

채무상환 유형별 패턴 및 행태 분석

- 머신러닝 방법론을 활용 -

이유정 한국자산관리공사
이혜진 한국자산관리공사
이현탁* 한국자산관리공사

요약 본 연구는 한국자산관리공사가 운영한 국민행복기금 데이터를 활용해 채무상환 유형별 패턴 및 행태에 대해 분석하였다. 본 연구에서 고려한 채무상환 4 가지 유형은 기한이익상실(실패)와 분할상환, 일시상환, 조기상환(성공)이며, 국민행복기금 약정체결 시 얻을 수 있는 정보를 활용해 다항분류(multiclass classification)를 수행하였다. 분석 결과, 실감면율과 약정횟수가 전체 채무상환 프로세스에서 가장 큰 영향을 미쳤으며, 4 가지 채무상환 유형별로 채무상환 결과에 미치는 결정요인이 상이한 것으로 나타났다. 본 연구는 약정건별 채무상환 유형별 확률을 도출하는 등 개인신용평가를 수행한데 기여했으며, 실감면율과 같은 정책변수를 시뮬레이션하여 기한이익상실 위험을 낮추고 완제가능성을 높이는 전략에 대해 논하는 등 실용적인 연구를 수행한데 의의가 있다.

주요단어 국민행복기금, 실감면율, 머신러닝, LightGBM, SHAP, 시뮬레이션

본 논문은 한국자산관리공사 조사연구보고서인 “채무상환 유형별 패턴 및 행태 분석” 내용을 바탕으로 작성되었습니다. 본 논문에 대해서 유익한 조언과 도움을 주신 김원대 원장님, 박정환 소장님, 이광제 팀장님, 박소연 팀장님 등 공사임직원 분들과 좋은 커멘트를 주신 부산대학교 윤성민 교수님께 감사드립니다. 본 논문의 내용 및 주장은 공사의 공식견해와 무관한 저자의 개인적인 견해를 밝힙니다.

* 교신저자. 주소: 48400, 부산광역시 남구 문현 금융로 40 부산국제금융센터(BIFC), 한국자산관리공사; E-mail: leeht@kamco.or.kr; Tel: +82-51-794-2935.

I. 서론

1997년 IMF 외환위기, 2008년 글로벌 금융위기와 같은 경제위기 상황에서 우리나라는 금융회사의 대규모 부실채권을 인수 및 정리하는 정책으로 대응하였다. 특히, 이 시기 동안 한국자산관리공사(이하, 공사)는 금융회사 부실채권 인수·정리 전문기관으로서 금융회사로부터 개인 연체채권을 인수하여, 채무상환에 어려움을 겪는 채무자의 채무부담 완화 및 경제활동 재기지원 역할을 수행하였다. 연체중인 개인무담보채권을 공사가 직접 인수하거나 국민행복기금과 같이 민간금융회사 등과 공동으로 설립한 기구를 통해 인수하는 등 정책적 채무조정 프로그램을 주도적으로 관리하였다. 공사가 인수한 채권을 채무자별 상환여건에 따라 감면율을 적용하고 장기분할 상환을 할 수 있도록 채무조정 제도를 운영함으로써, 채무자가 정상적인 금융활동 주체로서 조기에 복귀할 수 있도록 지원한다는 정책목표에 맞춰 그 역할을 수행하여 왔다.

본 연구의 목적은 국민행복기금 채무조정을 통해 축적된 데이터를 활용하여 보다 포용적 정책금융을 선도하기 위한 채무조정 관리방안을 고안하는 것이다. 또한, 4차 산업혁명 시대에 대내외적으로 빅데이터를 활용한 연구 수요가 증가하는 시대적 흐름 속에서 머신러닝(machine learning) 등 최신 기법을 활용한 착수 연구로서 의의가 있으며, 데이터 경제 활성화에도 이바지할 것으로 기대한다. 이에 본 연구는 과거 국민행복기금 약정자의 상환 패턴 분석을 통한 예측 모형을 구축하고, 상환 확률을 예측하는 한편 서민신용회복 효과 제고를 위한 가상의 시뮬레이션을 수행하는 등 데이터 기반 채권 관리방안을 제안하고자 한다.

구체적으로, 본 연구는 국민행복기금 약정자의 채무상환 패턴 분석을 통한 ①종결유형 예측, ②완제 확률 제고를 위한 채권관리 방향 제안에 중점을 둔다. 약정채결자의 상환행태를 분석하여, 4가지 종결유형(일시/조기상환, 분할상환, 기한이익상실) 결정요인을 분석하고(유형별 구체적 정의는 II.1절 참조), 완제 성공 또는 실패에 영향을 미치는 주요 변수의 크기 및 방향성을 탐색한다. 분석 결과를 바탕으로 ①약정채결자 종결유형의 예측, ②감면율 등 정책 또는 약정조건 변화에 따라 완제종결 확률 변화의 시뮬레이션을 수행할 것이다.

연구대상은 국민행복기금이 출범한 2013년 5월 이후 채무조정약정을 체결한 개인·일반채무자이며, 국민행복기금 출범 이후의 개인·일반 채무조정약정 종결(완제·기한이익상실) 데이터에 대하여 Microsoft社의 LightGBM 알고리즘을 적용하였다. 분석 결과, 실감면율과 약정횟수가 전체 채무상환 프로세스에서 가장 큰 영향을 미쳤으며, 4가지 채무상환 유형별로 채무상환 결과에 미치는 결정요인이 상이한 것으로 나타났다. 본 연구는 약정건별 채무상환 유형별 확률을 도출하는 등 개인신용평가를 수행한데 기여했으며, 실감면율 시뮬레이션을 통해 기한이익상실 위험을 낮추고 완제가능성을 높이는 전략에 대해 논하는 등 실용적인 연구를 수행한 것에서도 의의가 있다.

개인채무자의 채무조정 데이터에 관한 연구는 그 동안 자료 접근성 문제 등의 이유로 거의 진행되지 못했던 분야 중 하나다. 기존 채무조정에 관한 국내외 연구도 채무조정 이용자의 특성, 정책효과 등에 관한 기초 통계 연구 위주로 이루어져 왔다(남주하 외 3인, 2014; 박정수&남주하, 2017; 이수진 외 2인; 2019). 본 연구와 가장 관련성이 높은 선행연구는 김형준 외 2인(2019), 김원대&정영우(2020) 및 김원대(2021)가 있으며, 과거 국민행복기금 약정정보를 활용하여 기한이익상실 가능성에 미치는 이항로지분석을 수행해 결정요인을 추정하고 그 영향력을 분석하였다. 분석 결과, 국민행복기금 채무조정 차주의 기한이익상실 가능성은 약정체결금액, 실감면율 및 재약정 여부 등에 통계적으로 유의한 영향을 받는다고 주장했다. 본 연구는 이전의 선행연구와 비교할 때 기한이익상실뿐 아니라 일시상환, 분할상환 및 조기상환 등 완제범주를 세분화하여, 유형별 결정요인을 분석하고 그 영향력을 확인했다는 점에서 차별점이 존재한다. 그리고 상환패턴 및 행태를 분석하고 시뮬레이션을 수행함으로써 유형별 영향력 차이 유무를 확인하고자 했다는 점에서 본 연구의 특징을 찾아볼 수 있겠다.

본 연구의 구성은 다음과 같다. II장은 본 연구에서 활용한 데이터 및 알고리즘에 대해 설명하고, III장은 분석 결과에 대해 살펴본다. IV장은 실감면 및 약정체결금액을 조정하여 기한이익상실(실패) 위험을 낮추고 상환 가능성을 높이는 시뮬레이션 결과를 탐구하고, V장은 주요 결과를 요약하고 향후 연구방향에 대해 논한다.

II. 데이터 및 알고리즘

1. 데이터

본 연구는 2013년 5월부터 2020년 5월까지 체결된 국민행복기금 채무조정약정 중 완제 또는 기한이익상실로 종결된 282,794건의 개인채무자 약정 정보 샘플을 대상으로 한다. 종속변수는 채무상환 종결유형으로 기한이익상실 120,649건, 분할상환 95,121건, 조기상환 17,964건 및 일시상환 49,060건으로 구성되어 있으며, 유형별 정의는 다음과 같다.

