

2022
빅콘테스트

대출 목적별 대출신청여부 예측 및 서비스 메시지 제안

Team 수DA쟁이
이수빈, 최규진, 이상목

INDEX



01

주제 선정 배경

02

데이터 전처리

03

알고리즘 설명

04

예측 결과

05

서비스 메시지
제안

06

활용방안

01

주제 선정 배경

04 - 06

"대출 쓰기 겁나요"...빚 빨리 갚고 안 빌린다

금리 인상 기조가 지속되는 상황에서 신규 가계대출이 줄어드는 상황을 갈아타려는 대환 수요는 급증하고 있다. 앞으로도 계속 이자 부담이 점차 불어날 것으로 예상되면서 보다 낮은 금리나, 조금 높더라도 고정금리를 찾으려는 차주들이다.

일례로 국내 한 대형-
반기 3370건에서 하
반기에는 9월27일까

금리상한형 주담대를
올라도 연간 금리 상
한한다.

**"대출 환승
성공하실분
모십니다,"**

2022.09.30까지 핀다 앱에서 확인!

finda



대환대출에 대한 수요가
증가하는 가운데,
finda는 대환대출을
주력으로 삼고 있음

다양한 대출목적

생활비
주택구입
대환대출
...

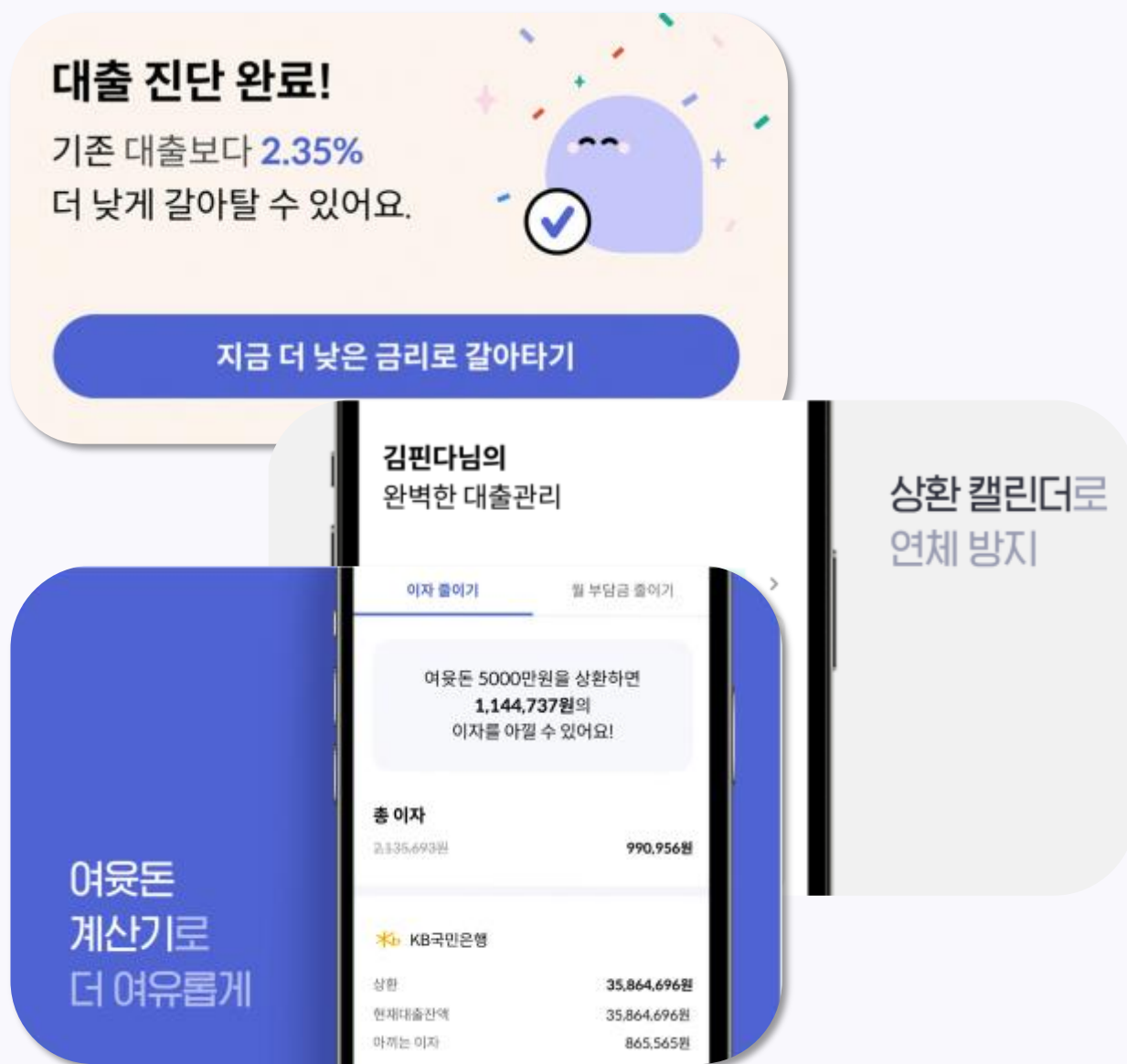


finda

**대출목적에 따른
대출신청여부 예측**

01

주제 선정 배경



여윳돈 계산기, DSR 계산기

...

finda가 제공하는
다양한 서비스 활성화

02

데이터 전처리

08 - 19

데이터 소개

User_Spec

user 신용정보

user 신용점수
고용형태
대출목적
...

