

신용 데이터의 이미지 변환을 활용한 합성곱 신경망과 설명 가능한 인공지능(XAI)을 이용한 개인신용평가*

원종관** · 홍태호*** · 배경일****

〈목 차〉

- | | |
|-----------------|----------------|
| I. 서론 | III. 연구 프레임워크 |
| II. 이론적 배경 | IV. 실험 및 실험 결과 |
| 2.1 개인신용평가 | V. 결 론 |
| 2.2 합성곱 신경망 | 참고문헌 |
| 2.3 설명 가능한 인공지능 | <Abstract> |

I. 서론

신용평가는 채무자가 은행, 증권사, 카드사 등 금융 기관에서 개인의 신용을 적절하게 평가하고 대출금 및 이자의 상환 여부를 정확하게 예측하는 것은 해당 기관의 영업 이익과 직접적으로 관련되기 때문에 중요한 의미를 갖는다(Bae and Kim., 2015). 최근 핀테크의 일종으로 주목 받고 있는 P2P 대출은 유연한 선정산 서비스를 가능하게 하고, 은행이 인수하지 않는 차입자에게도 대출을 공급할 수 있다는 특징을 가지고 있으며(박태현과 류두진, 2020), 관련 규제가 미비하고 진입장벽이 낮다는 점에서 급

속한 성장세를 보이고 있다. P2P 대출 분야에서는 상대적으로 신용도가 낮은 고객에게 대출을 제공한다는 점에서 정확한 할인을 산정하는 비즈니스 모델 구축에 있어 중요한 역할을 하고, 이를 위해서는 적절한 신용평가가 선행되어야 한다(최수만 외, 2020). 또한 COVID-19 팬데믹 현상이 지속됨에 따라 각 국가에서 경제적 피해를 입은 개인의 생계를 지원하기 위한 대출 지원이 추진되고 있는 상황에서(강구상 외, 2020), 신용 평가의 정확도를 향상시키는 것은 더욱 중요한 의의를 가진다.

통계적 기법 및 인공지능을 활용한 연구는 개인신용평가를 비롯해 다양한 금융 분야에 적

* “본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음” (IITP-2020-0-01797)

** 부산대학교 경영학과, jongkwan1@naver.com(주저자)

*** 부산대학교 경영학과, taeho.hong@gmail.com(교신저자)

**** (주)아톰릭스랩, k@atomrigs.io(공동저자)

용되고 있는 한편(이성욱과 노태협, 2007; 유환과 이영재, 2019), 딥러닝은 이미지, 음성, 영상, 자연어 처리 등 다양한 분야에서 괄목할만한 성과를 보여주며 성장하고 있다. 하지만 순환신경망을 주로 사용하는 주가 예측과 같이 시계열 데이터를 다루는 분야가 아닌 다른 금융 분야에서는 판별분석, 로지스틱 회귀분석과 같은 통계적 기법 및 서포트 벡터 머신, 인공신경망과 같은 머신러닝 기법을 활용한 모형이 주를 이루었으며(Abdou and Pointon, 2011), 딥러닝을 사용한 연구는 거의 존재하지 않는다(Hosaka, 2019). 최근 기업 신용평가, 주가 예측 등 금융 분야의 연구에서 딥러닝을 적용하는 연구가 늘어나고 있지만(Mai et al., 2019; Selvin et al., 2017; 이현상과 오세환, 2020), 그럼에도 불구하고 개인신용평가에 딥러닝을 적용하는 연구는 극히 드물다. 하지만 개인신용평가와 성격이 유사하다고 볼 수 있는 기업도산 예측에서, Hosaka(2019)는 수치형 데이터를 이미지 데이터로 변환해 이미지 처리에 특화된 딥러닝 기법인 합성곱 신경망(Han et al., 2018)에 적용할 경우 분류 정확도를 향상시킬 수 있음을 실증 연구를 통해 주장했다. 따라서 금융 데이터에서도 적절한 변수 및 데이터 구조 설정을 통해 딥러닝을 적용할 경우 예측 성과를 높일 수 있고(Ding et al., 2015), 개인신용평가에도 합성곱 신경망을 포함한 딥러닝을 적용한다면 예측 정확도를 향상시킬 수 있는 가능성이 존재한다. 따라서 본 연구에서는 이미지 처리에 우수한 성능을 보이는 딥러닝 기법인 합성곱 신경망을 개인신용평가 모형에 적용하기 위해 개인신용 데이터를 이미지 데이터로 변환하고 이전까지 사용해 왔던 통계적 기법 및 머

신러닝 기법을 적용한 모형들과 성과를 비교해 보고자 한다.

한편 딥러닝을 활용한 분류 및 예측 모형은 예측 성과에 있어 다양한 연구에서 머신러닝 및 통계적 기법을 활용한 모형들보다 탁월한 성능을 보여 왔지만, 왜 특정 예측 값을 산출하는 지에 대한 설명력이 부족하다는 점을 지적받고 있으며 따라서 실무에 활용하기에는 한계가 있다는 비판을 받아왔다. 특히 규제 산업으로 분류되는 금융 분야에서는 내부 직원의 이해도를 높이고 고객들에게 충분한 설명을 제공하기 위해 설명력을 가지고 있는 기법을 요구하고 있다(천예은 외, 2021). 이에 맞춰 최근 블랙박스의 특징을 가지는 딥러닝 기법 및 머신러닝의 갖는 한계를 보완하기 위해 설명 가능한 인공지능(eXplainable Artificial Intelligence, XAI)관련 연구가 진행되고 있고(Kim et al., 2020), 이를 활용한다면 설명력을 필수적으로 요하는 금융권에서의 딥러닝 활용 가능성이 높아질 수 있다. 따라서 본 연구에서는 합성곱 신경망에 사용할 수 있는 XAI 기법 중 하나인 Layer-wise Relevance Propagation(LRP)을 적용해, 합성곱 신경망을 적용한 신용평가 모형의 설명력 및 해석 가능성을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 제 2장은 이론적 배경으로 먼저 통계적 기법, 머신러닝 기법, 그리고 딥러닝을 적용한 개인신용평가 관련 연구들을 살펴보고, 이어 본 연구에서 적용할 딥러닝 기법인 합성곱 신경망, XAI 기법인 LRP에 대해 알아본다. 제 3장은 연구 프레임워크로 수치형 데이터를 이미지 데이터로 변환하는 방법 및 이미지의 픽셀 별 밝기 결정 방법, 그리고 합성곱 신경망 및 LRP적용에 대해 설명

한다. 이어 제 4장은 연구 프레임워크에 따라 진행된 실험 및 실험 결과에 대해 알아보고, 제 5장은 결론 및 연구의 의의와 한계 및 추후 연구 방향을 제시한다.

Ⅱ. 이론적 배경

2.1 개인신용평가

개인 신용평가는 구체적으로 직업, 주거 등 인구 통계학 정보와 금융 거래 이력 등 금융 정보를 활용해 개인의 미래 채무 상환 여부를 구별하는 분류 문제로 볼 수 있다. 전통적으로 금융 분야는 규제 산업으로 분류되어 설명력을 요구하고, 주로 수치형 데이터로 이루어진 데이터가 많아 예측 성과에 강점이 있는 딥러닝 기법보다는 통계 및 머신러닝 기법이 주로 사용되고 있었다. 하지만 알파고 쇼크 이후로 인공지능 및 세부 기술에 대한 사회적 수용도가 높아지고(조유정 외, 2021), 수치형 데이터에도 딥러닝을 적용할 수 있도록 하는 다양한 방법론이 등장하면서 금융 분야에서도 딥러닝을 활용한 연구가 늘어나고 있는 추세이다. <표 1>은 개인신용평가와 관련된 연구들을 정리한 것으로, 크게 통계적 기법을 적용한 연구와 머신러닝 기법을 적용한 연구, 그리고 딥러닝을 적용한 연구로 분류할 수 있다.

첫 번째 분류는 통계적 기법을 이용한 개인 신용평가로, Durand(1941)가 고객의 신용정보 데이터로 판별분석(Linear discriminant analysis)을 이용한 모형을 구축한 것을 시작으로 본격적으로 도입되기 시작했다. 또한

Orgler(1970)은 선형회귀분석(Linear Regression)과 유동성, 수익성, 위험성에 대한 변수를 이용해 상업적 대출에 대한 비용을 계산하는 모형을 구축했다. 이후 선형회귀분석보다 개인신용평가와 같은 이진 분류에 더 적합하고(Hand and William., 1997), 관련된 변수들에 대한 해석이 가능하다는 이점을 가지고 있는 로지스틱 회귀분석(Logistic Regression)을 개인신용평가에 적용한 연구가 진행되었다(윤종식 외, 2007). Wiginton(1980)은 로지스틱 회귀분석을 사용한 개인신용평가 모형과 선형회귀분석을 사용한 모형간 비교를 통해 로지스틱 회귀분석의 우수성을 주장했고, 최근에도 머신러닝 기법 및 딥러닝 기법을 활용한 연구의 성능 비교를 위한 벤치마크로 사용되고 있다(Triki, 2016; Nehrebecka, 2018). 따라서 본 연구에서는 통계기법 중 대표적인 로지스틱 회귀분석을 적용한 모형을 구축해 합성곱 신경망을 적용한 모형과 성과를 비교한다.