- (기한이익상실) 분할상환금의 상환을 6개월 이상 연체한 경우 등 기한의 이익상실사유가 발생하는 경우
- (일시상환) 채무조정 약정 체결 후 6개월 이내에 채무부담액 전액 상환하는 경우
- (분할상환) 채무조정 약정 체결 후 최장 10년 이내 매월 또는 분기별 균등 분할하여 전액상환하는 경우
- (조기상환) 채무조정 약정 체결 후 1년 이상 경과, 12회차 이상 납부한 이후 잔여 채무부담액을 일시에 모두 상환하는 경우

참고로 분할상환, 일시상환, 조기상환은 채무를 갚은 완제한 경우로 ‘성공’으로 구분할 수 있고, 기한이익상실은 ‘실패’로 구분할 수 있다. 독립변수는 무소득여부, 약정체결금액 및 감면율 등 약정당시 얻을 수 있는 데이터 17개로 구성되어 있다(<표 1> 참조).

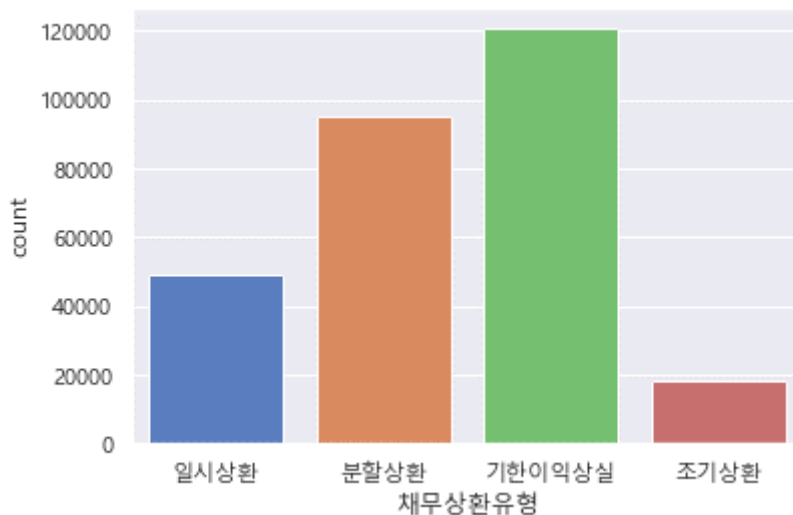
<표 1> 설명변수 기술통계량

변수	평균	표준편차	중앙값	최대값
계좌개수	2.5	2.0	2	14
금융기관수	1.7	1.2	1	10
약정시점연령	43.0	10.3	44	75
회생생계비	1,342,218	601,848	974,899	3,428,415
약정횟수	1.5	0.9	1	7
성별	0.6	0.5	1	1
채약정여부더미	0.3	0.5	0	1
무소득여부	0.6	0.5	1	1

감면율정책변화터미('16~)	0.2	0.4	0	1
시효연장횟수	0.1	0.3	0	2
약정채결금액	3,589,369	4,652,209	2,005,680	4,366,743
채무조정대상원금	6,871,369	8,861,767	3,790,859	8,399,991
회수예상가	377,395	2,786,628	0	47,947,938
부양가족수	1.8	1.2	1	6
연체기간	89.7	54.2	83	280
총채무액	10,706,394	13,151,062	5,830,470	13,899,790
실감면율	46.9	12.2	50	92

<그림 1>의 국민행복기금 채무상환 종결유형 건수를 보면, “기한이익상실”이 가장 많은 비중을 차지함을 확인할 수 있다. 국민행복기금 채무조정은 금융회사에서 상각한 장기연체채권을 대상으로 하는 만큼, 시간이 흐를수록 상환가능성이 낮아지는 부실채권에 대한 우려된 시각이 존재하였다. 그럼에도 불구하고, “일시상환”, “조기상환” 등을 포함한 완제종결 건수가 162,145건에 달하는 것은 소득이 없거나 낮아도 채무자의 상황을 고려하여 다양한 채무조정제도를 운영함으로써, 완제가능성을 높였다는 것을 시사한다. “조기상환”의 경우 데이터 분석 상 결과값이 상대적으로 적어 완제에 미치는 요인들을 탐구해 일반화하기 어려운 측면이 있어, 향후 결정요인 분석을 위한 보완이 필요할 것으로 보인다.

<그림 1> 국민행복기금 채무상환 종결유형 건수(2020년 5월 기준)

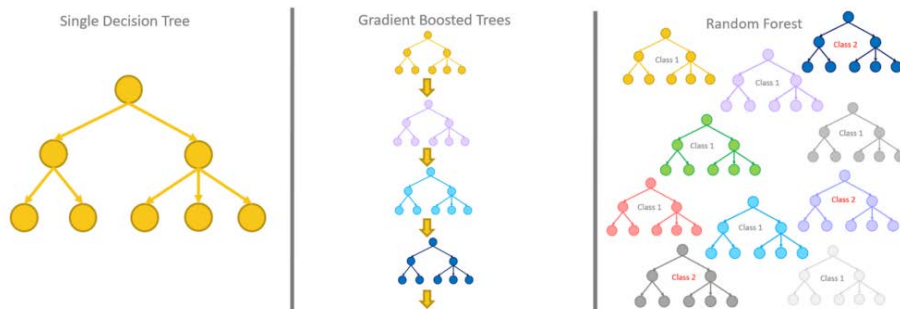


2. 알고리즘

본 분석은 Microfost사에서 개발한 LightGBM 알고리즘을 활용해 채무상환 유형을 판별하는 다항분류(multiclass classification)를 수행하였다. LightGBM 알고리즘은 기본적으로 데이터 또는 변수를 랜덤하게 선택해서 개별 의사결정트리(decision tree) 모형을 만들고, 이 개별 예측결과의 편차를 줄여가면서 예측력을 제고하는 대표적인 부스팅 계열의 앙상블(ensemble) 방법 중 하나이다.¹ LightGBM은 대용량 데이터 처리가 가능하고, 다른 GBM모형 대비 속도가 빠를뿐만 아니라, GPU와 같은 하드웨어 성능 지원도 가능한 장점을 가지고 있는 반면, leaf-wise라는 독특한 추정 방식으로 인해 과적합이 나타나는 단점을 가지고 있다.

LightGBM은 모수 추정하기 전에 셋팅하여 모형의 성능을 끌어올리는데 사용될 수 있는 초모수(hyperparameter)가 존재한다. 본 연구는 LightGBM에서 제공하는 다양한 초모수 중에서 ‘class_weight’, ‘boosting_type’, ‘num_leaves’, ‘max_depth’, ‘max_bin’, ‘learning_rate’, ‘subsample_for_bin’, ‘feature_fraction’, ‘min_data_in_leaf’, ‘lambda_l1’, ‘lambda_l2’, ‘min_child_weight’, ‘extra_tree’ 를 사용하였다.²

<그림 2> 의사결정트리 방법론



주) 좌측은 기본적인 의사결정트리, 중간은 앙상블 방법 중 부스팅 계열, 오른쪽은 배깅 계열을 묘사하였다.
자료) <https://laptrinhx.com/ensemble-models-bagging-boosting-237685625/>

¹ 기타 배깅(bagging) 방식의 앙상블 방법론이 있는데, 대표적인 알고리즘으로는 랜덤포레스트(random forest) 알고리즘이 있다.

² 각 초모수에 대한 설명은 lightgbm.readthedocs.io/en/latest/를 참고하길 바란다.

3. 데이터전처리

본 연구는 크게 세 가지의 데이터전처리를 수행하였다.

첫째, 채무상환 유형별 패턴 분석에 활용한 독립변수(<표 1> 참조)는 백분위(percentile) 0.1% 이상 및 99.9%이하 값으로 이상치를 보정하였다. 국민행복기금을 통해 약정을 체결하게 되면, 실무자가 약정정보를 전산에 입력하게 되는데 그 과정에서 이따금씩 입력오류가 발생할 수 있기 때문이다(예: 부(-)의 약정금액). 입력오류로 인한 이상치(outlier)를 조정해줌으로써 추정결과의 신뢰도를 높일 수 있는 장점이 존재한다.

둘째, 채무상환 종결유형의 불균형(imbalance)를 해결하기 위해 SMOTE(synthetic minority oversampling technique)을 활용하였다. SMOTE 알고리즘이란 건수가 적은 소수(minority) 종결유형의 건수를 샘플링(sampling)을 통해 가상의 합성(synthetic) 유형을 만들어 건수가 많은 다수(majority)의 개수와 맞춰주는 작업이라고 말할 수 있다. <그림 1>에서 유추할 수 있는 바와 같이, 본 연구 또한 데이터 구조의 불균형이 심하기 때문에 SMOTE 알고리즘을 활용하지 않게 되면, 건수가 가장 많은 ‘기한이익상실’은 많은 학습이 이루어져 예측결과가 우수한 반면, 건수가 가장 적은 ‘조기상환’은 예측결과가 저조하게 도출될 가능성이 크다.