Loan_result

사용자가 신청한
대출별 금융사별
승인결과

한도조회 일시
승인 금리
신청 여부
...

Log_data

finda 앱 로그정보

행동 명
행동일시
일 코드
...

불필요한 변수 제거 및 변수 생성

User_Spec

application_id, user_id : 테이블 병합을 위해 사용

birth_year, gender : 삭제

insert_time(생성시간), company_enter_month(입사연월) : 근속년수를 생성

personal_rehabilitation(개인회생자)

personal_rehabilitation이 결측 → 개인회생자와 무관하다고 판단

personal_rehabilitaion이라는 열을 생성해

0 : 개인회생자 납입중, 1 : 개인회생자 납입완료, 2 : 해당없음으로 분류

불필요한 변수 제거 및 변수 생성

User_Spec

existing_loan_cnt(기대출수), existing_loan_amt(기대출금액)

기대출수가 결측이면 기대출금액또한 결측이고, 기대출수는 0의 값을 가지지 않음

따라서 기대출수의 결측치를 0으로 대체함

불필요한 변수 제거 및 변수 생성

Log_Data

event(행동명) 한도조회 인트로 페이지 조회, 한도조회 시작하기 버튼 누르기, 본인인증완료, 한도조회 결과 확인
대출관리 서비스 이용, DSR 계산기 서비스 이용, KCB 신용정보 조회, 여윳돈 계산기 서비스 이용
회원가입, finda 앱 실행, finda 앱 로그인

user_id를 기준으로 one-hot-encoding

	EndLoanApply	UseLoanManage	UseDSRCalc	GetCreditInfo	user_id	timestamp	date_cd
0	0	0	0	0	576409	2022-03-25 11:12:09	2022-03-25
1	0	0	0	0	576409	2022-03-25 11:12:09	2022-03-25
2	1	0	0	0	72878	2022-03-25 11:14:44	2022-03-25
3	0	0	0	0	645317	2022-03-25 11:15:09	2022-03-25
4	0	1	0	0	645317	2022-03-25 11:15:11	2022-03-25
...
17843988	0	0	0	0	242442	2022-05-27 16:08:21	2022-05-27
17843989	1	0	0	0	242442	2022-05-27 16:08:21	2022-05-27
17843990	0	0	0	0	593062	2022-05-27 16:19:10	2022-05-27
17843991	0	0	0	0	593062	2022-05-27 16:19:15	2022-05-27
17843992	0	1	0	0	593062	2022-05-27 16:19:19	2022-05-27

Loan_Apply = 한도조회인트로페이지 조회

Loan_Apply + 한도조회 시작하기 버튼 누르기

Loan_Apply + 본인인증완료

Loan_Apply +한도조회 결과 확인

불필요한 변수 제거 및 변수 생성

예금은행 대출금리

한국은행 경제통계시스템(ECOS)에서 제공하는 **예금은행 대출금리**(신규취급액 기준, 잔액 기준)

가계대출	생활비, 대환대출, 사업, 기타
주택담보대출	주택구입
보증	전월세보증금, 자동차구입
신용	투자
ECOS 제공	

danger_score

‘loan_rate - 신규취급액금리’를 [0, 100]의 값으로 표준화하여 이용자의 위험점수를 생성

불필요한 변수 제거 및 변수 생성

DTI(Debt To Income)

DTI를 대출목적별로 구분하여 대출의 위험도 판단 및 서비스 메시지 제안에 활용

$$\text{주택대출의 DTI} = \frac{1}{\text{소득}} \left\{ \frac{\text{총대출금액}}{\text{대여년수}_{(20)}} + \text{기대출금액} \times \text{기존이자율} + \text{신규대출금액} \times \text{신규이자율} \right\}$$

$$\text{일반대출의 DTI} = \frac{1}{\text{소득}} \left\{ \frac{\text{총대출금액}}{\text{대여년수}_{(5)}} + \text{기대출금액} \times \text{기존이자율} + \text{신규대출금액} \times \text{신규이자율} \right\}$$

$$\text{대환대출의 DTI} = \frac{1}{\text{소득}} \left\{ \frac{\text{기대출금액}}{\text{대여년수}_{(5)}} + (\text{기대출금액} - \text{신규대출금액}) \times \text{기존이자율} \right. \\ \left. + \text{신규대출금액} \times \text{신규이자율} \right\}$$

$$\text{전월세보증금대출의 DTI} = \frac{1}{\text{소득}} \left\{ \frac{\text{기대출금액}}{\text{대여년수}_{(20)}} + \text{기대출금액} \times \text{기존이자율} + \text{신규대출금액} \times \text{신규이자율} \right\}$$

불필요한 변수 제거 및 변수 생성

Diff(Desired_amount - Loan_limit)

‘대출희망금액 - 대출한도’를 계산하여, 고객이 생각하는 ‘금융사별 대출조건’을 고려

$$\text{Diff} = \max(\text{대출희망금액}(Desired_{amount}) - \text{대출한도}(Loan_{limit}))$$


