



제 3회 대구 빅데이터 분석 경진대회

경제 변수를 활용한 초개인화 서비스 제안

팀명 GoDart

INDEX

01

기획배경

02

분석방법 계획

03

결과활용 및 시사점

01 기획배경

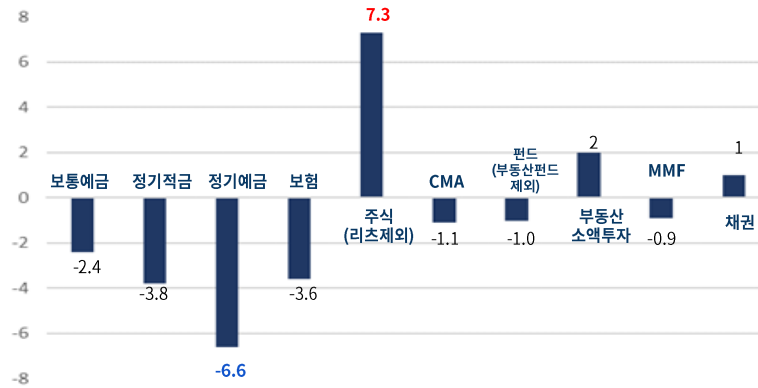
1. 금융상품 활용 증가

오랜시간 지속되어온 저성장, 저금리 환경



출처: 한국은행, 통계청

금융상품 이용률('19 대비 '20, p%)¹⁾



출처: 오픈서베이

- 단순 예금 상품보다 기대수익률이 높은 금융상품 활용 증가
- 코로나19 이후 현상 심화

01 기획배경

2. 방대한 양의 금융고객데이터 활용 가능성

마이데이터 사업

: 마이데이터란 본인 정보에 대한 개인의 권리를 보장하고, 정보 주체인 개인의 동의에 따라 본인의 데이터를 개방, 활용하는 것을 의미

- ✓ 오픈 API를 기반으로 방대한 양의 금융 고객데이터 활용이 가능하며 마이데이터 사업을 실시함으로써 새로운 비즈니스와 혁신의 기회가 도래할 것으로 예상
- ✓ 한국에서는 2022년 1월 시행을 목표로 준비 중
- ✓ 금융분야에서 마이데이터는 분산되어 있는 개인금융정보의 통합조회 및 관리, 맞춤형 데이터 분석과 이에 근거한 **금융상품 자문, 추천서비스 등 데이터 기반의 금융서비스 창출 가능**

