

# Credit score

EXCELLENT

961 - 999

GOOD

881 - 960

FAIR

721 - 880

POOR

561 - 720

개인 고객 대상,  
점수제 신용 평가 모델 개선을 통한  
리스크 관리 및 매출 증대

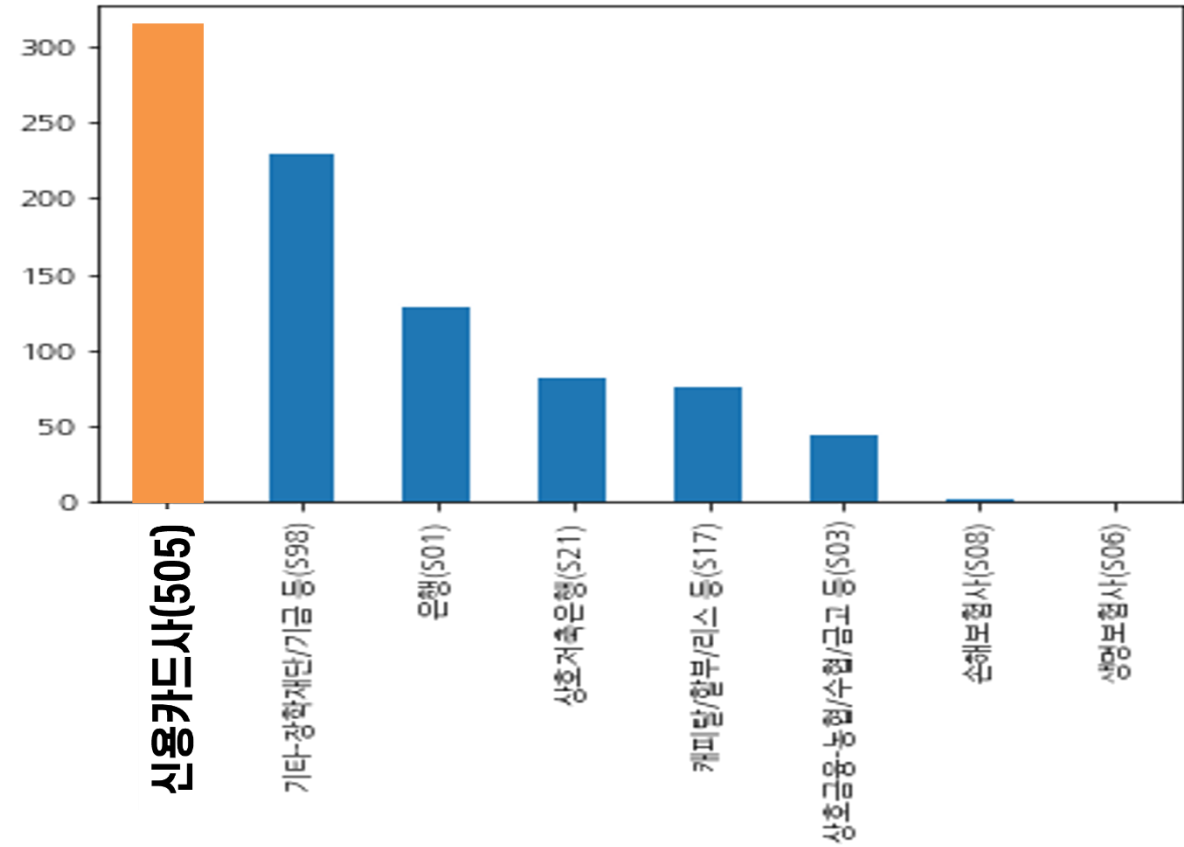
양진미

# Contents

- 1 추진배경
- 2 현상 및 개선기회
- 3 분석계획
- 4 분석결과
- 5 개선안 및 적용방안
- 6 Learned Lessons
- 7 참고

추진배경 다른 업권에 비해 연체율이 높은 신용카드사, 리스크 관리 필요

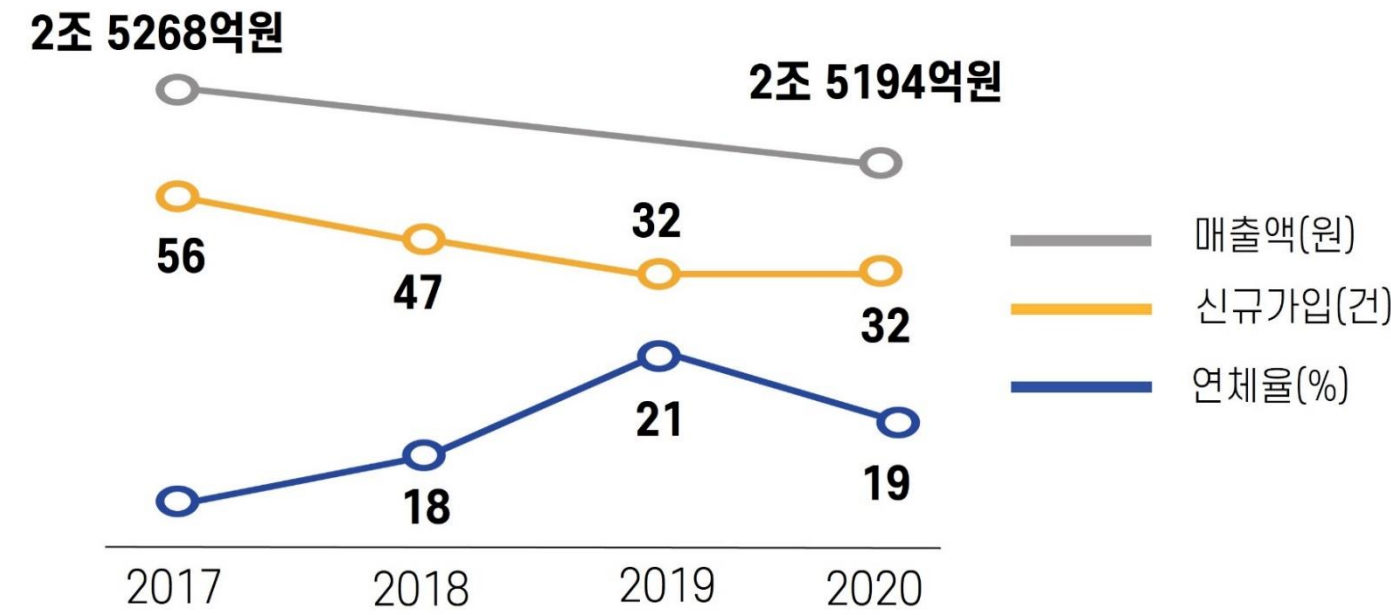
<업권별 연체 건수>



- 신용카드사의 연체 건수 가장 높음
- 다른 업권에 비해 리스크\* 관리 미흡

\* 리스크(risk) : 신용거래에서 발생하는 잠재적인 손실위험(e.g., 대출 및 신용카드 결제 연체)