대출희망금액
1억원

A금융사 대출한도 : 1억원, 금리 5%

B금융사 대출한도 : 8000만원, 금리 6%

C금융사 대출한도 : 9000만원, 금리 5%



A금융사 선택

결측치 제거

1. loan_limit와 loan_rate가 없는 행은 금융사에서 값을 보내주지 않은 경우로, 제거
2. application_id가 loan에는 있지만, user에는 없는 경우는 기간 차이로 인해 누락된 데이터로 해당 행은 제거
3. yearly_income의 결측치는 중앙값(2000만)으로 대체
4. yearly_income, existing_loan_amount에 log-scale

```
application_id    0
user_id          0
insert_time      0
credit_score      81767
yearly_income     1
income_type      0
company_enter_month 92311
employment_type  0
houseown_type    0
desired_amount   0
purpose          0
existing_loan_cnt 0
existing_loan_amt  78752
personal_rehabilitation 0
dtype: int64
```

결측치 제거 - MissForest

MissForest는 2011년에 도입된 RandomForest를 기반으로 한 결측치 제거 방법

Step00. 데이터 분류

Y_{mis} : 결측치

Y_{obs} : 결측치가 아닌 열의 데이터

X_{mis} : 결측치가 있는 행의 다른 특성들

X_{obs} : Y_{obs} 의 다른 특성들

X_{obs}	Y_{obs}	X_{obs}
X_{mis}	Y_{mis}	X_{mis}
X_{obs}	Y_{obs}	X_{obs}

결측치 제거 - MissForest

Step01. 평균/다른 방법으로 imputation

Step02. 얼마나 결측치의 수를 세어서 정렬

Step03. 결측치의 수가 적은 열부터 Y_{mis} 를

X_{obs} , Y_{obs} 와 RF를 이용해 보간

Step04. Step01 ~ Step03을 반복

Algorithm 1 Impute missing values with RF.

Require: \mathbf{X} an $n \times p$ matrix, stopping criterion γ

1. Make initial guess for missing values;
2. $\mathbf{k} \leftarrow$ vector of sorted indices of columns in \mathbf{X} w.r.t. increasing amount of missing values;
3. **while** not γ **do**
4. $\mathbf{X}_{old}^{imp} \leftarrow$ store previously imputed matrix;
5. **for** s in \mathbf{k} **do**
6. Fit a random forest: $\mathbf{y}_{obs}^{(s)} \sim \mathbf{x}_{obs}^{(s)}$;
7. Predict $\mathbf{y}_{mis}^{(s)}$ using $\mathbf{x}_{mis}^{(s)}$;
8. $\mathbf{X}_{new}^{imp} \leftarrow$ update imputed matrix, using predicted $\mathbf{y}_{mis}^{(s)}$;
9. **end for**
10. update γ .
11. **end while**
12. **return** the imputed matrix \mathbf{X}^{imp}

credit_score, existing_loan_amt, working(근속년수)에 대해 MissForest 사용

데이터 병합

Loan + User (LoanUser)

`Application_id`를 기준으로 데이터 병합 진행

LoanUser + Log

Pandas의 `merge_asof`를 이용하여 데이터 병합을 진행 :

기준이 되는 열 : `user_id`

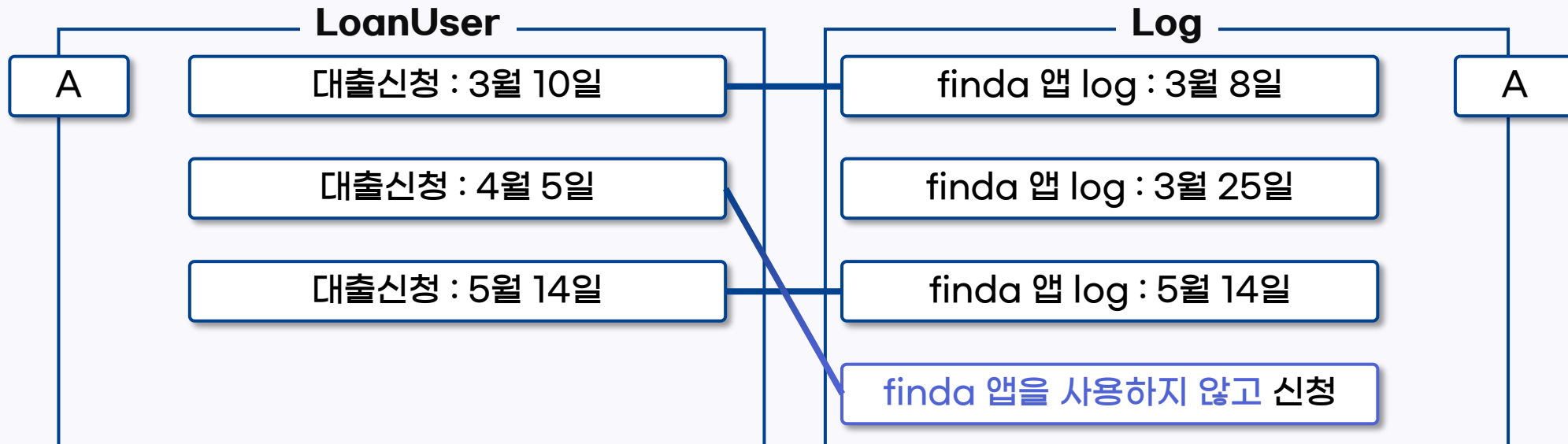
`left_on` : LoanUser 데이터의 `loan_apply_insert_time`(한도조회 일시)