두 번째는 머신러닝 기법을 적용한 개인신용평가이다. 지난 30여년간 의사결정나무(Decision Tree), 랜덤포레스트(Random Forest), 인공신경망(Neural Network), 서포트 벡터 머신(Support Vector Machine, SVM) 등 머신러닝 기법을 활용한 신용평가 모형을 구축하는 연구가 진행되어 왔다. 의사결정나무는 단순하고 예측 변수의 상호 작용 효과를 파악할 수 있기 때문에 통계적 배경이 많지 않은 실무자에게 설득력이 있는 것으로 알려져 있다(Sohn and Kim., 2012). Zhang et al.(2004)은 P2P 대출의 상환 여부를 예측하는 모형을 머신러닝 기법들인 의사결정나무, 로지스틱 회귀분석, 인공신경망을 사용해 구축하고 각 모형 간

성과를 비교했다. 또한 Ruyu et al.(2019)는 P2P 대출 플랫폼인 랜딩클럽(Lendingclub.com)의 데이터에 의사결정 나무, 나이브 베이스, 로지스틱 회귀분석을 적용해 대출 상환 여부에 대한 예측 성과를 비교했고, 의사결정 나무와 나이브 베이스를 적용한 모형의 경우 대출을 제때 변제한 고객들을 정확하게 분류하는 경향을 보였다고 주장했다. 랜덤포레스트는 Breiman(2001)에 의해 제안되었고, 의사결정나무 기법을 여러 개의 나무로 확장시킨 앙상블 모형이다. 랜덤포레스트는 개별 의사결정나무의 정밀도는 떨어질 수 있으나, 이들을 종합하여 예측에 사용해 정확도와 안정성이 높아질 수 있다는 점을 활용한 분류 기법이다(김성진과 안현철, 2016). 박성우(2019)는 개인신용평가를 위해 랜덤포레스트, k-최근접 이웃, 로지스틱 회귀분석 등 다양한 머신러닝 기법을 적용한 모형들을 구축했고, 성과 비교 결과 랜덤포레스트의 예측 성과가 가장 우수했다. 또한 Rao et al.(2020)은 랜덤포레스트와 엔트로피 가중치 방식을 함께 사용한 모형을 구축해 기존의 분류 기법인 로지스틱 회귀분석, 의사결정 나무를 적용한 모형보다 P2P 대출 데이터에서 우수한 성능을 보임을 실증 분석을 통해 주장했다. 인공지능망은 모형의 통계적 기법이 가지는 통계적 가정이 만족되지 않아도 사용할 수 있으며, 통계기반 기법보다 분류와 예측 문제에서 정확도 면에서 탁월한 성능을 보인다(이재식과 한재홍, 1995). 인공지능망은 개인신용 평가와 관련된 연구에 사용되며 우수한 예측 성과를 보이고 있는데, Hajek(2011)은 다층 피드포워드 인공 신경망, 확률 신경망, 계단식 상관 신경망, GMDH 다항식 신경망, 서포트 벡터 머

신을 활용해 개인신용평가 모형을 구축했고, 성과를 비교한 결과 확률 신경망이 가장 큰 성과를 보임을 주장했다. 또한 Byanjankar et al.(2015)은 P2P 대출에서의 채납, 변제를 예측한 연구에서 인공신경망을 사용한 모형이 통계적 기법인 로지스틱 회귀분석보다 더 우수한 성능을 보일 수 있음을 주장했다. 한편 서포트 벡터 머신은 Vapnik(1995)에 의해 개발된 분류 알고리즘으로, 선형 문제가 아닌 비선형문제를 고차원의 선형 문제로 대응해 나타냄으로써 수학적 분석이 용이하다는 장점이 있다(Hearst et al., 1998). 서포트 벡터 머신은 분류 문제에 탁월한 성능을 보이며 다양한 분야에서 사용되고 있는데, 금융 분야에서도 적용되고 있다. 윤종식 등(2007)은 소상공인 부실 예측 모형을 서포트 벡터 머신을 적용해 구축하고, 인공신경망, 의사결정나무와 같은 다른 머신러닝 기법에 비해 우수한 성능을 보임을 확인했고, Martens et al.(2007)은 개인신용평가에 서포트 벡터 머신을 사용했고, 실험에서 인공신경망을 사용한 모형보다 더 우수한 성능을 보였다. 머신러닝 기법 중에서 최근까지 가장 활발하게 연구되고 현장에서 활용되고 있는 방법으로는 인공 신경망과 서포트 벡터 머신이다(김성진과 안현철, 2016). 따라서 본 연구에서는 인공 신경망 및 서포트 벡터 머신을 적용해 모형을 구축하고, 로지스틱 회귀분석을 적용한 모형과 함께 합성곱 신경망을 이용해 구축한 모형과 성과를 비교한다.

마지막 세 번째는 딥러닝 기법을 활용한 개인신용평가로, 딥러닝은 이미지 인식(He et al., 2016), 음성 인식(Buyukyilmaz and Cibikdiken 2016), 영상 분석(Takahashiet al., 2017), 자연

어 처리(Mai et al., 2019) 등 다양한 분야에서 광범위하게 사용되고 있다. 최근 기업 도산 예측 및 개인신용평가 분야에서도 딥러닝이 사용되고 있는데, 김은미와 박지영(2019)은 온라인 P2P 환경에서 딥러닝을 적용한 다분류 기반 개인신용등급 예측 모형을 구축했고, 심층신경망(Deep Neural Network, DNN)이 서포트 벡터 머신이나 일반 다층 피드 포워드 신경망에 비해 우수한 성능을 보일 수 있음을 확인했다. Bayraci and Susuz(2019)는 심층신경망을 이용해 개인신용평가를 진행했고, 서포트 벡터 머신

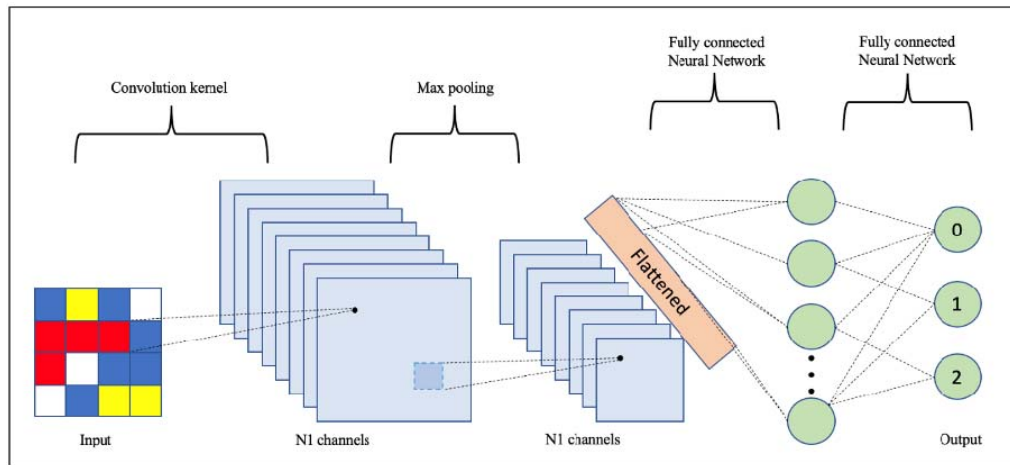
이나 나이브 베이스 등 다른 기법을 사용해 구축한 모형보다 월등한 성능을 보였다. 또한 Kvamme et al.(2018)은 합성곱 신경망을 데이터 형상을 보존하는 방식으로 모형을 구축했고, 일반적인 인공신경망이나 로지스틱 회귀분석을 사용한 모형보다 우수한 예측성적을 보였다.

2.2 합성곱 신경망

합성곱 신경망은 일반적인 인공 신경망의 단점인 공간적인 구조가 유실된다는 점을 해결하

<표 1> 통계적 기법, 머신러닝 및 딥러닝 기법을 활용한 개인신용평가 연구

분류	저자(년도)	적용 기법	연구 결과 및 의의
통계적 기법	Durand (1941)	판별분석	통계적 기법을 활용해 고객의 신용정보 데이터로 채무 불이행 여부를 최초로 예측함
	Orgler (1970)	선형회귀분석	유동성, 수익성, 위험성에 대한 변수를 이용해 상업적 대출에 대한 비용을 계산하는 모형 구축함
	Wiginton (1980)	로지스틱 회귀분석	로지스틱 회귀분석과 선형 회귀분석을 적용한 개인신용평가 모형을 비교해 로지스틱 회귀분석의 우수성을 주장함
머신러닝 기법	Zhang et al. (2004)	의사결정나무	의사결정나무, 로지스틱회귀분석 등의 기법을 사용해 P2P 대출의 상환 여부를 예측하는 모형을 구축함
	Ruyu et al. (2019)		의사결정나무, 로지스틱 회귀분석, 나이브 베이스를 적용한 P2P 대출 상환 여부 예측 모형을 구축함
	박성우 (2019)	랜덤포레스트	랜덤포레스트, k-최근접 이웃, 로지스틱 회귀분석을 적용한 개인신용평가 모형을 구축했고, 랜덤포레스트 기법의 예측 성과 우수성을 주장함
	Rao et al. (2020)		랜덤포레스트와 엔트로피 가중치를 결합한 신용평가 모형을 구축해 P2P 대출 예측에서의 성과를 높일 수 있음을 주장함
	Hájek (2011)	인공신경망	여러 종류의 인공신경망을 활용해 개인신용평가 모형을 구축하고 성과를 비교함
	Byanjankar et al. (2015)		인공신경망을 활용할 경우 통계적 기법을 활용한 모형보다 더 높은 예측 성과를 보일 수 있음을 주장함
	윤종식 (2007)		서포트 벡터 머신
	Martens et al. (2007)	서포트 벡터 머신을 적용한 개인신용평가 모형이 인공신경망을 적용한 모형보다 예측 성과가 더 좋을 수 있음을 주장함	
딥러닝 기법	김은미와 박지영 (2019)	심층신경망	심층신경망을 적용한 모형이 서포트 벡터 머신이나 인공신경망을 적용한 모형보다 성과가 우수함을 주장함
	Bayraci and Susuz(2019)		심층신경망을 적용한 개인신용평가 모형이 서포트 벡터 머신과 나이브 베이스를 적용한 모형보다 우수함을 주장함
	Kvamme et al. (2018)	합성곱 신경망	합성곱 신경망을 적용한 개인신용평가 모형이 일반적인 인공신경망이나 로짓 모형보다 우수함을 주장함



