

경제변수를 활용한 초개인화 서비스 제안

제3회 대구 빅데이터 분석 경진대회

TEAM GoDart

INDEX

1

기획배경 및 개요

1. 기획배경
2. 개요

3

분석결과

1. 분석결과

2

분석방법

1. 전체 프로세스
2. 전처리 방법
3. 모델링

4

활용방안

1. 초개인화 금융상품 추천서비스
2. 상품 및 정책 제안 서비스

1

기획배경 및 개요

기획배경

금융생활에 있어 거시경제의 영향력

자본시장 고도화, 기술의 성장으로 인해 세계경제 공동화 현상 심화

- ✓ 경기연구원, 황상연, '경기도 단기지역경제전망 모형 구축에 관한 연구(2010)'
: 금리, 환율, 유가 등 대내외 경제 충격이 다양한 지역경제변수에 영향을 미침을 보임
- ✓ KIET(산업연구원), '대내외 경제변수의 지역경제 영향 및 파급경로 분석(2017.12)'
: 내외 경제변수의 지역경제 영향 및 파급경로를 분석하며 경제의 연결성이 강화되는 상황에서 지역경제의 안정적인 운용을 위해서는 대내외 환경변화를 주시

금융상품 활용 증가

장기간 지속되어온 저성장, 저금리 환경(한국은행, 통계청)



코로나19 이후 자산가격 상승에 따른 금융상품 이용 증가

1

기획배경 및 개요

기획배경

“

이처럼 금융상품에 대한 소비자들의 관심이 증가하는 상황에서

개별 소비자의 욕구를 충족시키기 위해서는 **실물 경제상황 변화**에 따른

적절한 금융상품 추천 서비스가 필요”

금융상품을 찾고 선택하는데 있어 소비자의 기회비용을 줄여줄 수 있지만,
소비자 특성에 따른 분류단계에서 적절한 고민이 없다면 상품추천서비스는 오히려 소비자에게 혼란만 가져옴

소비자 맞춤 정교한 상품추천을 위해서는 고객의 특성을 잘 분류하는 것이 중요

1

기획배경 및 개요

개요



초개인화 서비스 제공 모델링

초개인화된 금융상품과 서비스를 추천 및 개발할 수 있도록
경제변수 변화에 따른 금융시장 참여자의 행동 변화를 예측, 분석하는 모델 개발



지역 경제 이바지

대구시 대내외 거시경제 변수와 대구은행 데이터를 통해 개인별 맞춤 서비스를 제공한다면
금융 서비스의 질적 수준을 높여 대구시민의 금융활용 능력제고에도 도움이 될 것임

2

분석방법

전체 프로세스

1 변수선택 및 데이터 전처리

- 변수선택
- 시계열 칼럼 생성

2 고객 별 민감도 레이블링

- 경제변수와 고객 칼럼 간 상관관계 파악
- 민감한 고객 기준 선택 (Elbow method와 유사한 방식)

3 고객 분류 모델링

- 오버샘플링
- 다양한 분류 모델 사용

4 성과 측정 및 활용방안 제안

- 성과측정
- 활용방안 제안

2

분석방법

활용데이터-변수선택

분류	데이터명	데이터기간	출처	형식
제공데이터	대구은행 고객데이터	2019.01~2021.04	대구은행	CSV
공공데이터	지역 대외변수	2018.01~2021.05	원/달러 환율	
	원/위안화 환율			
	국고채(3년)			
	회사채(3년, AA-)			
	대구시 제조업, 비제조업 BSI			
	지역 대내변수		고용률, 실업률	
	대구 소비자물가지수			
			한국은행	
			통계청	

2

분석방법

데이터 전처리-시계열 칼럼 생성

전체 칼럼 중 '삼개월', '전월' 데이터가 존재하는 칼럼 선택

삼개월	전월	현월
삼개월급여이체실적금액	전월급여이체실적금액	급여이체실적금액
삼개월신용카드사용금액	전월신용카드사용금액	신용카드사용금액
삼개월체크카드금액	전월체크카드금액	체크카드거래금액
삼개월수신평균잔액	전월수신평균잔액	수신잔액
삼개월현금서비스금액	전월현금서비스금액	현금서비스이용금액
삼개월대출월평균잔액	전월대출월평균잔액	가계자금대출잔액+주택담보대출잔액

6종류 X 3시점 = 총 18개의 시계열 칼럼 생성
그 후 생성된 칼럼들에 대하여 음수 값은 0으로 치환

2

분석방법

데이터 전처리-시계열 칼럼 생성

데이터 기준 시점이 12월일 때,

‘대출-2’의 경우 2달 전인 10월, ‘대출-1’의 경우 11월, ‘대출-0’의 경우 10월의 월 대출 데이터로 각각 정의

	대출-2	대출-1	대출-0	신용-2	신용-1	신용-0
count	5.000000e+05	5.000000e+05	5.000000e+05	5.000000e+05	5.000000e+05	5.000000e+05
mean	3.183177e+07	2.261979e+07	1.427833e+07	3.470278e+05	3.240125e+05	3.373513e+05
std	3.754073e+08	1.977700e+08	8.536451e+07	1.614006e+06	1.403046e+06	1.428156e+06
min	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
25%	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
50%	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
75%	0.000000e+00	0.000000e+00	0.000000e+00	3.000000e+05	3.000000e+05	3.000000e+05
max	7.660000e+10	3.830000e+10	1.500000e+10	5.940000e+08	2.000000e+08	2.000000e+08

