

# 핀다 앱 홈화면 진입 고객 기반 대출 신청 예측 및 군집 분석 : 맞춤형 서비스 메시지 제안

세상에 없던  
대출 비교 플랫폼  
**finda**



피난처머니(2조) : 권민재, 백은서, 전해빈



# CONTENTS

## 1. 개요

- 프로젝트 배경
- 목표 설정 및 분석 방향

## 2. 문제 수행 내용

- Dataset 소개
- EDA 과정
- 예측 모델 수립
- 고객 군집 분석 및 서비스 메시지

## 3. 주요 결과 및 시사점

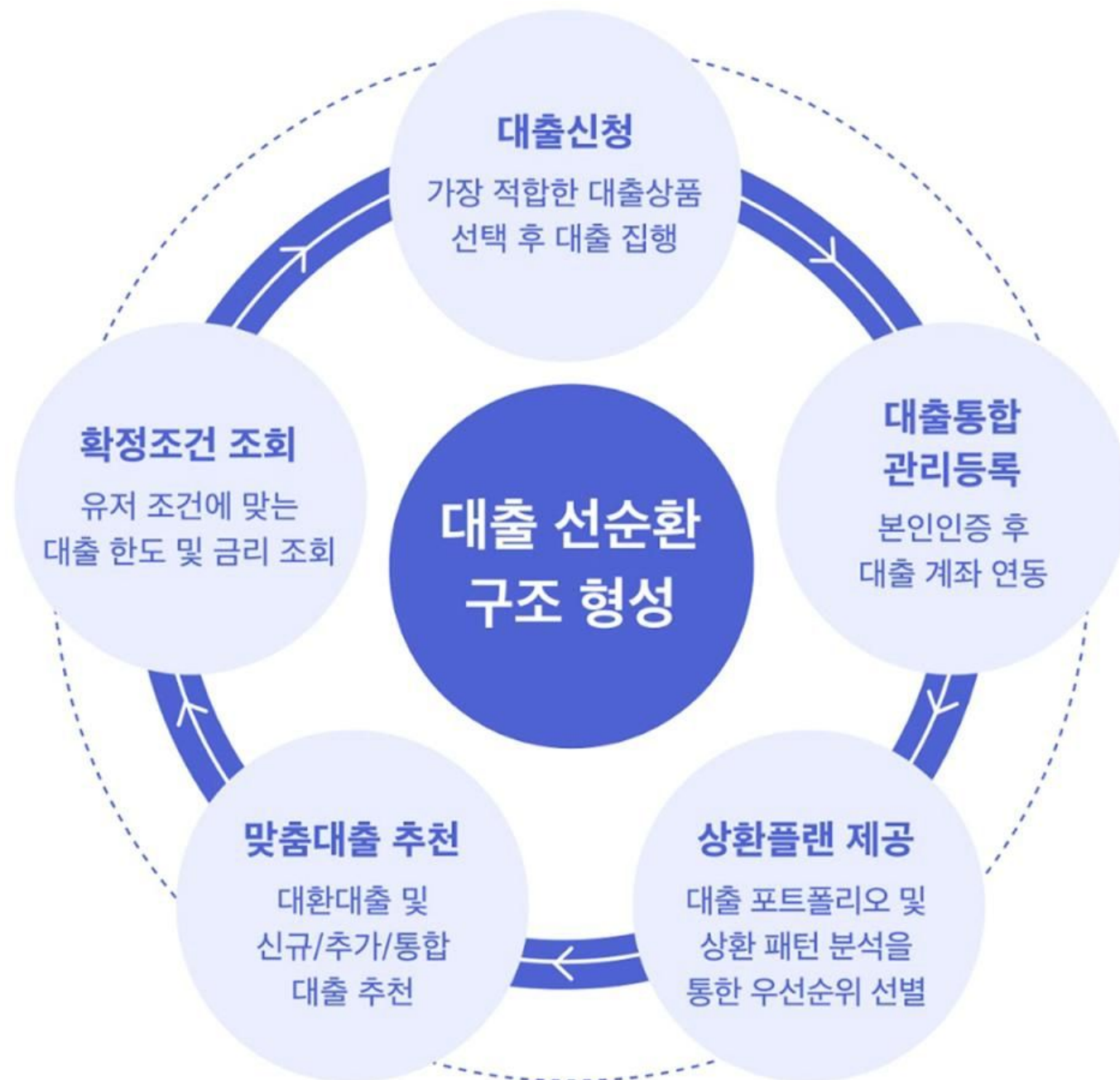
- 결과 요약
- 기대 효과 및 한계점

## 4. 개발후기 및 느낀점



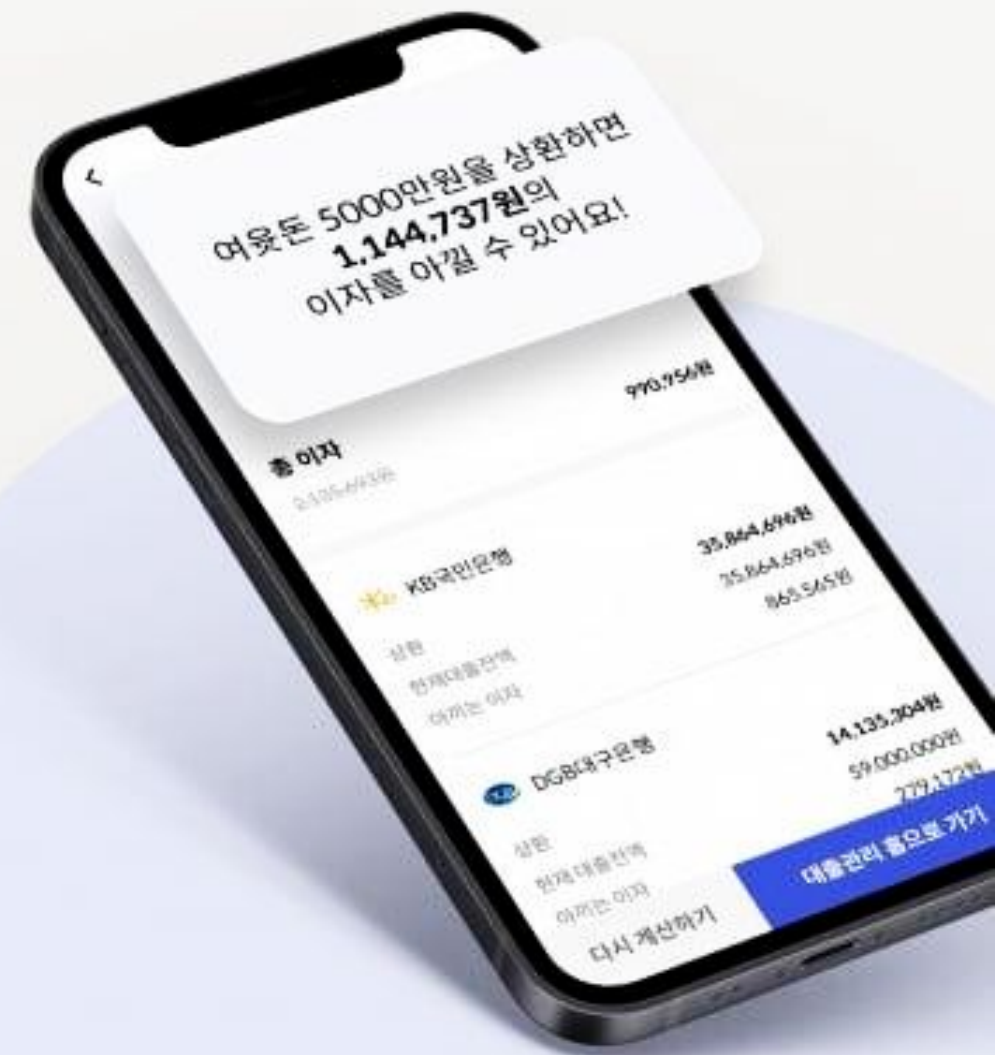


# finda

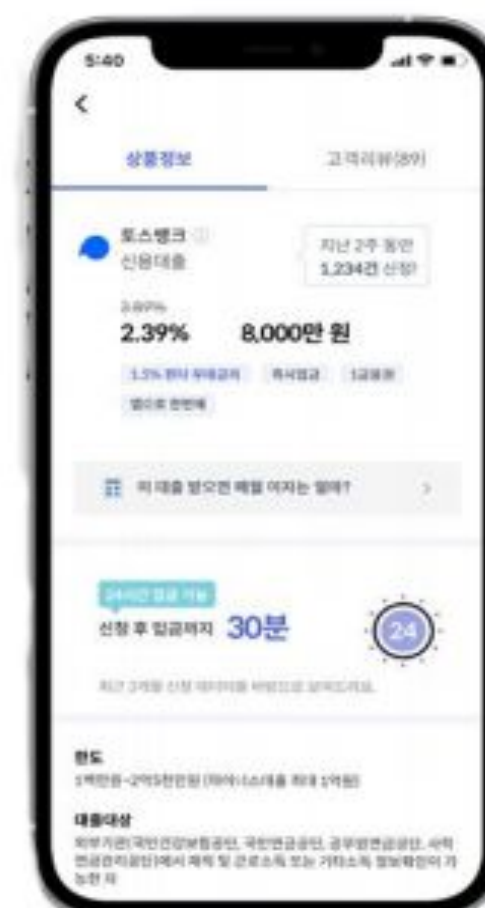


여윳돈 생기면  
어떤 대출부터  
갚을까?  
**핀다에서  
바로 확인!**

finda



5



# 대출관리 서비스

보유한 대출을 **통합적**으로 조회 가능하고, 대출 상환 과정에서 필요한 기능 제공



# 프로젝트 배경

## 업계 동향 및 필요성

디지털 금융 및  
대출 앱의 성장

금융 분야에서의 디지털 기술  
확장으로 대출 어플 및 온라인  
금융 서비스의 중요성이  
증가하는 추세

금융 분야의  
고객 유지

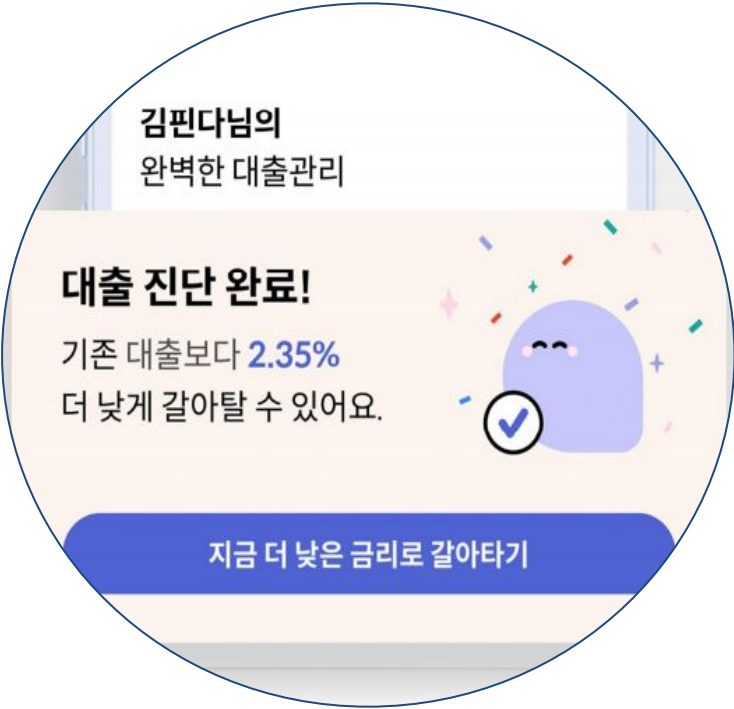
금융 분야에서의  
고객 유치와 유지의 중요성  
개인화된 메시지 제안 서비스는  
고객 충성도를 높일 수 있는  
효과적인 수단

프로젝트의  
필요성

고객들의 로그 데이터를 바탕으로  
