

# 개인신용평가의 미래

## 머신러닝과 대안정보를 활용한 신용위험 예측

박재영

국립부경대학교 디지털금융학과 조교수

jyp2503@pknu.ac.kr



### 배경

- 현대 사회에서 금융 서비스(대출, 신용카드 발급 등) 이용을 위해서는 금융기관의 신용평가가 필수임
  - 금융기관은 신용평가 시스템을 통해 차주의 신용위험을 분석(예측)하고, 이에 기반해 금융 서비스 제공 여부와 조건을 결정함
  - 이 과정에서 신용점수(Credit Score)는 대출 승인, 금리 결정, 한도 설정 등의 의사결정에 핵심적 지표로 활용됨
- 신용점수 기반의 전통적인 신용평가 방식은 금융거래 이력(과거 대출 상환 이력, 신용카드 이용 내역 등)에 의존하기 때문에, 금융거래 기록이 없는 개인에 대해서는 신용위험을 정확히 평가하기 어렵다는 한계를 가짐
  - 우리나라는 금융거래 이력이 부족한(또는 없는) 신파일러(Thin Filer)가 약 1,300만 명으로, 기존의 전통 모형은 이들의 신용위험을 정확하게 평가하기 어려움

- 사회초년생, 주부 등(즉 신파일러)이 상환 능력과 무관하게 대출 심사, 카드 발급 등에서 불이익을 받게 되는 금융 불평등이 발생하는 등 우리나라의 금융 포용<sup>1</sup>을 저해하는 결과를 초래함
- 누적된 신용 기록에 대한 정보가 부족한 경우, 낮은 신용점수로 인해서 신청이 거절될 수 있으며, 승인이 되더라도 높은 금리 또는 낮은 한도가 부여될 가능성이 높음
- 전통 모형의 한계점을 '보완'할 수 있는, 즉 금융소외계층의 신용위험을 정확하게 예측할 수 있는 새로운 모형이 필요함

● **규제 혁신 및 기술 발전으로 비금융정보(대안정보<sup>2</sup>)를 활용한 대안신용평가모형이 최근 주목을 받고 있음**

- 전통 금융사 및 핀테크(빅테크)에서 다양한 형태의 대안신용평가모형을 개발하고 있음
- Li et al.(2024)의 연구에 따르면, 대안 모형은 금융 불평등 해소뿐 아니라 금융 회사의 이익 증대에도 기여할 수 있는데, 해당 연구에서 개발한 대안 모형은 신파일러의 대출 승인율을 높이고(금융 포용 증진), 연체율을 낮추었음(이익 증가)

● **본고는 국내 주요 사례와 해외 문헌을 기반으로 대안신용평가모형의 현황을 살펴보고자 함**

**국내 주요 사례**

● **씨즈데이터는 비정형(텍스트) 데이터를 활용한 대안 모형 '다Zoom(DaZoom)'을 개발함**

- 다Zoom은 통장에 찍히는 수취인·송금인 정보나 이체 메모 같은 통장 적요 등을 활용함
- 하나은행은 다Zoom을 도입한 이후 대출 승인율이 16.4%에서 27.7%로 상승하였으며, 대출 불량률(대출 신청 1년 내 연체 60일 이상인 비율)은 0.88%에서 0.74%로 하락했다고 밝힘

● **통신대안평가<sup>3</sup>는 이동통신 가입자의 통신 데이터를 활용한 대안 모형 '이퀄(EQUAL)'을 개발함**

- 통신대안평가는 전 국민이 이용하는 통신 3사의 통신 데이터를 활용하여, 최고 수준의 데이터 커버리지를 통해 신파일러를 포함한 모든 국민에게 차별 없는 공정하고 평등한 신용평가를 제공할 것이라고 밝힘

1 금융 포용(Financial Inclusion)은 모든 개인과 기업이 합리적인 비용으로 자신의 필요에 맞는 금융 서비스(예금, 대출 등)에 접근하고 이용할 수 있는 상태를 의미하는데, 금융 포용의 관점에서 볼 때, 전통적인 신용평가모형은 다수의 개인이 주요 금융 서비스로 접근하는 것을 제한하는 장벽으로 작용하고 있음

2 대안정보란 기존 신용평가모형에서 자주 사용되는 금융거래 정보가 아닌 통신 데이터, 소셜미디어 활동, 소비 패턴 등과 같은 새로운 형태의 대안 데이터, 즉 비금융정보를 의미함

3 통신대안평가는 통신 3사(SK텔레콤, KT, LG유플러스)가 65억 원씩, KCB와 SG서울보증이 전략적 투자자로 27.5억 원씩 출자해 세운 합작법인

