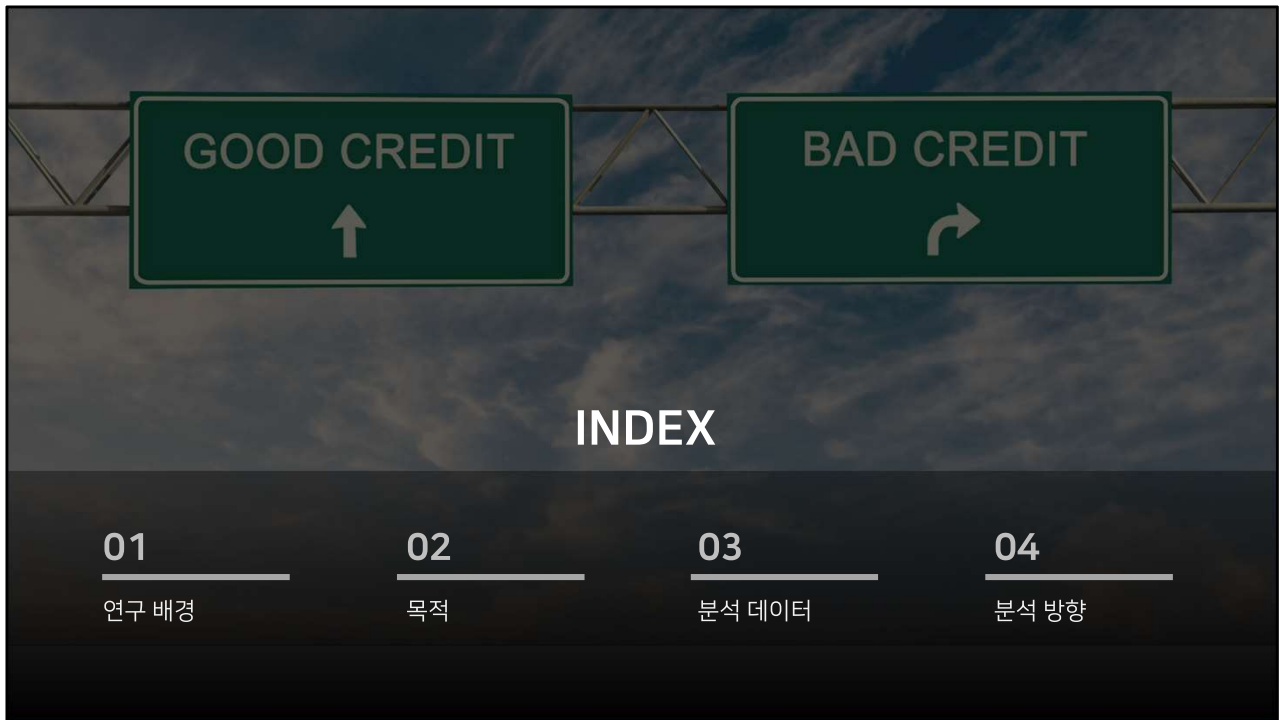


개인 신용도 예측 변수 분석

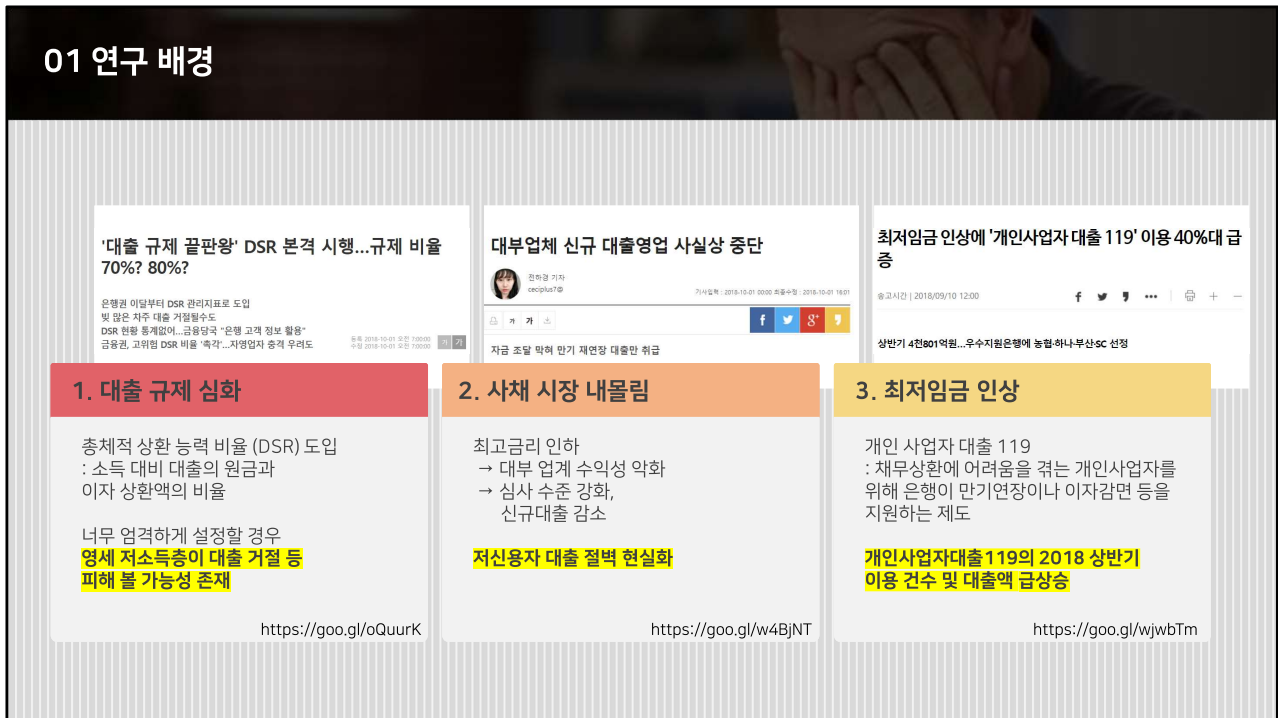
서울대학교 빅데이터 아카데미
2018-3 고급 빅데이터 분석 기법
BA 노은선 이현호 최의관

안녕하세요. 똑똑이 현호와 의관이와 함께 본 과제를 진행하게 된 노은선입니다.
저희는 이번 과제에서, 개인 신용도를 예측할 수 있는 다양한 변수에 대해 연구해보고자 합니다.



본 발표에서는 저희 연구의 배경, 연구 목적.
그리고 분석 데이터를 제공한 기업에 대한 정보,
마지막으로 분석방향에 대한 내용에 대해 간략히 말씀드리겠습니다.

01 연구 배경



저희는 최근 신용도와 관련하여 현실적으로 대두되는 다양한 이슈에 주목했습니다.

첫번째는 '총체적 상환 능력 비율, 일명 DSR 규제에 관한 기사입니다. 올 10월부터 은행권을 중심으로 도입되었으며, 점차 보험사와 내년 상반기에는 제 2금융권에서도 DSR이 활용될 예정입니다.

대출자의 소득 대비 모든 대출의 원금과 이자 상환액의 비율을 뜻하는 DSR을 관리 지표를 사용하게 되면
쉽게 말해 개인이 빚 갚을 능력만큼만 산정하여 대출해주게 됩니다.
이 기준이 너무 엄격하다면, 영세 저소득층이 피해를 볼 가능성이 커지게 되는 거죠.

--

두번째 기사는 대부업체 관련 기사인데요.
또한 최근 대부업체의 수익성이 약해지면서 심사 수준이 강화되고, 신규 대출이 감소되고 있다고 해요.
그래서 신규 대출을 희망하는 저신용자들은 상대적으로 대출절벽을 느끼고, 대부업체를 찾는 현상을 우려가 되고 있습니다.

마지막으로 최저임금이 인상함에 따라 연쇄효과로 대출 상황에 어려움을 느껴,
개인사업자대출119를 통하는 사업자들이 증가했다고 합니다.

