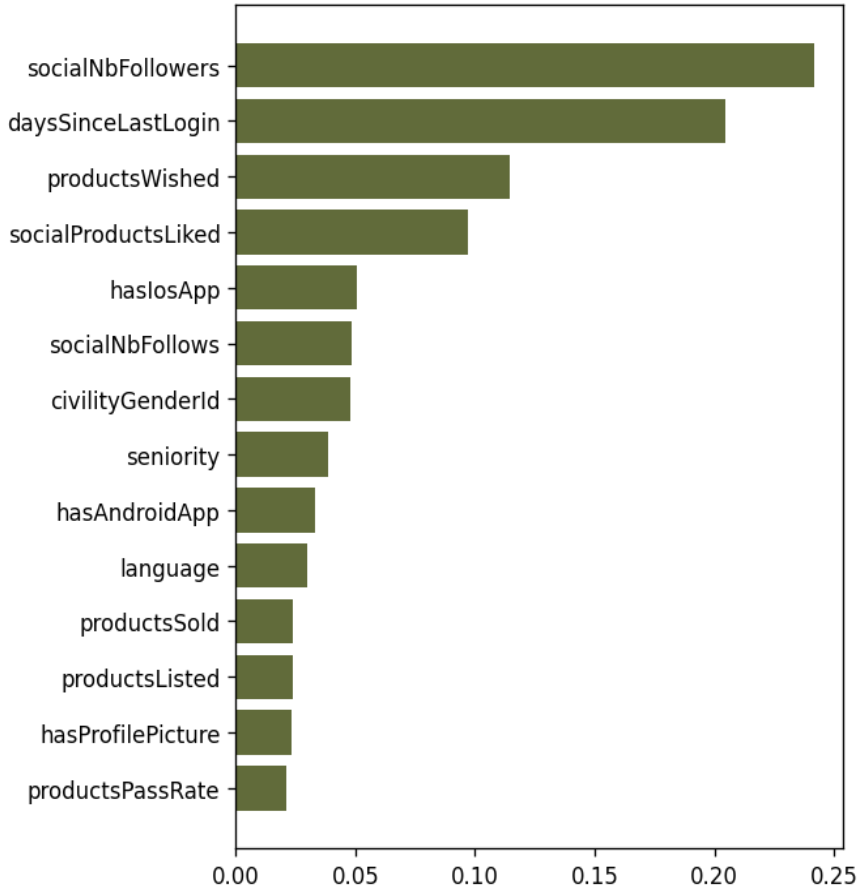


Gradient boosting decision tree
recall **0.76**

3. Interpretation

구매 확률과 관련도가 높은 특성 4가지

More relevant feature



Less relevant feature

특성중요도[1]를 보면, 'hasIosApp'와 'socialProductsLiked'의 차이가 두 배 수준의 급격한 차이가 남

따라서, 이번에는 구매 확률과 관련도가 높은 특성을 다음의 네 가지로 정함

- ✓ Number of users who follow this user's activity.
- ✓ Number of days since the last login.
- ✓ Number of products this user added to his/her wish list.
- ✓ Number of products this user liked.

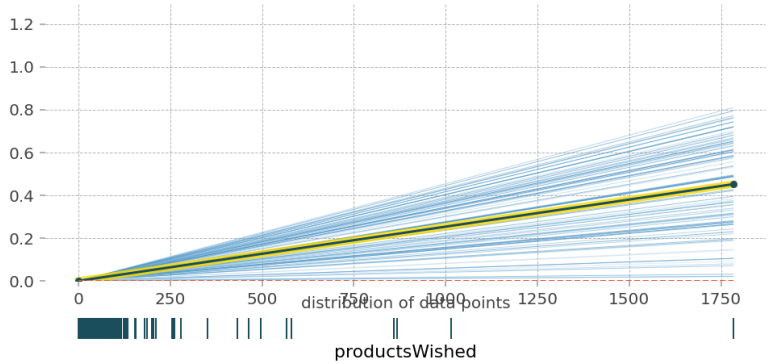
주의할 점: 양의 상관관계 또는 음의 상관관계 여부를 알 수 없고, 인과의 관계도 알 수 없음

1) Feature importance 는 특정한 특성이 없어지거나, 원래 데이터와 달라질 때, 모델의 성능이 얼마나 변하는지를 확인하여 특성이 얼마나 중요한 관련이 있는지 확인하는 기준

Partial dependence plot (PDP) and random 100 individual conditional expectation curves

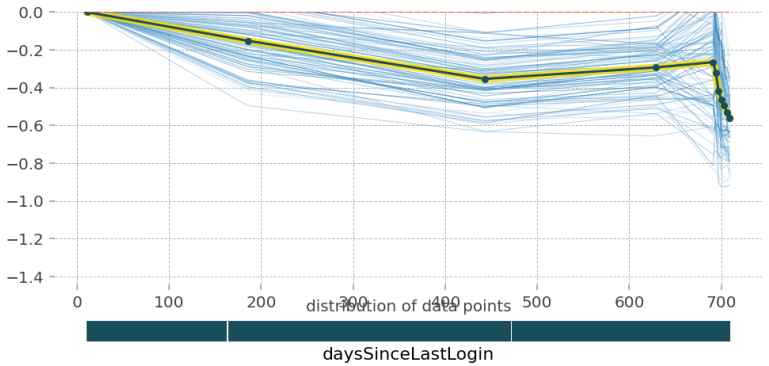
PDP for feature "productsWished"