셋째, 국민행복기금 채무상환 전체 데이터의 80%는 모형의 모수 추정에 활용하고, 나머지 20%를 모형의 검증에 활용하였다. 또한, 과적합(overfitting) 문제를 해소하기 위해 추가적으로 중첩교차검증(nested cross validation; nested CV)을 활용하여 최적의 모형 선택 및 초모수의 최적화를 수행하였다. 구체적으로, Nested CV를 구현하기 위해 외부루프는 RepeatedStratified 10-fold CV, 내부루프는 Stratified 5-fold CV를 수행했으며,³ 이 중첩교차검증 또한 5회 반복수행하면서 총 250개(=10 × 5 × 5)의 multiclass logloss값을 비교한다. 최종적으로 가장 작은 multiclass logloss를 도출한 최종 모형 및 초모수를 선택하였다.

³ Stratified K-fold CV는 전체 데이터의 회생종결 및 폐지 비율을 유지하면서 샘플의 fold를 나누는 방식이고, RepeatedStratified CV는 Stratified CV 과정 전체를 랜덤하게 반복 수행하는 방식이다. LightGBM의 초모수 최적화는 Python 패키지 hyperopt를 사용하였다.

4. 예측 결과

<표 2>는 훈련 및 시험데이터의 예측 결과를 보여준다. 훈련데이터(Panel A) 결과, 모형이 전체 채무상환 유형을 정확하게 판별한 비율(Accuracy)은 약 75% $(=(78,371+59,243+13,041+19,041+19,392)/226,235)$ 이며, 실제 기한이익상실이 발생한 경우를 기한이익상실로 예측한 비율(Recall)은 약 81% $(=78,371/96,671)$, 실제 분할상환 완제한 경우를 분할상환으로 예측한 비율은 약 78% $(=59,243/75,996)$ 이며, 실제 조기상환한 경우를 조기상환으로 예측한 비율은 약 91% $(=13,041/14,307)$, 실제 일시상환한 경우를 일시상환으로 예측한 비율은 약 50% $(=19,392/39,261)$ 로 추정되었다.

<표 2> 예측 정오표

Panel A: 훈련데이터

		예측				합계
		기한이익상실	분할상환	조기상환	일시상환	
실제	기한이익상실	<u>78,371</u>	16,642	191	1,467	96,671
	분할상환	12,651	<u>59,243</u>	164	3,938	75,996
	조기상환	600	587	<u>13,041</u>	79	14,307
	일시상환	3,560	16,269	40	<u>19,392</u>	39,261
	합계	95,182	92,741	13,436	24,876	226,235

Panel B: 시험데이터

		예측				합계
		기한이익상실	분할상환	조기상환	일시상환	
실제	기한이익상실	<u>18,260</u>	4,995	128	595	23,978
	분할상환	4,217	<u>13,217</u>	96	1,595	19,125
	조기상환	329	234	<u>2,988</u>	106	3,657
	일시상환	1,226	4,622	56	<u>3,895</u>	9,799
	합계	24,032	23,068	3,268	6,191	56,559

한편, 모형의 모수를 추정(또는 학습)하는데 사용되지 않아 모형의 진정한 성능을 평가하는데 활용되는 시험(Panel B) 결과를 살펴보면, 전체 채무상환 유형을 정확하게 판별한 비율(Accuracy)은 약 68% $(=(18,260+113,217+2,988+3,895)/56,559)$ 로 추정되었다. 실제 기한이익상실이 발생한 경우를 기한이익상실로 예측한 비율은 약 76% $(=18,260/23,978)$, 실제 분할상환 완제한 경우를 분할상환으로 예측한 비율은 약 69% $(=13,217/19,125)$, 실제 조기상환한 경우를 조기상환으로 예측한 비율은 약

82%(=2,988/3,657), 실제 일시상환한 경우를 일시상환으로 예측한 비율은 약 40%(=3,895/9,799)로 도출되었다.

전반적으로, 학습할 수 있는 경우의 수가 많은 기한이익상실의 예측성과가 가장 우수하며, 훈련 및 시험데이터에서 예측 성과가 비교적 고르게 나타난 것을 확인할 수 있다. 특히 일시상환 예측력이 가장 저조한 것으로 나타났는데, 이는 본 연구에서 고려한 17가지 독립변수로 완제종결의 3가지 유형을 구별하기 어렵기 때문인 것으로 추정한다. 향후 추가적인 변수 도입을 통해 3가지 완제종결 유형을 구분할 필요가 있다.

III. 분석 결과

1. 전체 변수중요도(feature importance)

본 절은 훈련데이터를 활용해 추정한 LightGBM 모형에 SHAP(Shapley Additive explanation) 방법론을 적용해 추정결과를 해석하고자 한다. 최근 머신러닝 및 딥러닝 방법이 널리 활용되면서 종속변수와 독립변수간 비선형 관계를 도출하는 등 모형의 복잡도(complexity)가 증가하고 있는데 반해, 결과가 도출되는 원인 및 해석이 어려워 블랙박스(black box)라고 불린다. 이러한 블랙박스 모형을 해석하기 위한 방법 중 하나인 SHAP은 변수중요도를 파악 위해 여러 특성들의 조합을 구성하고 해당 변수의 유무에 따른 총 한계효과(marginal effect)를 수치적으로(Numerically) 구한 값이다.

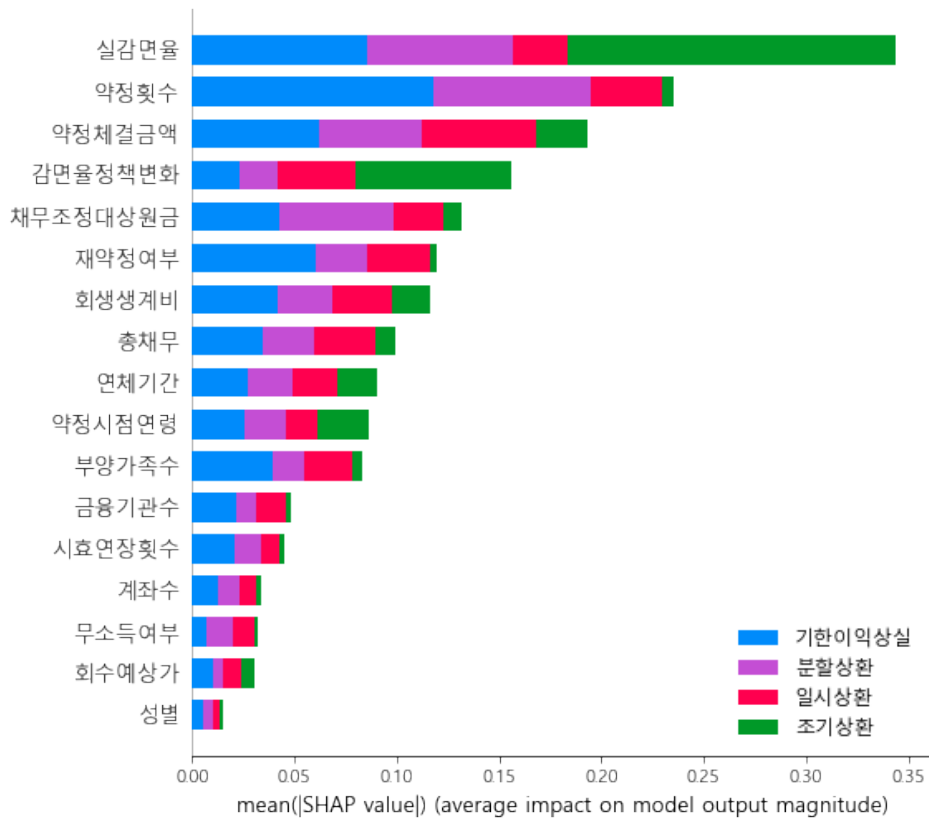
예컨대, 본 연구는 한 약정체결 건당 17개의 정보(<표 1> 참조)를 활용하므로 총 2^{17} 개의 하위(부분집합) 모형을 구성한다. 그 다음, 특정 정보(예: 실감면율)이 포함된 모형의 결과와 제외한 모형의 결과를 비교하여 궁극적으로 특정 정보가 결과에 미치는 영향(즉, 확률의 차이)을 수치적으로 구할 수 있다.⁴ 이에 반해, 전통적인 한계효과 분석은 최종 모형만을 고려하며 특정 정보의 유무에 따른 결과의 차이를 살펴보는 것이 특징이다. 본 연구는 2^{17} 개의 하위(부분집합) 모형에서 특정

⁴ SHAP에 대한 구체적 설명 및 예제는 <https://towardsdatascience.com/shap-explained-the-way-i-wish-someone-explained-it-to-me-ab81cc69ef30>을 확인하길 바란다.

