

정의 (Define)

1. 당위성 = 개인과 영세업자를 위한 금융
2. 담보주의의 폐해
3. 신용평가산업 독과점
4. 금융 3법
5. 설명력 이슈
6. 디지털 문맹과 싹파일러(thin-filer)

고객의 입장에서 공감 (Emphasize)

디지털 문맹

- 금융 경험 존재하나, "겁이 나요. 잘못 누르면 돈이 나간다고 그래서...", "간단한거 문자까지는 배웠는데 다른 것은 모르니까 답답한 거예요"
- 낮은 디지털 신뢰도 - 앱보다 창구직원을 신뢰하지만 창구직원또한 100% 신뢰할수는 없음
- 접근성 또한 부족 - 정부, 지자체 프로그램 (한정된 공급) 외 어떤채널도 믿음직스럽게 도와주지 못함

썬파일러

- 2년 신용카드 X, 3년 대출 X, 금융경험 X
- 은행 대출 어려움, 작은 사유로 7등급
- 높은 디지털 접근성과 신뢰성
- 금융 경험/이력 전무
- 개인정보 제공 민감
- 패배주의 - 자본 축적에대한 자신감 부족
- 막연한 낙관주의, "어떻게든 되겠지..?"

교집합

- “어렵다” (디지털경험/금융지식)
- 사회 소외계층들에게 보여지는 현실적인 부담요소들이 금융소외계층인 이들에게 필연적으로 존재
- 배우고자 하는 의지는 있지만 가르쳐 주는 채널이 부족 - 신뢰도 있는 채널이 부족
- 개인신용평가시스템(CSS)에 대한 인식은 전무 - 차후 머신러닝기술 설명력 이슈
- 복합적인 요소들로 인해 심리적으로 위축/방어적

머신러닝 유의미 차용 사례

농협은행: **통신사 정보** 등으로 신용·소득이 낮아도
상환능력이 있는 이를 골라낼 수 있어(?)
썬파일러에게 유리한 신용평가 모형을 개발

Fico Score XD2(미국): 금융소외자를 대상으로
비금융결제정보 Pay TV or cable, Utilities, Wireless
Phone, Landlines 활용, 신용측정이 불가능한 미국
성인 인구 20% 대상으로 개발, 현재 12개의 신용카드
발급 금융기관과의 시범 작업이 진행

Schufa(독일 대표 CB): 현재 은행, 특수신용기관,
기타 금융회사 등으로 구성된 시장에서 거의 100%의
시장점유율을 보유 - DB에 등록된 사람들 중 90%
이상 긍정적 정보를 보유 - 긍정적(미결제 신용)
정보를 제휴회사에 제공할 때에는 자료주체의 동의
필요 - Schufa의 Scoring 방법론은 비공개이나 **연체
사건의 발생 확률을 modeling 하는 logistic
regression**에 근거 - 독일 연방대법원은 2014년 1월
28일 Schufa의 신용평점 계산법이 영업비밀에
속한다고 판결

아이디어내기1 (Ideate)

- 대출결정(종속변수) = αx (정성적 능력=독립변수) + βx_1 (그릇 = 독립변수 2) + β (조정값)
- 기존 모델과 어떻게 다른가? - 즉 어떤 독립변수를 추가할 것인가?
- 유의미 비정형 데이터 종류: 심리테스트, 통신데이터, SNS 사용단어, 쇼핑내역, 도서구매, 도서관 이용, 중고거래, 유튜브 구독채널, 신문 구독여부, 동호회/단체, 사회공헌, 네비게이션 이력, 리더쉽 등
- 어떻게 수치화, Clustering 할 것인가?
- 긍정적(+), 부정적(-), 중립적(Δ) Clustering?
- SNS 단어 text mining?
- Feature Selection, SVM, Random Forest, Neural Network, 의사결정나무, RNN, CNN 등 중 어떻게, 무엇을 사용할 것인가?

아이디어내기2

- 인문학 - 심리학의 중요성 역설, e.g.
레몬향 연구, 인간심리라는 박스권
- 고객들의 심리적 방어태세를 깬 신뢰성을
확보한 채널 확보 - 어떻게 신뢰도 있는
채널을 형성할 것인가?
- 인문학 유튜브 - 인문학과 함께하는 금융
컨텐츠
- 믿을만한 사람이 앱 설치부터
사용방법까지 친절히 안내해준다면?
- 지자체 or 타금융권 회사들과의 협업?

Prototype 1 - R을 이용한 대출결정 모형

- 대출결정(종속변수) = $ax(\text{정형=독립변수}) + bx1(\text{비정형 = 독립변수2}) + \beta$
- 앞서말한, 비정형데이터중 10개 확보하고 데이터를 정제화, 수치화, 정규화 하였음을 가정
- **Variable Preselection** 과정속에서 **SVM** 등으로 긍정적, 중립, 부정적으로 **Classify** 할 수있음. 하지만, 부정적 데이터는 데이터 확보부터 큰 어려움이 따를것으로 예상되어서 제외
- 긍정적: 도서구매, 도서관 이용, 사회공헌, 신문 구독여부
중립: 심리테스트, 통신데이터, **SNS** 사용단어, 유튜브 구독채널, 동호회/단체 가입여부
- IV(Information Value)와 WOE(Weight of Evidence)를 이용한 변수선택법
- 가용 패키지과 함수 목록 : `creditR(IV.calc.data(), Information(create_infotables()), InformationValue(IV()), scorecard(iv()))` 등

이어서 Correlation Analysis과 Scorecard modelling

- Regerssion model estimation 변동성을 피하기위해
각각의 변수들은 낮은 correlation 유지해야함
- 방법론: 1. caret package에 findCorrelation(), 2.
corrplot package를 이용한 correlation matrix 시각화
등
- Variable Selection: Model's performance vs. the
number of free model parameters trade-off - BIC,
AIC, VIF 등을 이용하여 변수 유의미 결정
- 예를들어 BIC 변수 선택법에선 ($k=\log(\text{nrow}(\text{data}))$) 로
두고 glm 함수를 이용. 낮은 BIC값일수록 좋음.
- Scorecard 패키지에 있는 scorecard()함수를 이용하여
glm값을 frequencies값, default rates 도출
- scorecard2()를 이용해 최종 점수 도출.
- 일례로 `sc <- scorecard2(bins, train, y값, x 벡터)`
- 마지막으로 report() 함수 이용하여 summary값 완성
- Model 성능평가 과정 후략

Prototype 2 - 온라인 채널 확보

‘심리학으로 말하는 금융이야기’ 유튜브 채널

방법론:

특이점: 카뱅 공식채널이 아닌 개별채널,
회사대변 X

- 대본+나레이션+일러스트+애니메이션
- 메인 MC와 심리학, 금융 전문가와 함께하는 다양한 이야기

이유: 고객의 심리적 방어

컨셉: 본질은 심리학, 주로 금융투자과 인간의 심리에 대해 아주 다양한 주제

컨텐츠 예시: “동학 삼전운동, 나도 뛰어들어야 하나?” “왜 우리집은 투자에 실패했을까?”
“남자가 여자보다 바람필 확률이 높을까?”
“과학이 말하는 이별”

Prototype 3 - 오프라인 채널 확보

방법론:

기대효과: 디지털 금융 신뢰성 확보와 선점효과

- 오프라인 금융사와 제휴를 통해 오프라인 채널 확보
- 정부기관(지자체 등)과의 콜라보 - 금융교육 지원