행동 분석, 대출 신청 예측을  
비롯하여 고객에게 제공하는  
맞춤형 메시지의 제공은  
여러 이점 존재



# 목표 설정



핀다 홈화면  
진입 고객 기반  
특정 기간 안에  
대출 신청 고객 예측

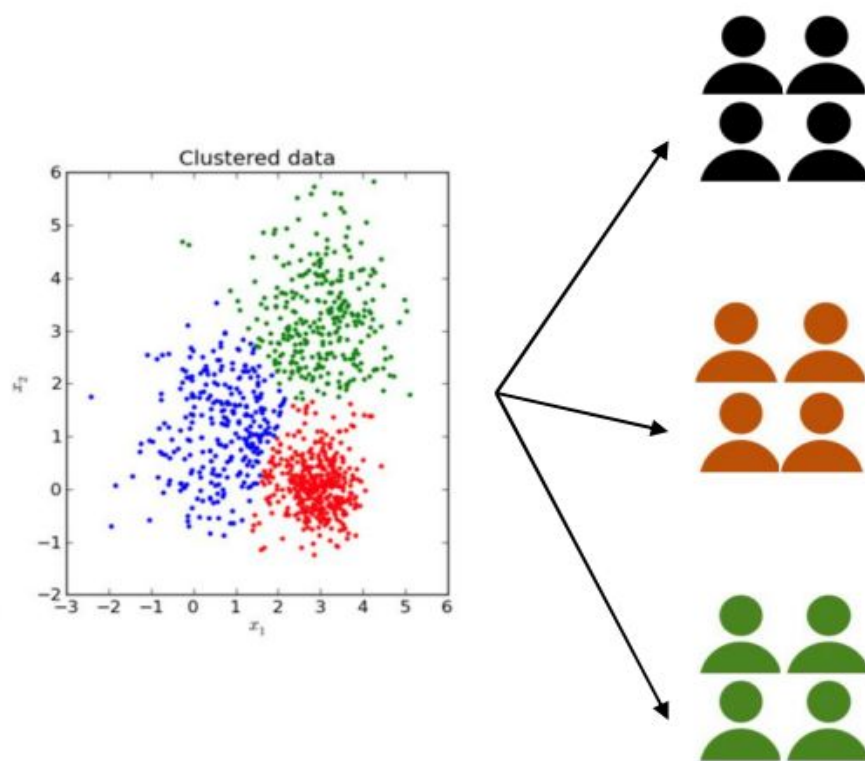
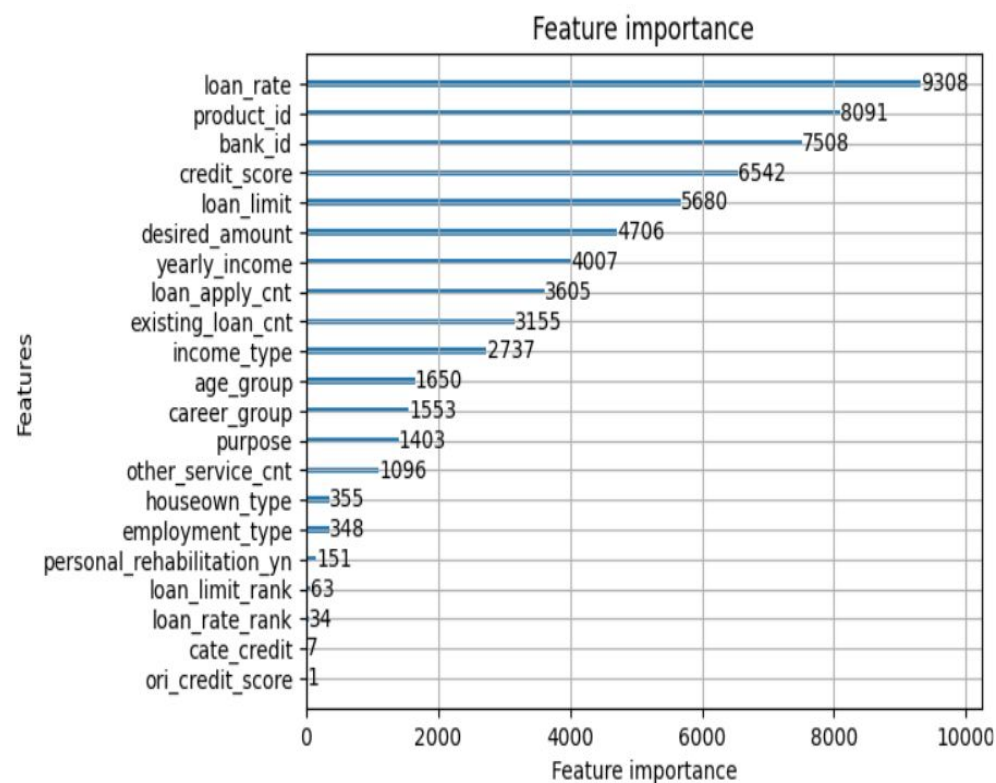


핀다 홈화면  
진입 고객 기반  
고객 군집 분석 및  
군집별 서비스 메시지 제안



두 분석을 바탕으로  
생성형 AI(chatGPT)를 활용한  
자동화 분석서 생성

# 분석 방향



군집 특성 및 메시지 제안
기대출이 있다는 점에서 승인 가능성이 증명되었으며, 더불어 신용등급이 높아 대출 승인이 비교적 유리한 군집
“다른 상품도 알아보셨나요? 대출 상품 비교해보자!”
기대출이 있지만 관리 서비스를 사용해보지 않은 군집으로 핀다의 대출관리서비스 사용을 유도
“관리는 처음부터 잘해야 하니까! 이런 서비스 어때요?”
신용 등급이 높아 기대출금이 비교적 고액, 효율적인 관리 방법 제시
“대출관리 잘하고 계신가요? 핀다에서는 고액 대출 관리도 쉽게!”
대출을 비롯한 핀다의 서비스를 모두 이용한 충성 고객 그룹
“쉽게 보는 합리적인 대출 상품과 쉽게 하는 현명한 대출 관리!”

## 예측 모델 목적 수립

고객의 대출 신청 예측을 가장 중요한 가치로 두고 예측모델을 수립한다.

## 예측 모델 특성 기반 군집 분석

대출의 신청 · 미신청 고객의 특성과 수립한 예측 모델의 중요 특성을 기반으로 군집화한다.

## 자동화 분석서

예측모델의 데이터와 분석 결과를 프롬프트에 녹여내어 자동화 분석서를 생성한다.