<그림 1> 합성곱 신경망 구조도

기 위해 등장했고 LeCun et al.(1998)에 의해 크게 발전되었다. 여기서 공간적인 구조란 이미지에서 거리가 가까운 픽셀끼리의 상관관계나 관련 정보를 의미한다. <그림 1>은 합성곱 신경망의 기본적인 구조도이다. 합성곱 신경망의 경우 크게 입력층(input layer)과 출력층(output layer), 하나 이상의 합성곱 계층(Convolution layer)과 풀링 계층(Pooling layer), 그리고 완전 연결 계층(Fully connected layer)으로 구성된다. 입력층에 이미지 형태의 데이터를 입력하고 합성곱 계층을 통해 필터링해 적절한 특징(features)을 추출한다. 합성곱 필터와 함께 사용되는 활성화 함수는 ReLU(Rectified Linear Unit) 함수로 그래디언트 소실(vanishing gradient)을 해결하기 위해 사용된다(이모세와 안현철, 2018). 한편 완전 연결 계층은 이후 분류 및 예측을 위한 일반적인 인공 신경망 구조를 가지게 된다. 여기서 일반적인 인공신경망은 과적합의 위험이 있기 때문에, 임의로 일정 비율의 뉴런을 제거하면서 학습하는 드롭아웃

(dropout)기법을 사용한다(Srivastava et al., 2014). 합성곱 신경망은 이미지 처리에 있어 다른 딥러닝 기법이나 머신러닝 기법에 비해 월등한 성능을 보인다(Lee and Kwon, 2017). 일반적인 인공 신경망의 경우 은닉층이나 은닉층의 노드 개수가 늘어날 경우 파라미터의 개수가 기하급수적으로 늘어난다. 이를 학습에 반영해 파라미터들을 튜닝하는 것에는 큰 비용이 발생하는데, 합성곱 신경망은 합성곱 층을 이용해 파라미터의 개수를 줄일 수 있음과 동시에 높은 예측성능 또한 보장할 수 있다.

그러나 합성곱 신경망이 금융 데이터에 적용되는 경우는 거의 존재하지 않는데, 그 이유는 합성곱 신경망은 재무 데이터나 신용 데이터처럼 수치형 데이터에 적용할 때보다 이미지 데이터에 적용할 때 더 좋은 성능을 보이기 때문이라고 판단된다. 하지만 재무 데이터나 신용 데이터와 같은 수치형 데이터도 이미지 데이터로 적절하게 변환할 수 있다면, 더 우수한 성능을 보일 수 있다(Hosaka, 2019; 김지윤과 조성

매, 2017). 따라서 본 연구에서는 합성곱 신경망을 적용하기 위해 수치형 데이터를 이미지 데이터 변환하는 방법을 활용하고, 이후 합성곱 신경망을 적용해 개인별 대출에 대한 채무 불이행 여부를 도출하는 과정을 제시한다.

2.3 설명 가능한 인공지능(eXplainable Artificial Intelligence, XAI)

설명 가능한 인공지능은 머신러닝 및 딥러닝을 포함한 인공지능 기술을 사용해 구축한 예측 모형들이 특정 결과 값을 산출할 때 어떤 근거로 해당 결과 값이 산출되는지를 알 수 있게 하기 위해 사용하는 기법이다(Adadi and Berrada, 2018). 최근 머신러닝 및 딥러닝 기법들을 활용해 모형을 구축할 때 예측 정확도 향상을 위해 하이퍼 파라미터 개수를 늘리거나 앙상블, 하이브리드 기법을 함께 사용해 복잡한 구조로 설계하는 경우가 많아지고 있고, 이러한

복잡성은 실무자로 하여금 해당 모형의 출력물 산출 과정의 해석을 불가능하게 만들었다. 설명 가능한 인공지능은 이러한 복잡성을 해소하기 위해 등장했고, 의사결정에 설명력이 필요한 다양한 분야에서 연구되고 있다(Arrieta et al., 2020; Böhle et al., 2019; Heyi et al., 2019; Sebastian et al., 2015). Shortliffe and Buchanan(1975)는 규칙 기반 의사결정 방식을 개발해 의사들이 환자의 병을 판단할 수 있게 했고 이후 Van Lent et al.(2004)은 인공지능 시스템은 복잡해져 가지만 자기 설명 기능에는 발전이 거의 없었다는 것을 지적하며 ‘설명 가능한 인공지능’의 개념을 주장했다. 이후 딥러닝과 비슷한 이유, 즉 컴퓨팅 파워 및 하드웨어적 한계로 제한된 범위에서 연구되어 왔지만 최근 머신러닝 기술의 발전 및 하드웨어 성능의 비약적인 성장으로 연구가 더 활발하게 진행되고 있다. 특히 최근에는 금융 분야에서도 설명 가능한 인공지능에 대한 연구가 진행되고

<표 2> 금융 분야에 적용된 설명 가능한 인공지능(XAI)

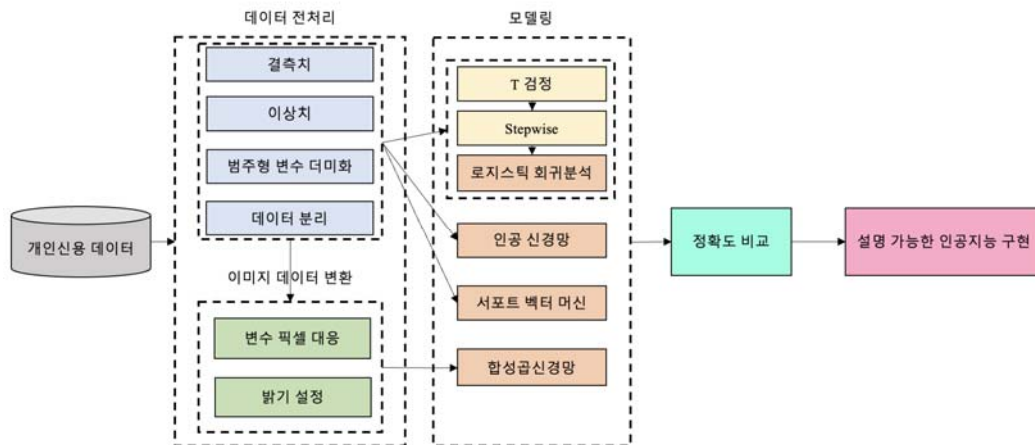
사용 알고리즘	저자(년도)	연구 결과 및 의의
Rule extraction (Neurorule; Trapen; Nefclass)	Baesens et al. (2004)	인공신경망을 활용해 우수한 성능을 가진 신용위험을 평가하는 모형을 구축하고, 신경망을 분해하는 Neurorule, 의사결정 나무를 이용하는 Trapen, 퍼지시스템을 이용하는 Nefclass를 각각 구축된 모형에 적용해 설명 가능성을 높이고 분류 모형이 신용 위험을 평가하는 기준을 제시
Rule extraction (SVM+Prototype; Fung et al., 2005)	Martens et al. (2007)	개인신용평가에서 SVM을 활용한 모형을 구축한 뒤, SVM을 분해하고 규칙을 만드는 알고리즘을 적용한 모형을 구축함. 이를 통해 개인신용평가에 적용될 수 있는 규칙을 생성했고, 기존에 주로 규칙을 만들던 기법인 의사결정나무 계열의 기법(C4.5, Trapen)들 보다 더 높은 분류 정확도를 보일 수 있음을 주장
LIME, SHAP value	Vincenzo Moscato et al. (2021)	LIME, SHAP 등 XAI를 random forest, logistic regression에 적용해 신용평가에 사용한 변수에 대한 신뢰성을 평가
SHAP value	Katsuya Futagami et al. (2021)	Lightgbm을 이용한 모형에 Shap value 개념을 도입해 어떤 변수가 기업 인수자와 병합자 쌍을 예측하는데 중요한 변수인지 정리
	Yang et al. (2021)	Xgboost에 Shap value를 적용해 기업 도산 예측에서 사전에 미리 예방할 수 있는 중요한 변수들을 추출할 수 있음을 주장

있는데, 금융 분야에서 설명 가능한 인공지능 및 이를 적용한 연구는 <표 2>에 정리했다. 한편 계층별 타당성 전파(Layer-wise relevance propagation, LRP)는 설명가능한 인공지능을 구현하기 위한 기법 중 하나로, 이미지 데이터에 적용하는 합성곱 신경망을 포함한 딥러닝 기법의 해석을 위해 사용된다(Bach et al., 2015). 타당성 전파(Relevance)와 분해(Decomposition)으로 구성되어 있고, 여기서 타당성 전파란 특정 결과값을 출력한 원인을 분해하고 그 비중을 분배하는 과정이며, 분해는 앞선 타당성 전파를 통해 얻어낸 원인을 가중치로 환원하고 해부하는 과정이다(Li et al., 2019). 즉, LRP를 활용하면 딥러닝 기법을 활용해 구축한 모형이 산출한 결과 값을 다시 사용하고 입력으로부터 구축한 모형의 각 계층을 역추적해 입력값인 픽셀 형태의 데이터에 어떤 부분에 큰 영향을 받아 결과 값을 산출했는지 나타내는 히트맵(Heatmap)을 출력하는 하고, 이를 통해 학습시킨 모형의 투명도 및 해석 가

능성을 높일 수 있다(Jung et al., 2021).