정보 유무가 미치는 SHAP(총 한계효과)을 추정하므로, 최종 모형만 고려한 전통적인 한계효과보다 더 일반화된 결과를 도출할 수 있는 장점이 있다.

본 연구는 분석 시간을 단축하기 위해 훈련데이터(Panel A)에서 소수의 샘플을 추출해 채무상환 유형별 패턴 및 행태에 대하여 살펴보고자 한다. 구체적으로, 총 226,235개의 훈련데이터(모수)의 분포(<그림 1> 참조)를 거의 동일하게 복제한 1,000개의 샘플을 추출했으며,⁵ 그 결과 채무상환 유형별 확률을 예측한 평균결과는 훈련데이터와 동일한 기한이익상실 44%, 분할상환 31%, 조기상환 7% 및 일시상환 17%를 가진다.

<그림 3> 채무상환 유형별 '평균' 변수중요도(내림차순)



주) 1,000건에 해당하는 약정건별 SHAP(총 한계효과)을 구하고 절대값을 취합후 전체 훈련데이터 평균을 도출해 변수중요도를 나타냈다. x축 막대의 크기는 채무상환 유형별 확률을 의미한다.

⁵ 본 연구의 SHAP 분석은 각 요인, 약정체결건, 채무상환 유형별로 SHAP을 17요인×4개 유형 ×1,000 건을 구했다.

<그림 3>은 채무상환 유형별로 미치는 영향도를 내림차순으로 나타낸 것이다. 즉, 채무상환에서 가장 큰 영향을 미치는 요인은 전체 막대의 크기 순서대로 실감면율, 약정횟수, 약정체결금액, ... , 회수예상가, 성별 순으로 나타났다. 그러나, 개별 채무상환 유형별로 약정요인이 미치는 요인은 상이한 것으로 추정되었다. 예를 들어, 약정횟수는 ‘기한이익상실’과 ‘분할상환’에 가장 큰 영향을 미치는 요인이고, 실감면율은 ‘조기상환’에 가장 큰 영향을 미치며, 약정체결금액이 ‘일시상환’에 가장 큰 영향을 미치는 것으로 나타났다(<표 3> 참조).

<표 3> 채무상환 유형별 영향 순위

	기한이익상실	분할상환	조기상환	일시상환
1st	약정횟수	약정횟수	실감면율	약정체결금액
2nd	실감면율	실감면율	감면율정책변화	감면율정책변화
3rd	약정체결금액	채무조정대상원금	약정시점연령	약정횟수
4th	재약정여부	약정체결금액	약정체결금액	재약정여부
5th	채무조정대상원금	회생생계비	연체기간	총채무
6th	회생생계비	총채무	회생생계비	회생생계비
7th	부양가족수	재약정여부	총채무	실감면율
8th	총채무	연체기간	채무조정대상원금	채무조정대상원금
9th	연체기간	약정시점연령	회수예상가	부양가족수
10th	약정시점연령	감면율정책변화	약정횟수	연체기간
11th	감면율정책변화	부양가족수	부양가족수	약정시점연령
12th	금융기관수	시효연장횟수	재약정여부	금융기관수
13th	시효연장횟수	무소득여부	금융기관수	무소득여부
14th	계좌수	계좌수	계좌수	시효연장횟수
15th	회수예상가	금융기관수	시효연장횟수	회수예상가
16th	무소득여부	성별	무소득여부	계좌수
17th	성별	회수예상가	성별	성별

2. 기한이익상실

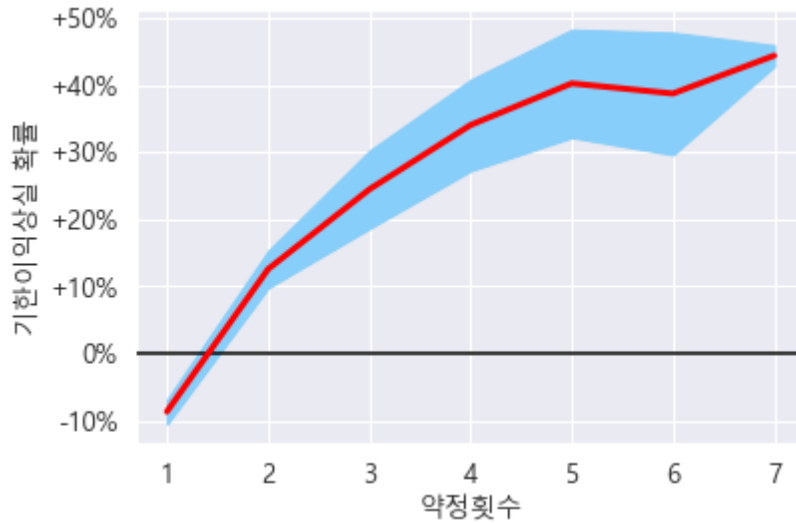
본 절은 기한이익상실에 영향을 미치는 상위 2개 요인에 대해서 구간별 SHAP에 대해 탐구한다(<그림 4> 참조). 참고로 분석에 활용한 1,000건의 평균 기한이익상실 확률은 44%이며, 44% 대비 각 약정횟수의 변화가 기한이익상실 확률을 얼마나 증가 또는 감소시키는지 계산한 총 한계효과(all marginal effects)를 보여준다.

Graph A는 약정횟수가 기한이익상실에 미치는 정도를 나타낸 그림이다. 전반적으로 약정횟수가 많을수록 기한이익상실 확률이 증가하는 것으로 나타나는데, 이는 채무를 갚지 못해 약정을 계속할수록 채무상환능력이 떨어짐을 의미한다. 구체적으로,

약정횟수가 1번인 채무상환 약정은 평균 44% 대비 약 9%p 정도 확률을 감소(=44%-9%)시키는 반면, 약정횟수가 6번일 경우에는 기한이익상실 확률이 약 40%p 증가(=44%+40%)하는 것으로 나타났다.

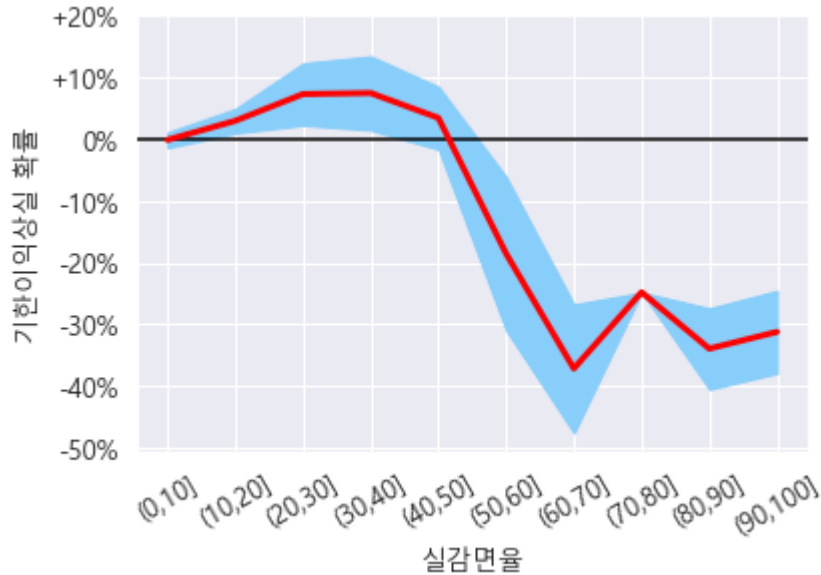
<그림 4> '기한이익상실'에 영향을 미치는 상위 2가지 요인

Graph A: 약정횟수(1st)



주) 0%축은 분석에 활용한 1,000개 약정건의 평균 확률 44%를 의미하며, 빨간색 선은 구간별 평균 값, 파란색 밴드는 표준편차를 의미한다.

Graph B: 실감면율(2nd)



Graph B는 실감면율의 구간대별 기한이익상실에 미치는 한계효과를 나타낸 그림이다. 실감면율이 클수록 기한이익상실 확률이 감소하는 경향을 보이는데, 실감율이 50%이하에서는 기한이익상실 확률이 최대 약 10%p 증가하는 반면, 50%초과에서는 최대 약 40%p 감소하는 것으로 나타났다.

종합하면, 위 2가지 요인이 기한이익상실 가능성을 최대 40~50% 증감시키는 것으로 추정되었다. 즉, 기한이익상실 가능성을 위해서는 약정 이탈을 방지하고 재약정을 최소화하는 일종의 전략이 필요할 것으로 판단된다.