# Dataset

## user\_spec

No.	컬럼ID	컬럼명	기타
1	application_id	신청서 번호	암호화 및 가명화
2	user_id	유저번호	암호화 및 가명화
3	birth_year	유저 생년월일	
4	gender	유저 성별	0: 여자, 1: 남자
5	insert_time	생성일시	
6	credit_score	한도조회 당시 유저 신용점수	10점 단위 반올림
7	yearly_income	연소득	100만 단위 반올림
8	income_type	근로형태	직장인(4대보험 O/X), 개인사업자, 전문직, 프리랜서, 기타
9	company_enter_month	입사연월	
10	employment_type	고용형태	
11	houseown_type	주거소유형태	
12	desired_amount	대출희망금액	100만 단위 반올림
13	purpose	대출 목적	
14	personal_rehabilitation_yn	개인회생자 여부	0: 개인회생자 X, 1: 개인회생자 O
15	personal_rehabilitation_complete_yn	개인회생자 납입 완료 여부	0: (개인회생자인 경우) 납입중, 1: 납입완료
16	existing_loan_cnt	기대출수	
17	existing_loan_amt	기대출금액	100만 단위 반올림

# Dataset

## log\_data

No.	컬럼명	컬럼 의미
1	user_id	유저 번호 ( 암호화 및 가명화 )
2	event	행동명
3	timestamp	행동일시
4	date_cd	일 코드



event_new	event description
SignUp	회원가입
OpenApp	핀다 앱 실행
Login	핀다 앱 로그인
ViewLoanApplyIntro	한도조회 인트로 페이지 조회
StartLoanApply	한도조회 시작하기 버튼 클릭
CompleteIDCertification	본인인증완료
EndLoanApply	한도조회 결과 확인
UseLoanManage	대출관리 서비스 이용
UsePrepayCalc	여윳돈 계산기 서비스 이용
UseDSRCalc	DSR 계산기 서비스 이용
GetCreditInfo	KCB 신용정보 조회



# Dataset

## loan\_result

No.	컬럼ID	컬럼명	기타
1	application_id	신청서 번호	암호화 및 가명화
2	loanapply_insert_time	한도조회 일시	
3	bank_id	금융사 번호	암호화 및 가명화
4	product_id	상품 번호	암호화 및 가명화
5	loan_limit	승인한도	십만단위 반올림
6	loan_rate	승인금리	둘째자리에서 반올림
7	is_applied	신청 여부(타겟)	예측레이블(NaN이 테스트 데이터)

# EDA 과정

---

## 3개의 데이터 Merge

주어진 데이터 셋에서 각각의  
id 값을 통해 Merge하여  
하나의 데이터로 통합



## 파생변수 생성

분석의 용이성과  
모델의 예측력을 높이기 위해  
기존 컬럼값을 활용하여  
8개의 파생변수 생성



## Scaling, Encoding

연속형 변수 -> RobustScaler  
범주형 변수 -> LabelEncoder



## 중요 Features 확인

앞선 전처리 진행 후,  
모델 학습을 위해  
필요한 컬럼만 사용하여  
중요 Features 확인



# EDA 과정

---

## 각 컬럼마다 존재했던 결측치는 어떻게 처리하였나?

### i. 값을 예측할 수 있는 컬럼

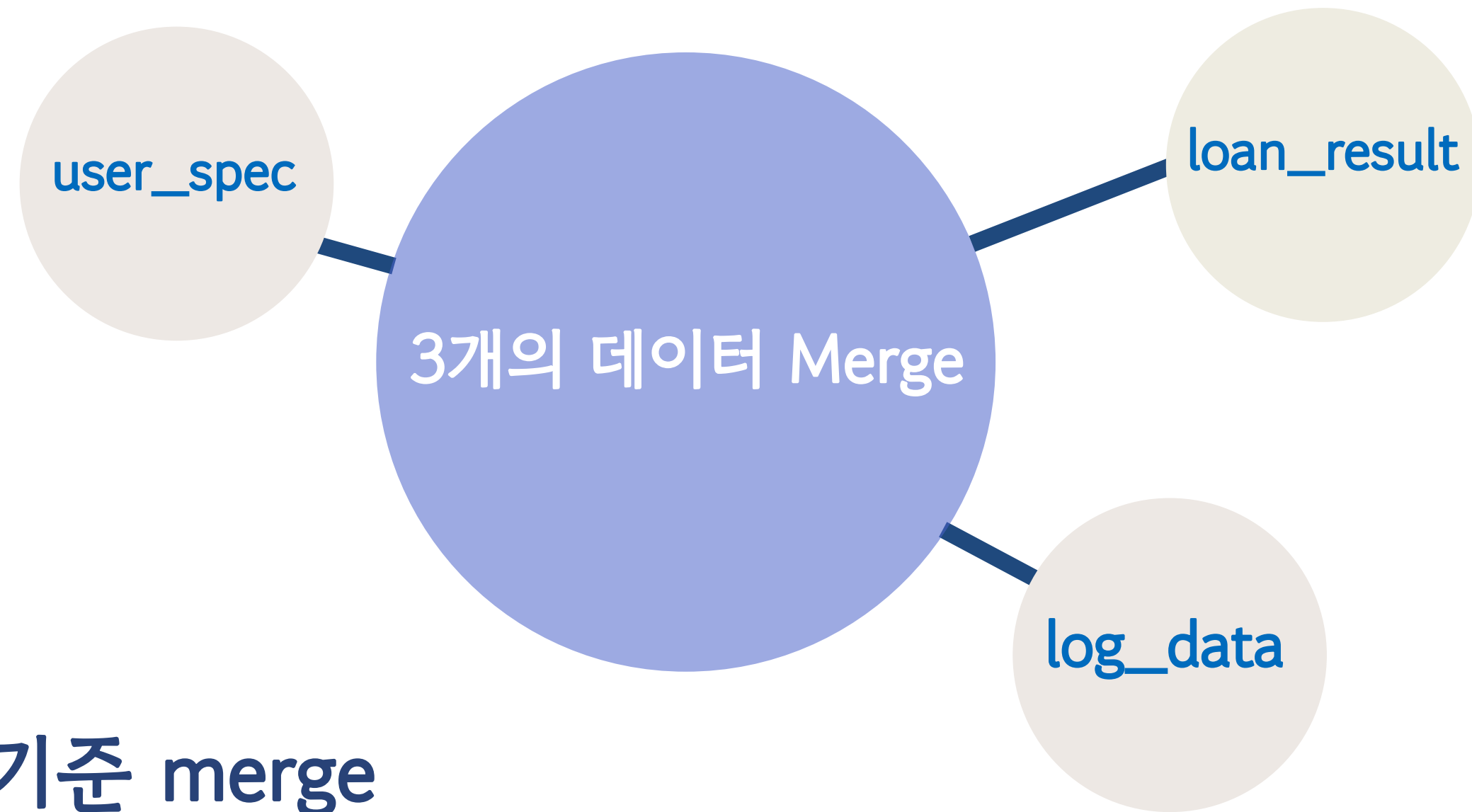
- [user\_spec] - personal\_rehabilitation\_yn(개인회생자 유무)  
2022-04-17 이전은 개인회생유무를 입력할 수 없어 -1로 대체
- [user\_spec] - personal\_rehabilitation\_complete\_yn(개인회생자 납입 완료 여부)  
위에서 대체한 개인회생자 유무를 알 수 없어 -1로 대체한 고객의 경우를 포함, 개인회생이 되지 않은 고객의 경우 납입 완료 여부를 확인할 수 없어 2로 대체
- [user\_spec] - existing\_loan\_cnt  
0으로 대체
- [user\_spec] - existing\_loan\_amt  
기대출이 0인 경우, 0원으로 대체
- [user\_spec] - desired\_amount  
purpose가 채워져 있는 경우, 대출을 희망하는 사람으로 볼 수 있어 반올림 기준 100만원으로 대체
- [log\_data] - loan\_apply\_cnt, other\_service\_cnt (파생변수)  
0으로 대체

### ii. 이외 컬럼

- [loan\_result] - loan\_limit, loan\_rate, credit\_score  
연속형 데이터는 전체 평균값으로 대체
- [loan\_result] - income\_type  
OTHERINCOME인 경우, company\_enter\_month를 999999로 대체  
OTHERINCOME 또는 FREELANCER가 아니면서, yearly\_income 0인 값, 50만원으로 대체
- [log\_data] - time\_stamp, event, mp\_os, mp\_app\_version  
분석에 필요하지 않다고 판단하여 drop

# EDA 과정

---



- 날짜 기준 merge
- log\_data

[본인인증] - [한도조회 시작 버튼 클릭] - [한도조회 결과 확인]

→ user\_id가 생성된 후 활동이 없는 user를 제외하기 위함



# EDA 과정

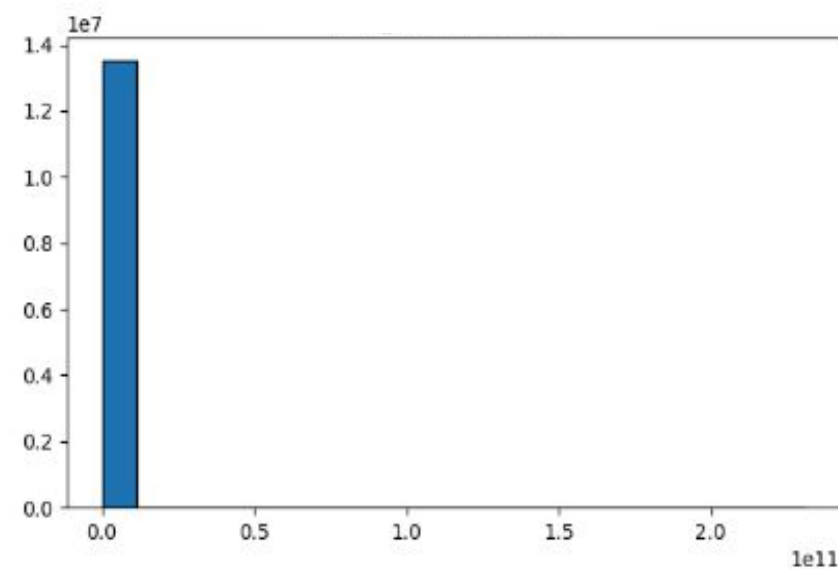
## 파생변수 생성

파생변수 컬럼명	설명
1 연령대(age_group)	: 기존 birth 컬럼을 바탕으로 연령 그룹화 연령대는 비슷한 사회적 지위 등의 특성을 공유하고 있다고 판단하여 생성함
2 경력(career_group)	: company_enter_month를 바탕으로 경력 그룹화 → 1년 미만, 1년 이상 3년 미만, 3년 이상 5년 미만, 5년 이상 10년 미만, 10년 이상 이후 연소득의 null 값은 경력을 바탕으로 값이 대체됨
3 승인한도 순위(loan_limit_rank)	: loan_limit의 순위
4 승인금리 순위(loan_rate_rank)	: loan_rate의 순위
5 신용등급(cate_credit)	: credit_score를 KCB 신용점수를 바탕으로 한 신용등급 실제로 쓰이고 있는 신용 등급을 사용하여 분석의 현실성을 높이고자 함
6 앱 실행 횟수(open_app_cnt)	: event 컬럼의 OpenApp 컬럼을 활용한 앱 실행 횟수 핀다 앱 이용의 활동성 정도를 파악하고자 함
7 대출 신청 관련 서비스 이용 횟수 (loan_apply_cnt)	: event 컬럼의 VeiwLoanApplyIntro, StartLoanApply, CompleteIDCertification, EndLoanApply 횟수를 더한 대출 신청과 관련된 서비스 이용 횟수 핀다 이용 고객들의 대출 신청에 대한 적극성을 파악하고자 함
8 기타서비스 이용 횟수 (other_service_cnt)	: event 컬럼의 UseLoanManage, UsePrepayCalc, UseDSRCalc, GetCreditInfo 횟수를 더한 기타서비스 이용 횟수 직접적으로 대출과 관련된 것 이외 기타 기능의 활용 정도를 파악하고자 함