<자사 당면 상황>



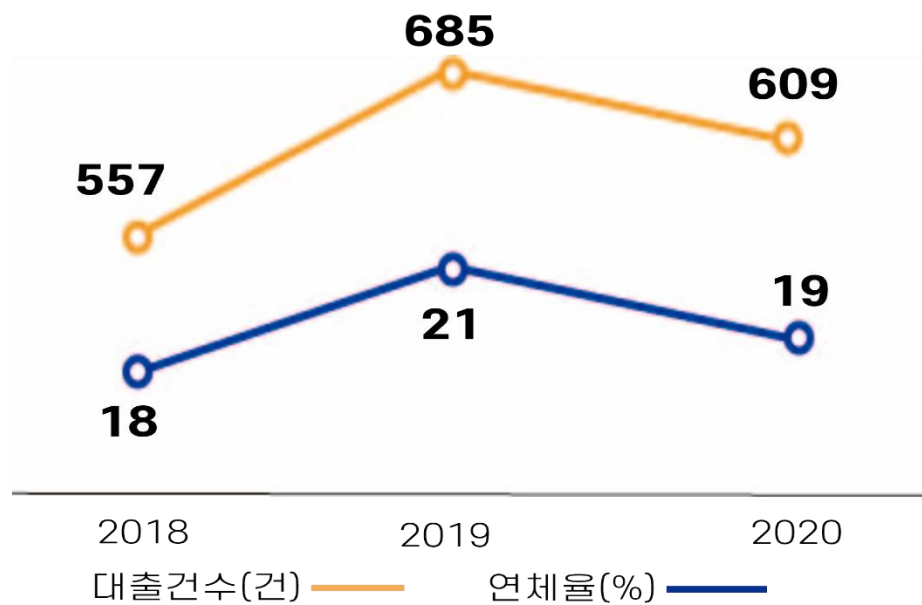
- 신규 가입 건수의 지속적 감소
- 증가하는 모습을 보이는 연체율
- 매출 감소 (▼ 0.3%)

매출 증대를 위한 신규 고객 유입과 리스크 관리의 필요성

## 현상파악

연체자의 증가와 높은 비율의 금융 소외계층을 관리하기 위해 세밀한 평가가 가능한 모델 개발이 필요

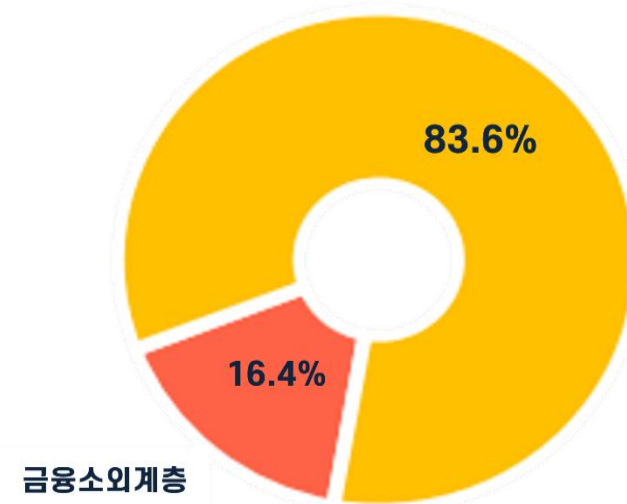
### ① 연체자 증가



대출건수 **약 8.5%**, 연체율 **약 1%** 증가

- 대출건수 증가에 따른 연체율 증가
  - 향후 지속적 연체율 증가 예상
  - 연체 가능성 점수 환산 모델 개발
- 연체자 리스크 관리

### ② 금융 소외 계층\* 유치 어려움



당사 고객 중  
금융 소외계층 비율 **16.4%**

- 평가 불가능한 금융 소외계층을 평가
- 신규 고객으로 유입
- 기존 고객에게 금융 기회를 제공함으로써  
당사 이익 창출

\* 금융 소외계층 : 신용평가지점 기준, 최근 2년 내 신용카드 사용 이력 및 3년 내 대출 경험이 없는 자(사회초년생·주부·고령층 등)

### ③ 세밀한 평가의 필요성

개인신용 평점, 어떻게 매기나		
구분	항목	가중치
차주특성	연령	50
	직업	100
	교육수준	100
	주거형태	100
	주택형태	50
	결혼상태	50
	수입 수준(가족)	100
	지출 수준	150
카드	개설 카드 수	50
대출	대출 건수	150
	대출 금액(합계)	150
연체	연체 건수	150
	연체 금액(합계)	250

등급	신용자 구분
1등급	우량 신용자
2등급	
3등급	
4등급	보통 신용자
5등급	
6등급	
7등급	저 신용자
8등급	
9등급	
10등급	

'올크레딧' 기준 신용등급평가표

자료:기존자사모델

- (기존) 부족한 평가 항목과 넓은 등급폭
- (개선) 세분화된 점수 구간 적용

# 개선기회

## <벤치마킹>

美 3대 신용조회회사, **FICO Score** 기반 세분화된 신용평가모형 운영



- 연체 가능성 수치화 → 신용평점 부여 (300~850점)
  - 통신료 정보, 공공요금 정보, SNS 데이터 등의 다양한 비금융정보 추가로 평가모형 예측력 향상
- 금융소외계층 중 70% 이상 금융서비스 접근성 개선

## 머신러닝 기반의 신용평가 모델 개발 선행연구

- Logistic Regression, Random Forest, Decision Tree, XGB, AdaB, SVM 등을 활용한 신용평가 모델 개발
- 기존 신용평가 모델보다 정교한 신용평가로 리스크 관리 효과 증대