2

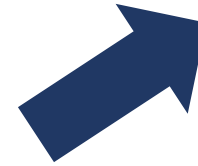
분석방법

민감도 레이블링-상관관계 파악

그랜저 인과검정을 통해 경제변수와
고객의 행동 패턴 간의 인과성 검증 시도



경제변수의 월간 변화와 고객 행동 변수(칼럼)의
월간 변화 평균 간의 상관관계 탐색을 통한
민감도 레이블링 시도



불연속적인 9개의 시점만으로
시계열 분석 부적합

“ 상관관계를 통해 레이블링 시도 시 경제변수와 고객 행동 변수(칼럼)간의 방향성 파악 용이 ”

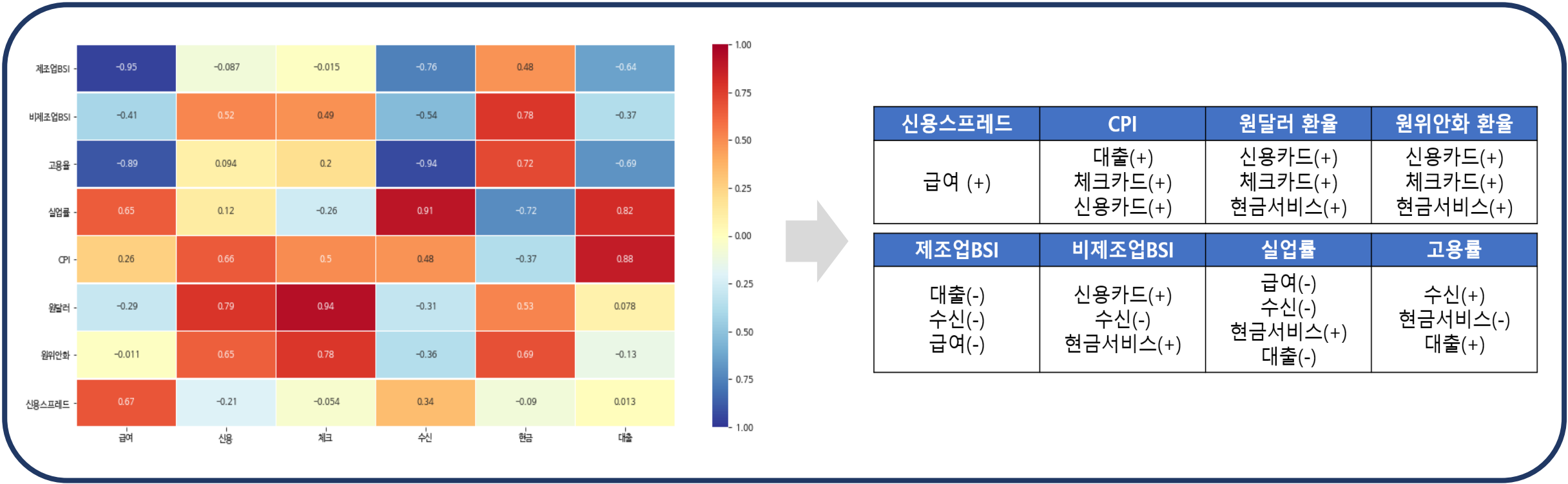
2

분석방법

민감도 레이블링-상관관계 파악

✓ 경제 변수의 월간 변화와 고객 행동 변수의 월간 변화 평균 간의 상관관계를 파악

Ex) CPI와 신용카드 사용량의 월별 변화 간의 높은 양의 상관관계가 존재
 -> CPI가 증가함에 따라 신용카드 사용량이 높게 증가한 사람들이 CPI에 민감하다고 판단



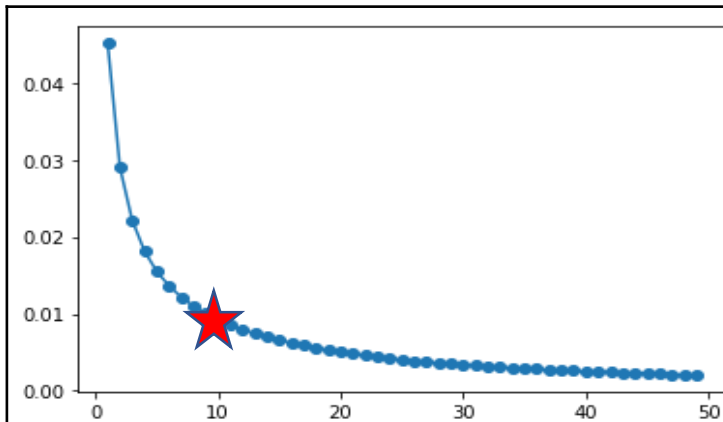
2

분석방법

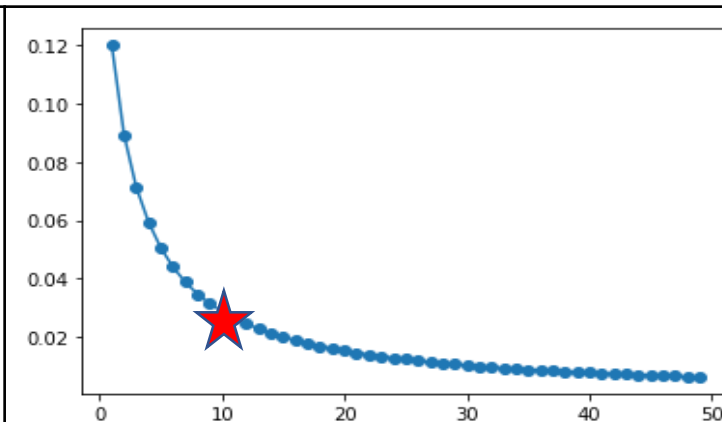
민감도 레이블링-기준 선택

상위 변화율 n%의 집단의 변화율 평균을 그래프로 나타내어 **민감도의 걱정 기준**을 선택함.