위 세 기사를 통해 알 수 있는 내용은,
최근 대출을 희망하는 사람들이 늘어나고 있지만, 대출규제가 심화되고 엄격해지면서
제 2금융권에서조차 대출을 받지 못해 사채를 찾는 사람들이 많아지고 있다는 사실입니다.

01 연구 배경

두루뭉실 신용평가 '등급제'로 240만명이 대출금리서 불이익



기존 모델의 한계

1. 일부 엄격한 기준
2. 금융 정보에 한정된 평가 기준

고객 : 상환 능력 저평가

기업 : 고객 유치 한계

**신용도 평가 측정 모델의
개선 필요성 시사**

금융전문가들은 현재 등급제 신용평가시스템은 등급구간의 한계에 있는 고객들이 자신의 신용상태보다 높은 금리를 적용받고, 특히 서민층 금융취약계층이 과도한 대출금리적용을 받게 되는 불합리한 구조라면서 진즉부터 평가시스템의 개편을 주장해왔다.

<https://goo.gl/rUyTBv>

저희가 초점을 맞춘 것은 일반 신용 불량자가 아니라,
'상환능력이 있음에도 대출을 거절당하는 억울한 사람들'입니다.

그들은 충분히 신용도가 더 올라갈 가치가 있음에도
현재의 엄격한 기준과, 금융 정보에만 한정된 평가 기준 탓에 실제 신용이 과
소평가되고 있습니다.

이는 개인에게는 상환능력이 저평가되어 실 가치보다 더 낮은 금융권을 찾게
될 것이고,
기업에 있어서는 상환 능력이 있는 고객을 유치하지 못합니다.

현재 신용평가는 금융정보만으로 사람을 평가하기 때문에
저희는 이에 상대적으로 불이익을 받는 사람들의 문제를 해결하고 싶었습니
다.

02 목적



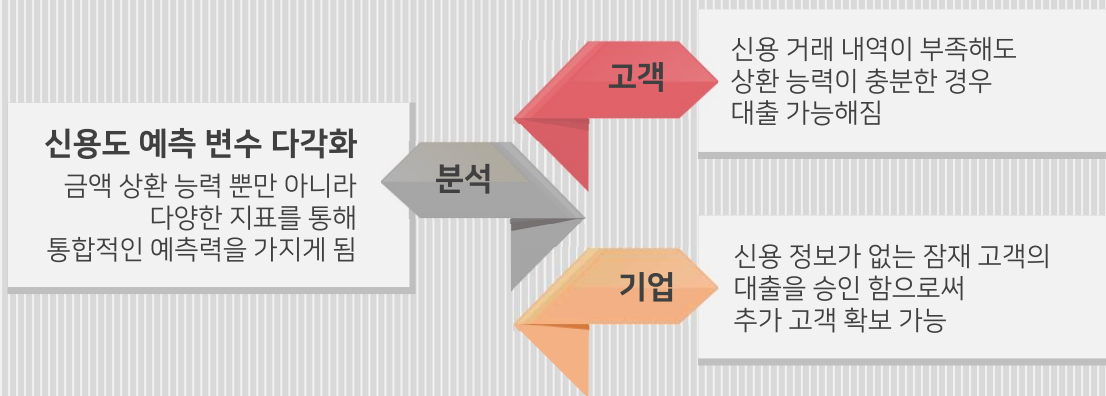
단순 부채 수준, 신용 형태, 거래 기간 뿐만 아니라,
고객의 신용에 기여하는 유의미한 특징을 찾을 수 있을까?

그래서 저희는 단순 부채 수준, 신용 형태, 거래 기간 뿐만 아니라
 고객에 신용 평가 지표에 기여할 수 있는 다양한 유의미한 특징을 찾아보고자
 했습니다.

금융 정보와, 개인이 가진 특징을 모두 아우르는 데이터를 통해 개인 신용도
 예측력을 높이하고자 하는 것이죠.

(천천히 읽기)

02 목적



신용도 예측 변수가 다각화 되어 통합적인 예측력을 가지게 되면,
고객과 기업에게 각각 긍정적인 기여를 할 수 있을 것입니다.