## 1. 머신러닝 기반 모델 개발

기존 신용 평가 모델보다 정교해진 신용 평가 모델

## 2. 등급제 → 점수제 변환

세밀한 리스크 관리

## 3. 비금융 데이터 추가

(SNS 데이터, 소비패턴, 고정지출 이력 등)

금융 소외계층의 금융 서비스 접근성 개선



**매출액 증대**

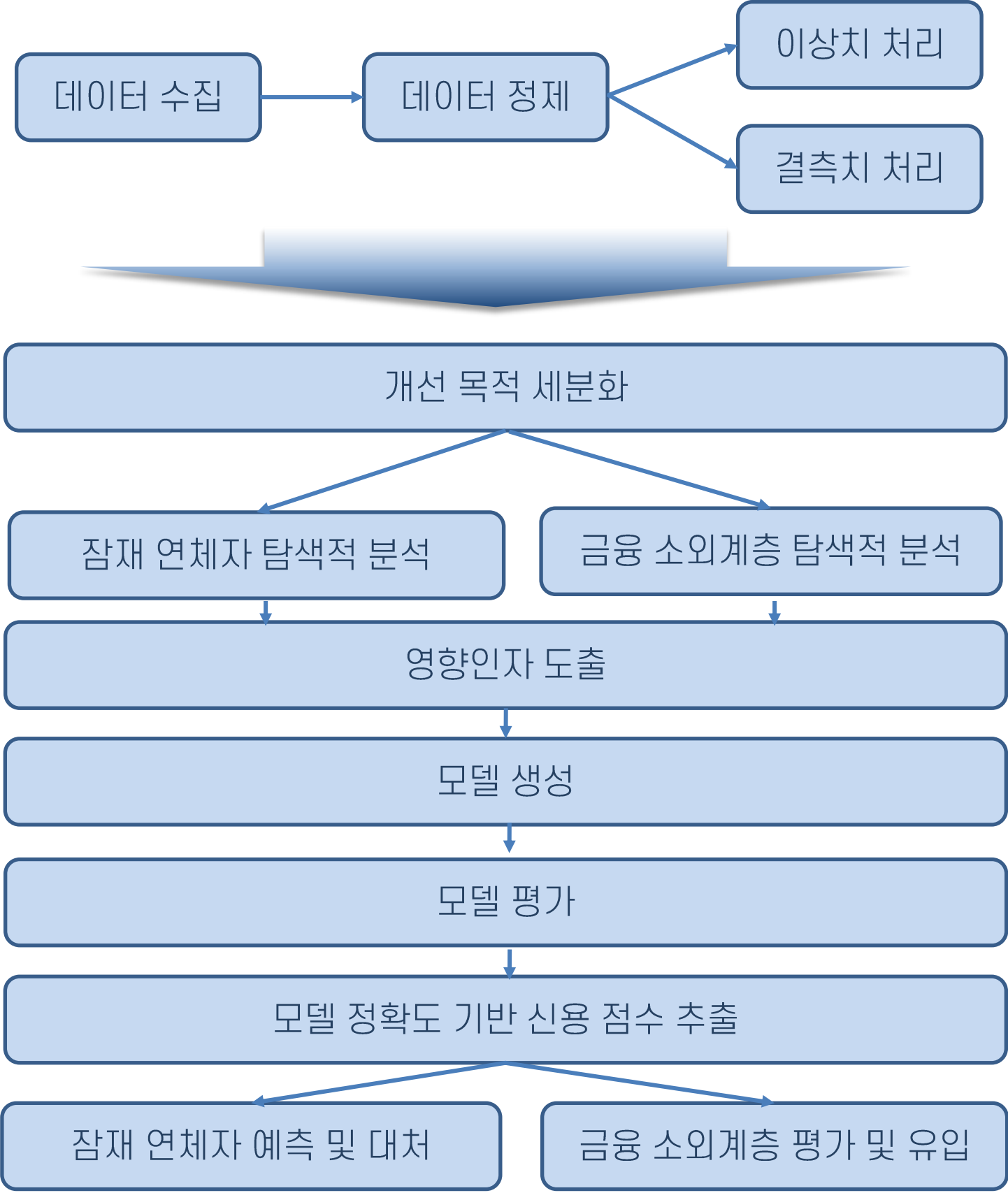
## <KPI>

측정 지표(KPI)	현수준	목표수준		
		2021년	2022년	2023년
매출액(원)	2조 5194억	2조 5268억	2조 5391억	2조 5510억
연체율(%)	19.33	18.00	16.50	15.00
금융 소외계층 신규 가입 수(건)	32	47	54	63

분석계획

분석 목적	분석 기법
데이터 시각화 및 분석	막대 그래프
	Box Plot
	Trend Chart
	히스토그램
	산점도
기술 통계	카이제곱 검정
	ANOVA
Vital few 도출	로지스틱 회귀분석
	의사결정나무
	랜덤포레스트
	그래디언트 부스팅
차주 정보에 대한 누락값 예측 모델 개발	로지스틱 회귀분석
	의사결정나무
	kNN
	그래디언트 부스팅
	SVC
신용점수 및 잠재 연체자 예측 모델 개발	나이브베이즈
	로지스틱 회귀분석
	의사결정나무
	랜덤포레스트
	그래디언트 부스팅
	인공신경망
	SVC
	나이브베이즈