`right_on` : Log 데이터의 `date_cd`(일 코드)를 사용함

데이터 병합

LoanUser + Log

한 사람이 여러 번 대출 신청을 시도할 수 있음. 따라서 user_id가 일치하는 행들에 대해서, `loan_apply_insert_time`과 `date_cd`의 차이가 3일 이내인 데이터들에 대해 병합을 진행
결측치의 경우, `finda` 앱을 사용하지 않았다고 판단해 log데이터의 항목을 0으로 대체

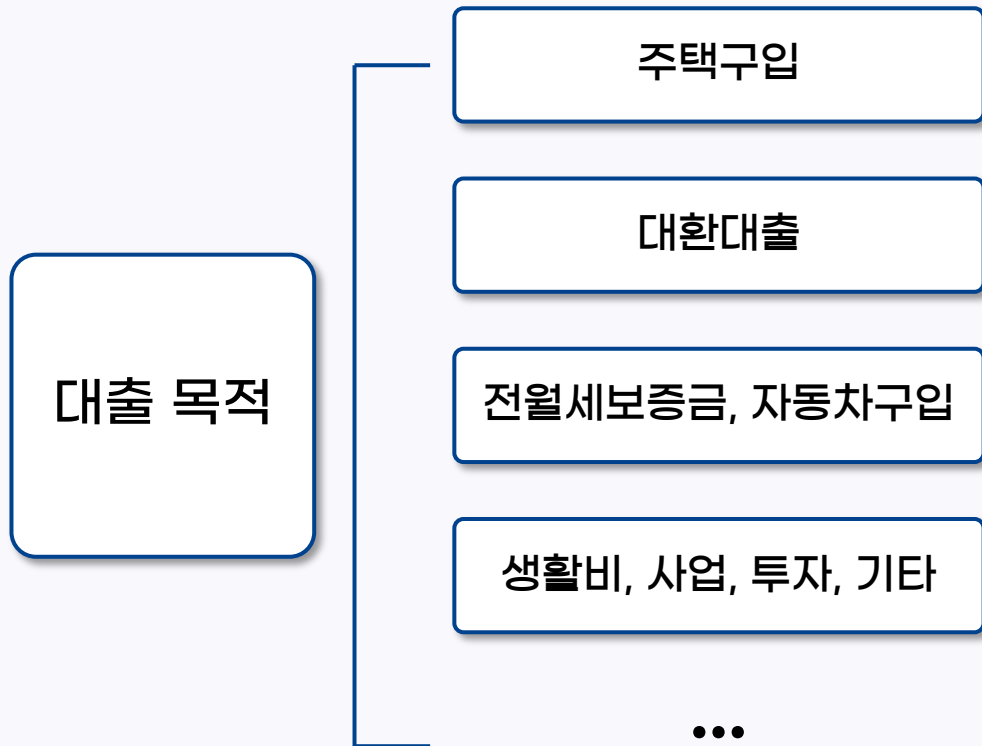


03

알고리즘 설명

21 - 24

대출 목적에 따른 대출 승인여부 예측



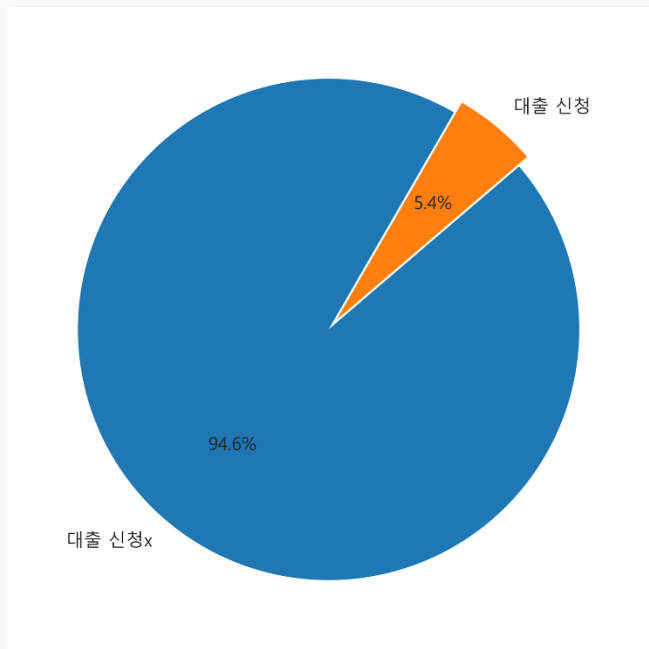
대출 목적별로 대출승인여부를
예측하는 모델 생성

대출 목적에 따른 대출 승인여부 예측

train-test split

각 목적에 따른 데이터 분리 후, train : test = 3 : 1로 데이터를 구분

불균형 데이터 문제 해결



전체 데이터를 살펴보았을 때,
대출신청여부에 대해서 불균형임을 확인할 수 있음!