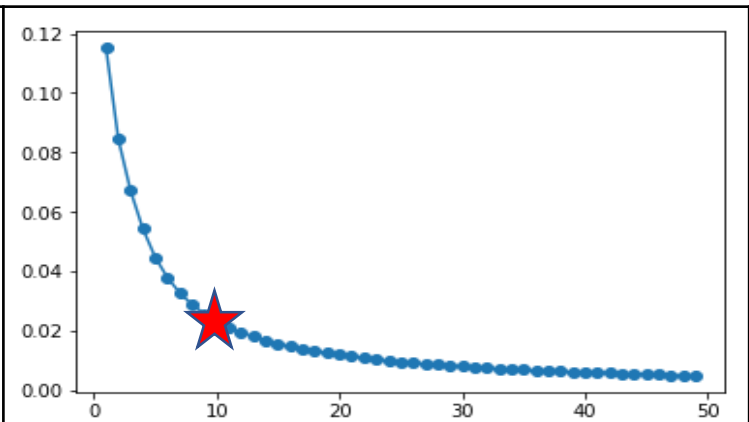
1. '고용률'에 민감한 집단을 레이블링 하기위해 지난 달 대비 고용률의 변화량을 구함
2. 수신, 현금서비스, 대출 금액의 지난 달 대비 변화량을 구하고, 세 변수의 단위 차이를 고려하여 Min-Max Normalization을 함
3. 세 변수의 변화량 평균을 개인 별로 구하고 상위 n%씩 끊어서 민감한 집단의 변화량 평균을 나타냄



환율



CPI

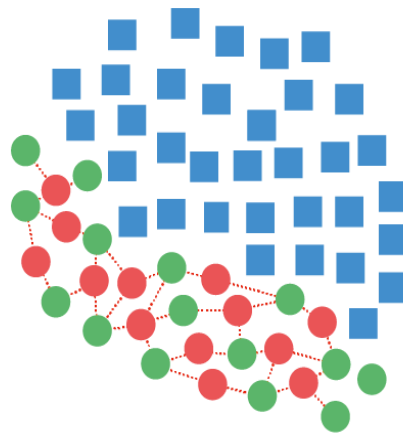


제조업

SMOTE (synthetic minority oversampling technique): 데이터의 개수가 적은 클래스의 표본을 가져온 뒤 KNN기법을 통해 임의의 값을 추가하여 over sampling하는 기법



Original Dataset



Generating Samples



Resampled Dataset

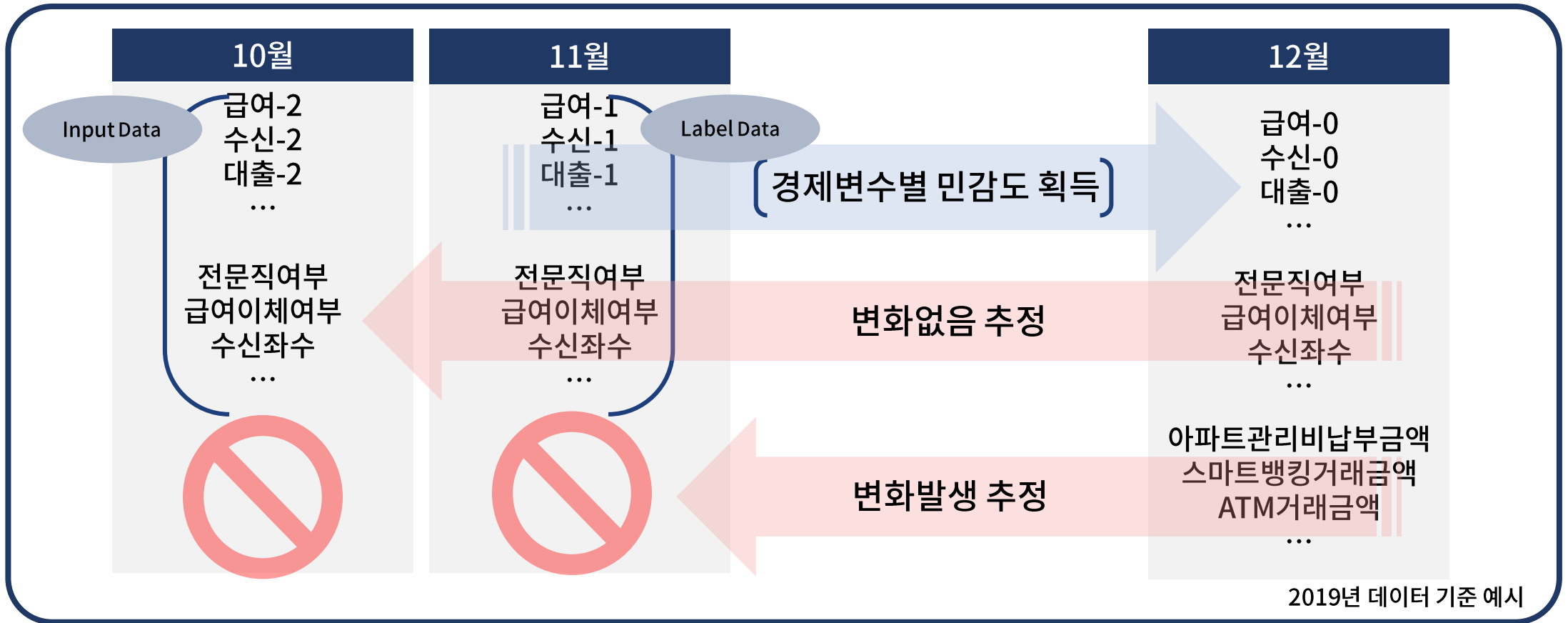
2

분석방법

모델링-문제 정의

문제정의

기준시점 1달 전까지의 데이터로 기준시점까지 1달간의 경제변수별 민감도를 분류하는 문제



“시계열 데이터를 활용하므로, 시간에 따른 데이터 축적 및 미래참조 문제 고려”