분석절차





**분석결과 -1**      그래디언트 부스팅 모델을 적용한 결측치(무응답) 처리

- 결측치 : 차주 정보 1,984건 중 604건 (약 30%)
- 내용 : 개인 평균 수입(월) 항목 ‘무응답’ 표기

<변수 중요도 측정표>

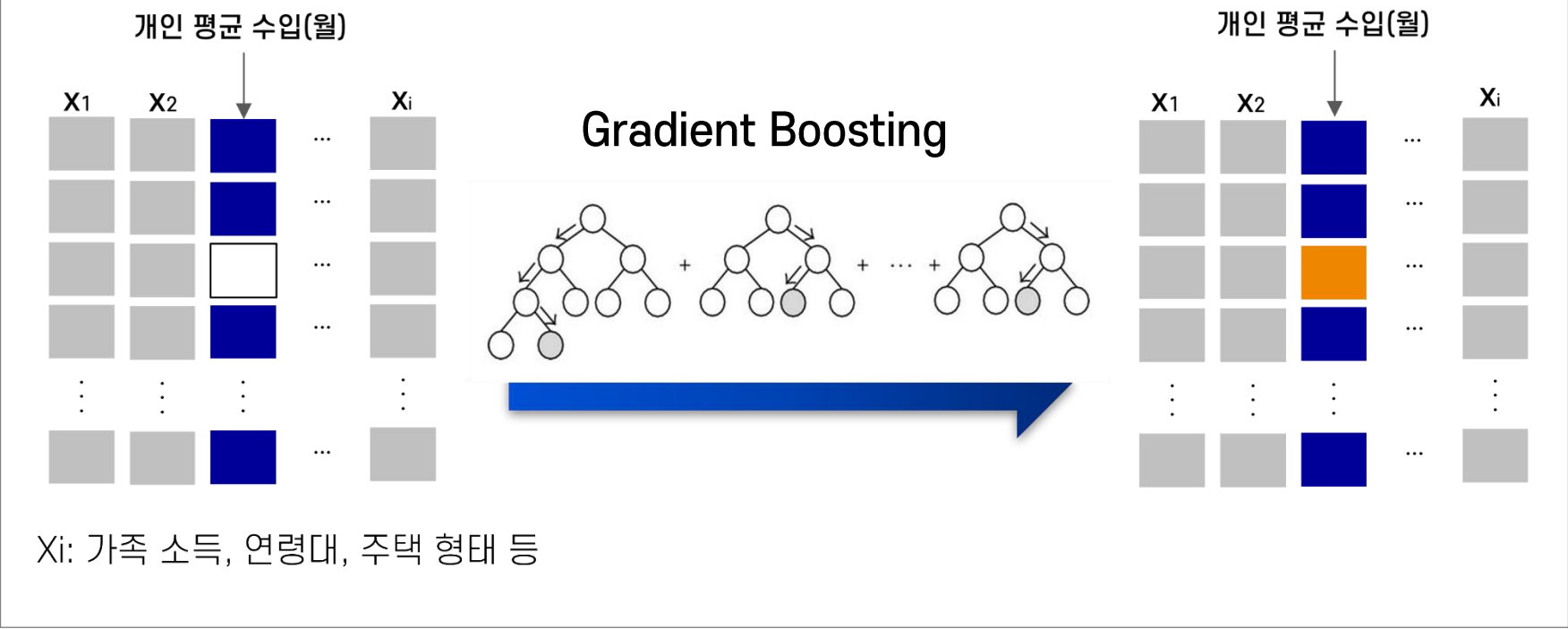
순위	로지스틱	의사 결정 나무	랜덤 포레스트	그래디언트 부스팅
1	주택형태	평균 대출 금액	평균 대출 금액	평균 대출 금액
2	교육수준	카드 보유 수	카드 보유 수	카드 보유 수
3	가족 평균 수입(월)	대출 수	대출 수	대출 수
4	<b>개인 평균 수입(월)</b>	연령대	연령대	연령대
5	주거형태	자녀 수 대비 가족 평균 수입(월)	자녀 수 대비 가족 평균 수입(월)	자녀 수 대비 가족 평균 수입(월)
6	연령대	<b>개인 평균 수입(월)</b>	가족 평균 수입(월)	가족 평균 수입(월)
7	자녀 수	교육수준	<b>개인 평균 수입(월)</b>	<b>개인 평균 수입(월)</b>
...	...	...	...	...

- 모델별 변수 중요도 파악 결과, ‘개인 평균 수입(월)’을 주요 영향 인자로 확인 ➡ 결측치 대체를 위해 기존 차주 정보로 모델 학습 진행

<모델별 정확도>

모델	로지스틱	의사 결정 나무	랜덤 포레스트	그래디언트 부스팅
정확도(%)	73.2	79.7	76.4	<b>85.1</b>

**<결측치 처리 프로세스>**



무응답 차주 수입을 예측 (정확도 85.1%)

분석결과 -2    각 영역별 변수 중요도

<당사 차주 정보>

차주 정보	연체 정보
성별	연체등록금액
연령	연체 기간
직업	연체 이유 개수
교육수준	
주거형태	대출 정보
주택형태	대출금액
결혼상태	평균 대출 금액
가족 평균 수입(월)	
개인 평균 수입(월)	카드 정보
개인 평균 지출(월)	카드 개수
자녀유무	보유 카드 업권 수
자녀 수	보유 카드 유형 수
개인사업자 여부	
연령별 결혼여부	
소득/가족구성원	
20대 결혼여부	
30대 결혼여부	

- 기존 평가 모델 변수
- 생성 모델에 사용하기 위해 수집된 차주 정보
- 파생 변수

<잠재 연체자 파악을 위한 중요 인자를 도출>

1. 비금융 데이터

순위	로지스틱	의사 결정 나무	랜덤 포레스트	그래디언트 부스팅
1	주택형태	평균 대출 금액	평균 대출 금액	평균 대출 금액
2	교육수준	카드 보유 수	카드 보유 수	카드 보유 수
3	가족 평균 수입(월)	대출 수	대출 수	대출 수
4	개인 평균 수입(월)	연령대	연령대	연령대
5	주거형태	자녀 수 대비 가족 평균 수입(월)	자녀 수 대비 가족 평균 수입(월)	자녀 수 대비 가족 평균 수입(월)
6	연령대	개인 평균 수입(월)	가족 평균 수입(월)	가족 평균 수입(월)
7	자녀 수	교육수준	개인 평균 수입(월)	개인 평균 수입(월)
정확도	0.7936	0.7147	0.7449	0.7919

2. 카드 데이터

순위	로지스틱	의사 결정 나무	랜덤 포레스트	그래디언트 부스팅
1	카드 보유 수	보유 카드 유형 수	카드 보유 수	카드 보유 수
2	보유 카드 유형 수	카드 보유 수	보유 카드 업권 수	보유 카드 유형 수
3	보유 카드 업권 수	보유 카드 업권 수	보유 카드 유형 수	보유 카드 업권 수
정확도	0.7936	0.7902	0.7885	0.7885

3. 대출 데이터

순위	로지스틱	의사 결정 나무	랜덤 포레스트	그래디언트 부스팅
1	대출 상품 수	평균 대출액	평균 대출액	평균 대출액
2	대출 건수	대출 건수	대출 건수	대출 건수
3	평균 대출액	대출 상품 수	대출 상품 수	대출 상품 수
정확도	0.7936	0.7449	0.76	0.7869