대출 목적에 따른 대출 승인여부 예측

UnderSampling

대출 목적이 **생활비**, **기타**인 경우 전체 표본의 50%가량을 차지

→ RandomUnderSampler를 통해 train데이터의 is_applied의 비율을

신청 : 미신청 = 1:1로 조정

OverSampling

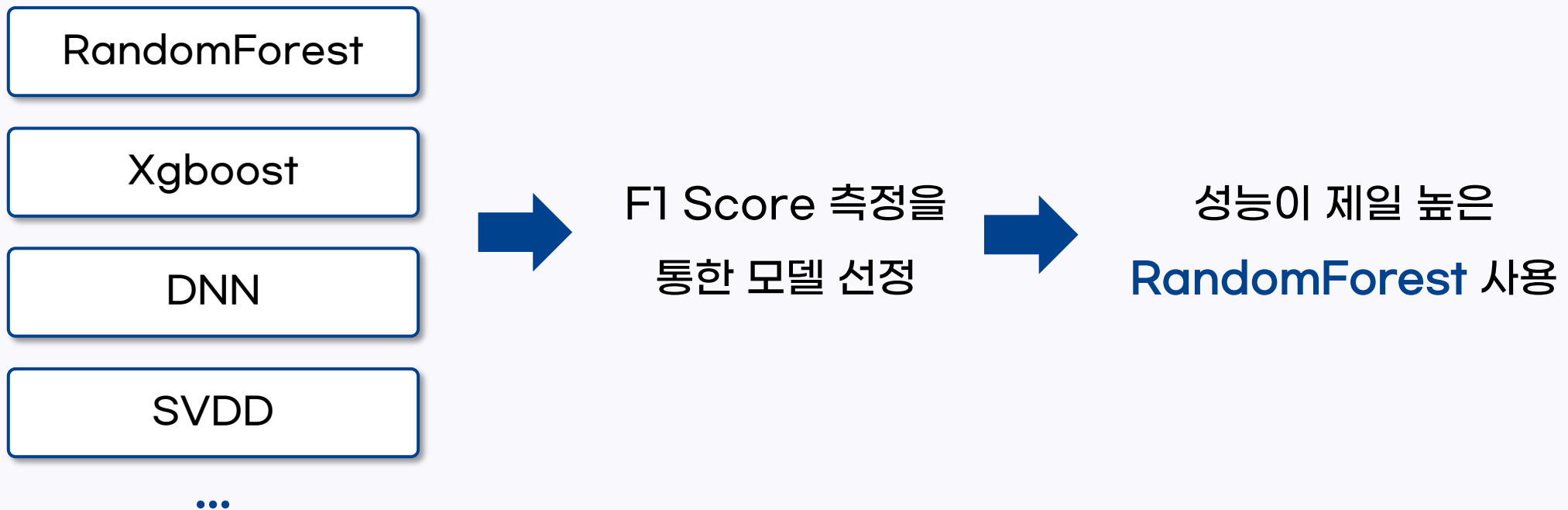
이외의 대출목적의 경우 SMOTE(Synthetic Minority Over-sampling Technique)를 통해 train

데이터의 is_applied의 비율을 **신청 : 미신청 = 1:1로 조정**

대출 목적에 따른 대출 승인여부 예측

RandomForest를 통한 대출 승인여부 예측

각 대출목적에 대해서 여러가지 머신러닝 모델을 통해 대출 승인여부를 예측 :



04

예측 결과

26 - 27

04

예측 결과

예측 결과

대출 목적	F1-Score
생활비 및 기타	0.3084
대환대출	0.3014
사업자금	0.3706
자동차 구입	0.3059
전월세보증금	0.2944
투자	0.2743
주택구입	0.2324