우선 고객의 경우, 신용 거래 능력이라고 평가되는 금융 부분이 부족하더라도,
다른 데이터에 기반하여 신용도가 충분하다고 평가 될 수 있다면
실 고객 신용 가치에 맞는 대출이 가능해질 것입니다.

그리고 기업은 신용 정보가 부족한 고객일지라도 다양한 분석을 통해
기업에게 도움이 될 수 있는 고객을 효과적으로 추가 유치할 수 있을 것입니
다.

03 분석 데이터 제공 기업 : Home Credit Group



저희는 저희가 생각하는 방향과 가장 유사한 방식으로 고객을 평가하는 금융
기업을 찾았고,
"홈크레딧" 이라고 하는 금융 회사의 오픈데이터를 사용하고자 합니다.
회사에 대한 정보에 이야기 하기 전에 광고로 먼저 소개 해드리고자 해요.

03 분석 데이터 제공 기업 : Home Credit Group

**HOME
CREDIT**

- 1997 년 체코에서 설립,
네덜란드에 본사 위치
- 현재 10개국에서 사업을 운영하는
국제적인 비은행 소비자 금융 기관

NO CREDIT CARD? NO PROBLEM!

NOW YOU CAN WITH
**HOME
CREDIT**

**SUPER ZERO%
INSTALLMENT**

20 MINS. APPROVAL*

TWO VALID IDs LANG!

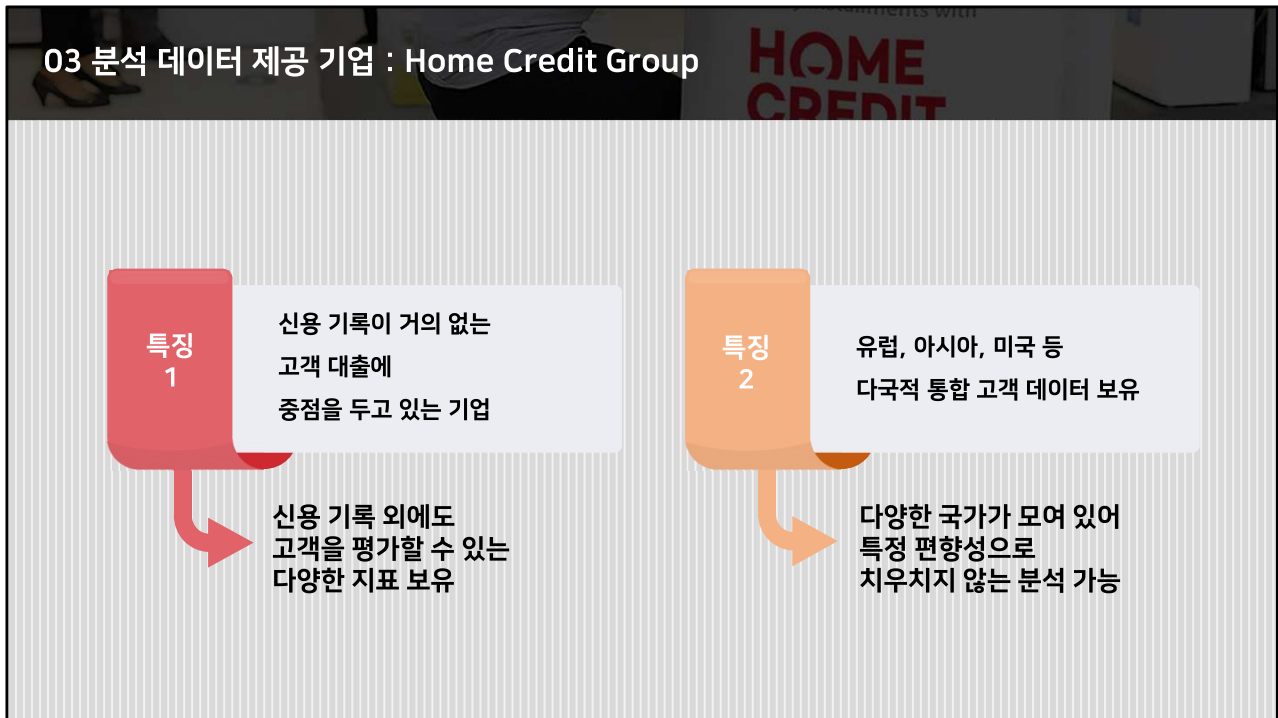
YOUR FAVORITES AT 0%!