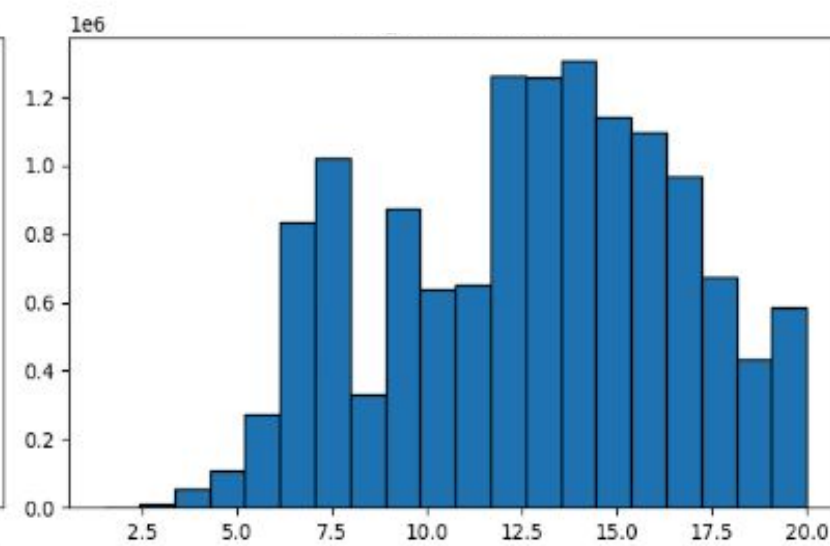
# EDA 과정

## Scaling, Encoding

Before



After



## 연속형 데이터 스케일링

loan\_result 데이터와 user\_spec 데이터의 연속형 변수 스케일링

히스토그램, Q-Q Plot을 통한 데이터 확인 결과 정규분포를 따르지 않아, RobustScaler 적용

## 범주형 데이터 인코딩

범주형 변수

'income\_type', 'employment\_type', 'houseown\_type', 'purpose', 'career\_group'에 대하여

one-hot encoding 진행



# EDA 과정

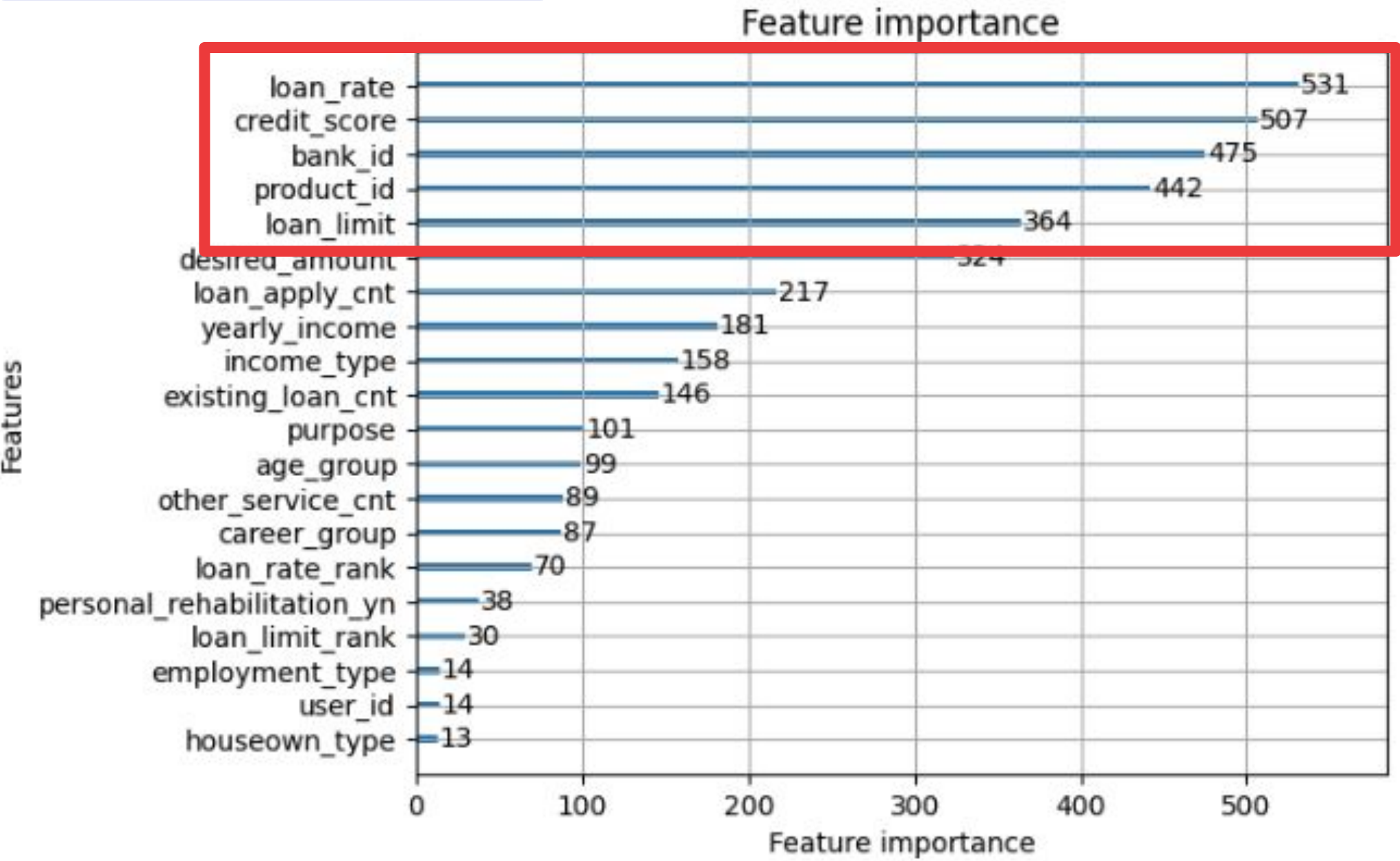
## 모델에 사용된 Features

변수명		변수명	
1	bank_id	11	desired_amount
2	purpose	12	existing_loan_cnt
3	loan_limit	13	yearly_income
4	loan_rate	14	<b>age_group</b>
5	credit_score	15	<b>career_group</b>
6	income_type	16	<b>loan_limit_rank</b>
7	employment_type	17	<b>loan_rate_rank</b>
8	houseown_type	18	<b>open_app_cnt</b>
9	product_id	19	<b>loan_apply_cnt</b>
10	personal_rehabilitation_yn	20	<b>other_service_cnt</b>