Ⅲ. 연구 프레임워크

본 연구의 프레임워크는 <그림 2>와 같다. 먼저 데이터 수집 및 전처리 단계에서는 결측치 및 이상치 처리, 더미 변수화 및 학습, 평가 데이터 분리와 같은 전처리의 경우 모든 모형에 사용 될 데이터에 공통적으로 진행하고, 합성곱 신경망을 이용한 모형의 경우 픽셀 대응, 밝기 설정 등과 같은 수치형 데이터를 이미지 형 데이터로 변환하는 과정을 추가적으로 진행한다. 이어 모델링 부분에서는 로짓, 인공 신경망, 서포트 벡터 머신, 합성곱 신경망을 적용한 모형들을 각각 구축하고, 다음 단계에서는 정확도와 맥니마 검정 결과를 바탕으로 성과 평가를 진행한다. 마지막으로 Layer-wise relevance propagation(LRP)를 적용해 설명 가능한 인공지능을 구현해 결과를 살펴본다.



<그림 2> 연구 프레임워크

3.1 데이터 수집 및 전처리

데이터 수집 단계에서 미국의 P2P 대출 업체인 랜딩클럽(Lendingclub.com)에서 년도, 분기별 개인 신용 데이터를 직접 내려 받아 통합한 후 개인 신용 데이터베이스를 구축한다.

각 모형들에 대한 공통적인 전처리 단계에서 결측치가 40% 이상 존재하는 변수는 제거하고, 관측치가 IQR방식에 의해 이상치로 선택될 경우 관측치를 제거한다. 또한 대출 승인 후에만 생성되는 변수는 출력 변수와 직접적인 상관관계가 있을 수 있어 예측 성과를 왜곡시킬 수 있기 때문에 제거한다. 이어 범주형 변수의 경우 범주 수가 1개일 경우 의미가 없고, 100개 이상인 변수는 차원을 과대하게 늘려 모형의 신뢰도를 떨어트릴 수 있기 때문에 각 변수들을 제거한다. 범주형 변수의 경우 더미화를 진행하고, 모형의 학습과 검증을 위해 데이터를 학습 데이터와 검증데이터로 분리한다.

이어서 합성곱 신경망을 적용하기 위한 모형을 위한 추가적인 전처리를 진행하는데, 입력 변수 픽셀 대응, 관측치 별 픽셀의 밝기 설정 순으로 진행된다. 본 연구에서 수치형 데이터의 이미지 데이터로 변환하는 방법론은 기업 도산 예측에 합성곱 신경망을 적용한 Hosaka(2019)의 방법론을 따른다. 먼저 입력 변수 픽셀 대응의 경우 입력 변수 개수와 대응되는 정사각형의 픽셀을 생성하고, 이후 무작위로 변수들을 픽셀에 배치한다. 하지만 합성곱 신경망에 투입되는 이미지의 형식이 무작위로 입력 변수가 배정된 형식일 때보다, 상관 관계가 높은 변수들이 밀집해 있는 형식일 때 예측 성과를 향상시킬 수 있다(Hosaka, 2019). 본 연구에서는

유사한 특성을 가지고 있는 변수들 간 픽셀 위치의 거리가 가까운 정도와, 상이한 특성을 가지고 있는 변수들 간의 픽셀 위치의 거리가 먼 정도를 CS(Coreelation Score)이라 정의하고, 아래 계산 식을 통해 수치화 한다. 여기서 i, j 는 임의의 선택된 픽셀이며, $x(i)$ 와 $y(i)$ 의 경우 픽셀 i 의 가로, 세로 축 위치이다. $R(i)$ 는 픽셀 i 에 위치한 변수를 의미하고, n 는 픽셀 i 와 픽셀 j 에 위치한 변수들간의 상관관계를 의미한다. 이어 임의로 픽셀에 배치된 두 변수를 선택하고, 두 변수의 위치를 변경했을 때 이전보다 CS의 값이 줄어든다면 픽셀을 위치를 바꾸는 방법을 통해 최적의 픽셀내 변수 배치를 결정한다.

$$CS = \sum_{(i,j) \in P} |c[R(i), R(j)]| * (x(i) - x(j))^2 + y(i) - y(j)^2$$

이어 관측치 별 픽셀의 밝기(BN) 설정의 경우 다음과 같이 설정한다. 은 관측치의 픽셀 i 에 위치한 변수 값을 의미하고, n 는 변수의 평균, 그리고 n 는 변수의 표준편차를 의미한다. 일반적으로 이미지 데이터의 밝기는 수치로 나타낼 경우 0부터 255까지로 실수로 나타내는데, 만약 계산식을 이용해 도출해낸 값이 0보다 작을 경우 0으로 변환하고, 255보다 클 경우 255로 변환한다.

$$BN = \frac{v[R(i)] - m[R(i)]}{\sigma[R(i)]} * 100 + 128$$

3.2 모형 구축

모형 구축 단계에서 로지스틱 회귀모형, 인공신경망 모형, 서포트 벡터 머신 모형, 그리고

합성곱 신경망 모형을 구축한다. 먼저 로지스틱 회귀분석을 적용한 모형의 경우, 사용할 수 있는 모든 변수를 사용하게 되면 모형의 성과 및 신뢰성에 좋지 않은 영향을 미칠 수 있어 영향력 있는 적절한 변수를 이용해 모형을 구축하는 것이 바람직하다(배재권 외, 2018). 변수 선정에는 다양한 방법론이 존재하지만, 본 연구에서는 t-검정과 logit stepwise를 사용해 로지스틱 회귀분석을 적용한 모형을 위한 최종 입력 변수를 선택한다. t-검정의 경우 입력 변수들과 종속 변수 간 5% 신뢰구간에서 유의미한 평균의 차이를 지니는 변수들로 선정했고, logit stepwise의 경우 전방 선택법(forward selection)과 후방 제거법(backward elimination)이 존재하는데, 본 연구의 경우 두 방법 모두 사용해 두 가지의 변수군을 만들고, 각각 로지스틱 회귀분석 기법을 적용해 모형을 구축한다. 전방 선택법을 통해 만든 변수군으로 구축한 모형은 F_Logit 이라고 칭하고, 후방 제거법을 통해 만든 변수군으로 구축한 모형은 B_Logit이라고 정의한다.

이어 머신러닝 기법인 인공신경망과 서포트 벡터머신을 적용해 모형을 구축할 때 로지스틱 회귀모형과 마찬가지로 과도한 변수는 예측 성과에 악영향을 줄 수 있다. 또한 인공신경망과 서포트 벡터 머신의 경우 다양한 종류의 하이퍼 파라미터가 존재하고, 적절한 하이퍼 파라미터를 설정할 경우 예측성고를 높일 수 있다. 변수 선정과 하이퍼 파라미터 설정에는 다양한 방법론이 존재하지만, 그중 본 연구에서는 유전 알고리즘(Genetic Algorithm, GA)을 사용해 변수를 선택하고 하이퍼 파라미터를 설정한다. 유전 알고리즘은 휴리스틱 기법을 활용한 최적화

알고리즘으로, 초기화(Initialization), 선택(selection), 진이(crossover and mutation), 중지(termination)의 단계로 이루어져 있고, 비용을 고려한 설계를 통해 인공지능 기법 예측 성과를 높이는 조합의 하이퍼 파라미터 튜닝이나 변수 선정에 사용하게 되면 예측 성과를 높일 수 있다(Pourbasheer et al., 2009; Sedighi and Afshari, 2010). 본 연구에서는 유전자 알고리즘을 적용한 인공신경망과 서포트 벡터 머신 모형을 구축하고, 각각 GA_NN, GA_SVM이라 정의한다.

마지막 합성곱 신경망 모형의 경우 전처리 단계에서 진행된 수치형 데이터의 이미지형 데이터 변환을 통해 생성된 데이터를 사용해 구축한다. 일반적인 합성곱 신경망의 구조와 같이 합성곱 계층(Convolution layer)과 풀링 계층(Pooling layer), 완전 연결 계층(Fully connected layer)을 차례로 쌓아 구축하고(이모세와 안현철, 2018), 합성곱 계층 및 풀링 계층, 그리고 완전 연결 계층의 노드 개수는 탐색 및 최적화 기법인 베이저안 최적화(Bayesian optimization)를 통해 탐색, 적용한다.

3.3 설명 가능한 인공지능(XAI) 구현

마지막 설명 가능한 인공지능 구현에서는 계층별 타당성 전파(LRP)를 활용해 모형 및 산출 결과에 대한 해석력을 제공한다. 계층별 타당성 전파의 기본적인 가정은 총 세 가지인데, 첫 번째는 각 뉴런은 어느 정도 출력 값에 기여하는 값을 가진다는 것이고, 두 번째는 타당성 점수 또는 기여도라 불리는 relevacne score는 출력 값으로부터 입력 층까지 재분배 된다는 것이다.