영업 이익	31.23 억 유로 (2017)
당기 순이익	€ 244 million (2017)
총 자산	215 억 3000 만 유로 (2017)
총 자본	20 억 3,000 만 유로 (2017)

거래 내역, 기존 신용도 등 내용 외에도 다양한 고객 신용 예측 변수를 찾기 위해,
해외에 위치한 Home Credit group에서 제공하는 데이터를 사용하고자 합니다.

Home Credit group은 체코에서 설립되고 네덜란드 암스테르담에 본사가 위치한 소비자 금융 기관으로,
10개 개국에서 사업을 운영하고 있는 규모가 큰 기업입니다.
이 분들 메인 카피라이트가 'Now you can'인데요. 이 카피라이트에 담긴 의미가 따로 있습니다.

03 분석 데이터 제공 기업 : Home Credit Group



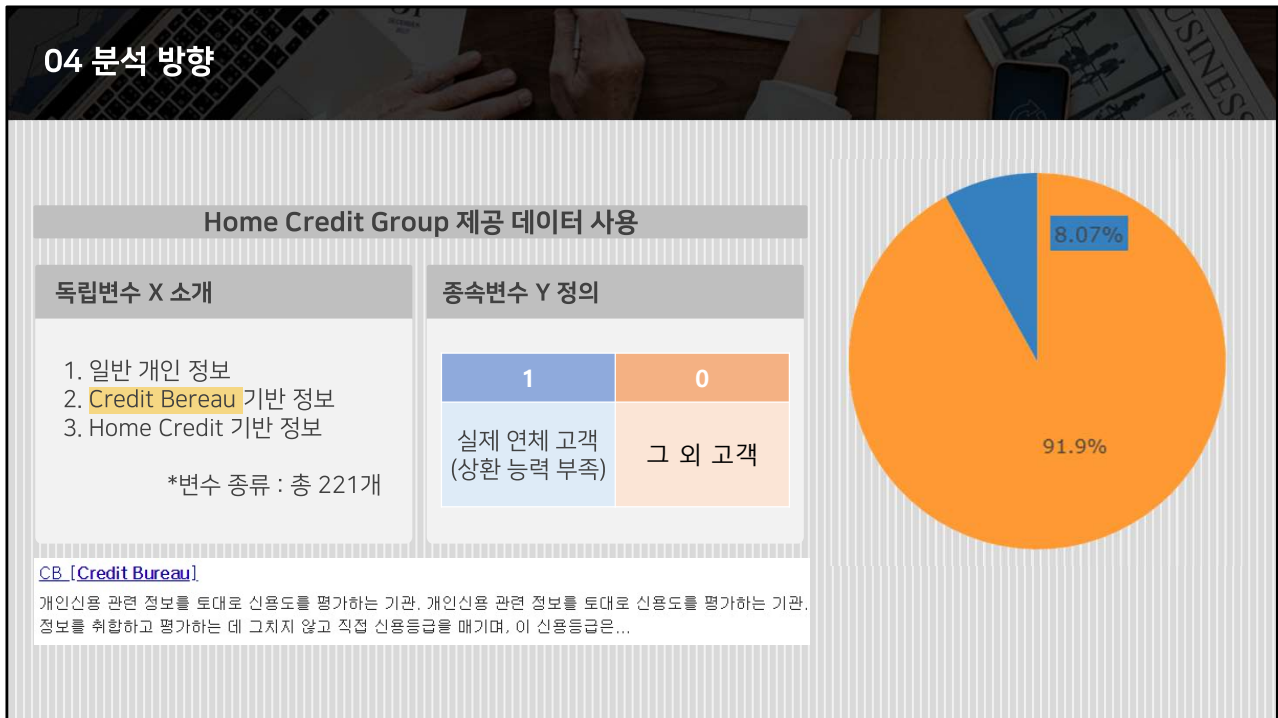
Home Credit group은 신용 기록이 거의 없거나, 전혀 없는 고객의 대출에 중점을 맞추고 있습니다.