14~20 : 파생변수

# EDA 과정

## 중요 Features



사용된 컬럼 중  
중요도가 높은 컬럼

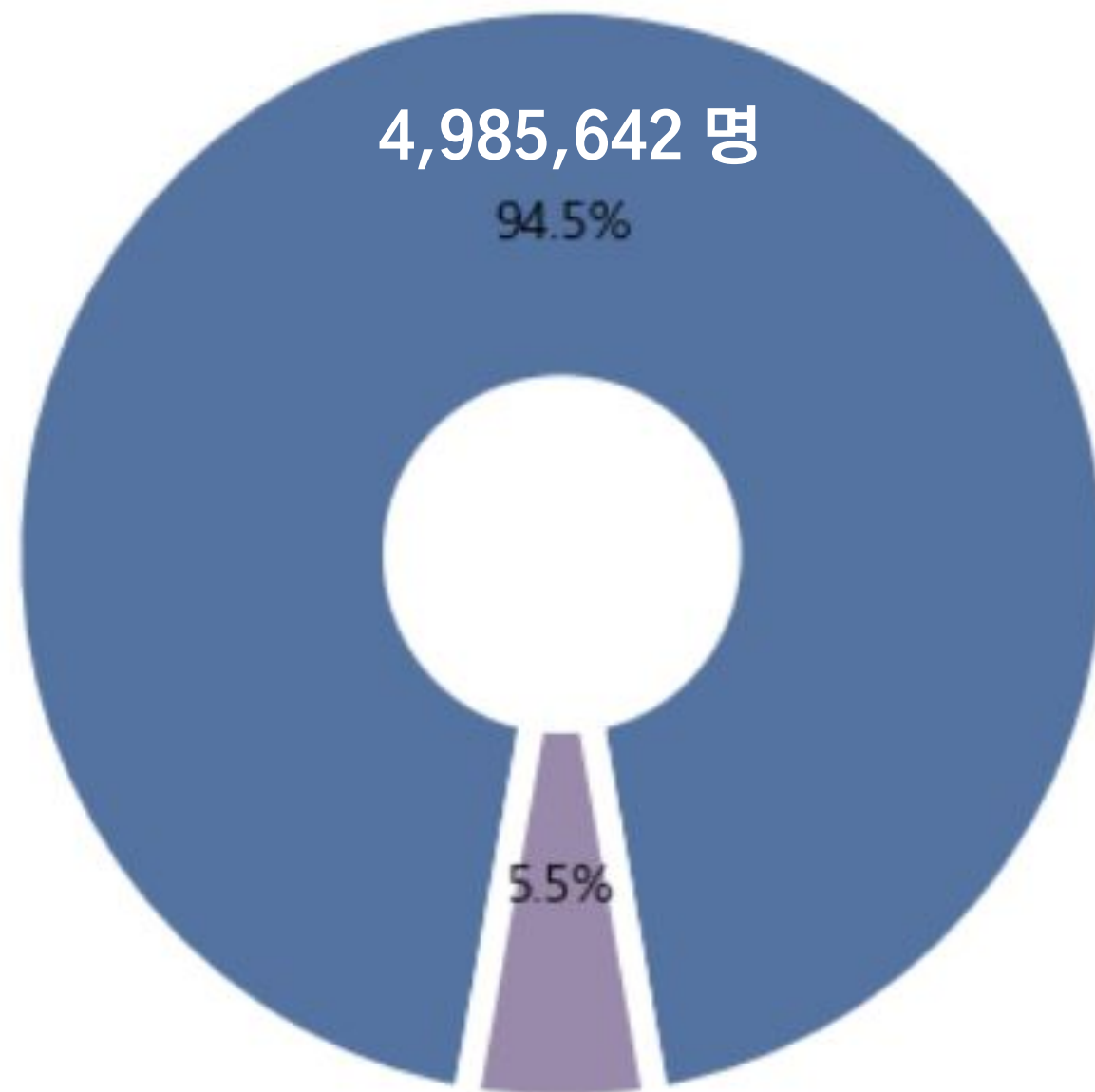
- loan\_rate (승인 금리)
- credit\_score (신용 점수)
- bank\_id (은행명)
- product\_id (대출상품명)
- loan\_limit (승인 한도)

# 신청 및 미신청자 간 데이터 불균형 해소

train : 3-4월 / val : 5월 / target : 6월

## 대출 신청 비율

대출 신청 N



대출 신청 Y 276,248 명

## ■ LightGBM 모델에서의 weight 기법

GOSS(gradient 기반 샘플링)와 EFB(특성이 다른 피쳐들로 학습)을 결합하여 학습시키는 기법이다.

신청자 예측에 실패할 때 발생하는 gradient에 미리 weight(가중치)를 주어 gradient를 크게 만든다.

이는 대출 신청자 데이터의 중요도를 높임으로써 대출 신청 예측 모델의 성능을 향상시킨다.

## ■ Undersampling

대출 신청 고객의 신용등급 비율을

대출 미신청 고객 데이터에 동일하게 적용한 후, 랜덤으로 샘플을 추출하여 데이터의 불균형을 해소한다.



0.15	precision	0.25
0.84	recall	0.66
0.55	F1-score	0.65
0.73	Accuracy	0.87
0.86	AUROC	0.89

## UnderSampling 기법

대출 미신청 고객 데이터에서 대출 신청  
고객 신용 등급 비율만큼 샘플링하여  
예측 모델 수립에 이용

## LightGBM 모델에서의 weight 기법

대출 신청 고객의 비율이 5.5%로  
매우 불균형한 데이터임을 확인하였고,  
데이터 불균형으로 인해 잘못 예측하는 경우를  
줄이고자 신청 class에 5배가중치 부여



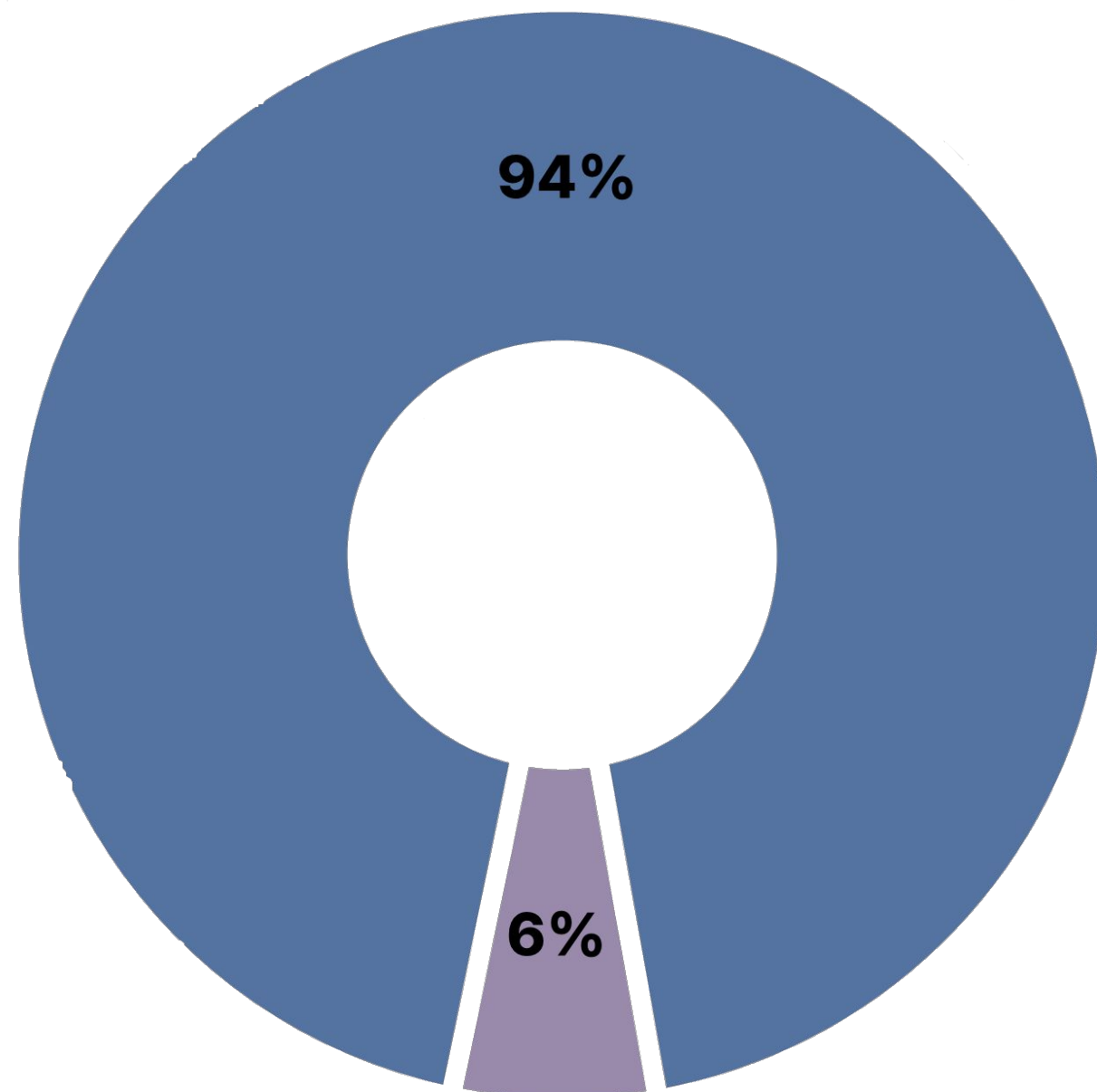
LightGBM 기법 간 모델 성능 비교

# 신청 및 미신청자 간 데이터 불균형 해소

train : 3-5월 / target : 6월

## 대출 신청 비율

대출 신청 N 3,066,159 명



대출 신청 Y 179,388 명

### ■ LightGBM 모델에서의 weight 기법

GOSS(gradient 기반 샘플링)와 EFB(특성이 다른 피쳐들로 학습)을 결합하여 학습시키는 기법이다.

신청자 예측에 실패할 때 발생하는 gradient에 미리 weight(가중치)를 주어 gradient를 크게 만든다.

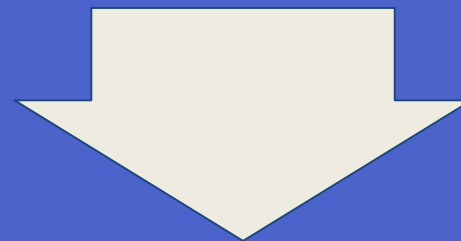
이는 대출 신청자 데이터의 중요도를 높임으로써 대출 신청 예측 모델의 성능을 향상시킨다.