마지막으로 세 번째는 입력 층으로 기여도를 재분배시 기여도의 총 합은 보존된다는 것이다. 본 연구에서는 합성곱 신경망을 활용해 구축한 신용평가 모형의 결과 값에 LRP를 적용해 개별 관측치별로 예측한 결과에 영향을 크게 미친 변수들을 탐색해 모형의 투명도(Transparency)를 높이는 과정을 제시한다. 본 연구에서는 Bach et al.(2015)가 제시한 방법론을 따라 타당성 점수를 산출하기 위해 Taylor series를 사용한다.

IV. 실험 및 실험 결과

4.1 데이터 및 전처리

본 연구에서 실험에 사용하는 데이터는 미국의 P2P 대출 기업인 랜딩클럽의 개인별 신용 데이터이다. 랜딩클럽의 홈페이지인 Lendingclub.com에서 데이터를 직접 다운로드 했으며, 실험에 사용된 데이터의 기간은 2014년 1월 1일부터 2019년 12월 31일까지이다. 개

인의 신용 및 신상 정보에 관련된 변수의 개수는 150개이며, 데이터는 2,694,817개의 관측치로 구성되어 있다. 주요 변수들의 기초 통계량의 경우 <표 3>과 같다.

전처리 단계에서 결측치가 40% 이상 존재하는 변수는 제거했고, <표 3>에서 확인할 수 있는 것처럼 Number of mortgage account(모기지 계좌의 수)와 같은 변수에서 최대 값인 94와 같은 값들은 IQR방식을 적용해 이상치라 판단하고 제거했다. 또한 개인의 채무 불이행 여부를 판단하기 때문에 '공동 대출 신청자의 등록 시 기입한 연간 소득 합계'와 같이 공동 대출자 관련 변수 도 함께 제거했다. 또한 '대출 금액 중 미납된 원금 잔액'과 같이 대출 초기에 알 수 없는 변수와, 변수의 범주 개수가 1개 혹은 100개 이상인 변수도 연구 프레임워크에 따라 제거했다. 총 150개의 변수 중 전처리 이후 61개의 변수가 남았으며, 종속 변수인 '미래 대출 상환 여부'의 경우 'loan status'를 활용해 생성했다. 랜딩클럽에 제공한 데이터의 loan status는 범주형 변수로, fully paid, charged off, current, late 등으로 구성되어 있고, 본 연구에서는 완납

<표 3> 주요 변수들의 기초 통계량

변수명	결측치 비율	평균 (건전 대출자)	평균 (부실 대출자)	표준편차	최대값	최소값
Interest rates	0.0	12.0	16.0	5.0	31.0	5.0
Monthly payment	0.0	437.0	478.0	275.0	1720.0	8.0
Payment to date	0.0	16970.0	8919.0	10429.0	71629.0	0.0
Loan amount	0.0	14512.0	16126.0	9574.0	40000.0	725.0
Number of trade	23.6	3.0	3.0	3.0	78.0	0.0
Number of mortgage account	0.0	2.0	1.0	2.0	94.0	0.0
Number of card account	0.0	8.0	8.0	5.0	89.0	0.0
Total upper credit limit	0.0	186934.0	150952.0	186176.0	9999999.0	0.0
Number of public records negatively impacting credit	0.0	0.0	0.0	8.0	13485.0	0.0
Annual income	0.0	80092.0	73196.0	115316.0	110000000.0	0.0

을 의미하는 Fully paid를 건전 대출자로, 납부 포기를 의미하는 Charged off를 부실 대출자로 지정해 종속 변수로 사용했다. 또한 건전 대출자와 부실 대출자의 비율은 80.5:19.5로 불균형한 분포를 보이고 있는데, 이러한 불균형 데이터를 활용해 모형을 구축할 경우 기계 학습의 성능을 저하시키는 요인이 될 수 있다(홍태호 외, 2019). 따라서 본 연구에서는 불균형 데이터를 균형 데이터로 만들기 위해 각 데이터로부터 10,000개씩 층화추출, 즉 언더샘플링을 진행했다. 이어 모형의 학습과 평가를 위해 데이터를 학습데이터와 평가데이터로 분리했는데, 일반적으로 개인신용평가의 경우 미래의 잠재적인 대출 신청자를 올바르게 분류하는 것이 중요하기 때문에(Mai et al., 2019) 본 연구에서는 2014년부터 2018년까지의 데이터를 학습 데이터로, 2019년 데이터를 검증 데이터로 설정한 후 실험을 진행했다.

이어 합성곱 신경망 기법 적용을 위해 기존 수치형 데이터를 이미지 데이터로 변환했다. 변수가 61개로 8×8 구조의 픽셀을 구축했고, 3.1.1에서 제시한 입력 변수 픽셀 대응 법과

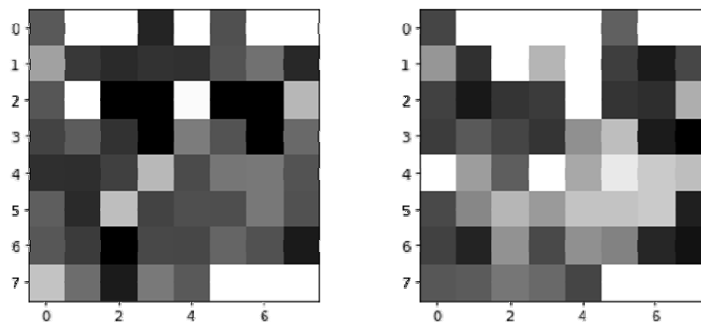
3.1.2에서 제시한 픽셀의 밝기 설정 법을 적용했다. 이어 유사한 특성을 가지고 있는 변수들 간의 픽셀 위치의 거리가 가까운 정도와, 상이한 특성을 가지고 있는 변수들 간의 픽셀 위치의 거리가 먼 정도인 CS(Correlation Score)를 최적화 방법인 몬테 카를로 시뮬레이션(Montecarlo simulation)을 사용해 최적화해 변수군의 픽셀 배치 조합을 탐색했다. <그림 3>은 변환된 개인별 신용 데이터를 이미지 형식으로 나타낸 한 예시이다.

4.2 예측 모형 구축

다음은 모형 구축 부분으로 통계적 기법인 로지스틱 회귀분석을 적용한 모형과, 머신러닝 기법인 인공신경망과 서포트 벡터 머신을 적용한 모형, 그리고 딥러닝 기법인 합성곱 신경망을 적용한 모형을 구축했다.

먼저 로지스틱 회귀분석을 적용한 모형의 경우, 변수 선정을 위해 t-검정을 실시해 5% 유의수준

에서 종속 변수와 유의미한 평균 차이를 보



<그림 3> 이미지 데이터로 변환된 신용 데이터 예시
(왼쪽부터 건전, 부실 대출자 데이터)

——신용 데이터의 이미지 변환을 활용한 합성곱 신경망과 설명 가능한 인공지능(XAI)을 이용한 개인신용평가

이는 입력 변수들을 선정했고, 이어 logit stepwise를 전방 선택법, 후방 소거법을 각각 적용해 총 2개의 변수군을 구축했다. 이어 로지스틱 회귀분석을 적용해 F_Logit, B_Logit 총 2개의 모형을 구축했다.

다음 유전 알고리즘을 변수 선정 및 하이퍼 파라미터 튜닝 기법으로 사용한 인공 신경망 모형의 경우, 인공 신경망의 노드의 개수와 Epoch, 활성화 함수 및 변수의 조합을 유전 알고리즘을 통해 탐색하고, 세대 마다 적합도를 평가용 데이터의 예측 성과로 설정해 다음 세대에 전이 되도록 설정해 GA_NN 모형을 구축했다. 마찬가지로 유전 알고리즘을 사용한 서포트 벡터 머신 모형의 경우에는 커널 함수, 감마, 그리고 C와 변수의 조합을 유전 알고리즘을 통해 탐색했고, 적합도를 평가용 데이터의 예측 성과로 설정해 GA_SVM 모형을 구축했다.

마지막으로 합성곱 신경망 모형은 4.1.에서 변환된 이미지형 데이터를 입력값으로 사용했으며, 이미지의 크기가 8x8 형식으로 일반적으

로 이미지 분류에 사용되는 형식보다 크기가 작기 때문에 합성곱 계층 및 풀링 계층은 차례로 두 번 쌓아서 진행했으며, 과적합을 방지하기 위해 Dropout층을 추가했다. 완전 연결 계층의 노드 개수는 베이지안 최적화를 통해 탐색, 적용해 CNN 모형을 구축했다.