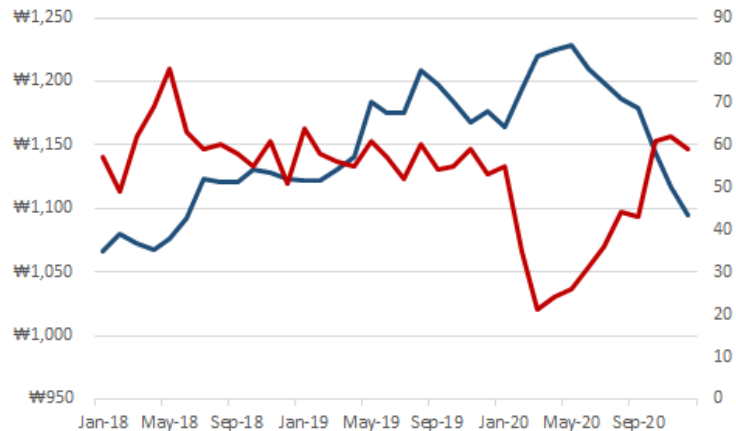
01 기획배경

3. 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 운영을 위해서는 대내외 환경변화를 주시

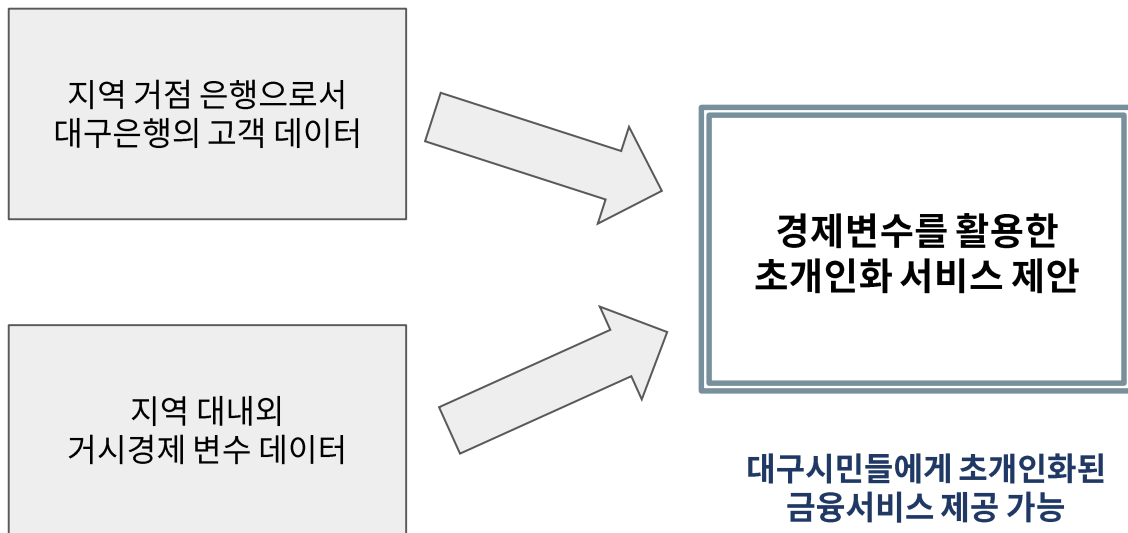
EX) 원/달러 환율과 대구 제조업 BSI



⇒ -0.72로 음의 상관관계가 존재함

01 기획배경

4. 주제선정



02 분석방법 계획

1. 사용데이터 소개

1.1 대구은행 고객데이터

■ 수정 전

- 시간 범위가 2018.01~2020.12 이기 때문에 12개월 X 3년 총 36개의 시점이 존재한다 생각
- 대구 은행 고객데이터의 전체 칼럼에서 정성적 판단으로 모델에 들어갈 칼럼 선택

■ 수정 후

- 연간 총 3개의 시점은 모델 학습을 위한 시계열 데이터로 사용하기엔 너무 짧은 기간
- 각 연도별 데이터 안에서 칼럼 간 변화를 알 수 있어 시계열 데이터로 활용할 수 있는 9개의 칼럼 선택

연도 별 시계열(4구간)