Number of unique grid points: 2



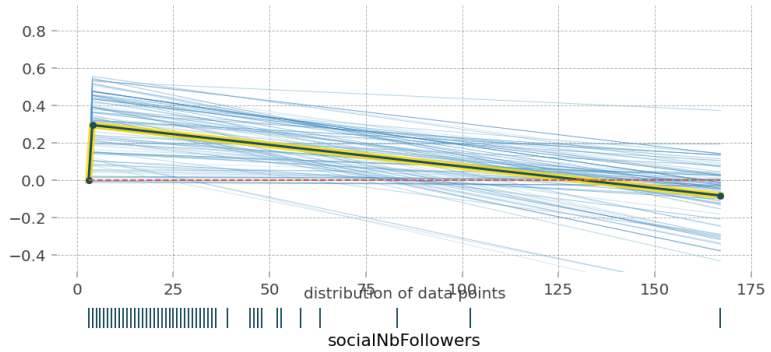
PDP for feature "daysSinceLastLogin"

Number of unique grid points: 11



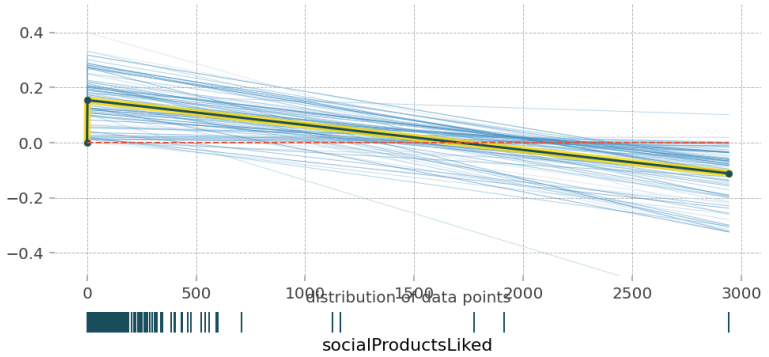
PDP for feature "socialNbFollowers"

Number of unique grid points: 3

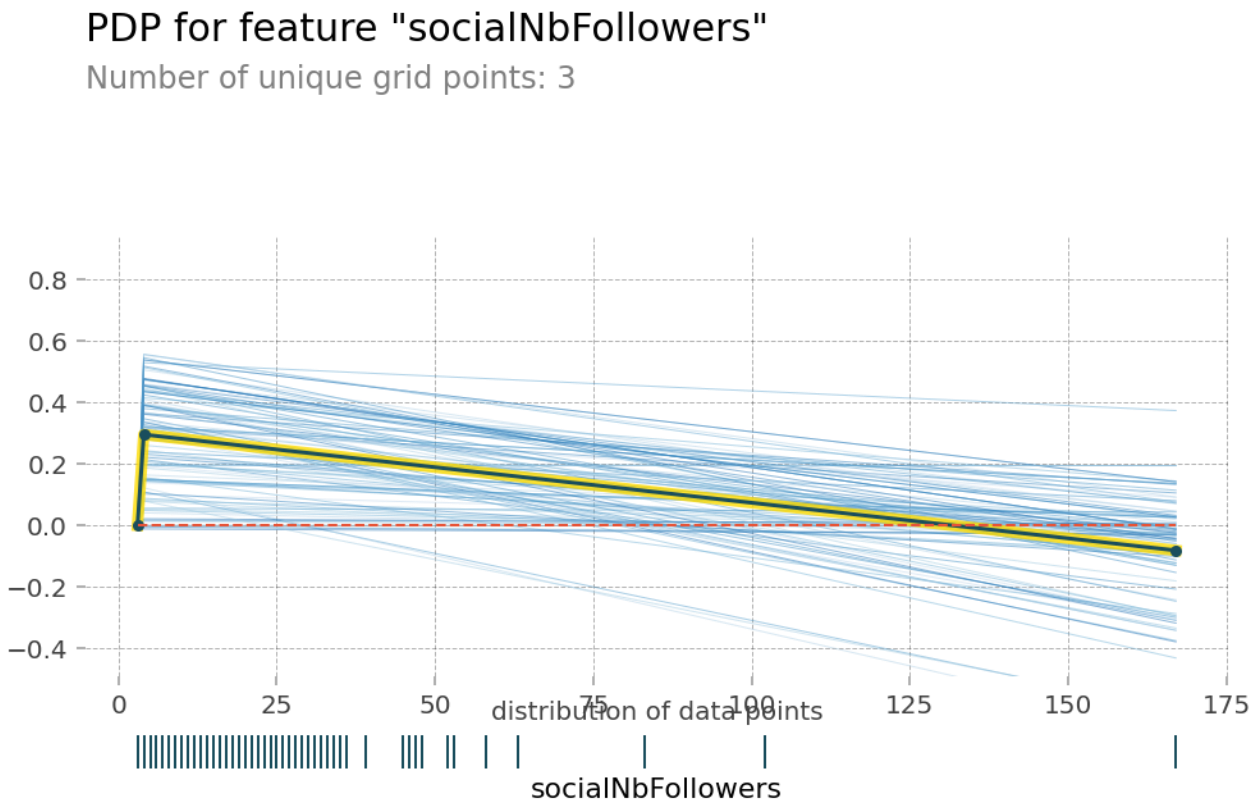


PDP for feature "socialProductsLiked"

