

# 가계대출 연체율 예측을 통한 시도별 정책 방향성 연구

이수진<sup>1</sup>, 원정인<sup>2</sup>, 강희용<sup>3</sup>, 이인성<sup>4</sup>, 김건<sup>5</sup>, 김진<sup>6</sup>

<sup>1</sup> 상명대학교 지능·데이터융합학부 학부생

<sup>2</sup> 성신여자대학교 수리통계데이터사이언스학부 학부생

<sup>3</sup> 상명대학교 경영학부 학부생

<sup>4</sup> 한국외국어대학교 통계학부 학부생

<sup>5</sup> 상명대학교 융합전자공학부 학부생

<sup>6</sup> 상명대학교 빅데이터융합전공 교수

owlemily01@naver.com, jeongin323@naver.com, heeyongyong83@gmail.com, lgtkpr4132@naver.com, kykg022@naver.com, jinkim@smu.ac.kr

## A Study of Policy Direction by City and Province through the Prediction of Household Loan Delinquency Rate

Su-jin Lee<sup>1</sup>, Jeong-in Won<sup>2</sup>, Hee-yong Kang<sup>3</sup>, In-seong Lee<sup>4</sup>, Gun Kim<sup>5</sup>, Jin Kim<sup>6</sup>

<sup>1</sup>Faculty of Artificial Intelligence and Data Engineering, Sangmyung University

<sup>2</sup>Dept. of Mathematics, Statistics and Data Science, SungShin Women's University

<sup>3</sup>Dept. of Business Administration, Sangmyung University

<sup>4</sup>Dept. of Statistics, Han-Kook University of Foreign Studies

<sup>5</sup>Dept. of Electronic Engineering, Sangmyung University

<sup>6</sup>Big Data Convergence Major, Sangmyung University

### 요 약

최근 경제침체로 인해 지속되는 연체율 상승의 원인을 지역별 및 시차별로 분석하였다. 독립변수를 가계대출변수, 부동산지수변수, 경제지표변수로 나누었고 통계적 모델링을 통해 총 19 가지 변수로 연체율을 예측하였다. 각 지역마다 상이한 결과가 도출되었는데 이를 바탕으로 지역별 연체율 감소 정책을 제안한다.

#### 1. 서론

최근 코로나 경제침체로 인하여 급격한 인플레이션이 발생했다. 이에 따른 고금리 정책으로 경기둔화와 가계대출 연체율의 상승세는 지속되고 있다. 현재, 전국단위로 연체율을 감소시키기 위한 정책이 시행되고 있다. 금융위원회에서는 저금리 대환 정책, 상생 금융 프로그램, 주택담보 대출의 만기 연장 정책 등을 시행 중이다. 하지만 전국단위 정책은 개별 지역의 특색을 반영하기 힘들다는 한계를 가진다. 지역별 맞춤형 정책 개선을 통해 더 효과적으로 연체율을 감소시킬 수 있다는 관점에 주목하여 연구를 진행하였다. 본 연구에서는 연체율 상승에 영향을 주는 원인과 타이밍을 분석하여 미래 연체율 상승 경로를 확인하였다. 이를 통해 시간적 여유를 가지고 지역별로 시행 중인 정책을 개선하여 안정적 지역 경제를 이루어 낼 수 있을 것으로 예상된다.

#### 2. 분석과정

##### 2-1. 독립변수 선정

본 연구에서 독립변수를 선정하는 과정에서 2023년 1월부터 8월까지의 ‘가계대출 연체’를 검색 키워드로 하여 네이버뉴스를 크롤링하였다. 크롤링 결과를 바탕으로 ‘담보 대출’, ‘부동산’ 등을 주요 용어로 선정하였다. 또한 선행된 연구[1]에서 도출된 독립변수를 가계대출변수, 부동산지수변수, 경제지표변수 3 분야로 나누어 <표 1> 같이 19 개의 독립변수를 선정하였다.