- 2,000여 개의 세부 데이터 항목을 토대로 고객의 ▲ 라이프스타일, ▲ 소비 패턴, ▲ 거래 능력, ▲ 연속성, ▲ 관리 성향 등 5개의 정보 영역을 심층적·다층적으로 분석함
- 이퀄은 기본적으로 통신 연체 이력에 통신 데이터 및 통화 사용 정보를 결합해 사용자가 1년 내 연체를 할 가능성이 있는지를 평가하는 구조임
- 예를 들어, 연체 이력이 없는 사용자가 자기의 데이터 사용량에 맞는 요금제를 쓰면서 다른 상품과 결합해 요금을 할인받을 경우 올바른 경제관념을 가진 것으로 판단하여 높은 점수를 부여함
- 또한, 시간대별 데이터 사용량이나 통화량을 분석해 규칙적으로 생활하는 사용자 역시 연체할 가능성이 낮다고 평가함
- 자체 테스트 검증 결과, 기존 전통 모형과 결합하여 사용하는 경우 시너지 효과가 발생하는 것으로 나타남
- 특히 신파일러와 외국인에 대한 변별력이 각각 20%, 46% 향상했으며, 불량률은 절반 수준으로 감소함
- 이퀄은 대출 심사, 카드 발급 등 다양한 금융 의사결정에 활용될 수 있음

#### ● 카카오뱅크는 2022년부터 자체 개발한 대안 모형 ‘카카오뱅크 스코어’를 운영하고 있음

- 카카오뱅크 스코어는 ‘선물하기’, ‘택시 승차 정보’, ‘도서 구매 이력’과 같은 비금융정보를 대출 심사에 활용함
- 교보문고·에스24·롯데멤버스·카카오모빌리티 등 11개 기관에서 가명 처리된 데이터를 제공받아서 개인신용과 유의미한 관계가 있는 변수들을 발굴함
- 저연령층 또는 신파일러인 경우, 기존 모형보다 카카오뱅크 스코어가 더 높은 변별력을 보임(즉, 금융 포용 향상에 기여)
- 카카오뱅크는 기존 모형으로 거절된 고객을 대상으로(즉 차주에게 유리할 때만 적용하는 포지티브 방식) 카카오뱅크 스코어를 활용하여 우량 고객 약 6.5만 명을 선별해 약 4,600억 원 규모의 중저신용 대출을 추가 승인함

#### ● 네이버페이는 나이스평가정보와 공동 개발한 대안 모형 ‘네이버페이 스코어’를 외부에 판매하여 수익을 창출하고 있음

- 네이버페이 스코어는 ▲ 네이버페이 이용 내역, ▲ 마이데이터 기반의 자산 데이터 등으로 신용위험을 평가하며, 차주에게 유리할 때만 적용되는 포지티브 방식임
- 관계자에 따르면, 단순히 결제액이 많다고 좋은 점수가 부여되는 것은 아니며, 소비 활동의 규칙성과 지속성, 꾸준한 투자 활동 등을 중요한 요소로 평가함
- 시뮬레이션 결과, 온라인 플랫폼 대출 비교 이용자의 신용평가에 네이버페이 스코어를 반영할 경우 신용평가모형의 변별력 지표가 기존 모형 대비 13.57%p 개선되었음

## 관련 해외 연구

- **대안정보를 활용한 신용평가모형 개발 및 검증에 대한 연구가 최근 활발하게 진행되고 있음**
  - Li et al.(2024)은 대안정보를 활용한 AI 모형의 금융 포용 증대는 모형의 향상된 정확성에서 기인하고, 정확성 향상은 대안정보의 사용과 고급 분석 기법(머신러닝 기법으로, 예를 들어 XGBoost)에서 기인하다고 하였음
  - 즉 대안정보의 사용과 고급 분석 기법이 신용평가 정확성을 향상시키고, 결과적으로 금융 포용이 증대되는 것임
  - Lee et al.(2024)은 식료품 구매 데이터를 활용하여 신용카드 연체 여부를 예측하는 모형을 개발하였으며, 신파일러 대상으로 기존 모형에 식료품 구매 데이터를 추가하는 경우 가장 우수한 성능이 나타남을 실증적으로 보임으로써, 식료품 구매 데이터의 가치를 확인함
  
- **한편, 금융 회사는 대출 심사 과정에서 수익성과 금융 포용 간의 상충 관계에 직면하게 될 수밖에 없음**
  - 대출 상환 가능성이 높은 사람에게 초점을 맞추면 수익성이 증가하지만 금융소외계층에게는 불이익을 주는 금융 불평등이 발생하며, 금융 포용을 달성하기 위해 대출 대상을 확대하면 연체율이 증가할 수 있으므로 수익성이 하락함
  