1~9월
10월
11월
12월

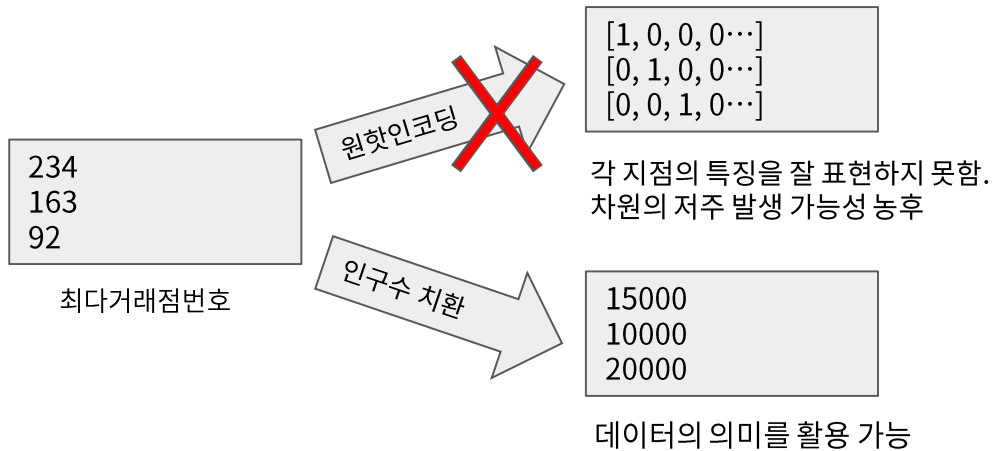
사용 칼럼

신용카드사용금액
일시불금액
할부금액
현금서비스금액
체크카드금액
급여이체실적금액
자동이체거래금액
수신평균잔액
대출월평균잔액

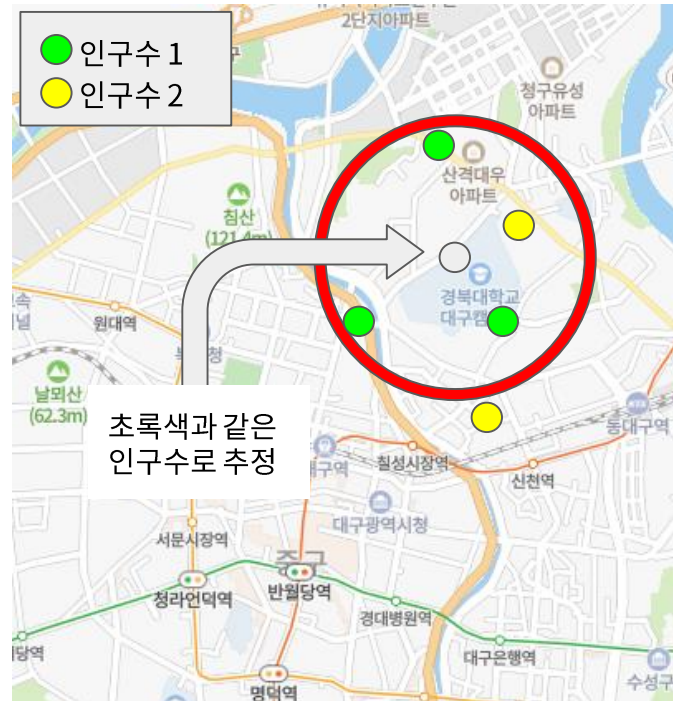
02 분석방법 계획

1. 사용데이터 소개

1.2 대구은행 고객데이터의 최다거래점번호 활용



최다거래점번호 치환 방안



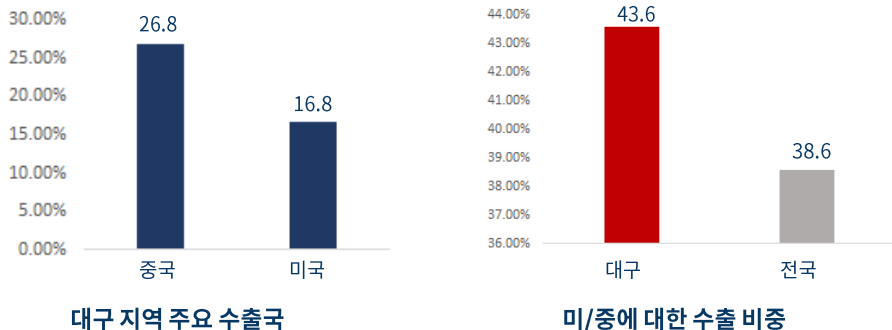
KNN을 활용한 추정 예시

02 분석방법 계획

1. 사용데이터 소개