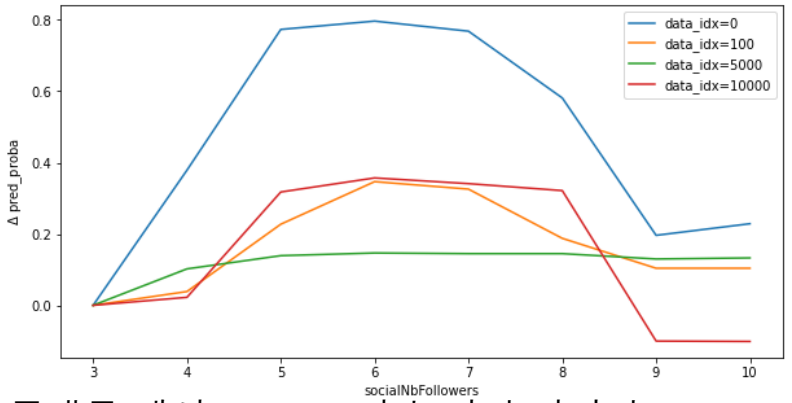
Number of unique grid points: 3



팔로워 수와 구매확률의 관계

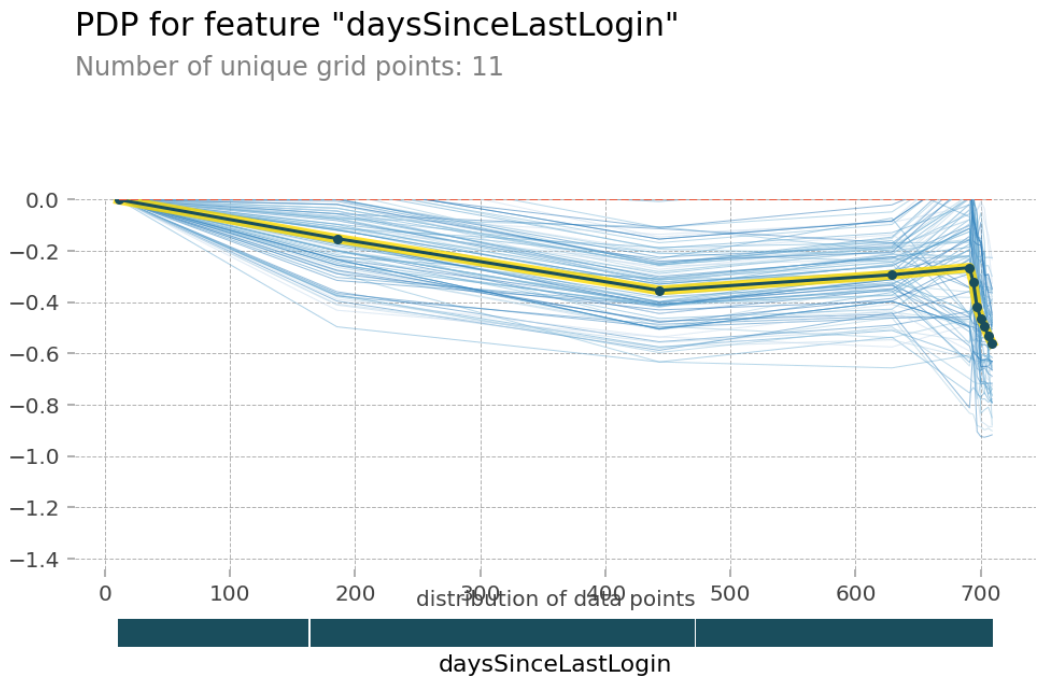


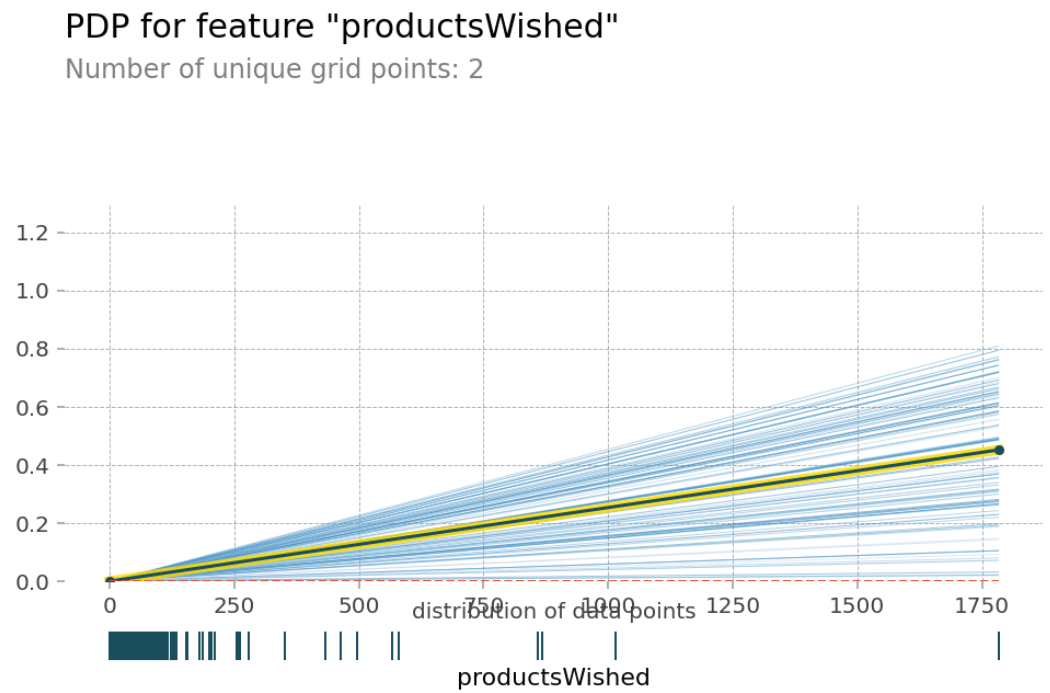
- 1~2명이 늘어날 때에만, 구매자일 확률이 늘어나고, 그 이후에는 감소함



- 구매를 해서 follower가 늘어난 것인지, follower가 늘어나서 구매가 늘어난 것인지 인과관계 불확실하여 추가 연구가 필요함
- 현재의 가설은 판매자가 구매자를 팔로우 하는 경우가 많다고 보고 있음. 대부분의 구매자가 구매량이 5개 미만이니까 판매자가 팔로우 하면 숫자가 근사함.