##### 2-2. 시차적용

모델링을 하기에 앞서, 19 개의 독립변수들이 선행되고 ‘가계대출연체율’이 후행으로 발생한다는 점을 감안하여[2] 가계대출연체율(종속변수)을 3 개월, 6 개월, 12 개월 미루어 상관분석 및 회귀분석을 진행하였다. 상관분석은 종속변수와 관련 있는 독립변수를 확인하기 위해 상관관계 절댓값이 0.5 이상인 변수만을 추출하였고, 회귀분석은 지역별 최적 시차 선정을 위해 해당 지역에 어떤 시차의 결정계수가 가장 높은 값으로 나타나는지 확인했다. 그 결과는 <표 2>에서 확인할 수 있다.

변수 분야	변수	출처	시차	지역	결정계수
가계대출 변수	기본 금리	한국은행 경제통계시스템	3개월	경기도	0.9160
	비은행 예금회계기준 가계대출금	한국은행 경제통계시스템		경상남도	0.8880
	예금은행 기타 가계대출금	한국은행 경제통계시스템		광주광역시	0.9580
경제지표 변수	예금은행 주택담보 대출금	한국은행 경제통계시스템	6개월	대구광역시	0.9640
	실업률	KOSIS		전라북도	0.9180
	전환 대비 이윤율	KOSIS		충청남도	0.8950
	고용률	KOSIS		경상북도	0.7970
	성장방식 수	대한민국 외교부의 주시계보		대전광역시	0.8910
	이동부도율	한국은행 경제통계시스템		울산광역시	0.8790
부동산 변수	소비자물가지수	한국은행 경제통계시스템	12개월	경상북도	0.9490
	교육비 지출 전항지수	한국은행 경제통계시스템		부산광역시	0.9840
	지가 변동률	한국은행 경제통계시스템		서울특별시	0.9530
	주택거래량	한국부동산원		인천광역시	0.9780
	물체수급충량	KOSIS		전라남도	0.9500
	매계수급충량	KOSIS		제주도	0.8490
	전주수급충량	KOSIS		충청북도	0.8250
	아파트매매가리지수	KOSIS			
	부동산지수	한국은행 경제통계시스템			
	부동산시장 소비지수	KOSIS			

<표 1> 독립변수 19 가지 데이터셋 <표 2> 시차적용 결과

### 3. 모델링

지역별 최종 변수는 다중공선성에 강한 ‘랜덤포레스트 회귀 모델’을 통해 선정하였다. 중요도가 높은 순으로 7 개를 선택하여 설명력이 약한 변수를 제거한 뒤 Z-score 정규화로 표준화하였다. 이후 변수간 다중공선성 존재 문제를 해결하기 위해 VIF 가 10 이상인 변수들을 제거한 뒤 최종 독립변수를 선택하였다. 또한, 시차별, 지역별로 나누어 분석하는 것이 통계적으로 유의미한 결과를 도출해 낼 것인가를 검증하기 위해 ANOVA 를 시행하였다. 시차별로 3 가지 지역을 선택하여 표본 수를 같게 만든 후, 1 차로 종속변수끼리, 2 차로 독립변수끼리 분석하여 집단끼리의 분포가 다른 것을 확인했다.

본 연구에서는 5 가지 모델(선형회귀모델, 릿지회귀 모델, 랜덤포레스트, XGboost, LSTM)분석을 통해 지역별 최적 모델을 선정하였다. 모델의 성능평가는 모델의 결정계수와 예측력을 평가하는 MSE 를 사용하였다. 지역별 모델링 분석 결과는 <표 3>과 같다.

[최적모델: LSTM]			[최적모델: 랜덤포레스트]		
지역	R <sup>2</sup>	MSE	지역	R <sup>2</sup>	MSE
부산광역시	0.87	0.000481	경기도	0.91	0.000153
충청남도	0.80	0.000127	광주광역시	0.84	0.000617
			대구광역시	0.84	0.000462
			울산광역시	0.84	0.000388
[최적모델: XGboost]					
지역	R <sup>2</sup>	MSE	지역	R <sup>2</sup>	MSE
강원도	0.64	0.000119	인천광역시	0.77	0.000254
경상남도	0.56	0.000792	전라북도	0.92	0.001868
서울특별시	0.94	0.000171	제주특별자치도	0.48	0.002733
			충청북도	0.14	0.000297

<표 3> 지역별 최적 모델과 결과

### 4. 결론 및 정책제안

최종 선정한 모델과 결과를 바탕으로 지역 및 시차별로 결론이 도출된다. 크게 각 시차별로 대표 1 개 지역의 맞춤 정책을 제안한다.