- **Lu et al.(2023)은 수익성과 금융 포용을 동시에 개선할 수 있는 방법을 제안하였는데, 구체적으로 다음과 같은 대안정보의 활용을 통해서 어떤 유형의 대안정보가 수익성 증대 및 금융 포용 개선에 효과적인지 살펴봄**
  - ① 모바일: 운영 체제, 통화 빈도, 연락처, 앱 사용, GPS 이동 경로 등 31개
  - ② 소셜미디어(웨이보): 팔로잉 수, 팔로워 수, 게시물 수, 댓글 수, 좋아요 수 등 15개
  - ③ 온라인 쇼핑: 구매 상품, 구매 금액 등 40개
  
- **이 연구는 핀테크 기업과 협력하여 일정 기간 동안 모든 대출 신청을 승인하는 메타실험 설계를 사용함**
  - 즉 거절되었을 고객의 데이터도 수집함으로써, 모든 경우의 데이터를 학습에 사용할 수 있었음
  - 이러한 접근 방식은 기존 학습 방법에서의 가장 큰 이슈 사항인 학습 데이터의 편향(bias)을 해결할 수 있음
  
- **분석 결과, 대안정보의 유형마다 서로 다른 결과가 나타났는데, <표 1>과 <표 2>에서 보면 알 수 있듯이 전통적 데이터와 비교하여,**
  - ① 모바일: 수익성 향상 및 금융 불평등 완화에 상당한 기여

- ② 소셜미디어(웨이보): 수익성 향상 및 금융 불평등 완화에 기여, but 모바일보다 약한 기여
- ③ 온라인 쇼핑: 수익성 향상에 기여, but 금융 포용 개선에 전혀 기여하지 않음

● 이 연구는 대안정보를 활용한 개인신용평가모형이 수익성 향상뿐 아니라 금융 불평등 해소에도 도움을 준다는 사실을 발견하였음

- 특히, 대안정보의 유형에 따라 성과에 기여하는 바가 서로 다를 수 있다는 점에 주목할 필요가 있는데, 온라인 쇼핑 데이터는 소득 수준과 강한 상관관계를 가지므로, 전통 데이터와 유사한 대출 승인이 발생한 것으로 보임

■ 표 1 - 수익성 분석: 다양한 대출 승인 기준 및 데이터셋에 따른 수익성 변화

(단위: 천 달러)

데이터셋	대출 승인 기준(왼쪽으로 갈수록 엄격)								
	5%	10%	15%	...	40%	45%	50%	...	70%
1. 전통	1.65	3.67	5.83		12.99	13.92	<b>14.73</b>		-5.65
2. 온라인 쇼핑	2.26	4.86	7.35		15.12	<b>15.75</b>	15.15		-6.01
3. 모바일	2.47	5.44	8.05		16.23	<b>16.96</b>	16.90		-6.87
4. 소셜미디어	2.46	5.41	7.96		16.00	<b>16.45</b>	16.20		-6.49

참고1. 대출 승인 기준 10%라고 하면, 대출 신청자 중에서 상위 10%에게만 대출 실행.

참고2. 거의 모든 데이터셋에서 승인 기준 45%일 때 수익이 가장 높음.

참고3. 서로 다른 대안정보 중에서 모바일이 가장 효과적: 수익 22% 향상((16.95-13.92)/13.92).

■ 표 2 - 금융 포용 분석: 데이터셋에 따른 대출 승인 고객의 특성 비교(대출 승인 기준 45% 대상)

	데이터셋			
	1. 전통	2. 온라인 쇼핑	3. 모바일	4. 소셜미디어
거주 도시 소득(단위: 달러)	7,805	7,650	6,833	6,972
개인 소득 수준	5.28	5.20	4.55	4.69
소득 대비 대출	1.18	1.20	1.33	1.31
교육 수준	4.10	4.05	3.88	3.90
주택 소유	0.50	0.49	0.42	0.42

참고1. 전통 데이터와 온라인 쇼핑 데이터를 사용했을 때, 대출 승인 고객의 특성이 거의 동일하게 나타났는데, 이것은 금융 포용 측면에서 온라인 쇼핑 데이터의 활용 가치가 낮다는 것을 시사함. 반면, 모바일과 소셜미디어의 경우에는 전통 데이터와 비교하여 고객 특성의 차이가 존재하며, 이것은 다른 유형(저소득층)의 고객에게 대출을 승인해준다는 것을 말함. 즉 포용적 대출이 이루어짐.

참고2. 개인 소득 수준: 1=150달러 미만; 2=150~300달러; 3=300~450달러; 4=450~600달러; 5=600~750달러; 6=750~900달러; 7=900~1,050달러; 8=1,050~1,200달러; 9=1,200달러 초과.