4.3 예측 성과 비교

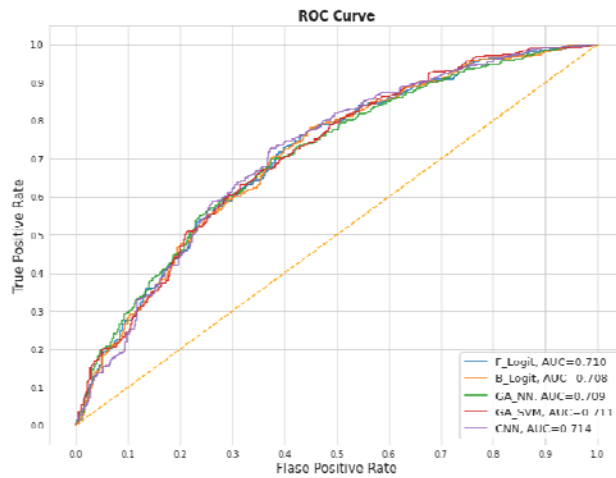
<표 4>는 각 모형의 예측 성과를 나타낸다. 학습(Train) 및 평가(Validation)데이터의 경우 2014년부터 2018년도의 개인별 신용 데이터를 사용했고, 검증(Test)데이터의 경우 2019년도의 데이터를 사용했다. 딥러닝 기법을 적용한 CNN 모형이 통계 기법을 적용한 F_Logit, B_Logit 모형, 그리고 유전 알고리즘과 머신러닝 기법을 적용한 GA_NN, GA_SVM 모형 보다 예측 성과가 우수하게 나타났다. <그림 4>의 ROC(Receiver Operating Characteristic) 곡선과 AUC(Area Under ROC Curve) 점수를 통

<표 4> 모형 별 예측 정확도

모형 \ 데이터셋	F_Logit	B_Logit	GA_NN	GA_SVM	CNN
학습 데이터	65.26	65.34	65.66	65.80	67.68
평가 데이터			66.86	66.20	66.89
검증 데이터	65.23	65.47	65.70	65.97	67.30

<표 5> 맥니마 검정 결과

	B_Logit	GA_NN	GA_SVM	CNN
F_Logit	0.06 (0.803)	74.21*** (0.000)	2.83* (0.09)	110.62*** (0.000)
B_Logit		70.83*** (0.000)	2.25 (0.13)	107.08*** (0.000)
GA_NN			44.08*** (0.000)	7.56*** (0.005)
GA_SVM				74.41*** (0.000)



<그림 4> 모형 별 ROC 곡선과 AUC 점수

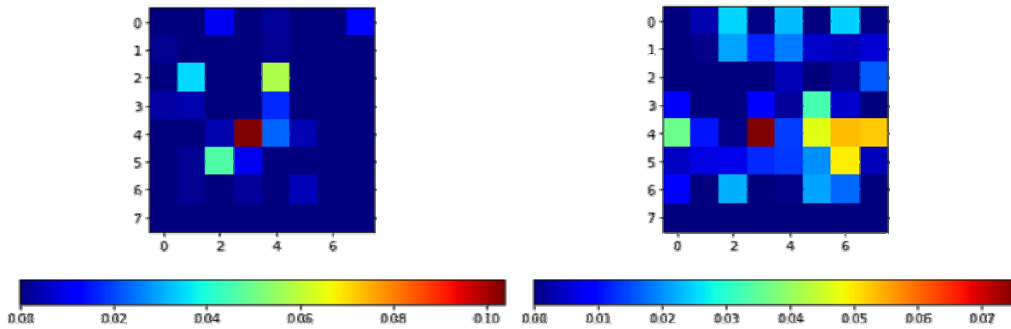
해서도 성과를 확인할 수 있는데, 일반적으로 ROC 곡선 아래 면적을 계산한 AUC의 값이 클수록 더 좋은 분류기로 판단된다(홍종선과 장동환, 2020). CNN 모형의 AUC 값이 0.714로 다른 모형들의 AUC 값보다 높아 더 좋은 분류 모형으로 평가할 수 있다. 또한 본 연구에서는 모형 간 예측 성과의 차이가 유의미함을 확인하기 위해 두 집단 사이의 평균 차이를 검정하기 위해 사용되는 맥니마(McNemar) 검정을 실시했다. <표 5>는 맥니마 검정 결과이다. CNN 모형은 F_Logit, B_Logit, GA_NN, GA_SVM 모형 모두와 유의수준 1%에서 통계적으로 유의미한 차이를 보여 예측 성과의 차이가 통계적으로 유의미하다고 볼 수 있다.

4.4 설명 가능한 인공지능 구현

<그림 5>는 구축한 신용평가 모형에 LRP를 적용해 <그림 3>의 데이터를 분류한 모형의 관측치 별 히트맵을 나타낸다. LRP를 활용하면

각 픽셀에 해당하는 변수가 결과에 미치는 영향력을 파악할 수 있고, 이를 통해 합성곱 신경망을 적용한 딥러닝 모형의 블랙박스 문제를 해결하는 데 도움을 줄 수 있다. 먼저 건전 대출자의 히트맵을 나타내는 좌측 그림을 살펴보면, 좌측 중앙에 위치한 변수들의 중요도가 높은 것을 확인할 수 있다. 해당 관측치에서 영향력이 큰 변수들은 거주 주택의 자가 여부, 지난 12개월 동안 개설된 계좌 수, 지난 24개월 동안 개설 계좌 수, 연체가 없는 계좌 비율이고, 해당 관측치의 경우 주택이 자가이며, 지난 12, 24개월 동안 개설된 계좌 수가 상대적으로 많으며, 연체가 없는 계좌의 비율이 상대적으로 높기 때문에 모형이 건전 대출자로 분류했다고 해석할 수 있다. 또한 부실 대출자의 사례를 나타내는 우측의 그림을 살펴보면, 중앙에 위치한 변수들의 중요도가 높은 것을 확인할 수 있다. 해당 관측치에서 영향력이 큰 변수들은 모기지 계좌 수, 지난 24개월 동안 개설된 계좌 수, 대출 범위에 있는 차용자의 FICO가 속한 상위 경제

——신용 데이터의 이미지 변환을 활용한 합성곱 신경망과 설명 가능한 인공지능(XAI)을 이용한 개인신용평가



<그림 5> LRP를 각 관측치에 적용해 산출한 히트맵(왼쪽부터 건전, 부실 대출자 데이터)

범위, 리볼빙 계좌 수, 활성화된 리볼빙 계좌 수, 사용 금액이 있는 리볼빙 계좌수, FICO 속 한 하위 경계 범위이고, 모기지 계좌수가 많고, 리볼빙 계좌수가 상대적으로 많으며, 신용점수를 나타내는 FICO 점수가 높지 않아 부실 대출자로 분류했다고 해석할 수 있다. 이처럼 합성곱 신경망을 사용해 구축한 모형에 LRP를 적용할 경우 각 개별 관측치 별로 예측 모형이 건전 대출자와 부실 대출자로 판단한 이유에 대한 투명성을 높일 수 있고 현업에 종사하는 신용평가 담당자의 경우 대출 거절에 대한 이유를 고객들에게 설명할 수 있다. 또한 영향력이 크다고 선정된 변수들은 선행 연구에서도 개인 신용평가에 유의미한 영향을 미친다고 판단되고 있는데, 예를 들어 Emekter et al.(2015)은 FICO 점수가 채무 불이행 여부를 구분하는 중요한 변수가 될 수 있다고 판단했고, 배재권 등(2018)은 주택의 자가 여부, 모기지 계좌 수, 연체가 있는 계좌 수를, 홍태호 등(2019)은 리볼빙 계좌 수를 개인신용평가 모형에 입력 변수로 사용했다. 이는 LRP를 활용해 탐색한 변수가 건전, 부실 대출자 구별에 주요한 영향을 미친 변수라고 판단할 수 있다.

V. 결론

최근 COVID-19 팬데믹이 지속됨에 따라 경제적 타격을 입은 가계를 지원하기 위한 대출 지원 정책이 추진되고 있는 상황에서 신용 평가 모형의 정확도를 높이는 것은 팬데믹이 끝난 이후에도 은행, 카드사를 포함한 금융기관의 건전성 위해 중요한 의의를 가질 수 있다. 본 연구는 최근 이미지 인식, 이미지 분류, 자연어 처리 등 다양한 분야에 활용되고 있는 딥러닝 기법인 합성곱 신경망을 개인신용평가에 적용했다. 이미지 분류에 우수한 성능을 보이는 합성곱 신경망을 적용하기 위해 수치형 데이터인 개인별 신용 데이터를 이미지 데이터로 변환했으며, 변환된 이미지 데이터에 합성곱 신경망을 적용해 대출 상환 여부를 예측하는 모형을 구축했다. 실험 결과 합성곱 신경망을 적용한 모형이 기존 신용평가 연구에서 사용되는 통계적 기법 및 머신러닝 기법인 로지스틱 회귀분석, 인공 신경망, 서포트 벡터 머신을 적용한 모형보다 우수한 예측 성과를 보였으며, 맥니마 검증을 통해 예측 성과의 차이가 통계적으로 유의미하였다. 또한 합성곱 신경망을 적용한 모형

에 계층별 타당성 전과(LRP)를 활용해 딥러닝 모형으로 개인 채무자의 미래 대출 상환 여부를 판단할 때 어떤 변수들이 분류에 큰 영향을 미치는지 파악하는 방법을 제시했다.