#### 4-1.3 3개월\_전라북도

전라북도는 2023 년 5 월 기준 타지역보다 연체율이 압도적으로 높았다. 현재 전북은행에서는 저신용자임에도 불구하고 중금리로 대출을 해주는 상품을 판매 중이다. 이로 인해 취약차주의 대출이 용이해져 연체율을 증가 원인으로 해석될 수 있지만, 반대로 불법사금융으로 빠지는 경우를 막아주는 프로그램이다. 그래서 해당 상품과 대출 전 주의사항, 신용개선 방안

등 금융상품에 대한 충분한 교육을 결합한다면, 더 좋은 효과를 볼 수 있을 것으로 예상된다. 또한 신용점수를 등급에 따라 금리에 차등을 두는 상품도 추천한다.

#### 4-2. 6개월\_경상북도

경상북도는 전국에 비해 가계 소득 및 안정성이 높은 편이다. 그러나 고위험 대출자 비율 또한 높다. 이 지역 대출자는 금리 1% 상승 시 대출상환 부담이 1.5%로 전국에 비해서도 낮은 반면 고위험대출자는 대출상환 부담 4.5% 증가로 타 지역에 비해 금리 변화에 민감하다. 따라서, 경상북도는 고위험 대출자를 타겟으로 리스크 관리 방안을 마련하는 것을 제안한다. 기타가계대출금의 시차를 분석하여 대출 건수의 변화를 예측 후 부도위험이 높은 가계에 대한 규제를 강화하는 직접적인 정책이 필요하다.

#### 4-3. 12개월\_서울특별시

서울특별시는 현재 ‘특례보금자리론’ 시행 후 ‘내 집 마련’을 목표로 하는 2030 세대가 증가하며 이들의 아파트 매수 비중이 30%대를 유지 중이다. 정부가 LTV 를 완화하는 정책을 펼친 것도 영향을 미친 것과 분석결과로 보아 부동산 관련 요인과 고금리, 2030 세대의 과도한 지출이 연체율 상승을 가져왔다. 따라서 서울특별시는 장기적인 교육 프로그램을 추진해야 한다. 경제 및 금융에 대한 지식이 부족한 청년들의 상황을 고려해 국가적 차원의 교육을 할 필요가 있다. 2030 세대를 타겟으로 한 금융교육 및 부동산 투자 교육 진행은 향후 연체율 감소 효과를 가져올 수 있다.

#### 4-4. 한계 및 추후 연구

본 연구는 거시경제 변수를 바탕으로 지역별 연체율 증가 요인을 분석하였기에, 미시적 요소를 고려하지 못한 한계점이 있다. 따라서 후속 연구에서는 연구 외적 변수(개인 데이터)와의 상호작용을 반영하여 분석하는 것을 제안한다. 시간이 지남에 따라 발생하는 지역별 연체율 데이터가 추가된다면, 높은 신뢰성과 성능을 제공하는 모델을 개발할 수 있을 것이다.

이 논문은 2023 년도 과학기술정보통신부의 재원으로 한국데이터산업진흥원-2023 데이터청년캠퍼스 지원을 받았습니니다.

#### 참고문헌

- [1] 이수호, 김동일., 거시경제지표를 활용한 주택담보대출 연체율 결정요인 분석., 비즈니스융복합연구, 41-46, .2023.04.
- [2] 전홍배, 이정진, 최운열, 스트레스테스트를 활용한 은행 가계대출부문의 안정성 연구, 한국은행 금융안정분석국 리스크분석팀, 14 권 2 호, 71-100., 2008.06.03.