

Alternative 신용평가 모델 개발을 위한 대출 연체 예측 알고리즘 개발

2017. 8. 8

—
한화생명 빅데이터팀



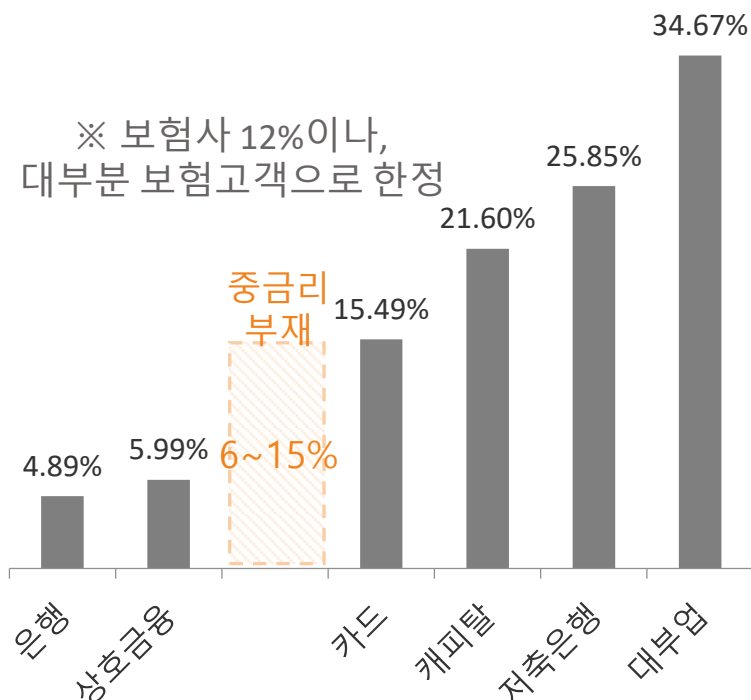
Contents

1. 배경 및 취지
2. Data Description
3. 평가 방식
4. [설명] 개인정보 비식별화
5. Q&A

배경 및 취지

국내 대출시장은 중금리수요 대비 공급이 부족하고, 은행과 비은행권 간의 금리 양극화가 존재해 왔음. 최근 들어 중신용자를 대상으로 한 시장확대 노력이 진행되고 있으나 기존의 신용평가 방식 활용만으로는 고객의 정확한 상환능력/의지를 파악하기에 한계가 있음