평가지표인 f1-score는 전반적으로 낮은 모습

		0 predict 1	
Real	0	91542	1272
	1	511	270

<사업자금>의 confusion matrix

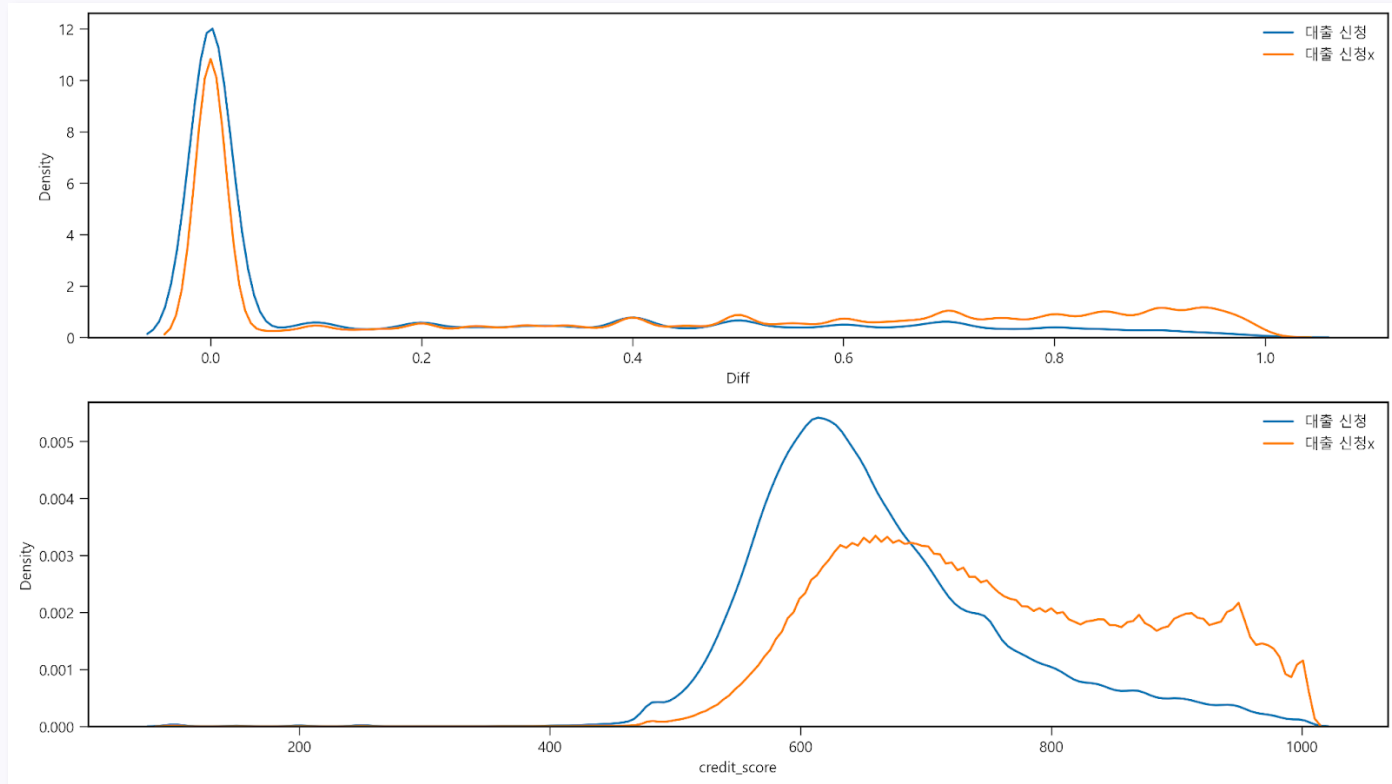
04

예측 결과

예측 결과

하지만 불균형 데이터 문제가 발생했고,

이용자가 금융사의 대출상품중 하나를 선택하므로, 두 집단간 차이가 크지 않다는 점을 고려



05

서비스 메시지 제안

28 - 38

① 연관 분석을 통한 대출 신청 유인 제공

연관분석(Association Analysis, A priori Algorithm)

상품이나 서비스를 구매하는 등 **일련의 거래나 사건 내에 존재하는 항목 간의 일정한 연관 규칙을 발견하는 분석**

연관규칙은 'if A, then B'와 같은 패턴으로 이루어져 있다.

① 연관 분석을 통한 대출 신청 유인 제공

지지도(Support)

$A \rightarrow B$ 의 지지도 $:=$ 전체 거래 중에서 A 와 B 가 동시에 판매되는 거래의 비율 $P(A \cap B)$,
1에 가까울수록 연관성이 높다

신뢰도(Confidence)

$A \rightarrow B$ 의 신뢰도 $:=$ A 의 거래 중에서 B 가 포함된 거래의 비율 $P(B | A)$,
1에 가까울수록 연관이 높다

향상도(Lift)

$A \rightarrow B$ 의 향상도 $:=$ A 를 구매할 때 B 를 구매할 구매확률의 증가 비율 $P(A \cap B)/P(A)P(B)$,
향상도가 1이면 두 품목은 독립이고, 1보다 크면 두 품목은 양의 상관관계를 가진다

① 연관 분석을 통한 대출 신청 유인 제공

Log_Data의 Event에 대해서 요인 분석 진행

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(is_applied)	(EndLoanApply)	0.500000	0.927321	0.475532	0.951064	1.025604	0.011872	1.485191
1	(EndLoanApply)	(is_applied)	0.927321	0.500000	0.475532	0.512802	1.025604	0.011872	1.026277
2	(is_applied)	(GetCreditInfo)	0.500000	0.540328	0.273983	0.547966	1.014136	0.003819	1.016897
3	(GetCreditInfo)	(is_applied)	0.540328	0.500000	0.273983	0.507068	1.014136	0.003819	1.014338
4	(is_applied)	(UseLoanManage)	0.500000	0.366793	0.195717	0.391435	1.067182	0.012321	1.040492
...
190	(is_applied, UsePrepayCalc, EndLoanApply, UseL...	(GetCreditInfo)	0.001577	0.540328	0.001577	1.000000	1.850729	0.000725	inf
191	(is_applied, UsePrepayCalc, GetCreditInfo, Use...	(EndLoanApply)	0.001578	0.927321	0.001577	0.999397	1.077725	0.000114	120.574364
192	(is_applied, UsePrepayCalc, GetCreditInfo, End...	(UseLoanManage)	0.001580	0.366793	0.001577	0.998194	2.721410	0.000997	350.585697
193	(is_applied, GetCreditInfo, EndLoanApply, UseL...	(UsePrepayCalc)	0.178633	0.003988	0.001577	0.008827	2.213599	0.000865	1.004883
194	(GetCreditInfo, UsePrepayCalc, EndLoanApply, U...	(is_applied)	0.003838	0.500000	0.001577	0.410905	0.821809	-0.000342	0.848759