2019, 2020년을 Train Set, 2021년을 Test Set으로 활용

2

분석방법

모델링-모델 선택

1 로지스틱 회귀분석

- 간단한 형태
- 타 모델과의 벤치마크로 활용

3 랜덤포레스트

- 트리 기반 앙상블(Bagging)
- 앙상블 방법에 따른 성능 차이 관측

5 딥러닝

- 신경망 기반 딥러닝
- 각기 다른 데이터의 특성 보존

2 XGBoost

- 트리 기반 앙상블(Gradient Boosting)
- 정형 데이터 분야에서 높은 성능

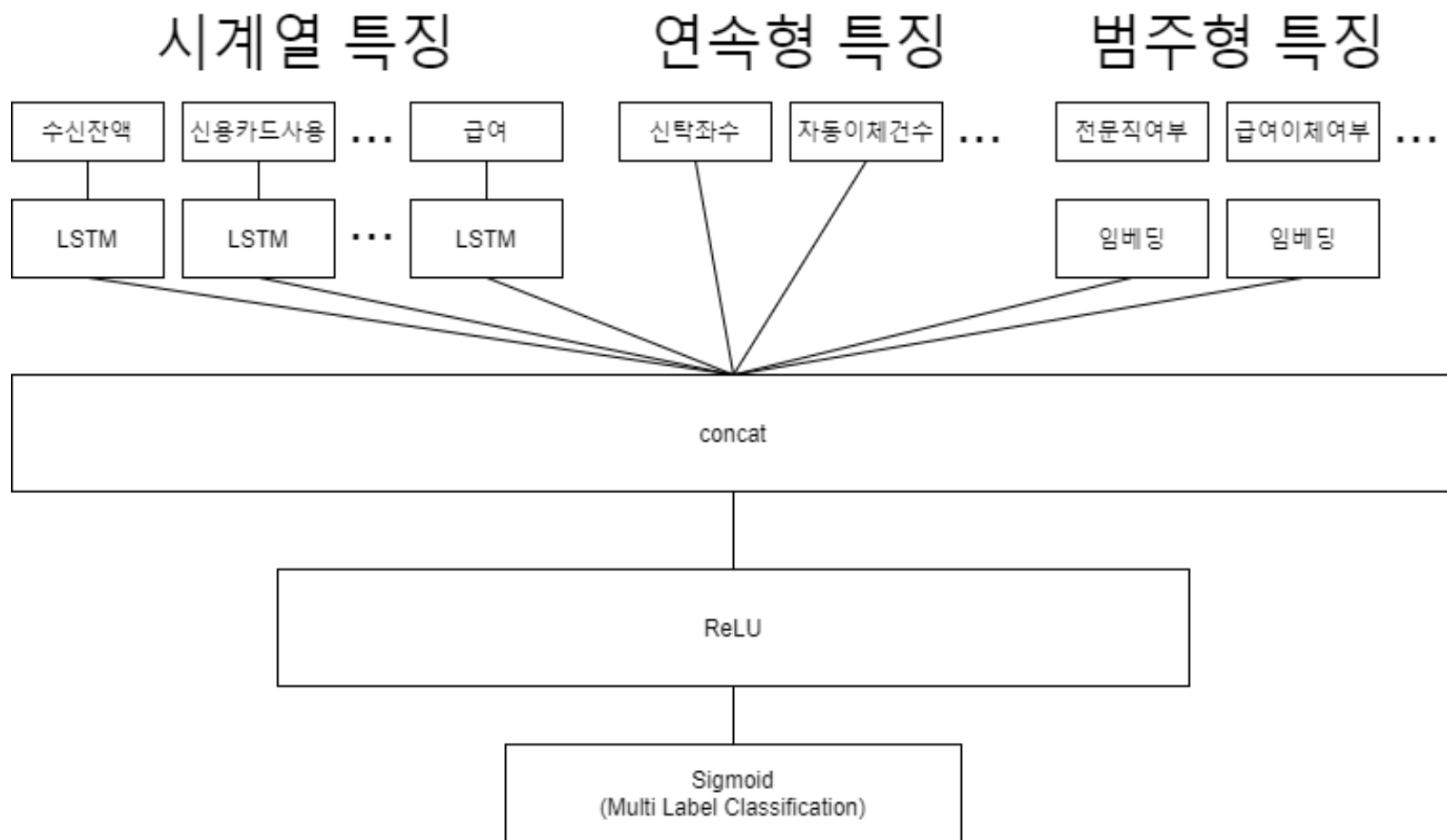
4 LightGBM

- 트리 기반 앙상블(Gradient Boosting)
- XGBoost 대비 빠른 연산 속도

2

분석방법

모델링-딥러닝 모델 구조



입력 데이터를 3가지로 처리

- 시계열
단방향 LSTM 사용

- 연속형
아웃라이어 winsorize

- 범주형
원 핫 인코딩

모든 계층의 결과를 이어 붙임

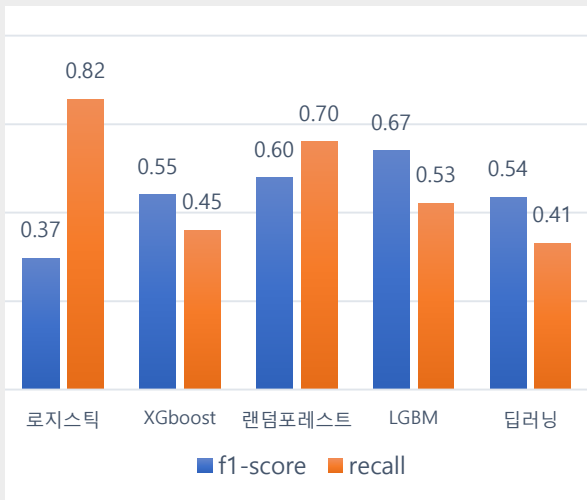
Sigmoid 활성화 함수를 통한
Multi Label Classification