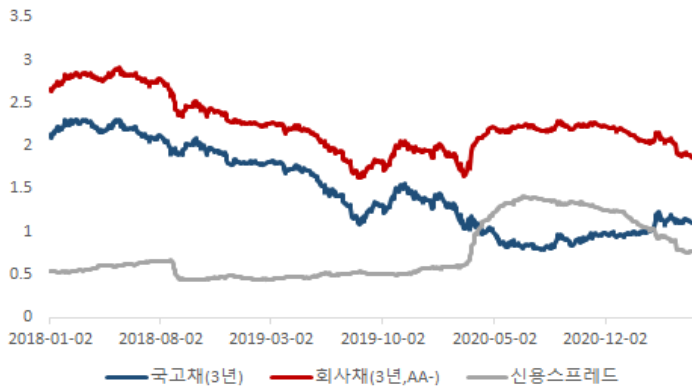
1.3 추가 활용할 공공/민간 데이터 - 지역 대외변수

(1) 원/달러, 원/위안화 환율²⁾



출처: 한국은행

(2) 신용스프레드(국고채 3년 - 회사채 AA- 3년)

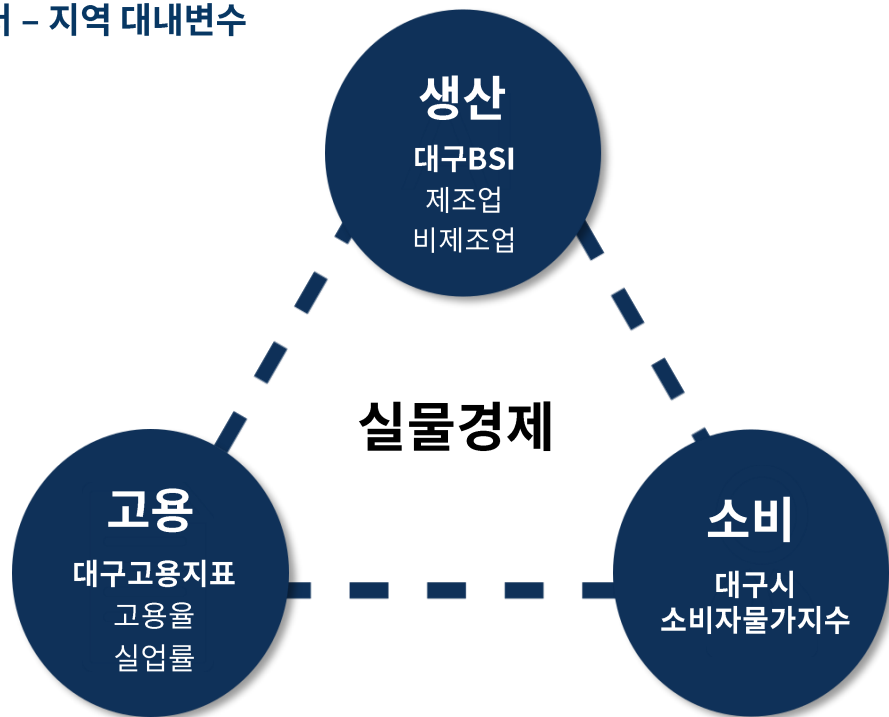


출처: 한국은행

02 분석방법 계획

1. 사용데이터 소개

1.3 추가 활용할 공공/민간 데이터 - 지역 대내변수



02 분석방법 계획

2. 레이블링 방법 - 그랜저 인과검정을 통한 타겟 변수 도출

그랜저 인과관계 검정은 두 시계열 간의 인과성의 존재 여부에 대해 검정할 수 있다.