3. 분할상환

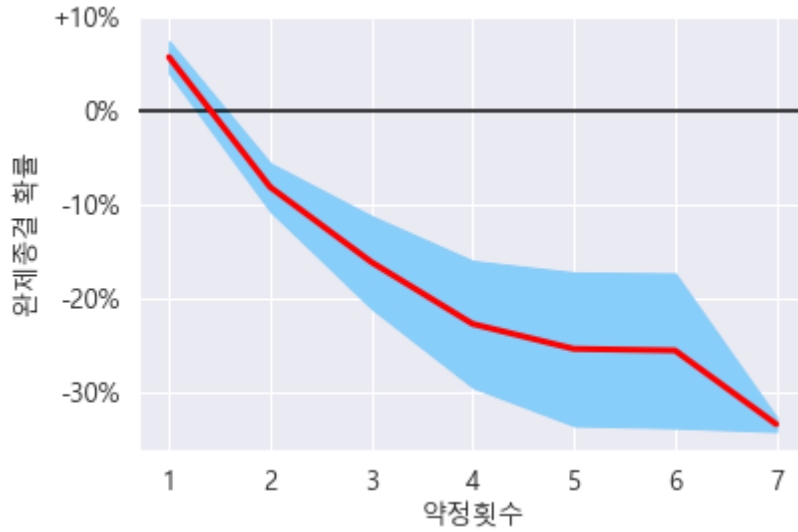
본 절은 분할상환에 영향을 미치는 상위 2개 요인에 대한 구간별 SHAP을 탐구한다(<그림 5> 참조). 분석에 활용한 1,000건의 평균 분할상환 확률은 약 31%이다.

Graph A는 약정횟수가 분할상환 확률에 미치는 영향을 나타낸 그림이다. 전반적으로, 약정횟수가 증가할수록 분할상환 가능성이 감소하는 것으로 나타났는데, 기한이익상실 결과(Graph A; <그림 4>)와 반대임을 확인할 수 있다. 예를 들어, 약정횟수가 1번인 약정은 평균 31% 대비 약 6%p 정도 분할상환 가능성을 증가시키는데 그치지만, 약정횟수가 최대 7번인 경우 분할상환 가능성을 30%p 이상 감소시키는 것으로 나타났다.

Graph B는 실감면율의 구간대별 분할상환 확률에 미치는 영향을 나타냈다. 직관적으로 채무감면을 많이 해줄수록 분할상환 가능성이 증가해야 하지만, 기한이익상실(Graph B; <그림 4>)의 결과와 유사하게 50% 이상에서 분할상환 가능성을 감소하는 결과가 도출되었다. 기한이익상실(실패)과 분할상환(성공)이 비슷한 추세를 보이는 주요 이유는 이 2가지 채무상환 유형 외에 채무를 조기에 갚는 조기상환 및 일시상환 유형 때문인 것으로 추정한다. 만약 기한이익상실과 완제종결(=분할상환+일시상환+조기상환) 2가지 상환 유형만 고려한다면, 실감면율이 분할상환에 미치는 영향은 ‘데칼코마니’처럼 정반대 결과를 도출해야 할 것이다.

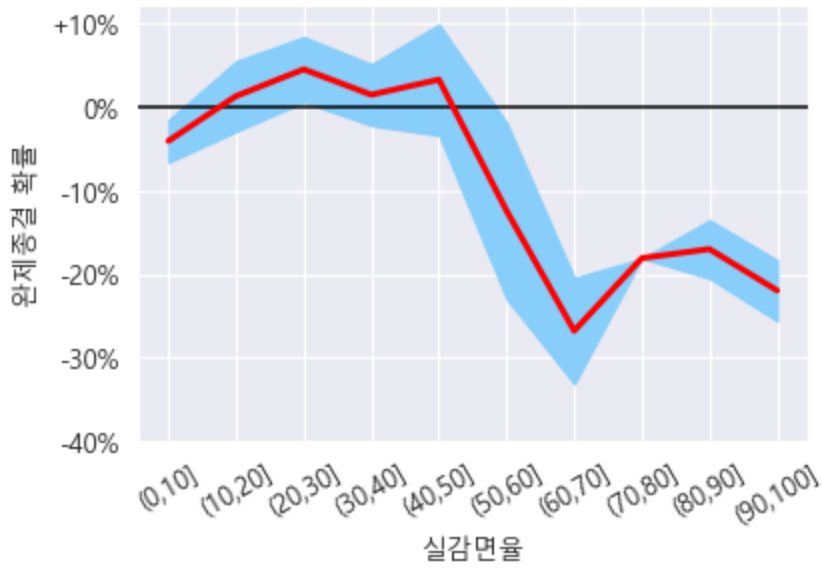
<그림 5> '분할상환'에 영향을 미치는 상위 2가지 요인

Graph A: 약정횟수(1st)



주) 0%축은 분석에 활용한 1,000개 약정건의 평균 분할상환 확률 31%를 의미하며, 빨간색 선은 구간별 평균값, 파란색 밴드는 표준편차를 의미한다.

Graph B: 실감면율(2nd)



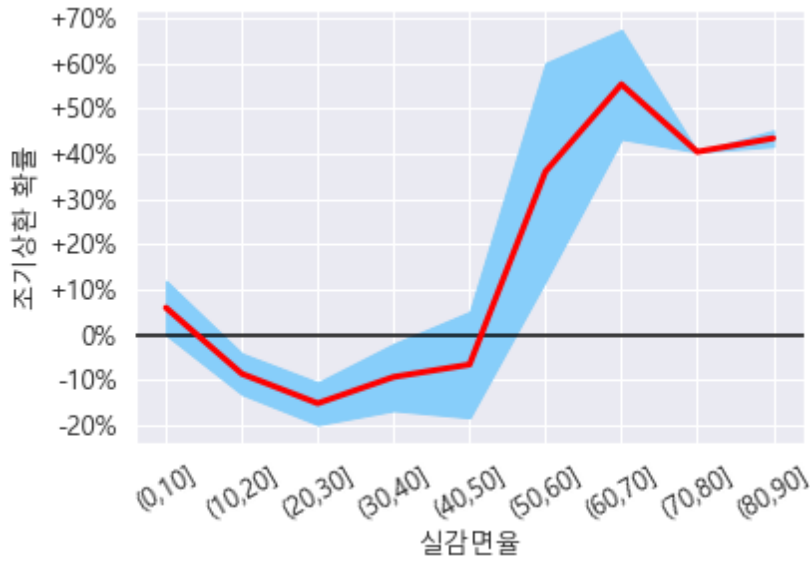
종합하면, 위 두 가지 요인은 분할상환 가능성을 각각 최대 30%p 감소시키는데 큰 영향을 미치는 요인으로 나타났다. 특히, 약정횟수가 기한이익상실과 분할상환을 구분하는 주요 결정요인으로 나타났으며, 기한이익상실의 시사점과 마찬가지로 재약정을 최소화하는 전략이 필요한 것으로 사료된다.

4. 조기상환

본 절은 조기상환에 영향을 미치는 상위 2개 요인의 구간별 SHAP에 대해 탐구한다(<그림 6> 참조). 분석에 활용한 1,000건의 평균 조기상환 확률은 약 7%이다.

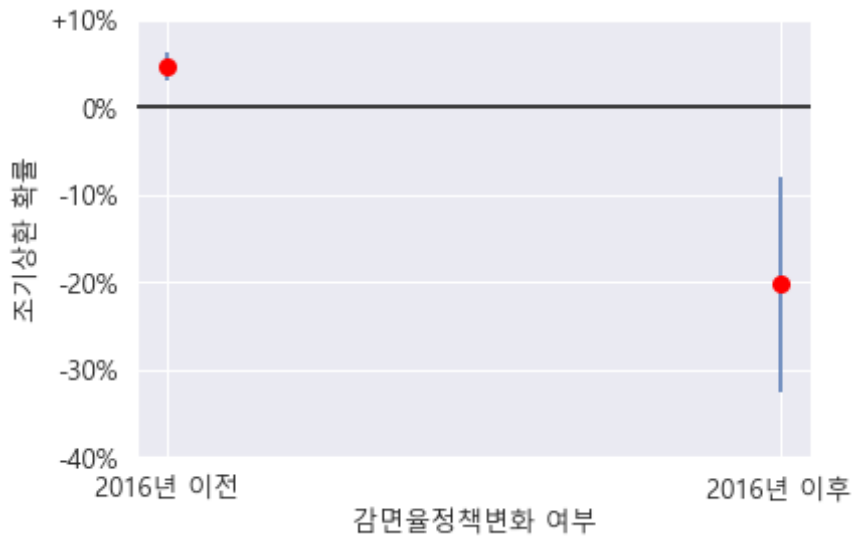
<그림 6> '조기상환'에 영향을 미치는 상위 2가지 요인

Graph A: 실감면율(1st)



주) 0%축은 분석에 활용한 1,000개 약정건의 평균 조기상환 확률 7%를 의미하며, 빨간색 선은 구간별 평균 값, 파란색 밴드는 표준편차를 의미한다.