## LightGBM 모델에서의 weight 기법

대출 신청 고객의 비율이 6%로  
매우 불균형한 데이터임을 확인하였고,  
데이터 불균형으로 인해 잘못 예측하는 경우를  
줄이고자 신청 class에 5배가중치 부여

데이터의 양은 **감소**.

precision, F1-score, Accuracy, AUROC **상승**



3-4월의 user의 행동과 5-6월의 user 행동이 달라

5월을 포함시킨 것이 성능을 향상시키는 것에 많은 영향을 준 것으로 추측

precision : **0.43**

< Validation Metrics >

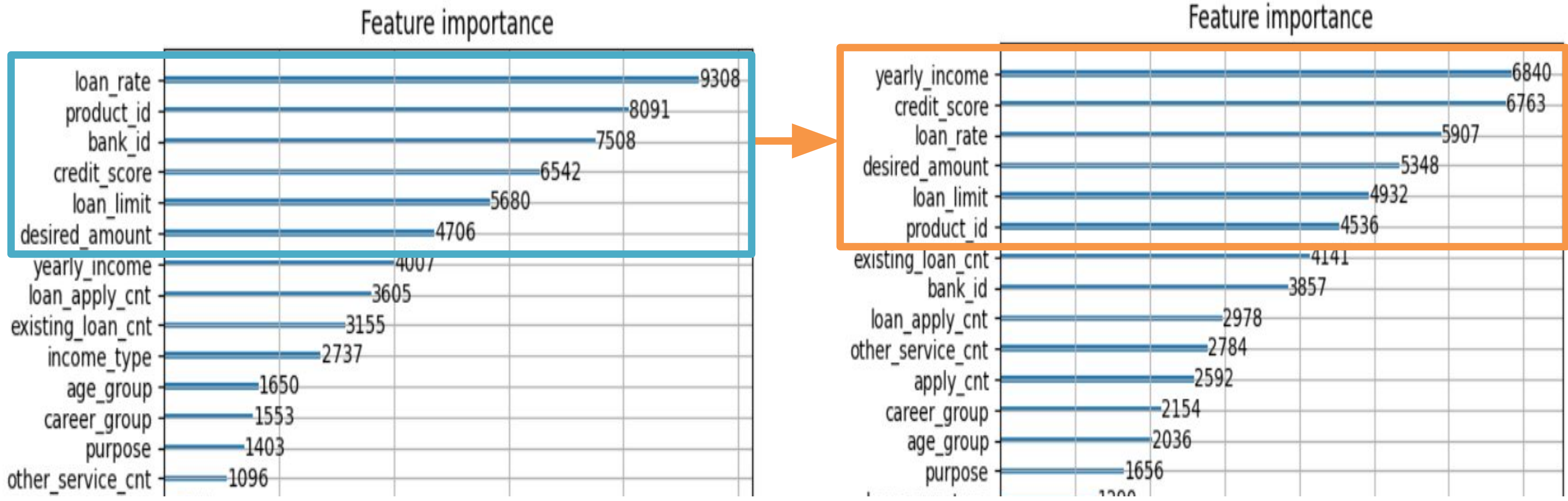
recall : 0.37

F1-score : **0.68**

Accuracy : **0.94**

AUROC : **0.9**

# Feature importance의 변화





# 고객 군집 분석과 군집별 서비스 메시지 제안

## 도메인 지식 및 데이터 기반 군집화

머신러닝 기법(k-means, DB Scan)을 사용하였으나 유의미한 결과를 얻지 못해  
데이터에 기반하여 유사하다고 판단되는 기준으로 군집화 진행

### 충성고객 기반 군집화

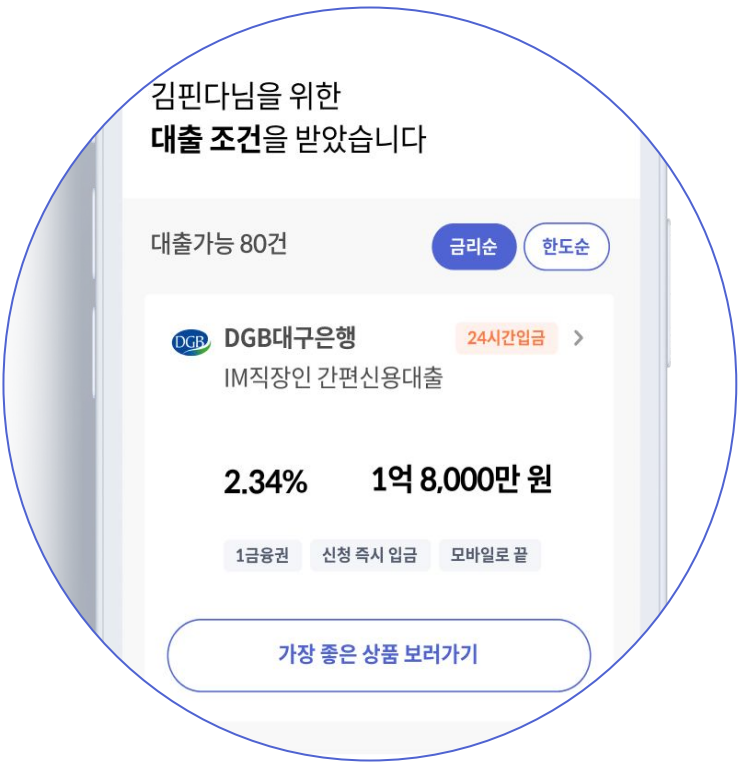
대출을 신청할 것으로 예측되는 고객  
즉, 충성고객을 기반으로 군집화가  
이루어질 수 있도록 방향 설정

### 대출 신청 / 미신청 고객 분류

각 군집별 메시지 내용에  
큰 차이를 두는 것을 목표로 수립,  
따라서 대출 신청 고객과 미신청  
고객으로 1차 분류 진행

# 고객 군집 분석

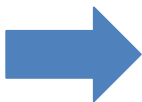
각 군집 형성을 위한 중요 요인 → **Feature importance**



“ 대출 신청과 서비스 이용 유도를 위한 기준을 세워  
고객들을 군집화하려고 함 ”

## 대출 신청 고객 군집

- 1. 신용등급
- 2. 서비스 이용 여부
- 3. 기대출 여부



대출 신청을 예측하는  
두번째 중요 요인인 신용 점수를 **신용 등급**으로 파생변수를 생성  
핀다 앱의 서비스 이용을 유도하기위해 **서비스 이용 여부를 선택**

# 고객 군집 분석

## 대출 신청 고객 군집

신용등급 최하위 고객 제외

신용등급이 높고 기대출 있는 경우.  
대출 신청 확률 증가 예상

	Group 1 (457명)	Group 2 (1,143명)	Group 3 (33,864명)	Group 4 (100,227명)
군집 설명	신용등급이 높고 기대출 1건 이상, 서비스 이용이 없는 고객	신용등급이 중간이며, 기대출 1건 이상, 서비스 이용이 없는 고객	신용등급이 높고, 기대출 1건 이상, 서비스를 이용한 고객	신용등급이 중간이며, 기대출 1건 이상, 서비스를 이용한 고객
신용 등급	가장 높음(1~4)	중간(5~6)	가장 높음(1~4)	중간(5~6)
서비스 이용 여부	x		1건 이상	
기대출 여부	1건 이상			

# 서비스 메시지

## 대출 신청 고객 군집에게 보내는 메시지

	군집 특성 및 메시지 제안
<b>Group1(457명)</b>	기대출이 있다는 점에서 승인 가능성이 증명되었으며, 더불어 신용등급이 높아 대출 승인이 비교적 유리한 군집
신용등급이 높고, 기대출 1건 이상, 서비스 이용이 없는 고객	“다른 상품도 알아보셨나요? 대출 상품 비교해보자!”
<b>Group2(1,143명)</b>	기대출이 있지만 관리 서비스를 사용해보지 않은 군집으로 핀다의 대출관리서비스 사용을 유도
신용등급이 중간이며, 기대출 1건 이상, 서비스 이용이 없는 고객	“관리는 처음부터 잘해야 하니까! 이런 서비스 어때요?”
<b>Group3(33,864명)</b>	신용 등급이 높아 기대출금이 비교적 고액, 효율적인 관리 방법 제시
신용등급이 높고, 기대출 1건 이상, 서비스를 이용한 고객	“대출관리 잘하고 계신가요? 핀다에서는 고액 대출 관리도 쉽게!”
<b>Group4(100,227명)</b>	대출을 비롯한 핀다의 서비스를 모두 이용한 충성 고객 그룹
신용등급이 중간이며, 기대출 1건 이상, 서비스를 이용한 고객	“쉽게 보는 합리적인 대출 상품과 쉽게 하는 현명한 대출 관리!”



# 고객 군집 분석

각 군집 형성을 위한 중요 요인 → **Feature importance**

“ 대출 신청과 서비스 이용 유도를 위한 기준을 세워  
고객들을 군집화하려고 함”

핀다 앱에서 대출 신청을 유도할 수 있도록

신용 등급, 연소득 변수를 선택

또한 핀다 앱에서 **대환 대출**을 신청하도록 기대출 횟수 변수 선택

핀다 앱의 서비스를 통해 **대출 관리**를 하도록

대출 관련 서비스 이용 횟수 변수 선택



대출 미신청 고객 군집

1. 신용등급
2. 연소득
3. 대출 관련 서비스 이용 횟수
4. 기대출 횟수



# 고객 군집 분석

## 대출 미신청 고객 군집

기대출 및 대출관리서비스 이용도 없는 고객은  
진입 이후 빠져나간 경우가 다수

기대출 및 기대출관리서비스 특성을 이용하여  
군집 간 메시지에 큰 차이를 두고자 함

기대출 일정 이상, 낮은 연소득의 고객은  
낮은 소득에 대한 우려와  
여러 대출 관리를 함께 고민할 것이라 판단

가진 대출을 관리해주는 방향이라면  
핀다앱 이용 지속에 좋은 영향

	Group 1 (7,904명)	Group 2 (7,116명)	Group 3 (5,138명)	Group 4 (7,220명)
군집 설명	신용등급 가장 높고, 기대출 없는 고객	연소득 높고 기대출이 많으며, 대출관리서비스 미이용 고객	연소득 높고 기대출이 없으며, 대출관리서비스 이용 고객	연소득 낮고 기대출이 많으며, 대출관리서비스 이용 적은 고객
신용 등급	가장 높음(1~2)	-		
연소득	-	상위 30% (4,500 초과)		하위 10% (2,300 미만)
대출 관리 서비스 이용 여부	-	x	1회 이상	5회 미만
기대출 여부	x	5건 이상	x	5건 이상

# 서비스 메시지

## 대출 미신청 고객 군집에게 보내는 메시지

	군집 특성 및 메시지 제안
<b>Group1(7,904명)</b>	신용등급이 높아 대출 승인 확률이 높지만, 기대출이 없어 핀다에서의 첫번째 대출을 유도할 수 있는 군집
신용등급 가장 높고, 기대출 없는 고객	“당신의 첫번째 대출, 핀다가 추천해줄게요!” “누구보다 많은 선택지, 보고 결정하자!”
<b>Group2(7,116명)</b>	기대출이 많고 연소득이 높아 대출 상환 능력이 기대되지만, 관리서비스 미이용 고객으로 대출 관리 서비스를 유도할 수 있는 군집
연소득 높고 기대출 많으며 대출관리서비스 미이용 고객	“어떤 대출 먼저 상환해야 할까? 흩어져 있는 당신의 대출, 모아서 알려줄게요!”
<b>Group3(5,138명)</b>	기대출이 없음에도 대출관리서비스를 이용하며 대출에 관심이 있다고 여겨지는 군집
연소득 높고 기대출 없으며 대출관리서비스 이용 고객	“핀다만의 보기 좋은 대출 비교! 궁금하지 않아?”
<b>Group4(7,220명)</b>	연소득이 낮아 대출 상환 방향에 어려움을 겪을 수 있어, 대출 상환일이 가까워졌을 때 잔액을 점검해 알려주는 핀다의 기능을 적극 활용할 수 있는 군집
연소득 낮고 기대출 많으며 대출관리서비스 이용 낮은 고객	“상환일이 가까워지고 있어요! 가장 먼저 상환해야 할 대출 알려줄게!”