지역대외변수
지역대내변수



신용카드사용금액
일시불금액
할부금액
현금서비스금액
체크카드금액
급여이체실적금액
자동이체거래금액
수신평균잔액
대출월평균잔액

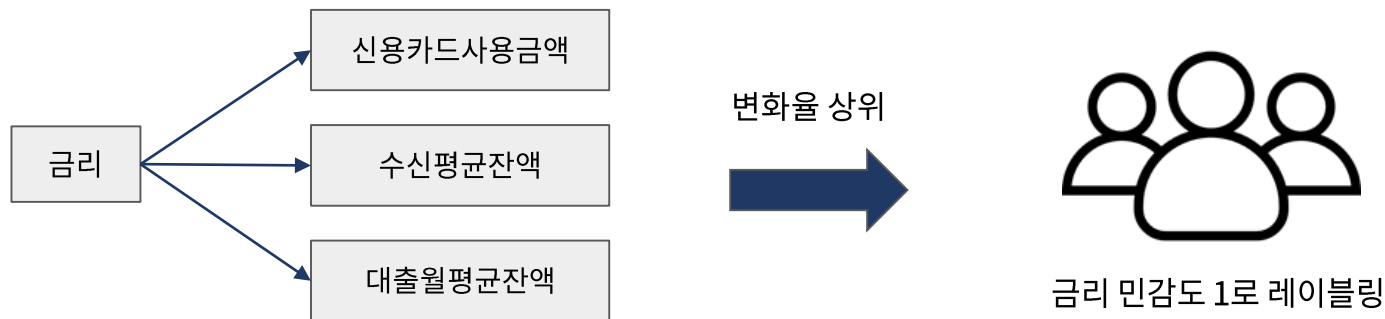
Ex. 금리에 대한 그랜저 인과관계 검정

- 1) 금리를 Y로 두고 행동 패턴 변수를 X로 두어 X와 Y 사이에 그랜저 인과관계가 존재하는지 검정함.
- 2) 인과관계가 존재하는 X 변수를 금리 민감도 측정에 활용.

02 분석방법 계획

3. 레이블링 방법 - 민감도 레이블링

그랜저 인과검정을 통해 도출된 행동 패턴 변수의 변화를 이용하여 개인별 민감도 레이블링



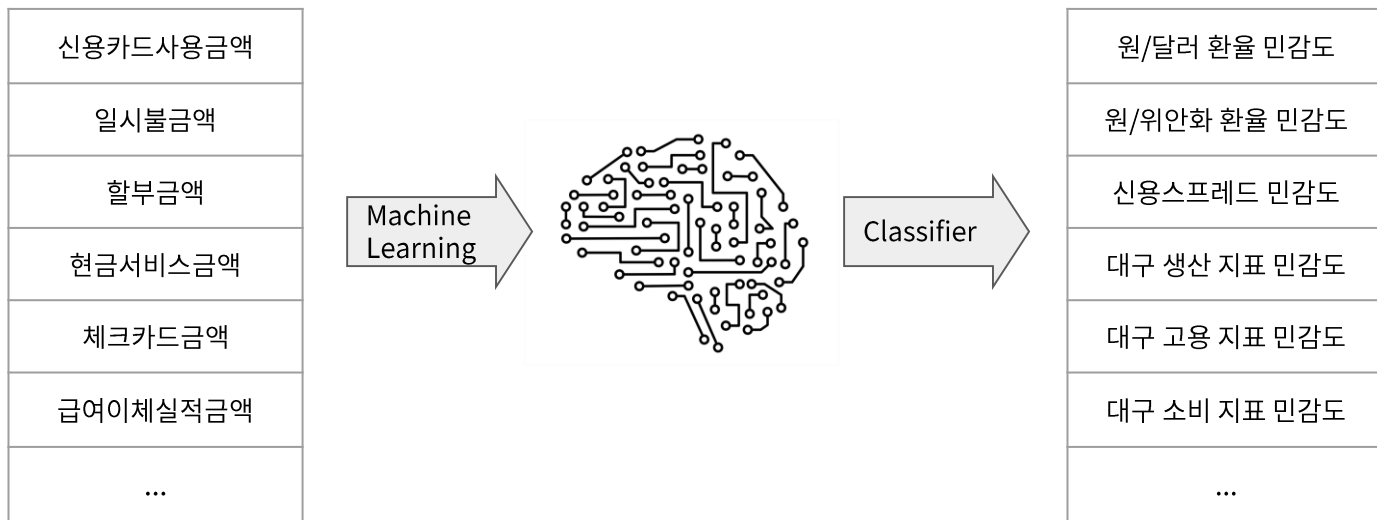
일정 임계치를 넘어서면 민감한 사람으로 분류하도록 레이블링하여
outlier의 영향을 최소화 할 수 있도록 설정할 예정