Graph B: 감면율정책변화(2nd)



Graph A는 실감면율의 구간대별 조기상환에 미치는 영향을 나타낸 그림이다. 여기서 나타난 한 가지 특징은 조기상환 결과가 기한이익상실의 결과(Graph A; <그림 4>)와 분할상환의 결과(Graph A; <그림 5>)와 반대되는 경향을 보인 점이다. 즉, 실감면율이 증가하면 증가할수록, 채무를 조기 상환하고자 하는 동기가 크게 증가하는 것으로 해석할 수 있다.

Graph B는 2016년 감면율정책 전후로 추가 감면을 실시한 결과를 나타낸다. 2016년 이전 정책적인 추가 감면이 없는 경우 조기상환 가능성이 평균 7% 대비 약 5%p 증가(=7%+5%)했지만, 2016년 이후 정책적 추가 감면시 조기상환 가능성이 약 20%p 감소한 것으로 나타났다.⁶ 이 결과가 의미하는 바는 추가 감면 수행 시 채무상환을 12회차 이상 납부하고 이후 잔여 채무부담액을 일시에 상환하는 (중기)조기상환보다 6회차라는 짧은 기간동안 채무를 상환하는 (단기)일시상환에 더 큰 영향을 미치는 것으로 예상된다.

종합하면, 실감면율 증가는 기한이익상실 및 (장기)분할상환보다 조기에 채무를 변제하고 신용회복하려는 동기에 영향을 주는 것으로 나타났다. 이에 대한 논거는 실감면율에 대한 SHAP(Graph A; <그림 6>)이 기한이익상실 및 분할상환 SHAP(Graph B; <그림 4> & <그림 5>)과 반대인 경향을 보여주기 때문이다. 특히, 실감면율 50% 이상인 구간에서 조기상환 가능성이 크게 증가하는 사실은 (장기)분할상환보다 채무를 조기에 상환하고자 동기를 유발할 수 있음을 시사한다.

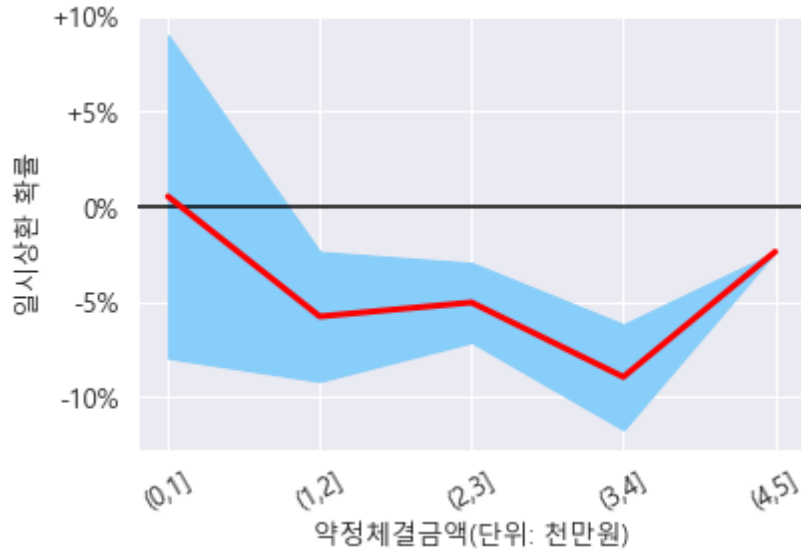
5. 일시상환

본 절은 일시상환에 영향을 미치는 상위 2개 요인의 구간별 SHAP에 대해 탐구한다(<그림 7> 참조). 분석에 활용한 1,000건의 평균 일시상환 확률은 약 17%이다.

⁶ 평균 조기상환 확률 7%에서 20%p 하락하면 부(-)의 확률이 도출되는 것처럼 보이나, 다른 요인들 또한 정(+)의 영향을 미쳐 최종적으로 확률의 범위인 0과 1사이에 위치하게 된다.

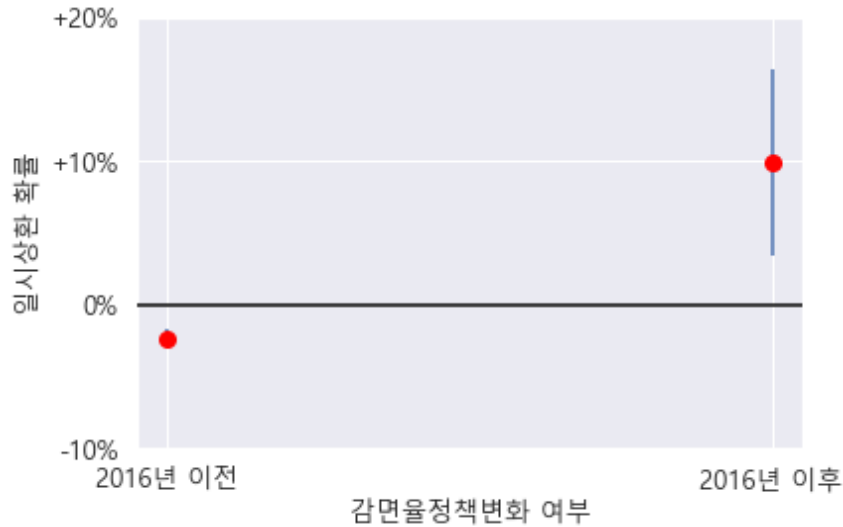
<그림 7> '일시상환'에 영향을 미치는 상위 2가지 요인

Graph A: 약정체결금액(1st)



주) 0%축은 분석에 활용한 1,000개 약정건의 평균 일시상환 확률 17%를 의미하며, 빨간색 선은 구간별 평균값, 파란색 밴드는 표준편차를 의미한다.

Graph B: 감면율정책변화(2nd)



Graph A는 약정체결금액이 일시상환 가능성에 미치는 구간별 영향을 나타낸 것이다. 전반적으로 약정체결금액이 커질수록 일시상환 가능성이 감소하는 것으로 나타났는데, 이는 변제해야할 채무가 많으면 일시에 갚기 어렵다는 직관과도 부합하는 결과이다. Graph B는 2016년 감면율정책변화로 인한 일시상환 영향도를 나타낸 것이다. III.4절에서 예상한 바와 같이, 조기상환의 결과와 그림이 반대되는 것을 확인할 수

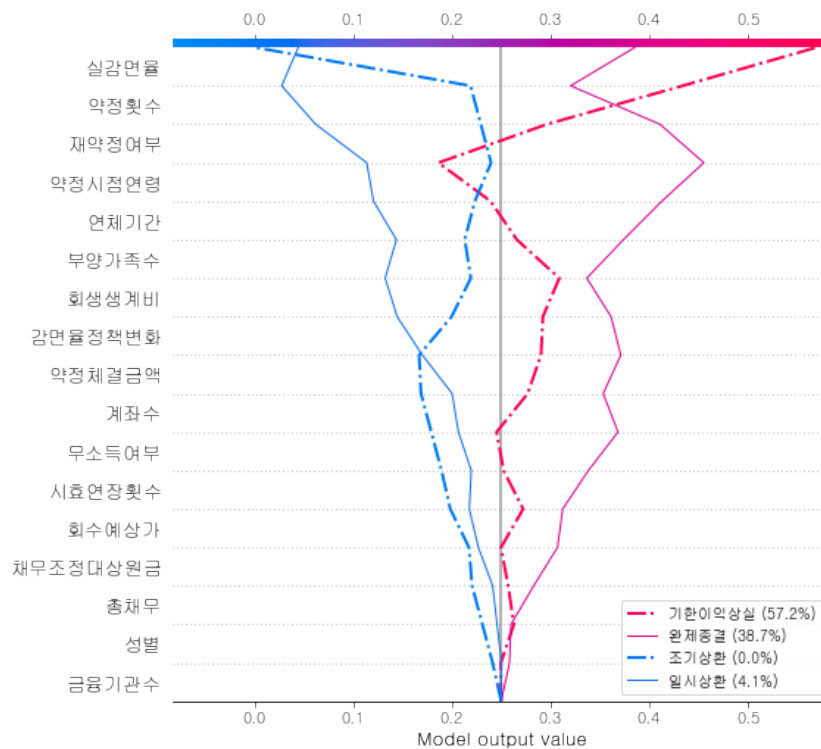
있다. 예컨대, 조기상환은 2016년 이후 정책적 추가 감면시 조기상환 가능성이 감소했지만, 일시상환은 오히려 평균 17% 대비 약 10%p 증가한 것을 확인할 수 있다.