<최종 사용 파생변수>

- 자녀수/가족수입
- 대출 건수
- 대출 상품수
- 카드 업권수
- 카드 유형수




분석결과 -3    모델 생성 방법

기존 모델 변수	Feature Selection
연령	성별
직업	연령대
교육수준	직업
주거형태	최종학력
주택형태	주거형태
결혼상태	주택형태
가족 평균 수입(월)	결혼상태
개인 평균 지출(월)	가족 평균 수입(월)
연체등록금액	개인 평균 수입(월)
대출금액	개인 평균 지출(월)
	자녀 유무
	개인사업자 여부
	평균 대출 금액
	카드 개수

- 
1. Dummy 변수 생성  
2. 7가지 모델 생성(의사결정나무, 그래디언트 부스팅, 랜덤포레스트, 로지스틱 회귀, 다항분포나이브베이즈, 베르누이나이브베이즈, KNN)  
3. 정확도 비교

**비금융 : 랜덤 포레스트**  
**금융+비금융 : 의사 결정 나무**

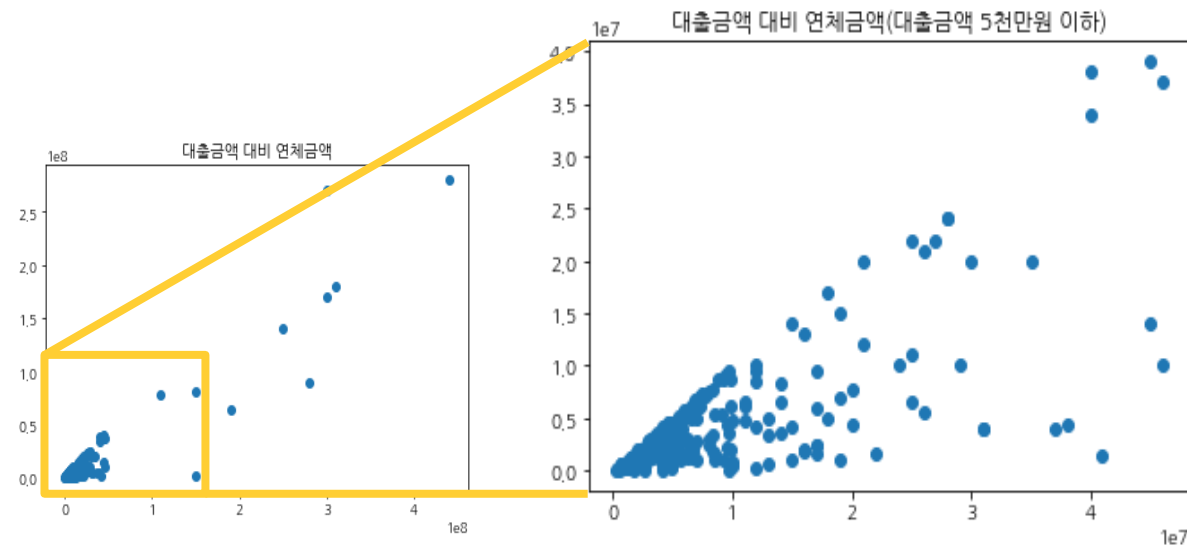


연체 확률 예측

**신용 점수 = (1 – 연체 확률 ) \* 1000**

## 분석결과 -4 리스크 대응 방안 도출 및 모델 개선 결과

<리스크 대응 방안 도출>

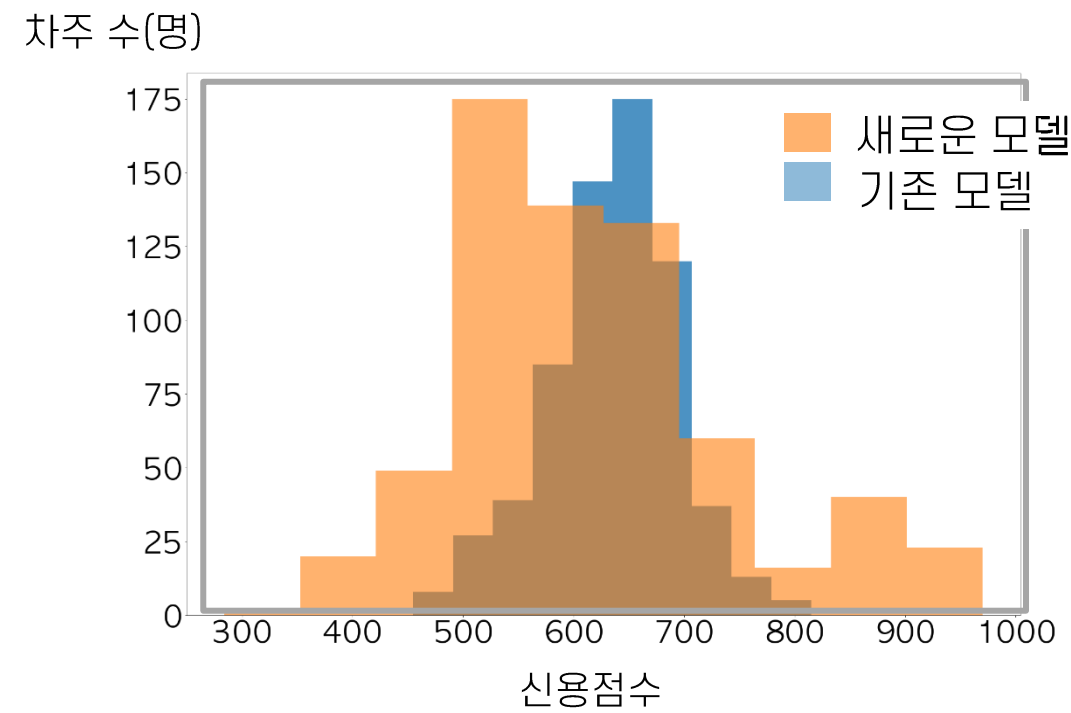


- 대출금액과 연체금액이 선형관계를 보임
- 연체 방지를 위한 대출 금액 한도 제한 필요