참고3. 소득 대비 대출: 대출 금액(달러)/(개인 소득 수준×1,000-500)

참고4. 교육 수준: 1=중학교; 2=전문직업학교; 3=고등학교; 4=고등기술학교; 5=대학교; 6=대학원.

참고5. 주택 소유: 1=보유; 0=미보유.

## 결론

- **전통적인 신용평가모형의 한계를 극복하기 위해 대안정보와 머신러닝을 활용한 새로운 형태의 신용평가모형이 등장하고 있음**
  - 선행 연구들은 웹 브라우징 데이터(Rozo et al., 2023), 이메일 패턴(Djeundje et al., 2021), 배달 앱 거래 데이터(Chioda et al., 2024), 온라인 쇼핑 내역(Lu et al., 2023) 등 다양한 형태의 대안정보를 활용하여 신용평가모형을 개발하고 성능을 검증함
- **국내 사례와 해외 문헌을 통해 비금융 데이터를 활용한 대안신용평가모형이 금융 포용과 금융 회사의 수익성 향상에 기여할 수 있음을 확인함**
  - 국내 사례(씨즈데이터, 통신대안평가, 카카오뱅크, 네이버페이)는 비정형 데이터, 통신 데이터, 소비 및 생활 패턴 정보 등을 활용하여 전통 모형 대비 높은 변별력을 보였으며, 신파일러와 금융소외계층의 신용 접근성을 확대하는 성과를 나타냄
  - 국외 연구에서도 모바일 데이터와 소셜미디어 데이터가 금융 포용 및 수익성 측면에서 긍정적 효과를 보였으며, 특히 모바일 데이터의 성과가 가장 우수했음
- **향후 신용평가의 미래는 머신러닝과 다양한 비금융 대안정보 활용의 확대가 예상됨**
  - 전통 금융 회사 및 핀테크 기업은 혁신적인 대안신용평가모형을 개발하여 전통적인 신용평가 방식으로는 평가가 어려운 비신용자(Unscored Consumers) 또는 저신용자(Underserved Consumers)에게 신용을 제공함으로써, 수익성을 높일 수 있음
  - 즉 신용위험을 효과적으로 정확하게 예측하여 대출 승인율을 높이고 연체율을 낮춤으로써 이자 수익을 극대화할 수 있음
  - 대안정보를 사용하는 경우 프라이버시 이슈가 발생할 수밖에 없으며, 사업자는 법적·윤리적 사항을 충분히 고려할 필요가 있음. 예를 들어, 데이터 수집 및 활용 과정에서 정보 주체의 동의를 명확히 받는 절차를 마련해야 함

## 참고문헌

- Chioda, L., Gertler, P., Higgins, S., & Medina, P. C.(2024), FinTech lending to borrowers with no credit history(NBER Working Paper No. 33208), National Bureau of Economic Research
- Djeundje, V. B., Crook, J., Calabrese, R., & Hamid, M.(2021), Enhancing credit scoring with alternative data, Expert Systems with Applications, 163, 113766
- Li, C., Wang, H., Jiang, S., & Gu, B.(2024), The Effect of AI-Enabled Credit Scoring on Financial Inclusion: Evidence from an Underserved Population of over One Million, MIS Quarterly, 48(4), 1803-1834
- Lu, T., Zhang, Y., & Li, B.(2023), Profit vs. Equality? The Case of Financial Risk Assessment and A New Perspective on Alternative Data, MIS quarterly, 47(4)
- Lee, J. Y., Yang, J., & Anderson, E. T.(2024), Using grocery data for credit decisions, Management Science
- Rozo, B. J. G., Crook, J., & Andreeva, G.(2023), The role of web browsing in credit risk prediction, Decision Support Systems, 164, 113879
- 남미래(2024.10.30.), 소외계층 금융사다리 된 신평일러 솔루션…시중은행도 ‘러브콜’, 머니투데이, <https://news.mt.co.kr/mtview.php?no=2024090616355755112>
- 남지현(2024.02.06.), “책 구매..쇼핑 내역도 분석해 대출 신용평가하죠”, 한겨레, <https://www.hani.co.kr/arti/economy/finance/1127322.html>
- 송기영(2024.08.12.), SBI저축은행·롯데카드, 통신데이터 신용평가 ‘이퀄’ 도입, 조선일보, <https://www.chosun.com/economy/money/2024/08/12/VQTEFTVZQA22ZOIMVAXKBHVLDA/>
- 정다은(2024.03.21.), 네이버페이, 대안신용평가모형 ‘네이버페이 스코어’ 개발…케이뱅크·SBI저축은행 도입, 전자신문, <https://www.etnews.com/20240321000041>