종합하면, 정책적인 추가 감면은 채무를 6개월 안에 갚도록 유인하는 (단기)일시상환에 가장 큰 영향을 미치며, 여기에 약정체결금액까지 적을 경우 채무상환 동기가 더욱 커지는 것으로 추정할 수 있다. 또한, 실감면율을 상향하면 분할상환 같은 완제 가능성에도 영향을 주는 것을 고려해보았을 때, 성실상환자에게 추가감면 가능 여부 등 현재 약정을 계속 유지하도록 유인책도 고려해볼직 하다.

IV. 사례분석 시뮬레이션

본 장은 앞서 추정한 모형을 바탕으로 실무적인 활용 가능성을 타진하기 위해 실제 기한이익상실로 종료된 약정체결 건에 대해 시뮬레이션을 수행한다.

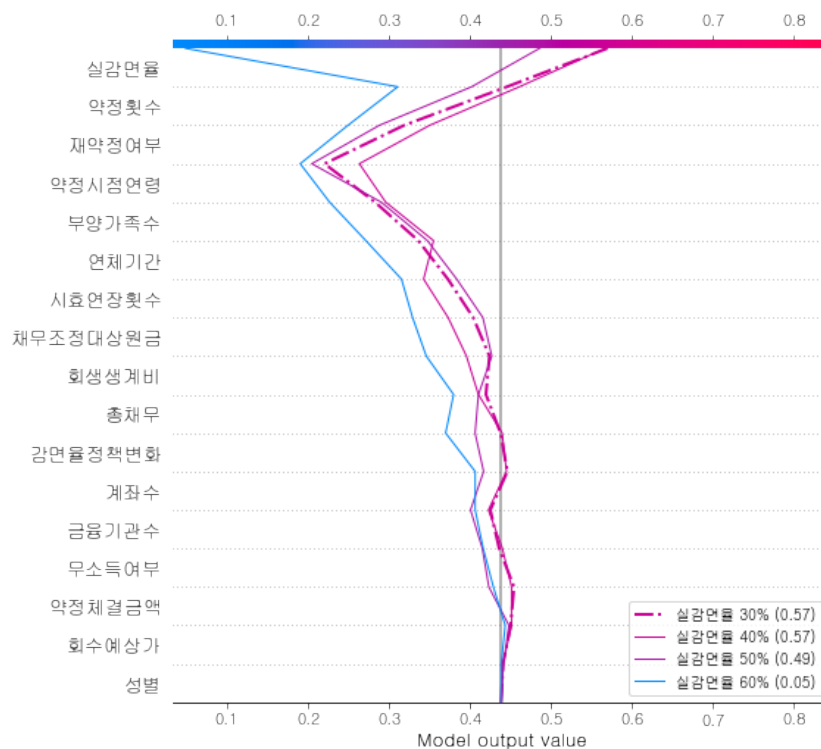
<그림 8> 약정체결 A에 채무상환 유형별 추정 확률 결과



여기서 개인정보보호 문제로 실감면율이 30%라는 정보 외에는 구체적인 약정체결 정보는 공개하지 않을 것이다. 구체적으로, 실감면율을 가상으로 최대 60%까지 적용하면서 약정체결금액을 이에 따라 조정해주는 경우 4가지 채무상환 유형 확률이 어떻게 변하는지 탐구하는 것이 목적이다.

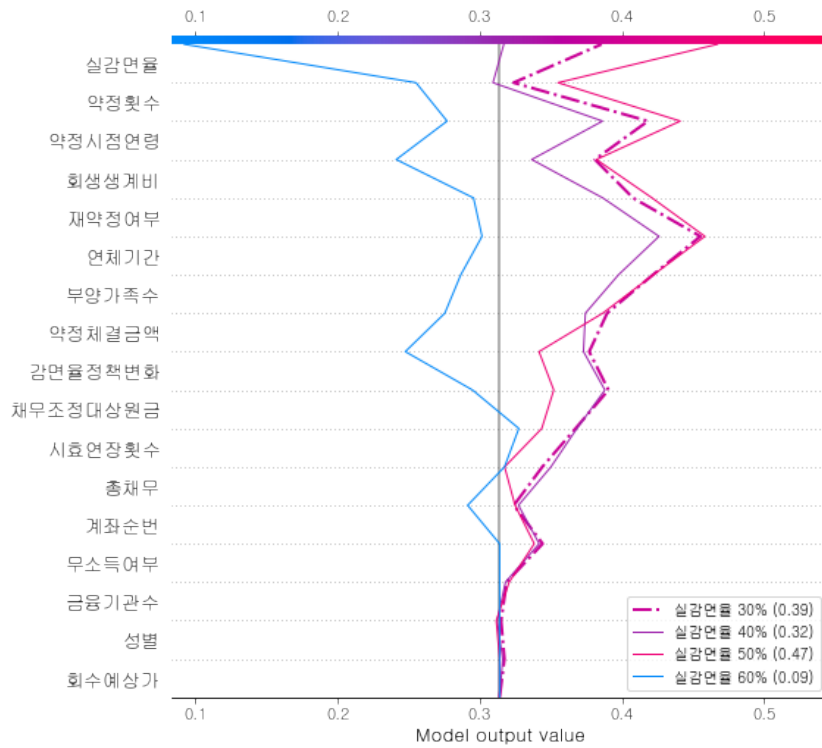
먼저, 임의의 약정체결 A에 대한 정보를 모형에 입력시 나타나는 4가지 채무상환 유형별 확률을 살펴보자(<그림 8> 참조). 그림 하단 25%(=100%/4가지 유형)을 시작으로 각 약정체결 정보 입력시 오른쪽으로 이동하면 해당 유형별 확률이 증가했음을 의미하고, 왼쪽으로 이동하면 해당 유형별 확률을 감소했음을 의미한다. 예를 들어, 그림의 가장 위 3가지 정보(실감면율, 약정횟수, 재약정여부)는 기한이익상실 확률을 오른쪽으로 이동시키는데, 이 3가지 정보가 기한이익상실 가능성을 증가시켰음을 의미하며, 최종적으로 57.2% 기한이익상실 확률을 도출했다고 해석할 수 있다. 반대로, 완제종결의 3가지 유형인 분할·조기·일시상환은 오른쪽으로 이동시 완제로 종결되는 긍정적인 의미로 해석해야 한다.

<그림 9> 약정체결 A에 ‘기한이익상실’ 시뮬레이션 결과



<그림 9>는 약정체결 A의 실제 실감면율 30%에서 60%까지 변화시킬 경우 기한이익상실에 대한 시뮬레이션한 결과를 보여준다. 실감면율이 30%인 경우 기한이익상실 확률이 57%였는데, 채무조정대상원금을 50%까지 감면해줄 경우, 기한이익상실 확률이 49%로 약 8%p(=57%-49%) 감소했음을 확인할 수 있다. 여기에 채무를 최대 60%까지 감면해주는 경우 IV장에서 살펴본 바와 같이 채무를 조기에 갚고자 하는 동기가 유발되면서 기한이익상실 확률이 5%까지 하락하는 것으로 나타났다.