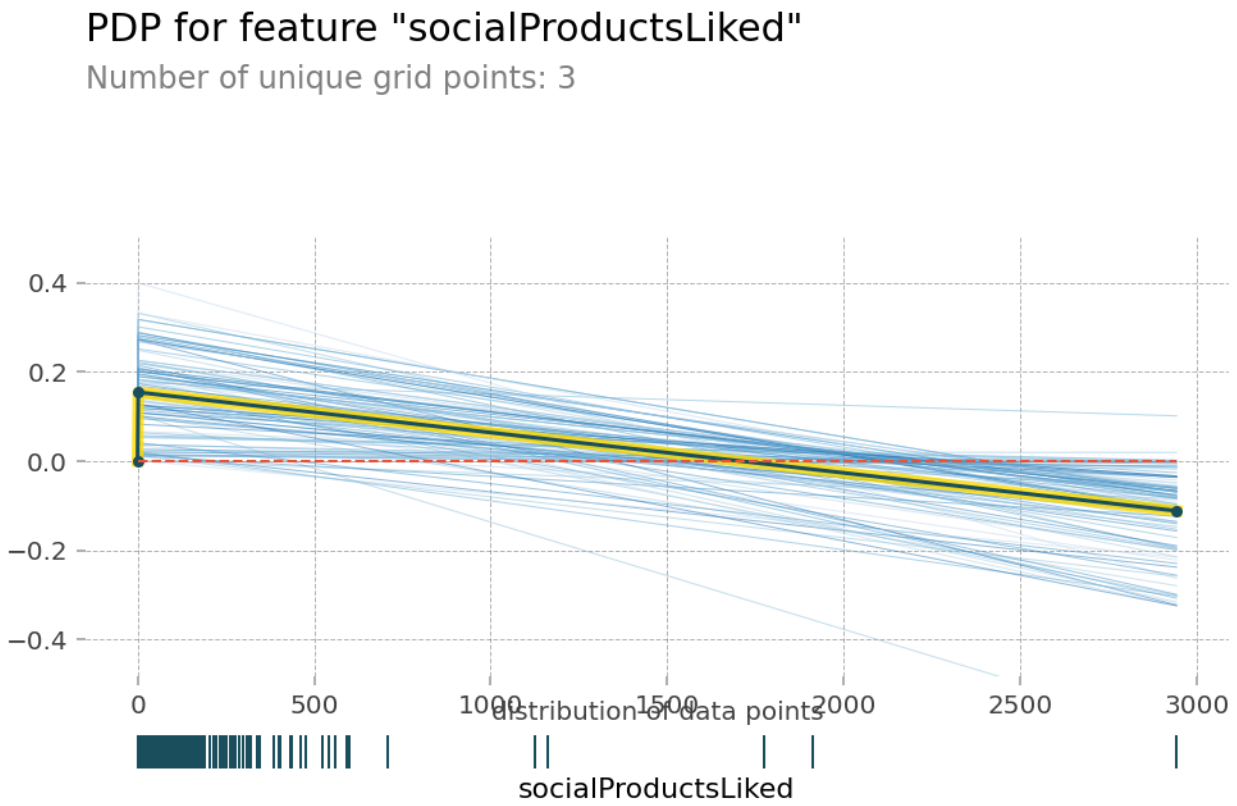
로그인 경과 일수, 관심상품 수와 구매확률 관계는 상식에 부합





최근에 로그인을 한 활동적인 사용자가 구매 확률이 높고, 관심 상품이 많은 사용자의 구매 확률이 높아진다는 것은 상식적

‘좋아요’를 한 제품 수와 구매확률 관계



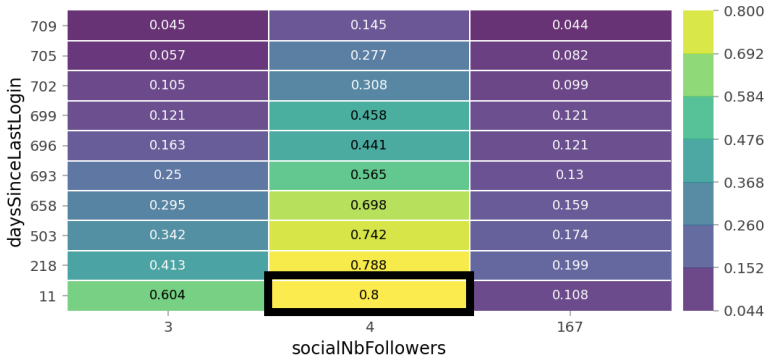
- 좋아요를 많이 누를수록 구매 확률이 낮아진다는 것은 상식에 반대 됨
- 좋아요가 무리하게 많은 경우는 불가능한 상황이라서, ICE가 잘못 예측한 결과로 보임
- 정확한 판단을 위해서는 추가 연구 필요

3-1. Feature interpretation

두 가지 특성의 PDP에서 구매 확률이 80%가 넘는 경우

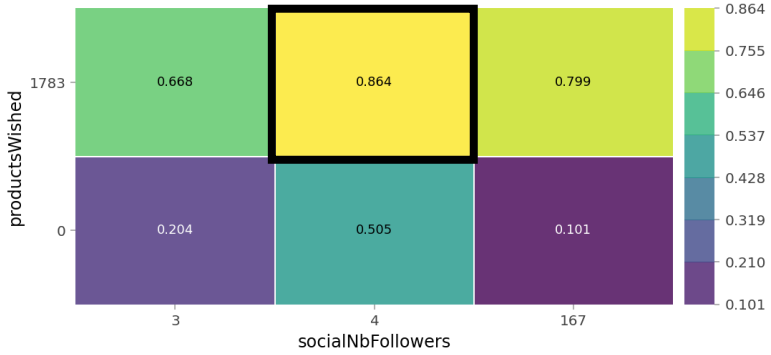
PDP interact for "socialNbFollowers" and "daysSinceLastLogin"