본 연구의 의의는 다음과 같다. 먼저 개인신용평가에 잘 사용되지 않았던 합성곱 신경망을 적용해 신용 평가 모형을 구축했다. 기존의 개인신용평가 모형과 관련된 선행 연구에서 합성곱 신경망이 잘 사용되지 않았던 이유 중 하나는 신용 평가 데이터는 수치형 데이터인 반면, 합성곱 신경망은 이미지 데이터를 처리할 때 우수한 성능을 보인다는 점이었다. 하지만 본 연구에서는 수치형 데이터를 이미지 형식의 데이터로 변환해 합성곱 신경망의 입력으로 사용할 경우 합성곱 신경망의 이점을 살릴 수 있다는 것을 실증 분석을 통해 보였다. 이를 통해 예측성과를 향상시킬 수 있었으며, 향상된 예측 성과는 은행 및 P2P 대출 업체가 대출시 건전 고객을 건전 고객으로, 부실 고객을 부실 고객으로 더 잘 분류할 수 있다는 뜻이고 이는 금융 기관의 영업 이익과 직접적 관련이 있기에 중요한 의미를 가진다. 두 번째로, 딥러닝 모형의 경우 구조가 복잡하고 그에 따라 모형이 판단하는 기준이나 해석이 불명확하다는 특징을 가지고, 예측력에서는 뛰어난 성과를 보이지만 실무에 직접 적용하기에는 해석력에 약점이 있다는 비판을 받아왔다. 본 연구에서는 딥러닝 모형의 구조를 금융 분야에서는 거의 사용되지 않았던 계층별 타당성 전과를 활용해 설명 가능한 인공지능을 구현했고, 블랙박스의 특징을 가지고 있는 딥러닝 모형에서 관측치 별 중요 변수를 파악하는 방법을 제시했다. 이를 활용해 금융 기관에서는 내부 직원의 모형에 대한 이

해도를 높이고, 외부 고객에게 더 설득력 있는 설명이 가능하도록 신용 평가 모형의 투명도를 높일 수 있다. 특히 EU에서는 2018년 딥러닝을 포함한 알고리즘을 활용해 자동화된 서비스를 제공하는 기업들에 소비자가 설명을 요구할 권리를 가지게 하는 개인정보 보호규정을 신설하는 등(천예은 외, 2021) 전 세계적으로 설명 가능한 인공 지능에 대한 필요성이 증대되고 있는 시점에서 중요한 의미를 갖는다고 볼 수 있다.

본 연구의 한계 및 추후 연구 방향은 다음과 같다. 첫째, 본 연구는 기업 도산 예측에 합성곱 신경망을 적용하는 선행 연구의 방법론을 바탕으로 수치형 데이터를 이미지 데이터로 변환했다. 하지만 기업 도산 예측과 개인신용평가는 금융 분야라는 공통점이 있지만, 변수의 특성 및 데이터의 구조에 차이가 존재한다. 따라서 추후 연구에서는 수치형 데이터를 이미지 데이터로 변환하는 더 다양한 방법을 탐구할 필요가 있다. 둘째, 본 연구에서 61개의 변수를 가지고 있는 수치형 데이터를 이미지 데이터로 변환했고, 8x8 형식의 픽셀을 합성곱 신경망의 입력으로 사용했다. 이는 합성곱 신경망을 적용하는 다른 선행 연구들 보다 작은 크기를 가지는 입력 데이터를 사용한 것인데, 이러한 이유 때문에 GoogLeNet이나 ResNet과 같이 기존 이미지 분류 문제에 다수 사용되고 우수한 성과가 증명된 모형들을 적용해보지 못했다. 따라서 추후 연구에서는 개인신용평가 도메인에서의 파생 변수 생성 및 변수 조합법을 더 연구해 픽셀의 크기를 확대시킬 필요가 있다. 마지막으로 설명가능한 인공지능(XAI)의 경우 LIME (Local Interpretable Model-agnostic

Explanations)이나 SHAP(SHapley Additive exPlanations) 등 다양한 방법을 활용해 구현할 수 있다. 본 연구에서는 합성곱 신경망을 해석하는 데 가장 보편적으로 활용되고 있는 계층별 타당성 전파를 활용해 구현했지만, 다양한 방법론을 활용해 중요 변수 집합의 공통점을 분석하는 등 더 심층적인 분석이 필요하다.

참고문헌

- 강구상, 최원석, 김종혁, 오태현, 이현진, 김승현, 박나연, “코로나 19 대응 주요국의 재정 및 통화금융 정책”, [KIEP] 세계경제 포커스, 제20권, 제11호, 2020, pp. 1-19
- 김성진, 안현철, “기업신용등급 예측을 위한 랜덤 포레스트의 응용”, 산업혁신연구, 제32권, 제1호, 2016, pp. 187-211.
- 김은미, 박지영, “온라인 P2P 환경에서 딥러닝을 적용한 다분류 기반 개인신용등급 예측모형”. 인터넷전자상거래연구, 제19권, 제4호, 2019, pp. 43-57.
- 김지윤, 조성배, “컨볼루션 신경망의 소셜대출상환예측을 위한 요인분석 군집화기반 2D 입력배열 방법”, 한국정보과학회 학술발표논문집, 2017, pp. 1337-1339.
- 박대현, 류두진, “P2P대출과 공급망 금융”, 재무연구, 제33권, 제3호, 2020, pp. 439-463.
- 박성우, “개인 신용 평가 예측에 대한 다양한 머신러닝 기법 연구”, 정보 및 제어 논문집, 2017, pp. 291-292.
- 배재권, 이승연, 서희진, “인공지능기법을 이용한 온라인 P2P 대출거래의 채무불이행 예측에 관한 실증연구”, 한국전자거래학회지, 제23권, 제3호, 2018, pp. 207-224.
- 유환, 이영재, “은행 금융상품에서 프라이빗 뱅커의 전문투자형 사모펀드 추천 의사결정”, 정보시스템연구, 제28권, 제4호, 2019, pp. 333-358.
- 윤종식, 노태협, 권영식, “신용카드 매출정보를 활용한 소상공인신용평가모형”, 중소기업연구, 제29권, 제2호, 2007, pp. 73-98.
- 이모세, 안현철, “효과적인 입력 변수 패턴 학습을 위한 시계열 그래프 기반 합성곱 신경망 모형: 주식시장 예측에의 응용”, 지능정보연구, 제24권, 제1호, 2018, pp. 167-181.
- 이성욱, 노태협, “통합 수리계획법을 이용한 개인신용평가모형”, 정보시스템연구, 제16권, 제1호, 2007, pp. 1-21.
- 이재식, 한재홍, “인공신경망을 이용한 중소기업 도산 예측에 있어서의 비재무정보의 유용성 검증”, 한국전문가시스템학회지, 제1권, 제1호, 1995, pp. 123-134.
- 이현상, 오세환, “시계열 예측을 위한 LSTM 기반 딥러닝: 기업 신용평점 예측 사례”, 정보시스템연구, 제29권, 제1호, 2020, pp. 241-265.
- 조유정, 손권상, 권오병, 인공지능의 사회적 수용도에 따른 키워드 검색량 기반 주가 예측모형 비교연구. 지능정보연구, 제27권, 제1호, 2021, pp. 103-128.

- 천예은, 김세빈, 이자윤, 우지환, “설명 가능한 AI 기술을 활용한 신용평가 모형에 대한 연구”, 한국데이터정보과학회지, 제32권, 제2호, 2021, pp. 283-295.
- 최수만, 전동화, 오경주, “P2P플랫폼에서의 대출자 신용분석 사례연구: 8 퍼센트, 렌딧, 어니스트 펀드”, 지식경영연구, 제21권, 제3호, 2020, pp. 229-247.
- 홍종선, 장동환, “ROC곡선 길이의 판단기준”, 한국데이터정보과학회지, 제31권, 제5호, 2020, pp. 851-863.
- 홍태호, 김성훈, 김은미, “GAN과 DNN을 활용한 딥러닝 기반의 지능형 개인신용 평가모형”, 인터넷전자상거래연구, 제19권, 제1호, 2019, pp. 1-16.
- Abdou, H. A. and Pointon, J., “Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature. Intelligentsystems in Accounting”, *finance and management*, Vol. 18, No. 2-3, 2011, pp. 59-88.
- Adadi, A. and Berrada, M., “Peeking Inside the Black-Box: a Survey on Explainable Artificial Intelligence (XAI)”, *IEEE access*, Vol. 6, 2018, pp. 52138-52160.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... and Herrera, F., “Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI”, *Information Fusion*, Vol. 58, 2020, pp. 82-115.
- Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., and Samek, W., “On Pixel-wise Explanations for Non-linear Classifier Decisions by Layer-wise Relevance Propagation”. *PloS one*, Vol. 10, No. 7, 2015, e0130140.
- Bae, J. K. and Kim, J., “A Personal Credit Rating Prediction Model using Data Mining in Smart Ubiquitous Environments”, *International Journal of Distributed Sensor Networks*, Vol. 11, No. 9, 2015, 179060.
- Baesens, B., Setiono, R., Mues, C., and Vanthienen, J., “Using Neural Network Rule Extraction and Decision Tables for Credit-risk Evaluation”, *Management science*, Vol. 49, No. 3, 2003, pp. 312-329.
- Bayraci, S. and Susuz, O., “A Deep Neural Network(DNN) based Classification Model in Application to Loan Default Prediction.”, *Theoretical and Applied Economics*, Vol. 22, No. 4, 2019, pp. 75-84.
- Böhle, M., Eitel, F., Weygandt, M., and Ritter, K., “Layer-wise Relevance Propagation for Explaining Deep Neural Network Decisions in MRI-based Alzheimer's Disease Classification”, *Frontiers in aging neuroscience*, Vol. 11, 2019, pp. 194.
- Breiman, L., “Random Forests”. *Machine learning*, Vol. 45, No. 1, 2001, pp. 5-32.