업권별 신용대출 평균금리



전통적인 신용평가 방식의 한계

대출기관



중금리 대출
경험 Data 부족

신용정보사 Data 및
자사 Data 활용

Rule 기반의
보수적인 신용평가

中 신용고객



금융거래 실적 부족

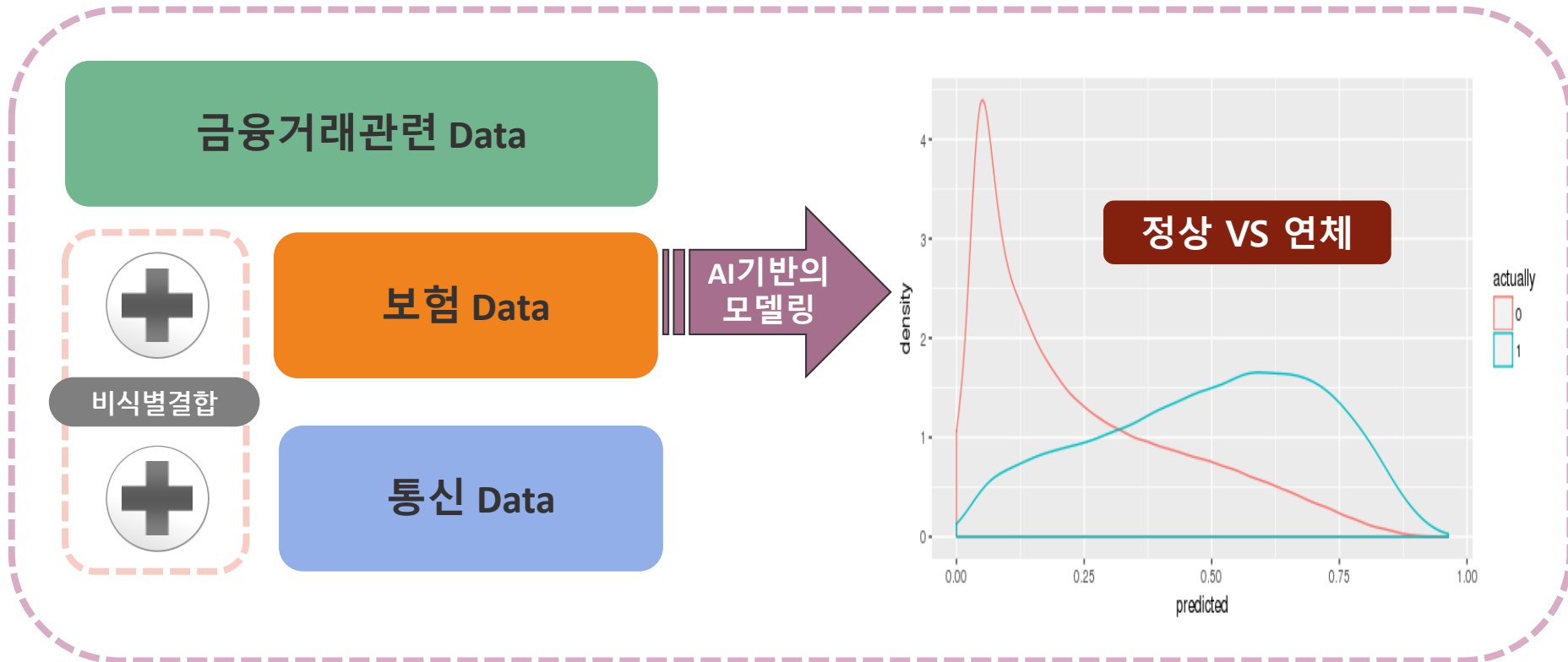
금융 이외 기관과의
다양한 거래

정교화 된 가치평가
(금리 인하)

배경 및 취지

빅콘테스트 2017에서는 기존 활용하던 금융거래관련 Data(SCI평가정보) 이외에 보험(한화생명) 및 통신(SKT) 데이터 등 다양한 산업의 Data를 활용하여 대출의 연체여부를 예측하는 알고리즘을 개발함으로써 Alternative 신용평가 모델 개발의 가능성을 검증하고자 함

Alternative 신용평가 모델 가능성 검증



Data Description

📦 Data 특성 개요 (세부사항은 '첨부' 참조)

1. 실제 Data를 기반으로 한 현실세계의 Data

- 실제 기업내부 Data 기반의 데이터 Sample
- 여러 회사(한화생명/SKT/SCI)의 Data를 실제로 접해볼 수 있는 기회
- 단, 파생변수등 고려시 다수 업종에 대한 종합적인 이해가 필요함

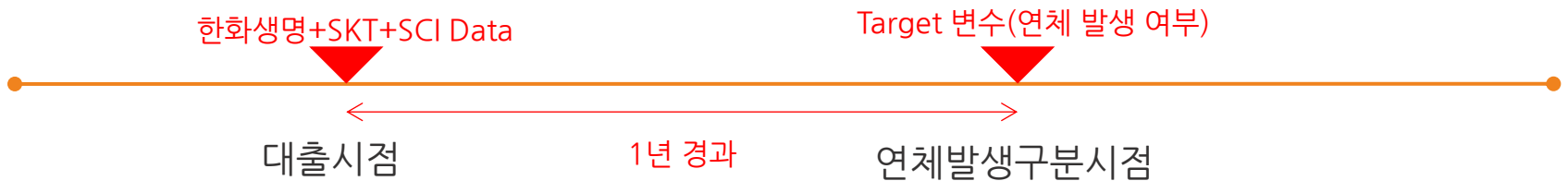
2. 비식별화 된 Data

- 일부 Data의 가공, 삭제, 마스킹 등을 통하여 주로 범주화 된 형태의 Data
- 실제 Data 대비 정보 손실로 인한 Data-quality는 상당히 낮아진 상태

3. Imbalanced Data

- Sampling을 통해 일부 비율의 조정을 했으나 Target의 반응비율(연체)가 매우 낮음(4%대)
- Imbalanced Data에 Handling이 관건이 될 수 있음

📦 Data별 산출 시점



평가 방식

구분		예측 (Predicted)	
		정상고객 (Predicted)	연체고객 (Predicted)
실제 (Actual)	정상고객 (Actual)	True/Negative	False/Positive
	연체고객 (Actual)	False/Negative	True/Positive

■ Precision : 예측한 실제 연체자 수 / 예측한 연체자 전체 수
 $\Rightarrow TP/(FP+TP)$

■ Recall : 예측한 실제 연체자 수 / 실제 연체자 전체 수
 $\Rightarrow TP/(FN+TP)$

■ F-measure : Precision 과 Recall의 조화 평균

\Rightarrow

$$\frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

개인정보 비식별화

개인정보 비식별화의 개념

기업이 보유한 고객 개인정보의 가공, 삭제, 마스킹 처리 등을 통해 특정 개인을 식별할 수 없도록 하는 기술적인 조치로써, 비식별화 이후 다른 정보와의 결합을 통해 재식별이 불가능한 상태를 비식별화로 정의함