Number of unique grid points: (socialNbFollowers: 3, daysSinceLastLogin: 10)



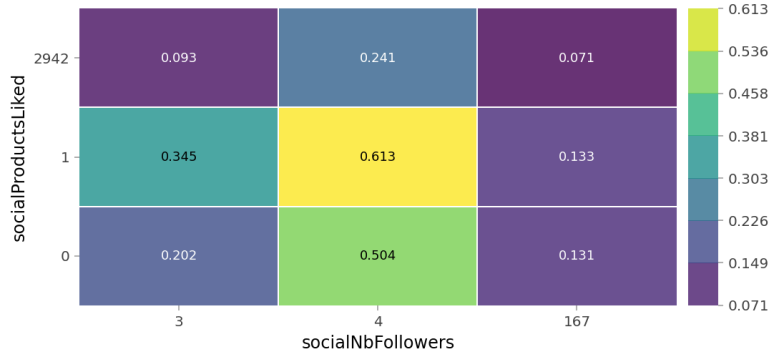
PDP interact for "socialNbFollowers" and "productsWished"

Number of unique grid points: (socialNbFollowers: 3, productsWished: 2)



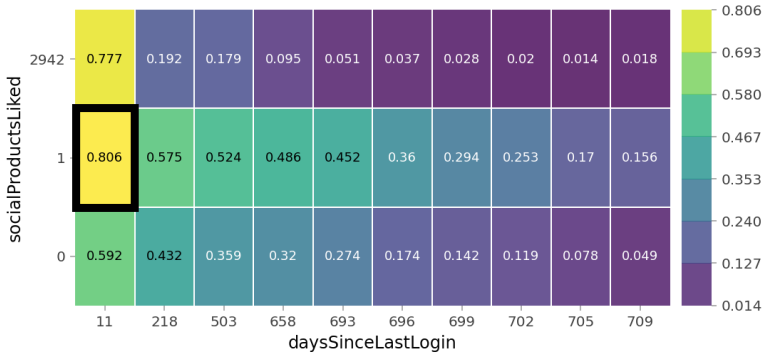
PDP interact for "socialNbFollowers" and "socialProductsLiked"

Number of unique grid points: (socialNbFollowers: 3, socialProductsLiked: 3)