3

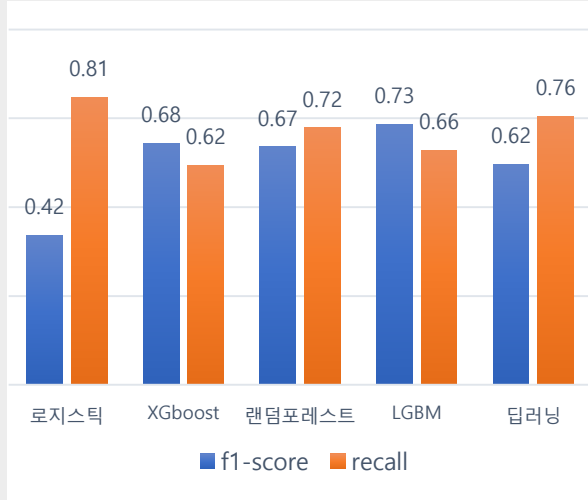
분석결과

성능 비교

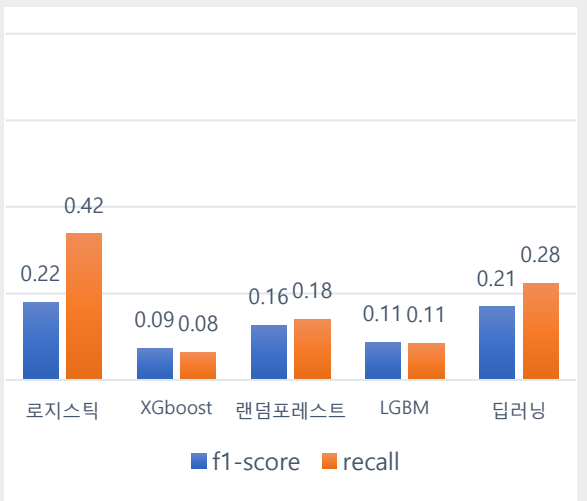
< CPI >



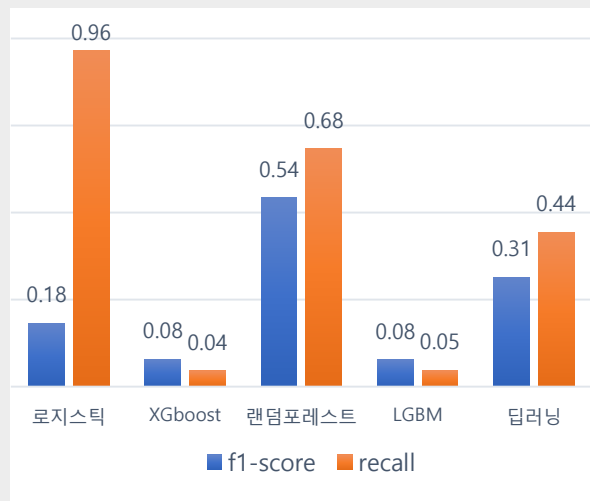
< 고용 >



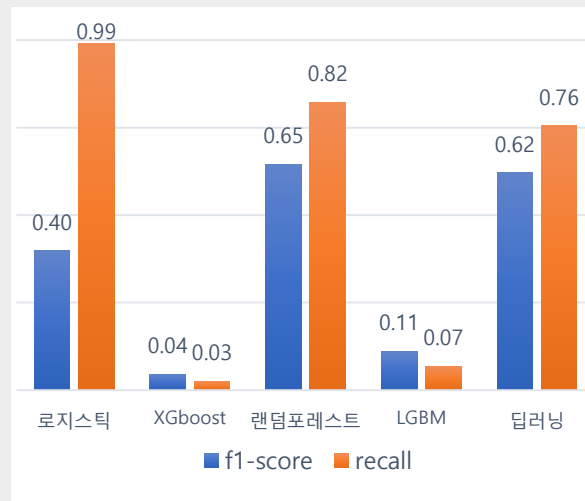
< 제조업 >



< 비제조업 >



< 환율 >



현재로서는 **랜덤 포레스트** 모델이 종합적으로 좋은 성능
-> 최종 모델로 랜덤 포레스트 선택

비제조업 및 환율의 경우 Gradient Boosting
방식의 모델 성능 저조
-> 앙상블 적용 시 부스팅 제외 필요

딥러닝 모델은 열악한 데이터에도 불구하고 좋은 결과
-> 실 데이터 활용 시 성능 개선 기대

4

활용방안

초개인화 금융상품 추천 서비스

| 고객별 맞춤 금융상품 추천 서비스 예시 Overview

STEP 1

모델에 의한 경제변수 별 민감고객 분류

STEP 2

경제변수 별 민감고객들의 다음 시기 금융행동 변화 예측

STEP 3

추천 상품군 정의

STEP 4

민감고객군 세분류 작업 진행

STEP 5

고객별 맞춤 상품 추천

4 활용방안

초개인화 금융상품 추천 서비스

고객별 맞춤 금융상품 추천 서비스 예시

STEP 1

모델에 의한 경제변수 별 민감고객 분류

랜덤포레스트	CPI	고용률	제조업BSI	비제조업BSI	환율
1	67007	53919	60626	75869	77439
0	432993	446081	439374	424131	422561