# 생성형 AI

## 핀다 홈페이지 진입 고객 기반 고객 군집 분석 및 군집별 서비스 메시지 제안

### Prompt

## Prompt 작성 기준

1. 클러스터링한 군집 데이터를 바탕으로 작성할 것
2. 생성형 AI가 군집의 특성을 고려하여 **군집명**을 짓도록 작성하고, 길지 않은 메시지로 표기할 것
3. 기업과 고객에게 각각 메시지를 나타내도록 구성할 것
4. 고객에게는 군집의 기본 정의를 기반으로 군집을 설명할 것
5. 기업에게는 고객에게 제공되는 메시지에 대한 근거를 제시할 것

```
def generate_prompt(group_info):
    prompt = f"""
    그룹 정보와 이 그룹에게 보낼 메시지에 맞게 그룹에 대해 예시처럼 RPG 게임 속 직업처럼
    창의적이지만 부정적이지 않은 별명을 지어주세요.
    이 그룹에게 보낼 메시지는 핀다앱의 서비스의 활발한 사용을 유도할 수 있도록 예시 속 메시지와
    {group_info}를 참고해서 간결하게 작성해주세요.
    메시지의 내용은 사용자가 핀다앱의 서비스에 대해 흥미를 유발할만한 메시지를 예시 속 메시지처럼
    최대한 간결하게 작성해주세요.
    회사에게 보고하기 위해 그룹 정보를 이용하여 그룹에 대한 설명을 자세하게 적고, 이 메시지의 효과도
    구체적으로 설명해주세요.

    ---
    예시)
    [ 고객에게 제공할 내용 ]
    당신은 "핀다앱의 금융 모험가"입니다.
    메시지: "최강의 동료, 핀다앱과 함께라면 더 큰 성장을 이룰 수 있어요!"

    [ 회사에 보고할 내용 ]
    그룹 정보 및 메시지 효과 : {group_info} 신용등급이 5등급으로 중간수준이고 대출을 받아본 경험이
    있지만 핀다앱 서비스 사용 경험이 없어서 핀다앱에서 서비스 활발하게 사용을 유도할 그룹
    - 이 그룹은 대출 경험이 있는 사용자로 신용 등급이 중간 수준입니다. 이들에게 핀다앱 서비스를
    활발하게 사용하도록 유도하는 새로운 메시지입니다.
    - 별명을 "핀다앱의 금융 모험가"로 정의하여, 사용자들에게 새로운 도전을 제안하는 메시지입니다. 이를
    통해 사용자들의 흥미를 유발하고 핀다앱의 서비스에 참여할 동기를 부여합니다.
    - 메시지는 간결하고 흥미로운 내용으로 작성되어 있어 사용자들에게 호감을 줄 것으로 예상됩니다.
    - 이 그룹의 대출 경험이 있어 핀다앱의 다른 유용한 서비스들을 사용할 수 있는 잠재력이 크기 때문에,
    이들의 참여를 유도하는 것이 중요합니다.
    - 이 메시지를 통해 신규 사용자들의 핀다앱 서비스 활발한 사용을 유도할 수 있습니다.

    ---
    """
    return prompt.strip()
```



finda



안녕하세요

김핀다 님

신용등급 : 2등급

기대출수 : 3건

서비스 이용 경험 0

당신은 “금융 마스터”입니다.

“ 핀다앱이 최적의 경로를  
안내해드립니다. “

[ 메시지 의미 ]

핀다앱의 서비스를 이미 사용한 경험이 있으며,  
큰 대출 금액을 대상으로 하는 고객입니다.

이들에게 별명을 “금융 마스터”로 정의하여,  
핀다앱을 통해 자신의 재무를 효과적으로 관리하고  
통제할 수 있음을 전달하는 메시지입니다.

# 생성형 AI

## 핀다 홈페이지 진입 고객 기반 대출 신청 고객 예측 자동화 분석서

### Prompt 작성 기준

1. 수립한 예측 모델 데이터를 기반으로 작성
2. 예측 모델에서 **중요 피처**로 드러난 승인 금리, 상품 번호, 금융사 번호, 신용 점수, 승인 한도를 우선으로 고려하여 분석서를 출력 할 수 있도록 작성
3. 정제된 보고서의 형식으로 받아볼 수 있도록 **보고서 출력 항목을 예시**로 작성
4. 전반적으로 현재 고객의 상황을 판단하되, 대출 실행 및 **강제성을 띄지 않도록** 유의

### Prompt

```
def generate_prompt(user_id, yearly_income, credit_score, cate_credit, purpose,
desired_amount, bank_id,
product_id, loan_limit, loan_rate, is_applied, feature_query):
    prompt = f"""
    Finda앱 사용 고객의 대출 신청 예측 결과에 대한 이유를 분석해서 보고하는 보고서 형식으로
    작성해주세요.

    고객 정보의 금융사명, 상품 번호, 대출 상품 한도, 대출 상품 금리는 고객이 Finda앱으로 대출 조회
    결과 중에 선택했을 것으로 예상되는 정보입니다.

    예측 모델이 고려한 중요 특징을 이용해서 모델이 예측한 신청 여부에 대한 이유를 고객 정보를
    자세히 분석해서 구체적인 내용을 보고서 형식으로 작성해주세요.

    ---
    [ 대출 신청 예측 결과 보고서 ]
    고객명 : {user_id}
    연소득 : {yearly_income}원
    신용 점수 : {credit_score}점
    신용 등급 : {cate_credit}등급
    대출 목적 : {purpose}
    대출 희망 금액 : {desired_amount}원
    선택 금융사명 : {bank_id}
    선택 상품 번호 : {product_id}
    대출 상품 한도 : {loan_limit}원
    대출 상품 금리 : {loan_rate}%
    신청 여부 예측 결과 : {is_applied}
    ---
    """

    return prompt.strip()
```

핀다 홈화면 진입 고객 기반 대출 신청 고객 예측 자동화 분석서 **결과**

[보고서]

1. 고객 정보 분석 결과

- 고객명 : 352393
- 연소득 : 60,000,000원
- 신용 점수 : 760점
- 신용 등급 : 5등급
- 대출 목적 : 사업자금
- 대출 희망 금액 : 30,000,000원
- 선택 금융사명 : 49
- 선택 상품 번호 : 136
- 대출 상품 한도 : 7,000,000원
- 대출 상품 금리 : 6%
- 신청 여부 예측 결과 : **신청**

2. 모델 분석 결과

- 예측 모델은 고객의 개인 정보, 금융사 및 상품 정보를 기반으로 대출 신청 여부를 예측하기 위해 사용되었습니다. 모델은 다양한 특징을 고려하여 학습되었으며, 이 보고서에서는 가장 중요한 특징들에 대해 분석 결과를 설명하겠습니다.

3. 신용 점수

- 해당 고객의 신용 점수는 760점으로 상당히 높은 수준입니다. 신용 점수는 고객의 신용 이력을 기반으로 산출되며, 높은 신용 점수는 신용 대출 신청 승인에 긍정적인 영향을 미칩니다. 따라서, 신용 점수가 높은 이유로 대출 신청이 예측되었습니다.

4. 신용 등급

- 고객의 신용 등급은 5등급입니다. 신용 등급은 신용 점수를 기반으로 산출되며, 신용 점수와 마찬가지로 높은 신용 등급은 대출 신청 승인에 긍정적인 영향을 미칩니다. 따라서, 신용 등급이 높은 이유로 대출 신청이 예측되었습니다.

5. 대출 목적

- 고객의 대출 목적은 사업자금입니다. 대출 목적은 대출 신청 여부를 예측하는데 있어서 중요한 특징입니다. 사업자금은 투자나 자금이 필요한 상황 등 경제적인 목적으로 사용되므로, 해당 고객이 대출을 신청하는 이유로 대출 신청이 예측되었습니다.

6. 대출 희망 금액과 상품 한도

- 고객이 대출 희망 금액으로 30,000,000원을 선택하였으나, 선택한 상품의 대출 한도는 7,000,000원으로 제한되어 있습니다. 대출 희망 금액과 상품 한도는 대출 신청 여부 예측에 영향을 미치는 중요한 특징입니다. 고객이 희망하는 대출 금액과 상품 한도의 차이로 대출 신청이 예측되었습니다.

7. 대출 상품 금리

- 선택한 상품의 대출 상품 금리는 6%입니다. 대출 상품 금리는 고객에게 부과되는 이자율을 의미하며, 낮은 금리는 대출 신청을 더욱 유리하게 만듭니다. 따라서, 상대적으로 낮은 대출 상품 금리로 인해 대출 신청이 예측되었습니다.