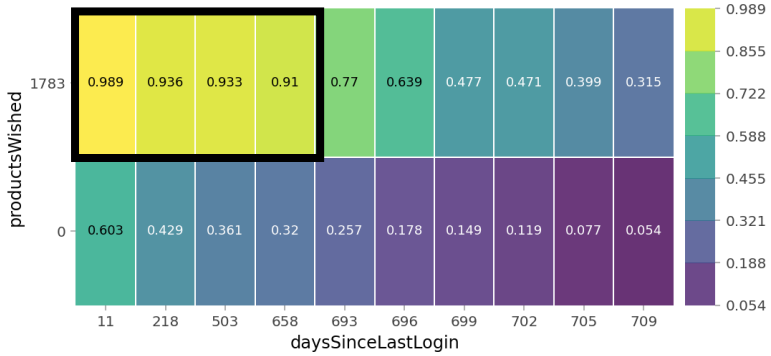
PDP interact for "daysSinceLastLogin" and "socialProductsLiked"

Number of unique grid points: (daysSinceLastLogin: 10, socialProductsLiked: 3)



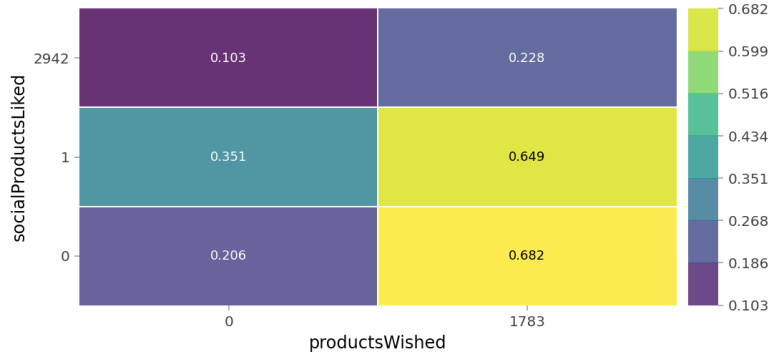
PDP interact for "daysSinceLastLogin" and "productsWished"

Number of unique grid points: (daysSinceLastLogin: 10, productsWished: 2)



PDP interact for "productsWished" and "socialProductsLiked"

Number of unique grid points: (productsWished: 2, socialProductsLiked: 3)

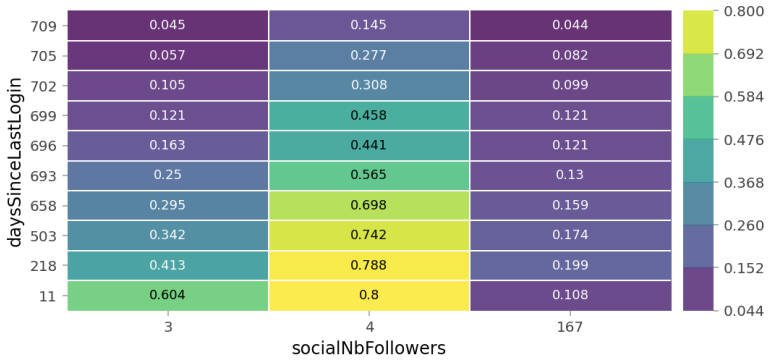


3-1. Feature interpretation

관심 상품으로 많이 넣을 수록 구매 확률이 증가하기 때문에, 고객의 참여도를 높여야 함

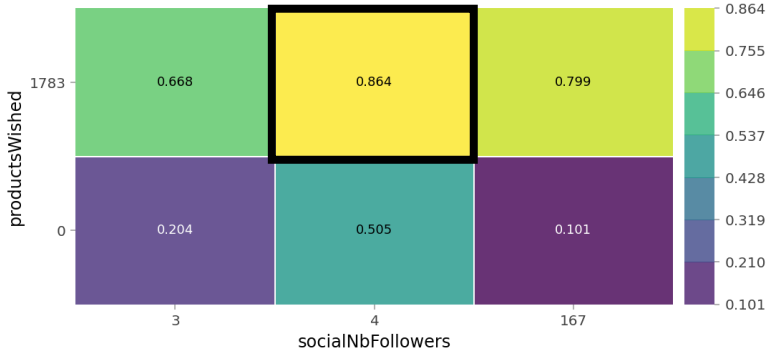
PDP interact for "socialNbFollowers" and "daysSinceLastLogin"

Number of unique grid points: (socialNbFollowers: 3, daysSinceLastLogin: 10)



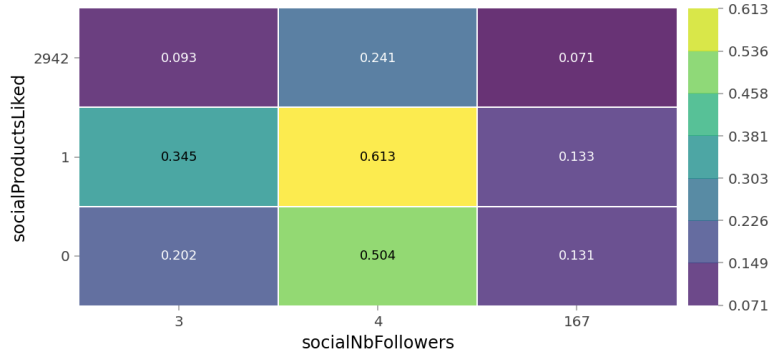
PDP interact for "socialNbFollowers" and "productsWished"

Number of unique grid points: (socialNbFollowers: 3, productsWished: 2)