① 연관 분석을 통한 대출 신청 유인 제공

‘신뢰도(Confidence)가 높다 ↔ 연관이 높다’ 이므로 is_applied와 신뢰도가 높은
연관규칙을 만들



① 연관 분석을 통한 대출 신청 유인 제공

즉, 신용정보 조회를 한 이용자가 대출 신청을
할 가능성이 높으므로 해당 이용자에게
‘신용정보 조회를 하시겠습니까?’와 같은
서비스 메시지를 출력하고 이용자의
대출 신청 유인을 제공

고객님과 비슷한 사용자는.. **finda**

기존 대출보다 **3.14%**
더 낮은 대출을 받았어요

더 좋은 대출 찾아보기

<예상 메시지>

대출 재신청을 위한 메시지

finda는 이용자가 정보를 입력하면, 해당 정보에 기초하여 금융사들이 대출조건을 제안하는 방식
이후 이용자는 다양한 대출조건 중 **가장 좋은 대출조건을 선택**

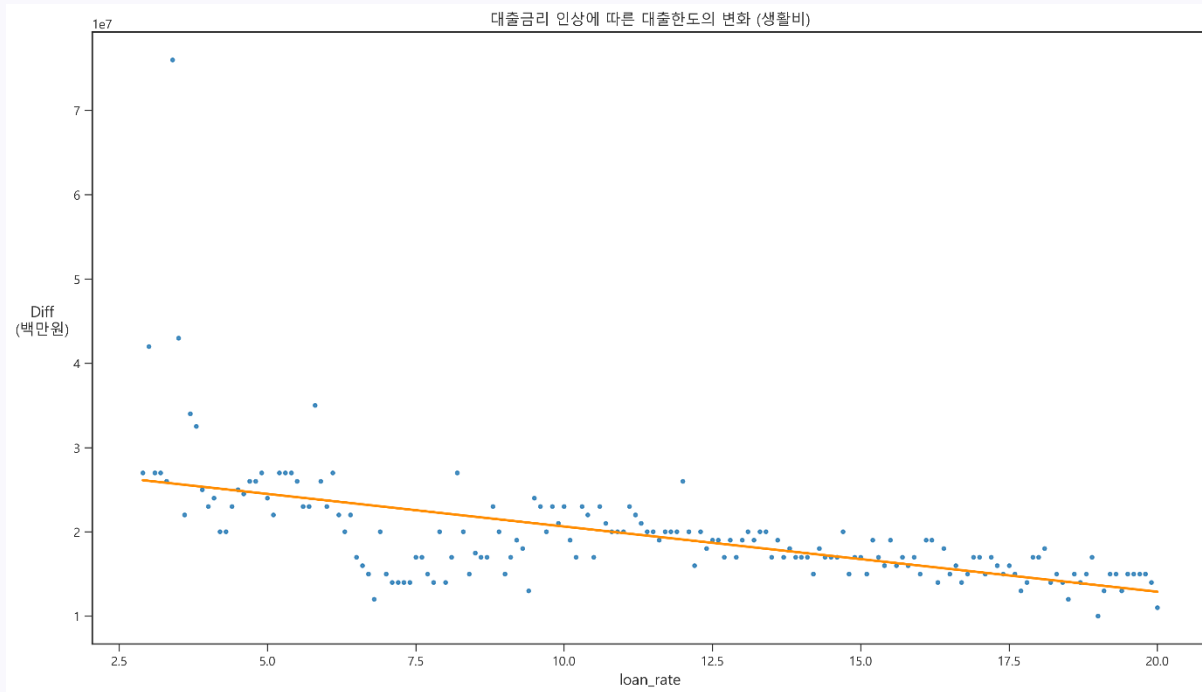
따라서 이용자가 대출을 거절했다면

1) 대출한도가 희망대출금액보다 낮아서 **2)** 대출금리가 원하는 수준보다 높아서 로 생각가능

05

서비스 메시지 제안

대출 재신청을 위한 메시지

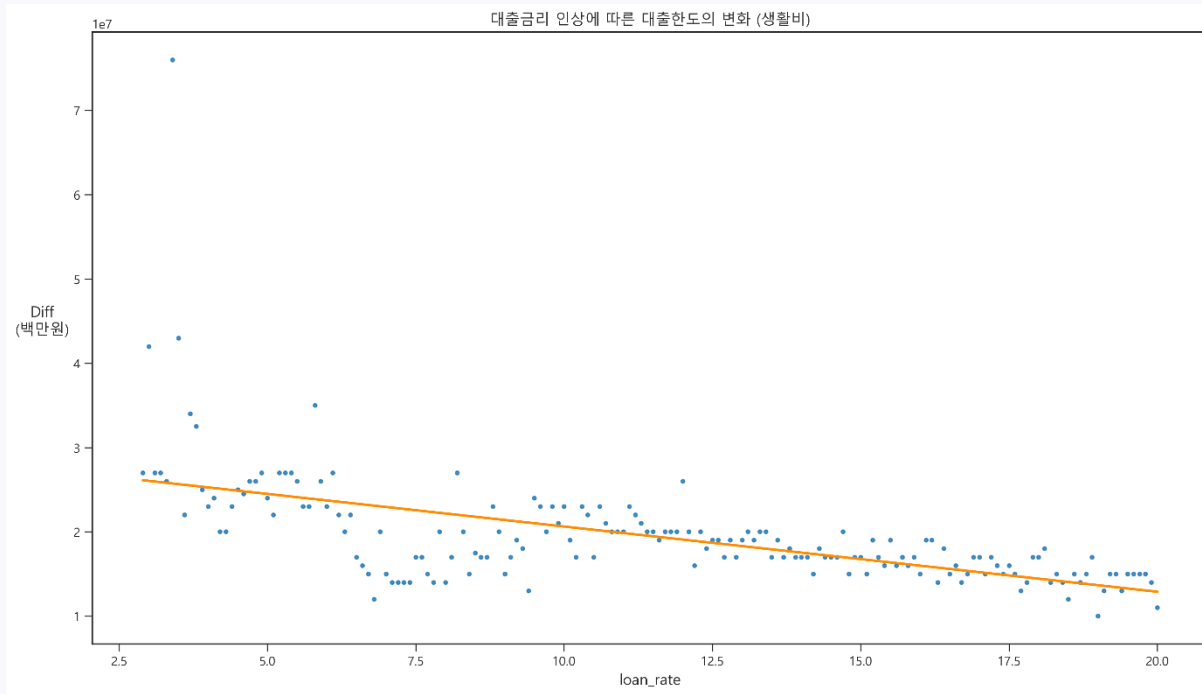


대출금리로 groupby를 하고, Diff의 중앙값을 모아놓은 데이터를 통해
‘대출금리 1%p 인상이 대출한도를 얼마나 높이는지’를 확인

05

서비스 메시지 제안

대출 재신청을 위한 메시지



선형회귀 결과, ‘생활비’을 목적으로 하는 대출미신청 고객은 ‘대출금리 1%p 상승이 대출한도를 77만원 높였다’라는 결과를 얻음

대출 재신청을 위한 메시지

따라서 대출한도에 불만족하는 고객의 경우

대출금리를 높여서 대출한도를 높이는 전략을 통해서 대출한도를 높여 대출을 재신청할 수 있음

Example) 대출희망금액이 500만원이지만, 대출한도가 300만원인 고객의 경우 대출금리 3% 인상을 통해서 희망하는 한도에 맞는 대출을 할 수 있을 것으로 기대

따라서 ‘고객님과 비슷한 사용자들은 원하는 한도로 대출받기 위해 대출금리를 3%p 올렸어요’와 같은 메시지를 출력해 사용자의 의사결정에 도움을 줌

대출 재신청을 위한 메시지



대출희망금액
500만원