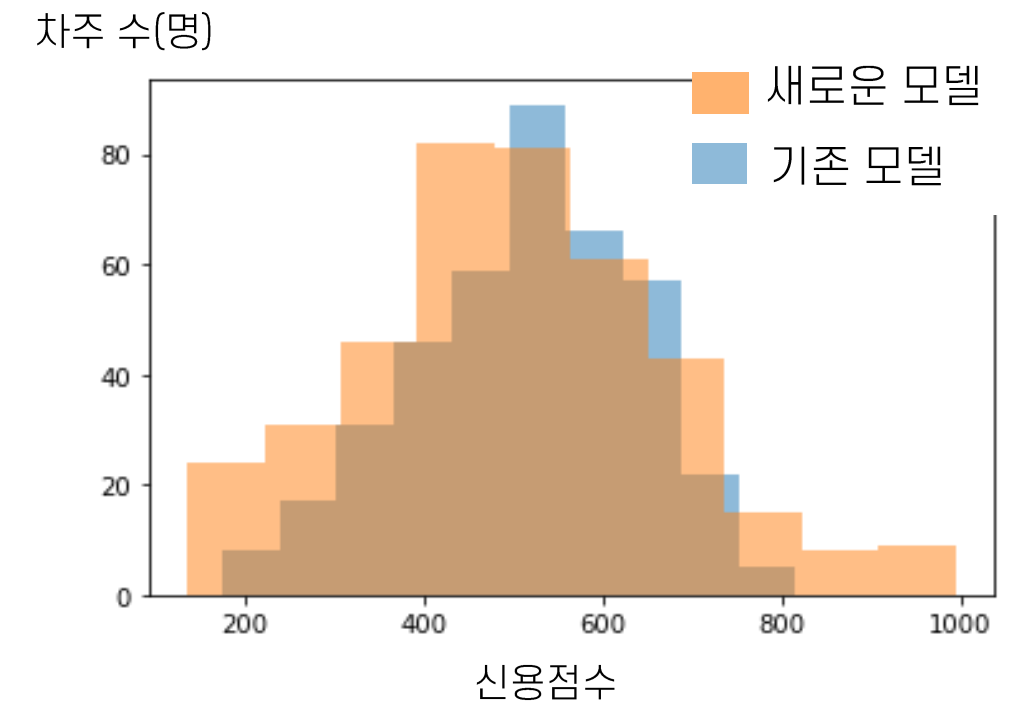


- 위험 점수가 높은 고객 : 높은 이율 적용,  
대출 금액 한도 하향 조정

<모델 개선 결과>  
-비금융 데이터



<모델 개선 결과>  
-비금융/금융 데이터



- ① 잠재적 연체자 추가 탐지 ➡ 사전에 대출한도 및 이자율 조정
- ② 금융소외계층의 금융접근성 향상 ➡ 신규 고객 유치 확대

## 개선안 매출 증대를 위한 실시간 반영 모델 개발과 리스크 관리

### <연체율 예측 모델 개발>

- 기존 데이터를 활용 연체 고객 특성을 파악
- 잠재 연체자를 예측하는 연체율 예측 모델 개발

### <비금융 데이터를 활용한 금융 소외계층 유치>

- 비금융 데이터(성별, 연령대, 자녀 유무, 자녀 수, 개인사업자여부 등) 활용을 통한 금융 소외계층 평가 모델 개발
  - 금융 소외 계층 평가를 통해 신규 고객 유치 가능
  - 금융 소외 계층에게 폭 넓은 금융 기회 제공

### < 등급제에서 점수제로의 변환 및 평가항목 추가 >

- 세분화된 점수제 신용 평가 전환 및 평가 항목 추가로 신규 고객 유치
  - 부채 상환 능력이 있지만 기존 폭넓은 등급제 신용 평가와 부족한 평가 항목으로 인해 불이익을 받는 고객 구제 및 유치 가능