8. 종합적 평가

- 위 분석 결과에 따르면, 해당 고객의 신용 점수와 신용 등급이 높으며, 대출 목적이 사업자금인 경우로 대출 신청이 예측되었습니다. 또한, 대출 한도와 대출 희망 금액의 차이가 존재하고, 선택한 대출 상품의 금리가 낮은 것도 대출 신청에 영향을 미친것으로 분석되었습니다. 종합적으로 고객의 개인 정보와 대출 관련 정보들이 대출 신청 여부에 영향을 미쳤다고 판단할 수 있습니다.

9. 결론 및 추천

- 위 분석 결과에 따르면, 해당 고객의 대출 신청 여부를 예측하는데 있어서 긍정적인 요인들이 존재하였습니다. 따라서, 이 고객의 대출 신청을 승인하는 것이 타당한 결정일 수 있습니다. 그러나 현재 선택한 상품의 대출 한도와 희망 금액 차이가 있으므로, 상품 변경 혹은 대출 희망 금액의 조정을 고려할 필요가 있습니다.

- 이 보고서의 분석 결과는 예측 모델을 기반으로 하였으며, 실제 대출 신청을 판단하기 위해서는 추가적인 검토와 판단이 필요합니다. 이 보고서는 참고 자료로 활용되어야 하며, 최종 결정은 신용기관이나 금융 기관의 정책과 기준을 따르는 것이 필요합니다.

- 추후 모델의 정확도 및 성능 향상을 위해 더 많은 데이터와 다양한 특징을 활용하여 분석을 진행함으로써 더 정밀한 예측을 이루어낼 수 있을 것으로 기대됩니다. 이를 위해 고객 데이터의 수집 및 관리에 주의를 기울이고, 모델 업데이트와 개선을 시행하는 것이 추천됩니다.

# 결과 요약

핀다 홈페이지 진입 고객 중 특정 기간 안에 대출 신청 고객 예측 모델 수립

예측 모델 1.  
LightGBM \_weight 기법

예측 모델 2.  
LightGBM \_undersampling 기법

- 정밀도, f1-score, AUROC 및 정확도에서 모델 1이 모델 2보다 뛰어난 성능을 보이고 있다고 판단
- 대출 신청 고객 예측이 프로젝트의 목적이므로, 대출 신청 고객의 정밀도가 높은 모델 1을 최종 예측 모델로 선정

	ls_applied = 1(신청자)		Validation Metrics		
	precesion	recall	F1-score	Accuracy	AUROC
모델 1 (weight)	0.25	0.66	0.65	0.87	0.89
모델 2 (undersampling)	0.15	0.84	0.55	0.73	0.86



# 결과 요약

## 핀다 홈화면 진입 고객 기반 고객 군집 분석 및 군집별 서비스 메시지 제안

### 군집 분석

대출 신청 고객 / 미신청 고객

도메인 지식과 데이터에 기반하여  
유사하다고 판단되는 기준으로 군집화 진행

- 군집별 특성을 두드러지게 나타내고자 군집의 1차 분류 기준을 대출 신청 여부로 설정
- 대출 신청 가능성이 높거나 대출에 대한 적극성의 척도로 삼을 수 있는 항목을 기반으로 군집 분류

	분류에 사용된 특성			
대출 신청 고객	신용 등급	연소득	대출관리서비스 이용 횟수	기대출 횟수
대출 미신청 고객		서비스 이용 여부	( 포함 )	기대출 여부

# 결과 요약

## 핀다 홈화면 진입 고객 기반 고객 군집 분석 및 군집별 서비스 메시지 제안

각 군집별 메시지 내용에 큰 차이를 두는 것을 목표로 수립,  
따라서 대출 신청 고객과 미신청 고객으로 1차 분류 진행

서비스 메시지

고객 맞춤형 메시지 제안

- ‘당신의 첫번째 대출’과 같은 표현으로 메시지를 받는 고객이 맞춤화된 관리를 받고 있음을 체감할 수 있게 함
- 고객에게 핀다가 제공하는 서비스 기능을 메시지에 녹여내어 고객의 서비스 이용 유도

	Group 1	Group 4
군집 설명	신용등급이 높아 대출 승인 확률이 높지만, 기대출이 없어 핀다에서 첫번째 대출을 유도할 수 있는 군집	연소득이 낮으면서 기대출이 많은 군집으로 핀다의 기능(상환일이 가까워지면 잔액 점검 후 알림)을 적극 활용할 수 있는 군집
서비스 메시지	“당신의 첫번째 대출, 핀다가 추천 해줄게요!”	“상환일이 가까워지고 있어요! 가장 먼저 상환해야 할 대출 알려줄게!”

# 기대효과 및 한계점

1

대출 신청 예측에 중요도를 두고 모델을 선정하여,  
프로젝트 목적에 보다 부합하는 예측 결과

기대효과 1



2

메시지가 불쾌한 광고로 느껴지지 않도록  
군집마다 별명을 제시하며 앱을 긍정적으로 볼 수 있게 함

기대효과 2



3

대출 중개 플랫폼의 다양화로  
맞춤형 마케팅을 통해 고객의 유치와 유지

기대효과 3



1

머신러닝 기법으로 군집화가 뚜렷하게 이루어지지 않아  
데이터 기반, 도메인 지식을 통한 논리로 군집화

한계점 1



2

Streamlit을 이용하여 시각화하고자 하였으나,  
예측 모델 수립에 대한 오랜 고민으로 서비스화하지 못함

한계점 2



# 개발후기 및 느낀점

---

**권민재** : 대출 서비스에 대해서 무지했던 나는 핀다라는 앱을 처음 알게 되었고, 기존 도메인 지식이 없는 상태에서 프로젝트를 진행하기란 다소 무리가 있다고 보였다. **전처리부터 약간의 문제가 있어서 생각보다 더딘 진행률을 보였지만, 결과적으로 생성형 AI까지 만들어 냈다는 것에 큰 의의를 두고 있다.** 프로젝트를 처음 진행할 때는 생각보다 쉬울 것으로 판단하였지만, 모델 형성, 클러스터 군집 등에서 막히는 모습을 보고 역시나 **데이터 분석은 예상치 못한 변수가 많다는 것을 느꼈다.**

**백은서** : 이번 프로젝트를 통해 **머신러닝에 대한 두려움을 깬 것이 가장 큰 수확이 아닐까 싶다.** 예측 모델을 수립하고 군집화하는 과정에서 주어진 데이터를 바탕으로 나만의 **논리를 세우는 연습**을 할 수 있었고, **대용량의 데이터를 다루어 보면서 Bigquery를 익힐 수 있었다.** 포트폴리오나 기획안이 아닌 ‘분석결과보고서’를 처음으로 작성해보며 구조와 흐름을 배우고, 또한 보다 **설득력 있는 글을 작성**하기 위해 사고하는 데이터 분석가의 자질을 기르게 된 좋은 기회였다고 생각한다.

**전혜빈** : 도메인 지식 부족으로 대출에 대한 핀다 앱을 직접 사용하며 필요한 정보를 찾아 전처리를 진행했다. 유저 데이터, 대출 관련 데이터 및 로그 데이터와 같은 큰 데이터셋을 처음 다뤄보았기 때문에 다양한 접근 방법을 시도했다. 이 과정에서 데이터 전처리를 위한 기준 설정 및 방법을 습득하게 되었고, **데이터의 불균형 문제를 해결하기 위해 가중치 조정과 언더샘플링을 적용해보며 LightGBM에 대해 공부할 수 있었다.** 군집화를 여러 가지 방법으로 시도해보았지만 쉽지않았다. 도메인 지식을 활용하여 군집화를 수행하게 되었지만 이 부분에서 아쉬움을 느꼈다. 향후 군집 고객에게 제공할 메시지 및 고객의 대출 신청 예측 모델과 관련된 자동화 분석서를 작성하기 위해 **생성형 AI를 활용하여 프롬프트를 개선하는 과정을 통해 많은 것을 배울 수 있었다.**



# 참고 자료

---

## LightGBM 공식문서

<https://lightgbm.readthedocs.io/en/stable/pythonapi/lightgbm.train.html>

## LightGBM 논문

<https://papers.nips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf>

### 3 Gradient-based One-Side Sampling

In this section, we propose a novel sampling method for GBDT that can achieve a good balance between reducing the number of data instances and keeping the accuracy for learned decision trees.

#### 3.1 Algorithm Description

In AdaBoost, the sample weight serves as a good indicator for the importance of data instances. However, in GBDT, there are no native sample weights, and thus the sampling methods proposed for AdaBoost cannot be directly applied. Fortunately, we notice that the gradient for each data instance in GBDT provides us with useful information for data sampling. That is, if an instance is associated with a small gradient, the training error for this instance is small and it is already well-trained. A straightforward idea is to discard those data instances with small gradients. However, the data distribution will be changed by doing so, which will hurt the accuracy of the learned model. To avoid this problem, we propose a new method called Gradient-based One-Side Sampling (GOSS).

GOSS keeps all the instances with large gradients and performs random sampling on the instances with small gradients. In order to compensate the influence to the data distribution, when computing the information gain, GOSS introduces a constant multiplier for the data instances with small gradients (see Alg. 2). Specifically, GOSS firstly sorts the data instances according to the absolute value of their gradients and selects the top  $a \times 100\%$  instances. Then it randomly samples  $b \times 100\%$  instances from the rest of the data. After that, GOSS amplifies the sampled data with small gradients by a constant  $\frac{1-a}{b}$  when calculating the information gain. By doing so, we put more focus on the under-trained instances without changing the original data distribution by much.



세상에 없던  
대출 비교 플랫폼

finda

Q & A