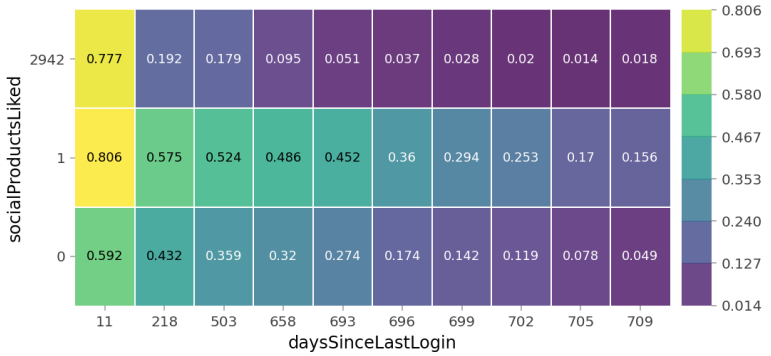
PDP interact for "socialNbFollowers" and "socialProductsLiked"

Number of unique grid points: (socialNbFollowers: 3, socialProductsLiked: 3)



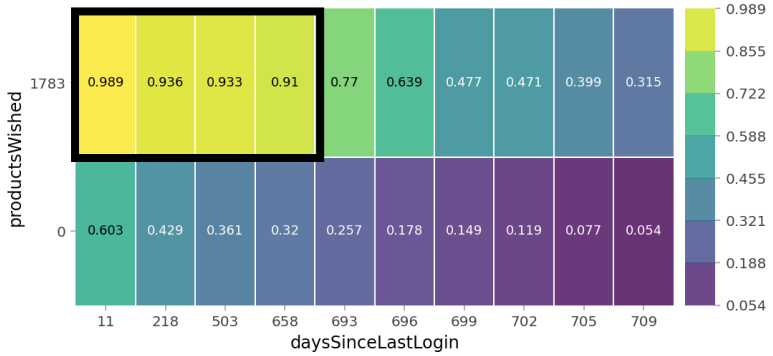
PDP interact for "daysSinceLastLogin" and "socialProductsLiked"

Number of unique grid points: (daysSinceLastLogin: 10, socialProductsLiked: 3)



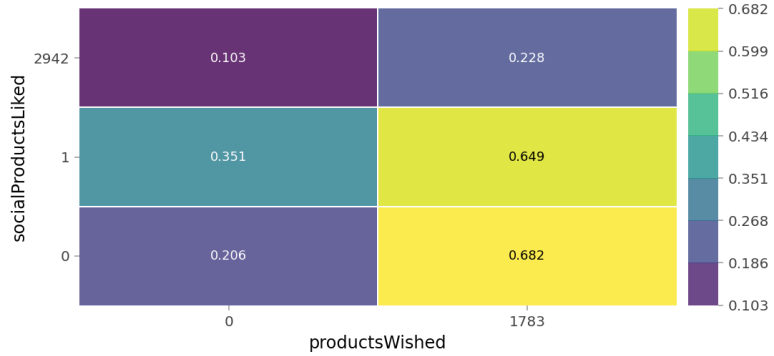
PDP interact for "daysSinceLastLogin" and "productsWished"

Number of unique grid points: (daysSinceLastLogin: 10, productsWished: 2)



PDP interact for "productsWished" and "socialProductsLiked"

Number of unique grid points: (productsWished: 2, socialProductsLiked: 3)

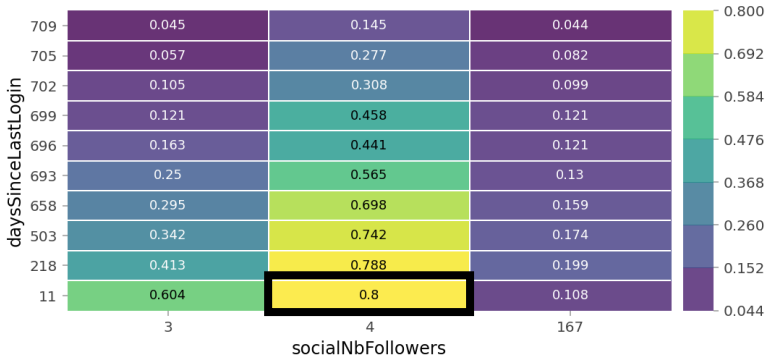


3-1. Feature interpretation

최근 로그인을 했을 수록 구매확률이 높아지기 때문에, 계정의 활동성을 높여야 함

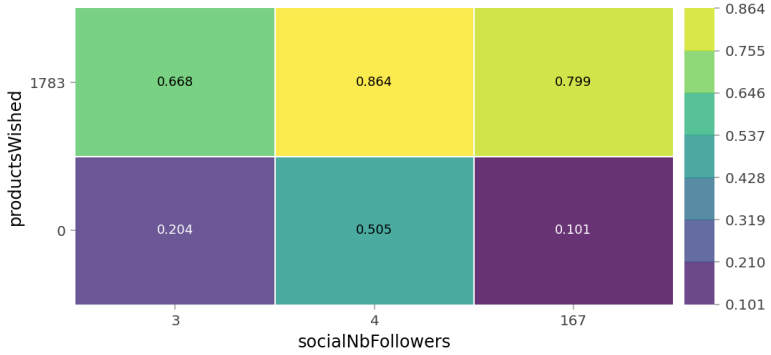
PDP interact for "socialNbFollowers" and "daysSinceLastLogin"