<그림 10> 약정체결 A에 '분할상환' 시뮬레이션 결과

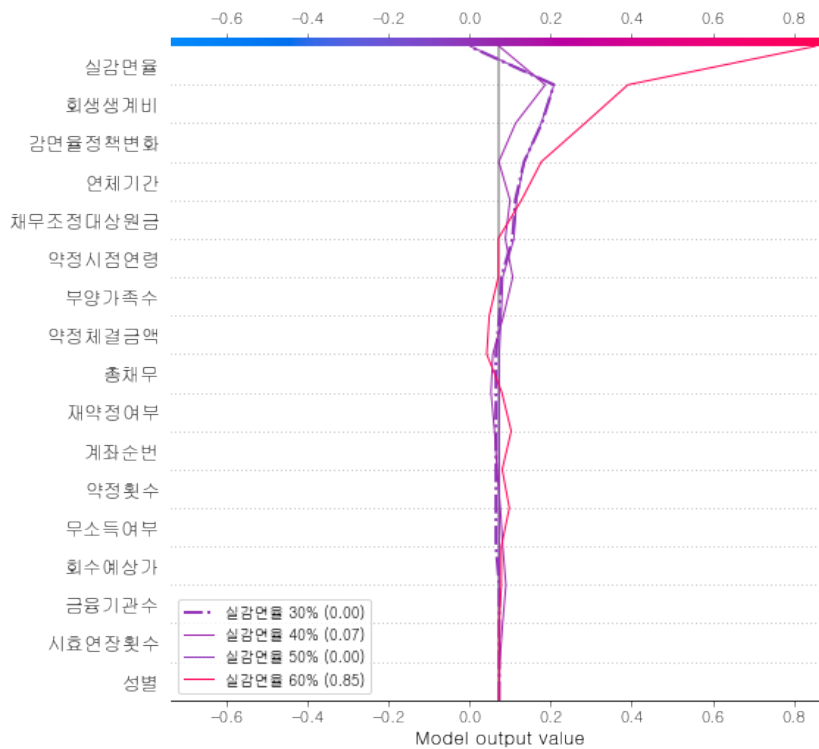


<그림 10>은 같은 시뮬레이션 조건 아래 약정체결 A의 분할상환 시뮬레이션 결과를 보여준다. 실제 실감면율이 30%인 경우 분할상환 확률이 39%였는데, 채무를 50%까지 감면시 확률이 47%로 약 8%p(=47%-39%) 증가했음을 확인할 수 있다. 즉, 50% 미만의 실감면율 구간에서는 기한이익상실 감소효과(<그림 9> 참조)가 분할상환 증가효과로 변한 것으로 해석할 수 있다. 여기에 최대 60%까지 채무를

감면해주면, 채무를 장기적으로 분할상환하기보다 비교적 이른 시간 안에 채무를 갚고자하는 동기가 발생할 수 있으며, 이로 인해 분할상환 확률이 9%까지 하락하는 것을 확인할 수 있다.

<그림 11>은 약정체결 A의 조기상환 시뮬레이션 결과를 보여준다. 실제 실감면율이 30%인 경우 조기상환 확률은 0%였으며, 실감면율을 50%까지 적용하더라도 확률이 0%대로 추정되는 등 채무감면 효과가 명시적으로 나타나지 않는 것이 특징이다. 그러나, 채무를 60%까지 감면해줄 경우 조기상환 확률이 85%까지 상승하는 것을 확인할 수 있는데, 이는 실감면율을 60%로 적용해줄 경우 약정체결 A(또는 유사 그룹)가 채무를 장기적으로 분할상환하는 것보다 조기에 상환하고자 하는 의지의 변화가 있을 수 있음을 암시한다.

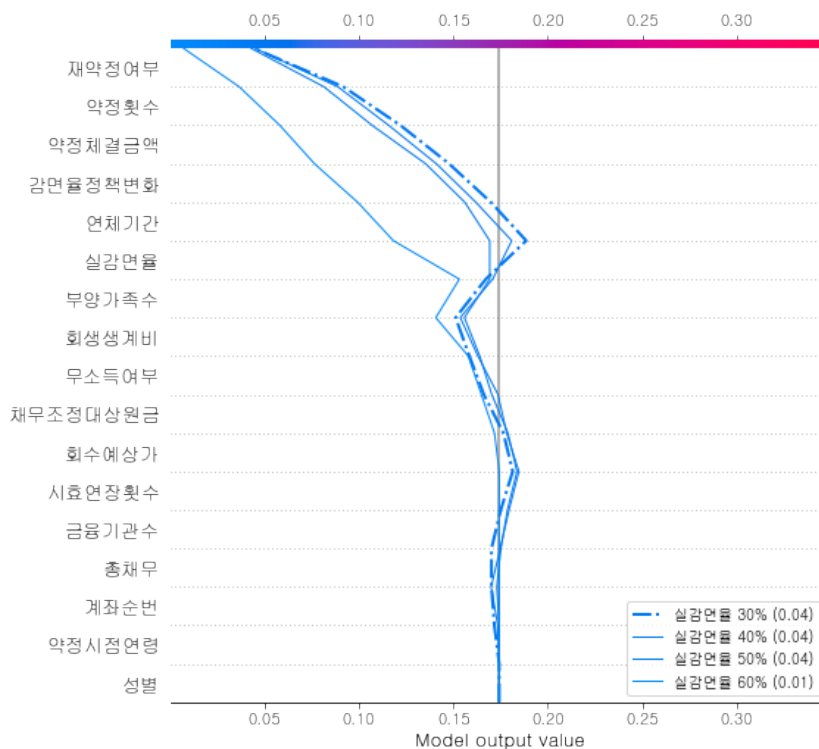
<그림 11> 약정체결 A에 '조기상환' 시뮬레이션 결과



마지막으로, <그림 12>은 약정체결 A의 일시상환 시뮬레이션 결과를 보여준다. 실제 실감면율 30%를 적용받을 경우 일시상환 확률이 약 4%였으며, 채무를

감면해준다 할지라도 시뮬레이션 확률의 변동폭이 크지 않는 것으로 나타났다. 이 결과는 앞서서도 언급한 바와 같이 약정체결 A(또는 유사 그룹)는 6개월 안에 채무를 갚는 (단기)일시상환보다 12개월 이후 채무를 갚고자 하는 (중기)조기상환을 선호하기 때문인 것으로 추정한다. 구체적으로 어떠한 원인으로 인해 일시상환 또는 조기상환 행태가 상이하게 나타나는 지는 향후 추가적인 연구를 통해 탐구해야할 것으로 보인다.

<그림 12> 약정체결 A에 '일시상환' 시뮬레이션 결과



V. 결론

본 연구는 한국자산관리공사에서 운영 및 관리한 국민행복기금 데이터 및 최신 머신러닝 기법을 활용해 채무상환 유형별 채무상환의 결정요인에 대해 탐구하였다. 또한, 약정체결 건별 유형별 확률을 추정해 개인신용위험을 평가하는 방안을 제시했을 뿐만 아니라, 실감면율 및 약정체결금액과 같은 정책변수를 조정하여 기한이익상실

위험을 줄이고 완제종결 가능성을 높이는 시뮬레이션 전략도 제안해보았다.

분석 결과, 실감면율과 약정횟수가 전체 채무상환 과정에서 가장 큰 영향을 미치며 개별 유형별로 미치는 영향이 상이한 것으로 나타났다. 한 가지 재미있는 결과는 실감면율이 약 50%미만인 경우 유의미한 채무상환 패턴을 보이지 않다가, 그 비율 이상인 경우 채무를 조기에 갚으려는 움직임이 발견된다는 점이다. 이러한 비선형 패턴은 “x가 증가(감소)하면 y도 증가(감소)한다”는 전통적인 선형분석 방법에서는 탐구할 수 없는 결과임에 틀림없다. 또한, 기한이익상실 가능성을 낮춰 한계차주의 신용회복 지원을 위해서는 재약정을 통해 약정횟수가 증가하지 않도록 지속적인 모니터링과 채무상환 유인전략이 필요할 것으로 추정한다.

향후 국민행복기금과 같은 다양한 채권매입형 신용회복지원 프로그램에 대한 빅데이터 분석을 고도화할 필요가 있으며, 이를 통해 채무상환에 대한 유의미한 패턴을 찾아내고 이를 신용회복 지원제도에 활용할 수 있을 것으로 기대한다. 예컨대, 채무자별 또는 약정체결건별로 동일한 집단을 묶어 코호트(cohort)분석을 수행한다면, 코호트별 맞춤형 채무조정 프로그램 전략을 세우는데 도움이 될 수 있을 것이다.

참고문헌

- 김원대. (2021). 국민행복기금 채무조정의 완제결정요인 분석. 부산대학교 박사학위논문
- 김원대 & 정영우. (2020). 채무조정 차주의 기한이익상실에 미치는 결정요인 연구: 국민행복기금을 중심으로. *경영경제연구*, 42(4), 45-74.
- 김형준, 류두진 & 조훈(2019). 개인부실채권 상환약정체결 및 완제 행태 분석. *채무연구*, 32(2), 187-219.
- 남주하, 박정수, 이정민 & 홍석철. (2014). 국민행복기금 채무감면율의 적정성에 대한 소고. *응용경제*, 16(2), 5-30.
- 민혜숙(2016). 캠프 채무조정 약정체결자 DB분석. 캠프연구소, 조사연구보고서.
- 박정숙 & 남주하. (2017). 개인 워크아웃제도의 유효성 분석. *한국경제의 분석*, 23(3), 1-53.
- 이수진, 구정환 & 이규복. (2019). 우리나라 다중채무자 특성 및 전환요인 분석. *한국경제의 분석*, 25(1), 137-187.
- Lundberg, S., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.