02 분석방법 계획

4. 모델링

4.1 문제 정의

- 고객 개인의 금융 상태 정보를 통해 개인마다 거시경제변수에 대한 민감도를 추출하는 문제
- Multi-label Classification. 하나의 sample이 여러 개의 label로 분류됨.

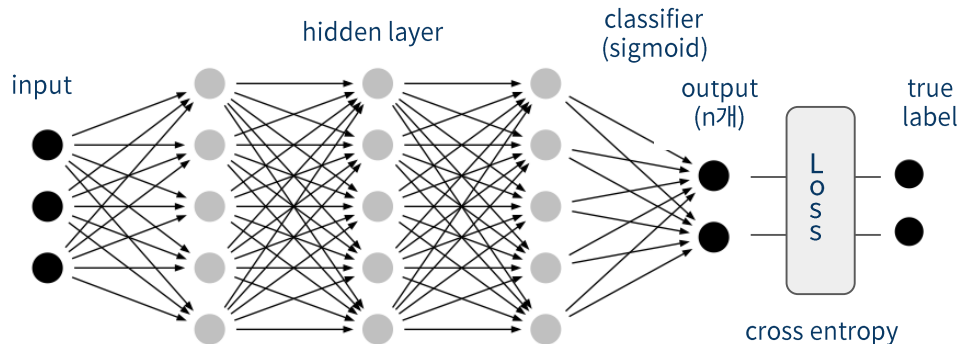


02 분석방법 계획

4. 모델링

4.2 베이스라인 소개

- 다층 신경망(MLP)를 통한 분류 모델의 적용 시도
- input은 한 시점에 대한 개인 금융 상태 정보
- output은 각 거시경제변수에 대한 민감도
- 민감도를 검출할 거시경제변수 수(n) 만큼 노드를 가지는 레이어(classifier)를 통과하여 분류
- 손실함수는 CategoricalCrossEntropy 사용

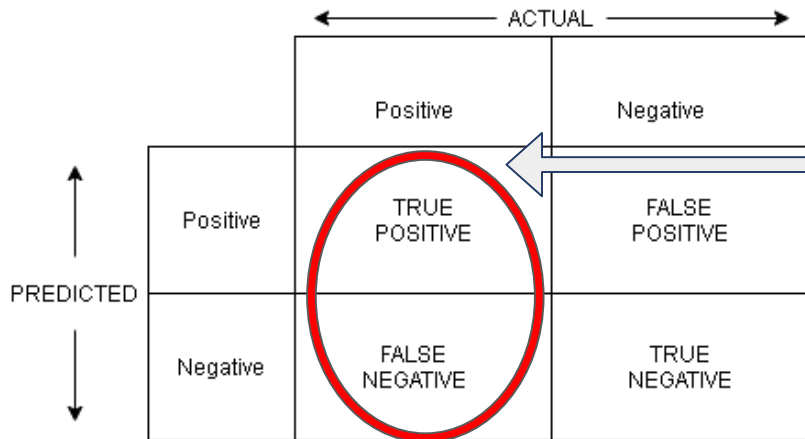


02 분석방법 계획

4. 모델링

4.3 평가 방법

- 경제변수에 민감한 사람을 검출하는게 목적으로, 최대한 많은 True Positive를 검출하는게 목표
- F1-score 및 재현율(Recall)로 평가
- 머신러닝 분류 모델도 적용하여 성능 비교 분석