Number of unique grid points: (socialNbFollowers: 3, daysSinceLastLogin: 10)



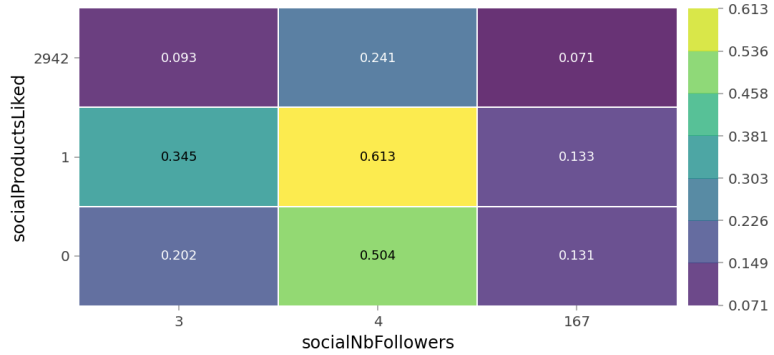
PDP interact for "socialNbFollowers" and "productsWished"

Number of unique grid points: (socialNbFollowers: 3, productsWished: 2)



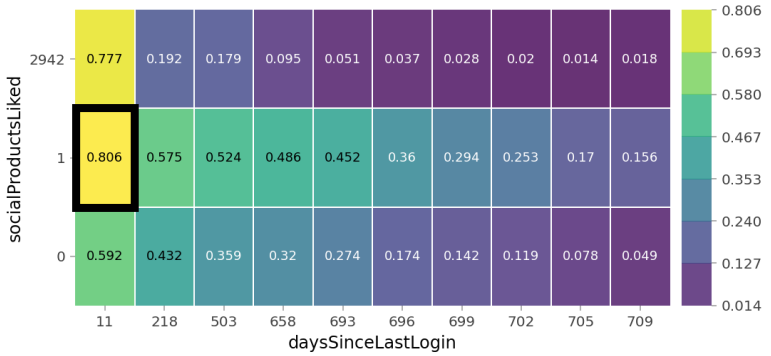
PDP interact for "socialNbFollowers" and "socialProductsLiked"

Number of unique grid points: (socialNbFollowers: 3, socialProductsLiked: 3)



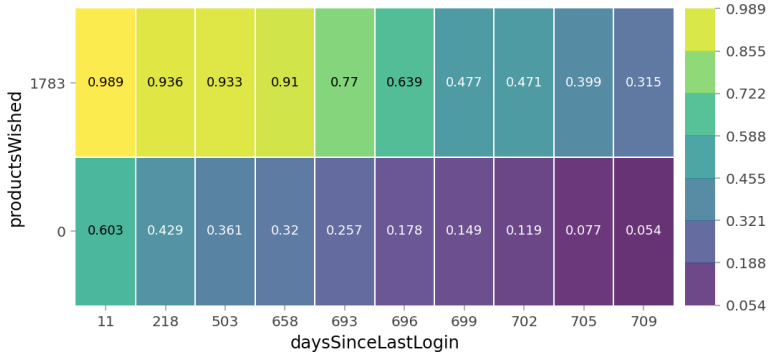
PDP interact for "daysSinceLastLogin" and "socialProductsLiked"

Number of unique grid points: (daysSinceLastLogin: 10, socialProductsLiked: 3)



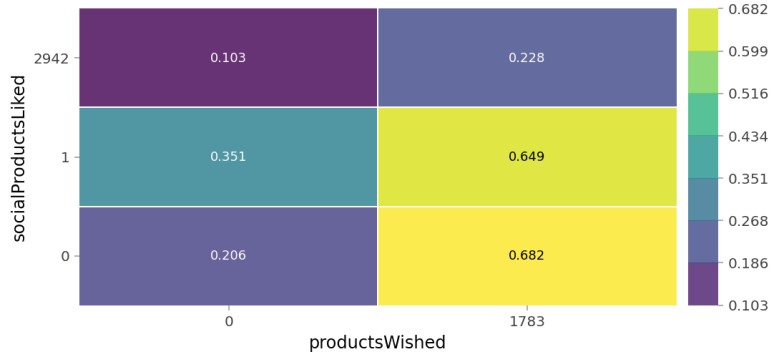
PDP interact for "daysSinceLastLogin" and "productsWished"

Number of unique grid points: (daysSinceLastLogin: 10, productsWished: 2)



PDP interact for "productsWished" and "socialProductsLiked"

Number of unique grid points: (productsWished: 2, socialProductsLiked: 3)



3가지 제안

1. Marketing funnel 에서 intent 단계로 진입시키기 위한 노력에 집중한다.

2. 흥미로운 컨텐츠를 제공하는 등의 방법으로 지속적으로 로그인하고 싶게 만든다.

3. Social product like가 많을 수록 구매확률이 계속 증가하는 것은 아니기 때문에, like가 구매 의도를 추측하는 지표가 될 수 있도록 재설계한다.