	전문직여부	세분화고수 신고객여부	세분화고소 득고객여부	실적기준고 객우대구분 코드	방카슈랑스 보유좌수	수익증권좌 수	신탁좌수	연간상품가 입건수	급여이체여 부	자동이체거 래건수	수신좌수
count	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000
mean	0.010015	0.295925	0.654204	1.239062	0.413305	0.664775	0.246351	8.195664	0.365715	18.097016	6.669115
std	0.099574	0.456462	0.475632	1.329100	1.312997	3.454236	0.992112	5.953490	0.481635	18.743281	6.998161
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	1.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	4.000000	0.000000	8.000000	3.000000
50%	0.000000	0.000000	1.000000	1.000000	0.000000	0.000000	0.000000	7.000000	0.000000	15.000000	5.000000
75%	0.000000	1.000000	1.000000	2.000000	0.000000	0.000000	0.000000	10.000000	1.000000	24.000000	8.000000
max	1.000000	1.000000	1.000000	5.000000	33.000000	242.000000	49.000000	83.000000	1.000000	1746.000000	260.000000

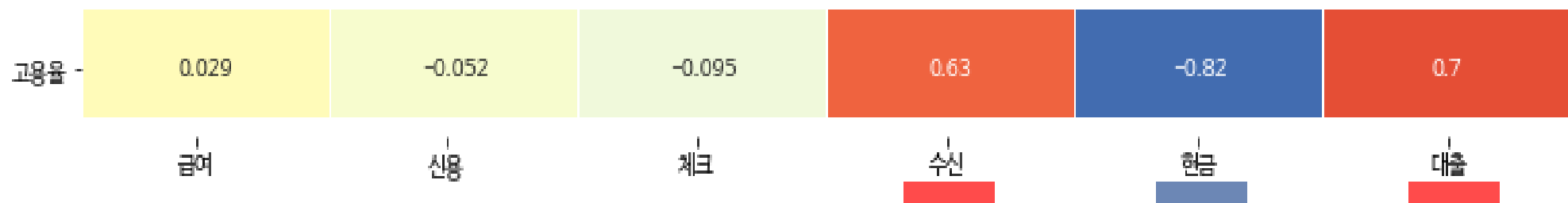
4 활용방안

초개인화 금융상품 추천 서비스

고객별 맞춤 금융상품 추천 서비스 예시

STEP 2

경제변수 별 민감고객들의 다음 시기 금융행동 변화 예측



“고용률이 상승할 시, ‘수신잔액’, ‘대출잔액’ 칼럼은 **상승**,

‘현금서비스 이용금액’ 칼럼은 **하락**”

4 활용방안

초개인화 금융상품 추천 서비스

고객별 맞춤 금융상품 추천 서비스 예시

STEP 3

추천 상품군 정의

CASE 1

수신잔액 ↑

저축을 위한 여유자금이 증가하는 고객이라고 가정

CASE 2

고용상황 ↑ 실물경기 ↑ 현금서비스 < 장기대출 ↑

경기가 좋을때는 단기대출보다 미래를 구체적으로 설계하여 장기대출을 늘릴 고객이 더 많을 것이라는 가정

< 추천 상품군 >

예•적금 상품군

대출 상품군

+ @

4 활용방안

초개인화 금융상품 추천 서비스

고객별 맞춤 금융상품 추천 서비스 예시

STEP 4

민감고객군 세분류 작업 진행

고용에 민감한 고객(57941명)

	전문직여부	세분화고수 신고객여부	세분화고소 득고객여부	실적기준고 객우대구분 코드	방카슈랑스 보유좌수	수익증권좌 수	신탁좌수	연간상품가 입건수	급여이체여 부	자동이체거 래건수	수신좌수
count	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000	53919.000000
mean	0.010015	0.295925	0.654204	1.239062	0.413305	0.664775	0.246351	8.195664	0.365715	18.097016	6.669115
std	0.099574	0.456462	0.475632	1.329100	1.312997	3.454236	0.992112	5.953490	0.481635	18.743281	6.998161
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	1.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	4.000000	0.000000	8.000000	3.000000
50%	0.000000	0.000000	1.000000	1.000000	0.000000	0.000000	0.000000	7.000000	0.000000	15.000000	5.000000
75%	0.000000	1.000000	1.000000	2.000000	0.000000	0.000000	0.000000	10.000000	1.000000	24.000000	8.000000
max	1.000000	1.000000	1.000000	5.000000	33.000000	242.000000	49.000000	83.000000	1.000000	1746.000000	260.000000