<비식별화 예시>

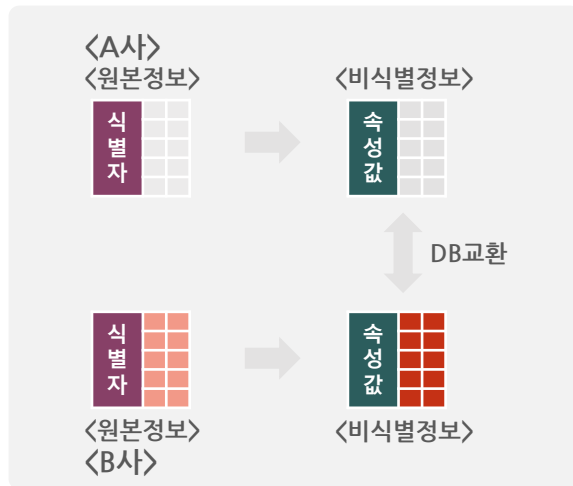
처리기법	Raw 데이터	비식별 데이터
가명처리	홍길동, 35세, 서울거주, 한화생명재직	임꺽정, 35세, 서울거주, 한화증권재직
총계처리	홍길동 35세, 임꺽정 40세, 황진이 45세	한화생명 나이 합 : 120세, 평균 : 40세
데이터삭제	주민등록번호 : 750101-1234567	75년생, 남자
데이터범주화	홍길동 35세	홍씨, 30~40세
데이터마스킹	홍길동, 35세, 서울거주, 한화생명재직	홍OO, 35세, 서울 거주, OO생명재직

개인정보 비식별화

개인정보 비식별화 활용 방법

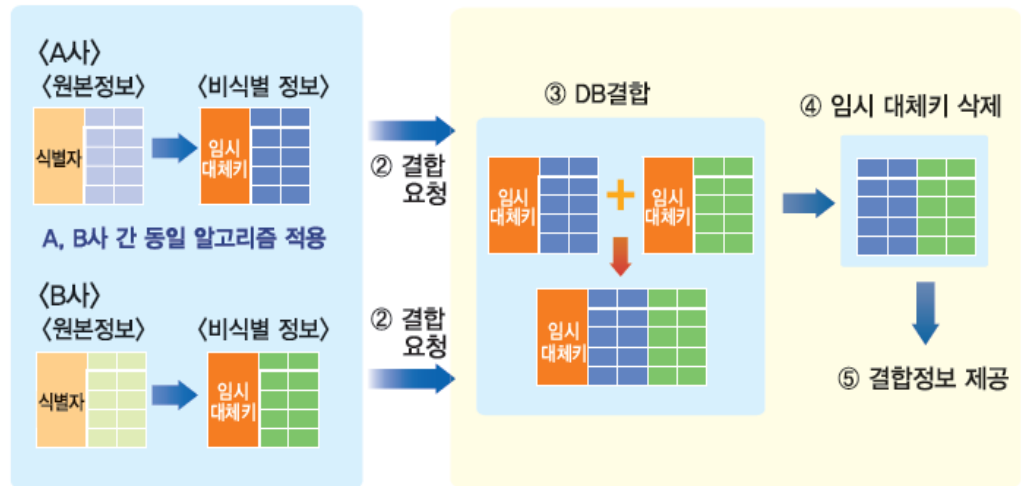
각 사별 식별정보 제거 후 교환·활용하는 방법과 전문기관을 통해 각 사별 식별화된 데이터 결합 후 비식별화 처리된 데이터를 활용하는 방법이 있음

① 식별자 제거 後 기업간 교환



- 식별정보 제거 후 회사별로 교환 가능
- 성/연령/지역/직업별 통계 등 분석가능

② 식별 데이터 결합 後 비식별 데이터 교환



- 개인식별정보(임시대체키) 수준의 데이터 융합 가능
- 양사 간 매칭되는 고객에 한해 분석 가능
- 융합 이후 식별정보 삭제로 인한 속성별 통계값 분석

개인정보 비식별화

개인정보 비식별화 활용 예시

동종 및 이종 산업간의 데이터 융합 및 분석을 통한 새로운 인사이트 발굴 및 상품/서비스 개발



Q & A