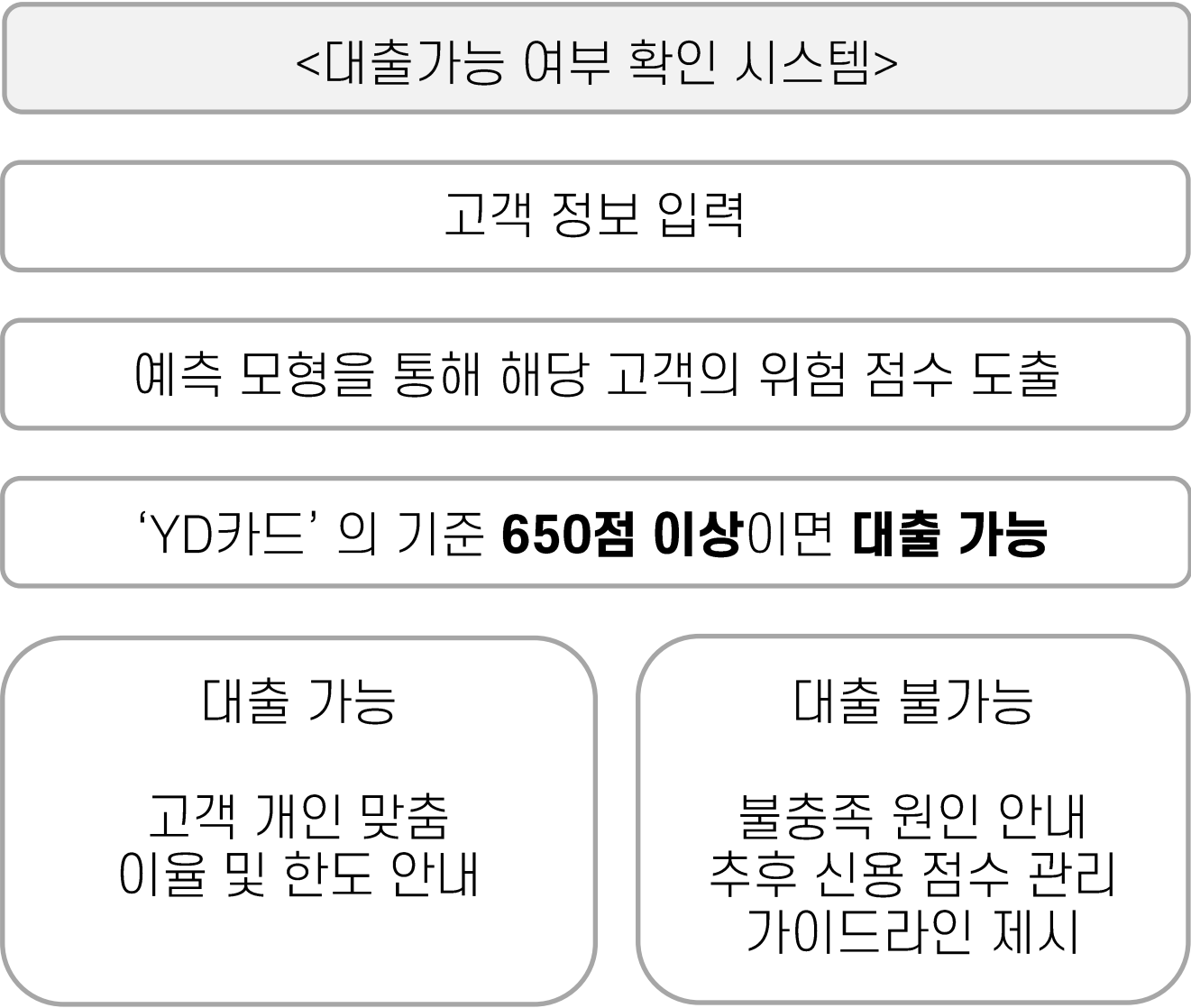


예측된 위험 점수에 따라 **이율과 대출 한도를 차등 적용**하여 리스크 관리

영향인자 변동 시 자동 update가 가능한 **실시간 반영 모델**

웹페이지 개발/운영 : 고객에게 **설명 가능한** 가이드라인 제시 가능

**개선안**    대출 가능 여부 확인 가능한 웹 서비스 제공



- 고객은 자신의 신용 점수를 바탕으로 이율, 한도 등을 확인하고, 필요시 패널티를 적용하는 등의 방법으로 금융 접근 기회 확보
- 당사는 보다 신뢰성 있는 모델을 통한 리스크 관리가 가능해지며 최종적으로 매출액 증대의 목표 실현 가능

## 발전방향

### <다양한 비금융 데이터 추가>

SNS 데이터, 공공요금 결제정보, 고정지출내역, 온/오프라인 소비패턴, 보험이력, 건강 정보 등의 비금융 데이터를 추가하여 향상된 정확도의 모델 생성 가능

### <가중치 부여 방법>

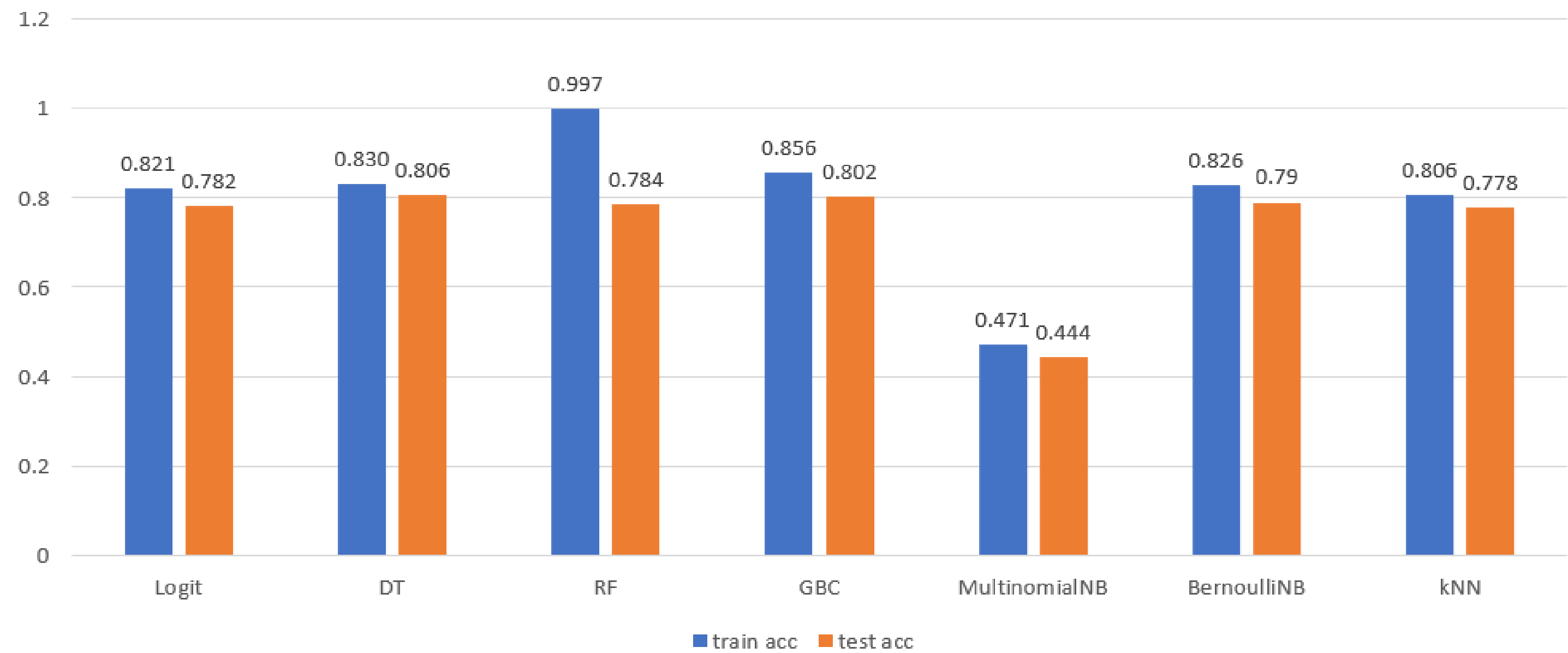
변수와 범주마다 가중치를 부여하여 모델링 하였으나 모델이 복잡해지고 해석하기 힘든 문제점 발생