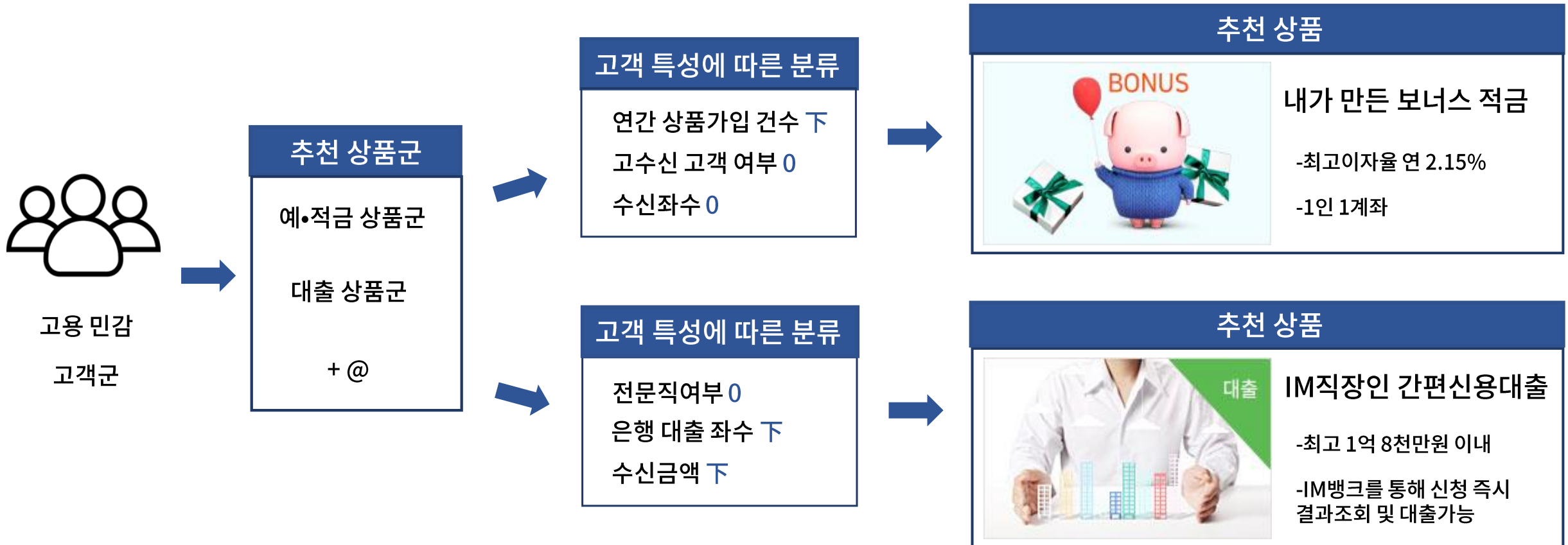
4 활용방안

초개인화 금융상품 추천 서비스

고객별 맞춤 금융상품 추천 서비스 예시

STEP 5

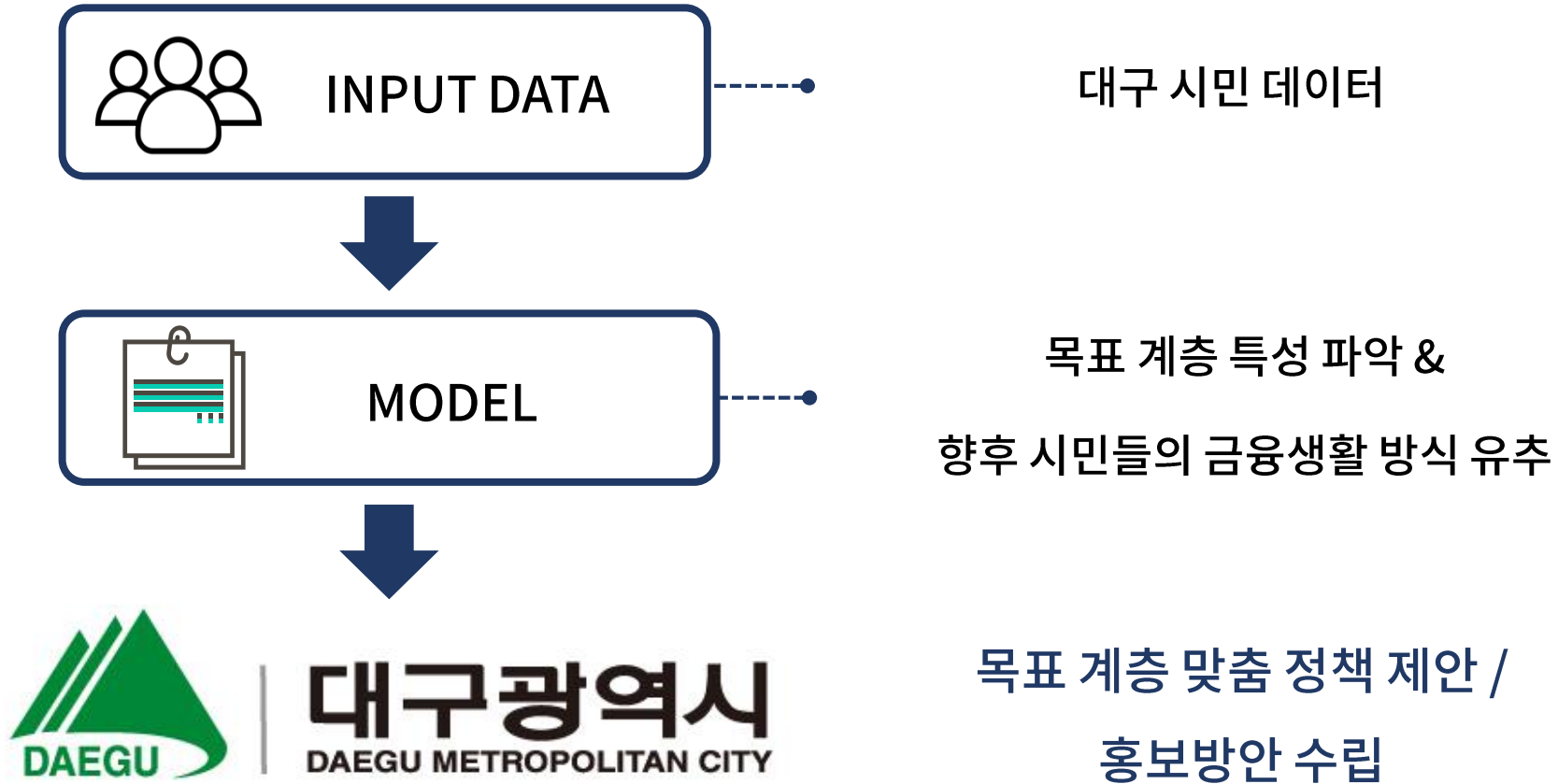
고객별 맞춤 상품 추천 예시



4 활용방안

군집 별 특성 파악을 통한 상품 및 정책 제안 서비스

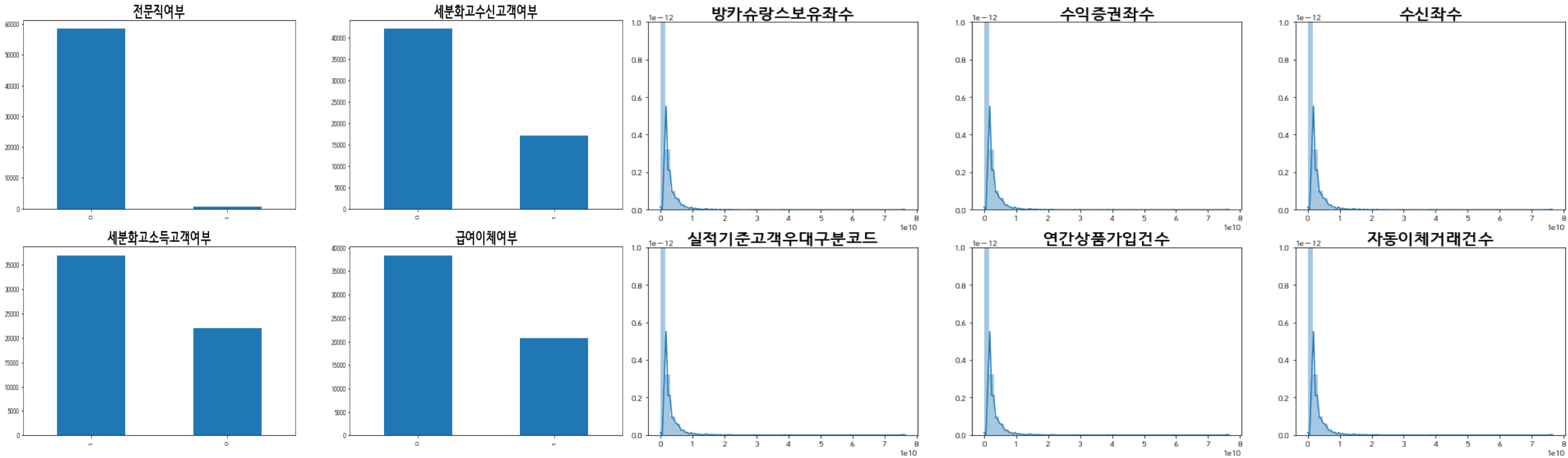
정책 수립 프로세스



4 활용방안

군집 별 특성 파악을 통한 상품 및 정책 제안 서비스

제공 데이터셋을 이용한 예시



INSIGHT