True Positive - 실제 True 이면서, True로 예측한 것
 False Negative - 실제 True 이지만, False로 예측한 것

$$\text{재현율(Recall)} = \text{TP} / (\text{TP} + \text{FN})$$

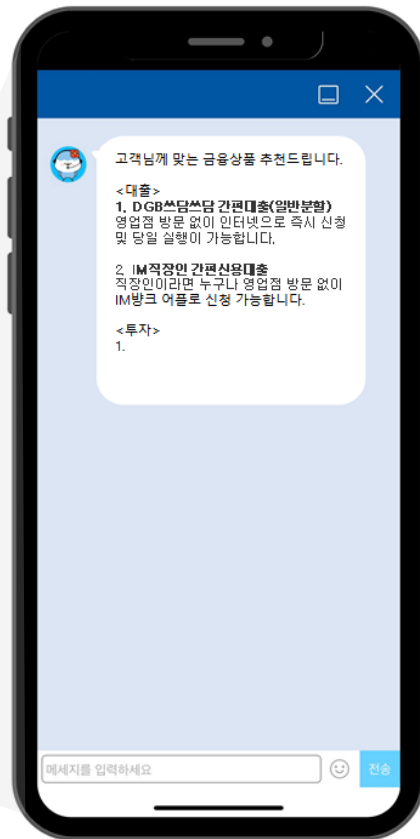
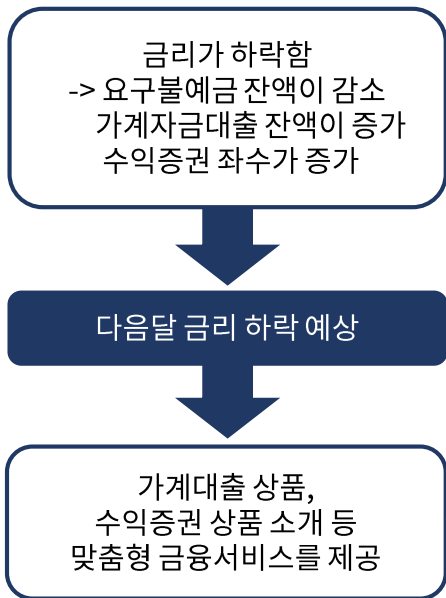
- 실제 True인 것 중, True로 예측한 비율

종합적인 성능(F1-score)을 약간 포기하더라도,
 높은 재현율(Recall)을 우선한다.

03 분석 결과 활용 및 시사점

1. 활용방안 - 맞춤형 상품추천

Ex. 금리에 민감한 집단



출처: 대구은행 IM샵

03 분석 결과 활용 및 시사점

1. 활용방안 - 맞춤형 상품추천

확장된 데이터들을 기반으로
더욱 다양하고 정교한 고객 행동 패턴을 찾아
초개인화된 금융서비스를 제공

시민 편의성 증대

마이데이터 사업

- ☑ 여수신
- ☑ 금융투자
- ☑ 카드
- ☑ 보험
- ☑ 통신

03 분석 결과 활용 및 시사점

2. 활용방안 - 그룹화를 통한 상품개발 및 정책수립



환율 민감도가 높은 집단



금리 민감도가 높은 집단



실업률 민감도가 높은 집단

각 집단의 특성 파악



- ✓ 은행: 타겟 집단에 따른 상품 개발 가능
- ✓ 대구시: 집단의 특성을 토대로 정책수립 가능

03 분석 결과 활용 및 시사점

2. 활용방안 - 그룹화를 통한 상품개발 및 정책수립

대구, 저소득 취약계층 '금융문맹' 퇴치 나선다

| 시, DGB금융그룹 등과 업무협약

DGB금융그룹-대구시 '취약계층 금융 신용관리' 업무협약 체결

금융취약계층의 특성을 파악하여 금융투자상품에 대한 이해가 부족한 취약계층을 도와
지역 경제 불평등 해소에도 기여할 수 있을 것으로 기대됨.

제 3회 대구 빅데이터 분석 경진대회

감사합니다

GoDart