→ 설명 가능한 모델 생성을 위해 변수에만 가중치 부여하는 비교적 간단한 방향으로 수정

➡ 향후, 변수와 변수를 포함하는 범주에도 가중치를 부여하는 방법으로 정확도가 향상된 모델 생성 가능

참고

잠재 연체자 예측 모델 정확도 비교





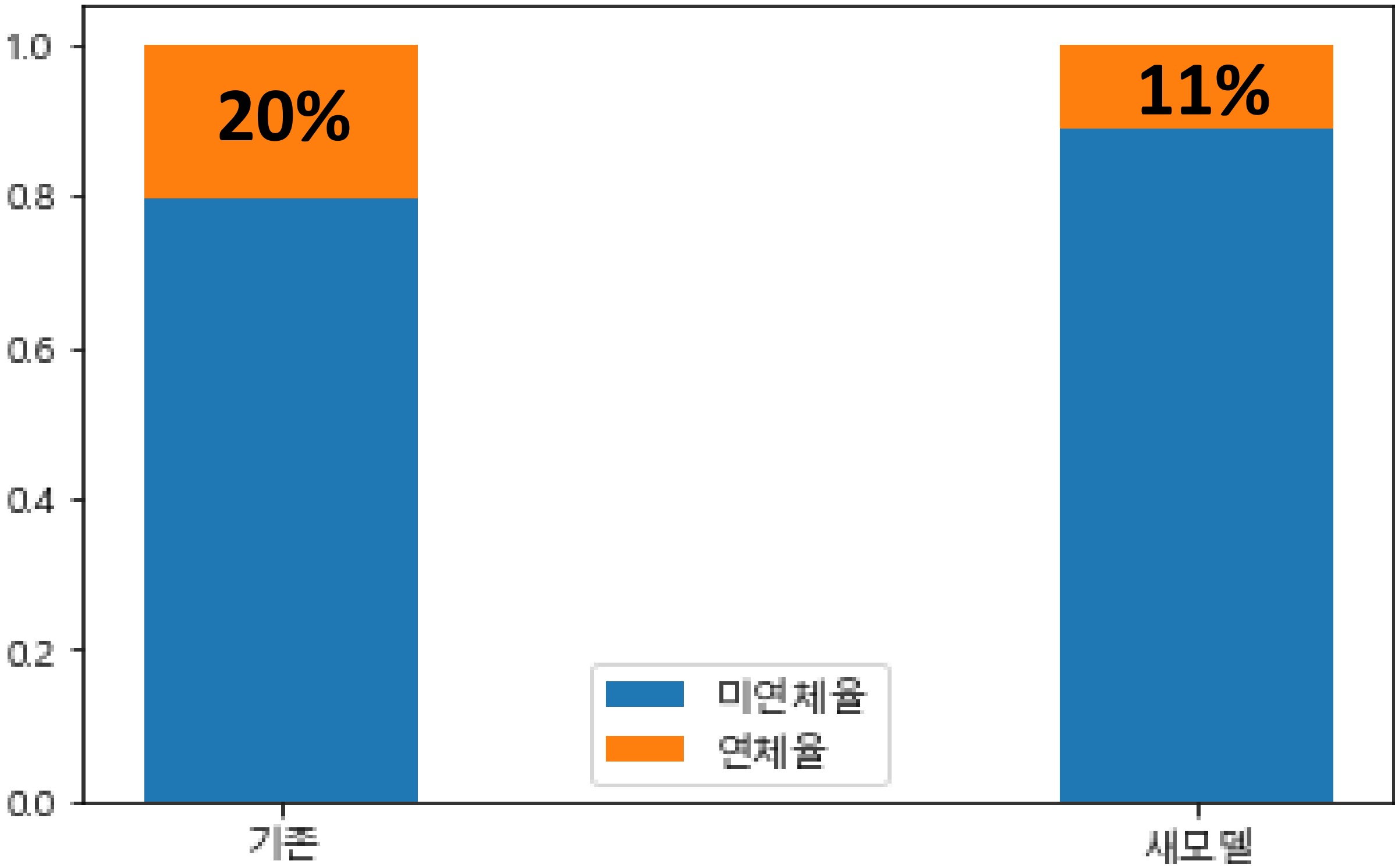
참고

overdue	chaju_id	gender	agegroup	job	...	card_cnt	20_married	30_married	Prob 0	Prob 1	y_pred
1	C-1001	남성	30대	기능/노무직	...	4	0	0	0.672	0.328	0
0	C-1002	여성	60대	전업주부	...	6	0	0	0.984	0.016	0
0	C-1003	남성	40대	판매/서비스직	...	2	0	0	0.896	0.104	0
...	...	...	...	...	...	...	...	...	...	...	...
0	C-2998	여성	50대	기능/노무직	...	5	0	1	0.997	0.003	0
0	C-2999	남성	40대	기능/노무직	...	2	0	1	0.862	0.138	0
0	C-3000	여성	50대	전업주부	...	4	0	1	0.797	0.203	0

Leaf Label	0	1
card_cnt>=1&avg_loan_amt>=405000&avg_loan_amt<=14750000&card_cnt>=3	0.9464	0.0536
card_cnt>=1&avg_loan_amt<405000&card_cnt<3&gender(남성)&agegroup(30대, 40대, 50대)&job(행정관리/전문직, 판매/서비스직, 기능/노무직)&edu(대학교 졸업, 전문대학 졸업, 대학원 졸업)	0.5525	0.4475
card_cnt>=1&avg_loan_amt<405000&card_cnt<3&gender(남성)&agegroup(30대, 40대, 50대)&job(행정관리/전문직, 판매/서비스직, 기능/노무직)&edu(고등학교 졸업)	0.9791	0.0209
card_cnt>=1&avg_loan_amt>=405000&avg_loan_amt<=14750000&card_cnt>=3	0.9464	0.0536
card_cnt>=1&avg_loan_amt<405000&card_cnt<3&gender(남성)&agegroup(30대, 40대, 50대)&job(행정관리/전문직, 판매/서비스직, 기능/노무직)&edu(대학교 졸업, 전문대학 졸업, 대학원 졸업)	0.5525	0.4475
card_cnt>=1&avg_loan_amt<405000&card_cnt<3&gender(남성)&agegroup(30대, 40대, 50대)&job(행정관리/전문직, 판매/서비스직, 기능/노무직)&edu(고등학교 졸업)	0.9791	0.0209

-> 카드 소유 개수 1-2개이며 평균 대출금액이 40만 5천원 이하, 남성이고 3-50대, 행정관리/전문직, 판매/서비스직, 기능/노무직이면서 고등학교 졸업자이면 97.91%로 미연체자

참고



**감사합니다.**