고용에 민감한 시민들의 경우 수익증권 좌수 및 상품가입이 없는 경우가 많다

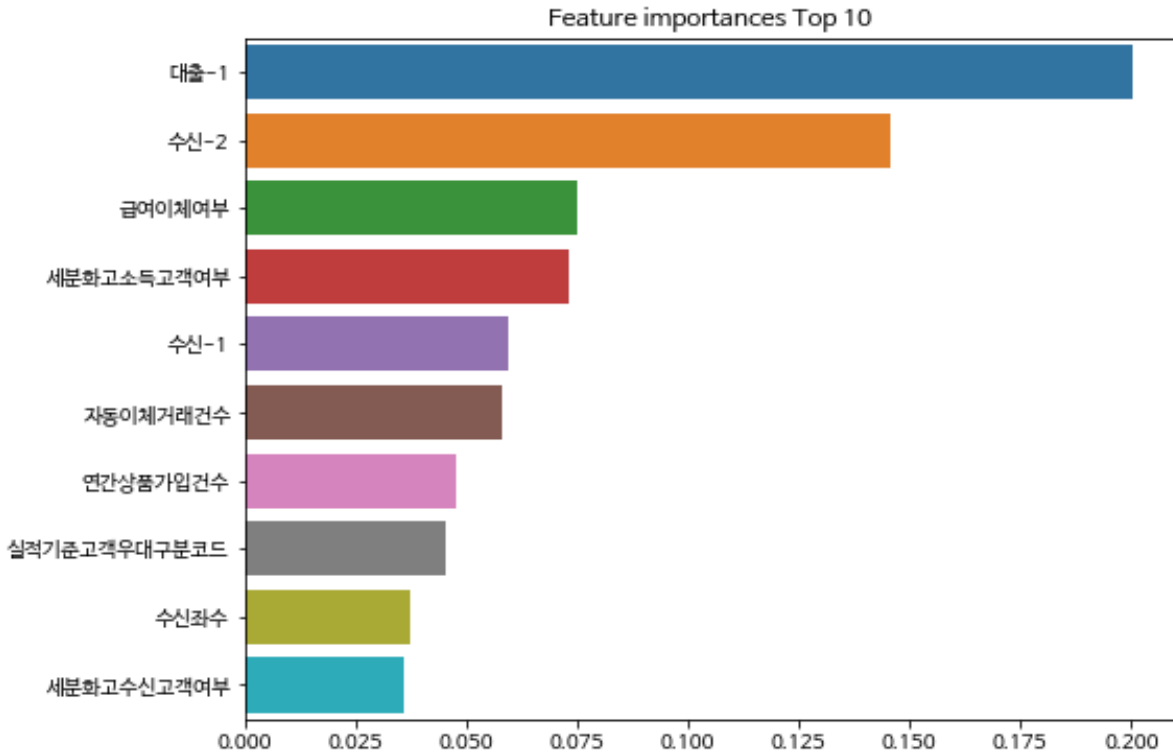
SOLUTION

해당 계층을 타겟으로 하여 금융상품 활용방안 교육정책 및 홍보 정책을 펼칠 수 있다.

4 활용방안

군집 별 특성 파악을 통한 상품 및 정책 제안 서비스

제공 데이터셋을 이용한 예시



Feature Importance를 활용하여 경제 정책 수립 과정에서

시민들의 금융행동에 주로 영향을 주는 변수들이

무엇인지 파악함으로써

효과적이고 유효성이 있는 적절한 정책 제안 가능

감사합니다