A금융사 대출한도 : 300만원, 금리 4%

B금융사 대출한도 : 200만원, 금리 6%

C금융사 대출한도 : 250만원, 금리 5%



대출신청하지 않음

고객님과 비슷한 사용자는.. **finda**

대출 금리를 **3%p** 정도 올려
원하는 한도의 대출을 받았어요

새로운 조건으로 대출 찾아보기

<예상 메시지>

06

활용방안

40 - 42

Application 1. 이용자의 대출 신청과 연관이 큰 서비스 파악

연관분석을 통해서 ‘신용정보 조회’를 한 사용자가 대출 신청을 할 가능성이 높다는 사실을 파악
즉, 다른 금융기관이 아닌 **finda**에서 신용정보를 조회할 경우, 대출 신청 가능성이 높음

반면 **finda**의 다른 서비스인 DSR 계산기, 여윳돈 서비스...는 지지도가 낮아 연관분석에서 제외
즉, DSR 계산기, 여윳돈 서비스를 단순 정보조회가 아닌 대출 신청과 연관짓는 서비스 고안

Ex) DSR 계산기 사용시 수수료 절감..

또한, 신용정보 조회를 한 사용자가 대출 신청 가능성이 높기 때문에, **신용정보 조회시
대출 신청에 이득을 주는 방안을 도입**하여 대출 신청 가능성을 높일 것으로 기대

Application 2. 사용자의 금융사별 대출 상품 선택을 바탕으로 한 상품 설계의 발전

사용자는 개인정보, 대출 희망 금액...을 입력한 후, 금융사들이 제안한 대출 조건 중에서 선택

새롭게 도입한 변수인 **Diff**는 **대출 희망 금액과 대출 한도 금액의 차이**로, 사용자가 선택한 대출 조건을 바탕으로 금융사는 **보다 우량한 사용자에게 혹은 보다 많은 사용자에게 대출**을 가능하게 할 상품을 설계할 수 있음

Application 3. 대출 금리와 대출한도를 바탕으로 한 고객의 선택 폭 확대

일반적으로 ‘**대출한도를 늘린다 = 위험이 커진다**’가 성립하므로, 대출한도의 증가는 대출금리 인상을 가져옴

대출 신청을 하지 않은 고객 중, 대출한도에 만족하지 않는 고객이 존재할 것이므로, 해당 고객에게 ‘**대출한도를 늘릴 때 금리가 얼마나 인상되는지**’를 보여줘 합리적인 의사결정에 도움을 주고자 함

2022
빅콘테스트

Thank You

Team 수DA쟁이
이수빈, 최규진, 이상목