그렇기 때문에 이들은 보다 다각화된 지표를 통해 신용도를 예측하고자 실제로 많은 노력을 하고 있구요,
바로 다음 페이지에서도 보실 수 있겠지만 실제로도 변수가 굉장히~ 많아 다양한 접근이 가능한 상태입니다.

또한, 이 기업은 미국, 중국, 러시아, 인도 등 다양한 대륙에 위치한 10개국에서 사업을 운영하고 있기 때문에,
해당 기업에서 제공하는 데이터는 한 국가만이 가질 수 있는 편향성이 두드러지지 않을 것이라 생각했습니다.

거래 내역, 기존 신용도 등 내용 외에도 다양한 고객 신용 예측 변수를 찾기 위해,
Home Credit group에서 제공하는 오픈 데이터를 사용하고자 합니다.

04 분석 방향



독립 변수로 사용하는 X는 크게
개인의 성별, 나이, 학력 등 일반적인 개인 정보,
Credit Bureau라는 신용도 평가 기관에서 온 일반적인 신용 관련 정보,
마지막으로 Home Credit group 자료로 구성된 3가지로 구분할 수 있습니다.

종속 변수 Y는 실제 연체 고객을 1로, 그 외 고객을 0으로 하는 binomial
distribution 을 기준으로 사용할 것이며,
실제로 오른쪽 파이차트처럼 전체 중 8% 가량이 상환 능력이 부족한 고객으
로 확인되고 있습니다.

04 분석 방향

Client data	application_{train test}	Train & Test data (ex) 성별, 자가유무, 자차유무, 자녀 수, 수입 등	124 columns	158MB / 24MB
Credit Bureau	bureau	고객 신용 데이터 (ex) CB 기반 신용 상태, 신용 유지기간, 빚, 신용 유형	17 columns	162MB
	bureau_balance	월별 Balance 데이터 (ex) 월별 대출 상태	3 columns	358MB
Home Credit	POS_CASH_balance	HC기반 (ex) 신용점수 유지 기간, 현재 계약 현황, 만기기한	8 columns	374MB
	credit_card_balance	기존 신용카드 대출 정보 (ex) 예전 신용 대출 계약 상태, 예전 채권 총액	22 columns	404MB
	previous_application	대출 status 데이터 (ex) 선금금, 이자, 지불 방법	25 columns	386MB
	installments_payments	대출 상환 관련 데이터 (ex) 대출 신청일, 할부 금액	7 columns	689MB
Total	7 files		221 columns	2.49GB

전체 7개 파일로 구성된 현재 파일을 앞서 말씀드린 3가지로 크게 구분해보았습니다.

Client data에서 성별, 수입 등 기본정보를 확인할 수 있구요,

Credit Bureau 데이터를 통해 기존에 규정해오던 신용도를 평가하는 다양한 기준들에 대해 볼 수 있으며,

Home Credit에서 가지고 있는 고객 데이터도 보실 수 있습니다.

CB와 HC 데이터와 큰 차이점이 있다면,

CB가 기존 신용도 평가를 구성하는 주요 지표라면,

HC는 그 정보 이상의 개인의 Action을 좀더 구체적으로 담았다 보실 수 있겠습니다.

저희가 선택한 데이터들이 용량이 너무 많고 변수도 많다 보니 걱정도 되긴 하지만,

아무쪼록 금전적인 자료 외에도 개인 신용도를 예측할 수 있는 유의미한 변수를 한 번 찾을 수 있도록 고민해보겠습니다.



Any Questions?

개인 신용도 예측 변수 분석
서울대학교 빅데이터 아카데미
2018-3 고급 빅데이터 분석 기법
BA 노은선 이현호 최의관

저희 발표는 여기까지입니다. 혹시 질문 있으실까요?

여기에서 얘기하지 못한 질문이나 다양한 의견 있으시면,
이번 자리가 아니더라도 나중에 저나, 현호와 의관이에게 꼭 말씀해주세요.
감사합니다.