- Buyukyilmaz, M. and Cibikdiken, A. O., “Voicegender Recognition using Deep Learning. In 2016 International Conference on Modeling”, *Simulation and Optimization Technologies and Applications (MSOTA2016)*. December 2016, pp. 409-411.
- Byanjankar, A., Heikkilä, M., and Mezei, J., “Predicting Credit Risk in Peer-to-peer Lending: A Neural Network Approach”, *In 2015 IEEE Symposium Series on Computational Intelligence*. December 2015, pp. 719-725.
- Ding, X., Zhang, Y., Liu, T., and Duan, J. “Deep Learning for Event-driven Stock Prediction”, *In Twenty-fourth inter national joint conference on artificial intelligence*. June 2015.
- Durand, D., “Risk Elements in Consumer Installment Financing”. *National Bureau of Economic Research*, 1941.
- Emekter, R., Tu, Y., Jirasakuldech, B., and Lu, M., “Evaluating Credit Risk and Loan Performance in Online Peer-to-Peer (P2P) Lending”, *Applied Economics*, Vol. 47. No.1, 2015, pp. 54-70.
- Fung, G., Sandilya, S., and Rao, R. B. “Rule Extraction from Linear Support Vector Machines”, *In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, August 2005, pp. 32-40.
- Futagami, K., Fukazawa, Y., Kapoor, N., and Kito, T., “Pairwise Acquisition Prediction with SHAP Value Interpretation”. *The Journal of Finance and Data Science*, Vol. 7, 2021, pp. 22-44.
- Grezmak, J., Zhang, J., Wang, P., Loparo, K. A., and Gao, R. X., “Interpretable Convolutional Neural Network through Layer-wise Relevance Propagation for Machine Fault Diagnosis”. *IEEE Sensors Journal*, Vol. 20, No. 6, 2019, pp. 3172-3181.
- Han, D., Liu, Q., and Fan, W., “A New Image Classification Method using CNN Transfer Learning and Web Data Augmentation”, *Expert Systems with Applications*, Vol. 95, 2018, pp. 43-56.
- Hand, David J., and William E. Henley. “Statistical Classification Methods in Consumer Credit Scoring: a Review.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* Vol. 160, No. 3 1997, pp. 523-541.
- Hájek, P., “Municipal Credit Rating Modelling by Neural Networks.” *Decision Support Systems*, Vol. 51, No. 1, 2011, pp. 108-118.
- He, K., Zhang, X., Ren, S., and Sun, J., “Deep Residual Learning for Image Recognition.”, *In Proceedings of the IEEE conference on computer vision*

- and pattern recognition*, 2016, pp. 770-778.
- Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., and Scholkopf, B., "Support Vector Machines". *IEEE Intelligent Systems and their applications*, Vol. 13, No. 4, 1998, pp. 18-28.
- Hosaka, T., "Bankruptcy Prediction using Imaged financial Ratios and Convolutional Neural Networks", *Expert systems with applications*, Vol. 117, 2019, pp. 287-299.
- Jung, Y. J., Han, S. H., and Choi, H. J., "Explaining CNN and RNN Using Selective Layer-Wise Relevance Propagation". *IEEE Access*, 9, 18670-18681.
- Kim, B., Park, J., and Suh, J., "Transparency and Accountability in AI Decision Support: Explaining and Visualizing Convolutional Neural Networks for Text Information", *Decision Support Systems*, Vol. 134, 2020, 113302.
- Kvamme, H., Sellereite, N., Aas, K., and Sjursen, S., "Predicting Mortgage Default using Convolutional Neural Networks", *Expert Systems with Applications*, Vol. 102, 2021, pp. 207-217.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., "Gradient-based Learning Applied to Document Recognition". *Proceedings of the IEEE*, Vol. 86, No. 11, 1998, pp. 2278-2324.
- Lee, H. and Kwon, H., "Going Deeper with Contextual CNN for Hyperspectral Image Classification", *IEEE Transactions on Image Processing*, Vol. 26, No. 10, 2017, pp. 4843-4855.
- Li, H., Tian, Y., Mueller, K., and Chen, X., "Beyond Saliency: Understanding Convolutional Neural Networks from Saliency Prediction on Layer-wise Relevance Propagation". *Image and Vision Computing*, Vol. 83, 2019, pp. 70-86.
- Mai, F., Tian, S., Lee, C., and Ma, L., "Deep Learning Models for Bankruptcy Prediction using Textual Disclosures", *European journal of operational research*, Vol. 274, No. 2, 2019, pp. 743-758.
- Martens, D., Baesens, B., Van Gestel, T., and Vanthienen, J., "Comprehensible Credit Scoring Models using Rule Extraction from Support Vector Machines", *European journal of operational research*, Vol. 183, No. 3, 2007, pp. 1466-1476.
- Moscato, V., Picariello, A., and Sperlì, G. "A Benchmark of Machine Learning Approaches for Credit Score Prediction", *Expert Systems with Applications*, Vol. 165, 2021, 113986.
- Nehrebecka, N., "Predicting the Default Risk of Companies. Comparison of Credit

- Scoring Models: LOGIT vs Support Vector Machines”, *Ekonometria*, Vol. 22, No. 2, 2018, pp. 54-73.
- Orgler, Y. E., “A Credit Scoring Model for Commercial Loans”. *Journal of money, Credit and Banking*, Vol. 2, No. 4, 1970, pp. 435-445.
- Pant, D. R., Neupane, P., Poudel, A., Pokhrel, A. K., and Lama, B.K., “Recurrent Neural Network based Bitcoin Price Prediction by Twitter Sentiment Analysis”, *In 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*, October 2018, pp. 128-132.
- Pourbasheer, E., Riahi, S., Ganjali, M. R., and Norouzi, P., “Application of Genetic Algorithm - Support Vector Machine (GA-SVM) for Prediction of BK-channels Activity”, *European journal of medicinalchemistry*, Vol. 44, No. 12, 2009, pp. 5023-5028.
- Rao, C., Liu, M., Goh, M., and Wen, J., “2-Stage Modified Random Forest Model for Credit Risk Assessment of P2P Network Lending to “Three Rurals” Borrowers“. *Applied Soft Computing*, Vol. 95, 2020. 106570.
- Roondiwala, M., Patel, H., and Varma, S., “Predicting Stock Prices using LSTM”, *International Journal of Science and Research(IJSR)*, Vol. 6, No. 4, 2017, pp. 1754-1756.
- Ruyu, B., Mo, H. and Haifeng, L., “A Comparison of Credit Rating Classification Models Based on Spark-Evidence from Lending-club”. *Procedia Computer Science*, Vol. 162, 2019, pp. 811-818.
- Van Lent, M., Fisher, W., and Mancuso, M., “An Explainable Artificial Intelligence System for Small-unit Tactical Behavior”. *In Proceedings of the national conference on artificial intelligence*, July 2004, pp. 900-907.
- Vapnik, V., “The Nature of Statistical Learning Theory”, *Springer science & business media*. 2013.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., “Dropout: a Simple Way to Prevent Neural Networks from Overfitting”, *The journal of machine learning research*, Vol. 15, No. 1, 2014, pp. 1929-1958.
- Sedighi, M. and Afshari, D., “Creep Feed Grinding Optimization by an Integrated GA-NN System”. *Journal of Intelligentmanufacturing*, Vol. 21, No. 6, 2010, pp. 657-663.
- Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K.,and Soman, K. P. , “Stock Price Prediction using LSTM, RNN and CNN-sliding Window Model”. *In 2017 international conference onadvances in computing*,

- communications and informatics (icacci*, September 2017, pp. 1643-1647.
- Shortliffe, E. H. and Buchanan, B. G., “A Model of Inexact Reasoning in Medicine”, *Mathematical biosciences*, Vol. 23, No. 3-4, 1975, pp. 351-379.
- Sohn, S. Y. and Kim, J. W., “Decision Tree-based Technology Credit Scoring for Start-up Firms: Korean Case”, *Expert Systems with Applications*, Vol. 39 No. 4, 2012, pp. 4007-4012.
- Srinivas, M. and Patnaik, L. M., “Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms”. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 24, No. 4, 1994, pp. 656-667.
- Takahashi, N., Gygli, M., and Van Gool, L., “Aenet: Learning Deep Audio Features for Video Analysis”. *IEEE Transactions onMultimedia*, Vol. 20, No. 3, 2017, pp. 513-524.
- Triki, I., “Credit Scoring Models for a Tunisian Microfinance Institution: Comparison between Artificial Neural Network and Logistic Regression”. *Review of Economics & Finance*, Vol. 6, 2016, pp. 61-78.
- Wiginton, J. C., “A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior”. *Journal of Financial and Quantitative Analysis*, 1980, pp. 757-770.
- Yang, H., Li, E., Cai, Y. F., Li, J., and Yuan, G. X., “The Extraction of Early Warning Features for Predicting Financial Distress based on XGBoost Model and Shap Framework”, *International Journal of Financial Engineering*, Vol. 8, No. 3, 2021, 2141004.
- Zhang, Y., Jia, H., Diao, Y., Hai, M., and Li, H., “Research on Credit Scoring by Fusing Social Media Information in Online Peer-to-Peer Lending”. *Procedia Computer Science*, Vol. 91, 2016, pp. 168-174.

원 종 관 (Won, Jong Gwan)



부산대학교 경영학과에서 학사학위를 취득하였다. 현재 부산대학교 경영학과 경영정보전공 석사과정에 재학 중이다. 주요 관심분야는 지능형 테크핀, 암호화폐 예측, 신용평가, 딥러닝, AI 등이다.

홍 태 호 (Hong, Tae Ho)



현재 부산대학교 경영학과 교수로 재직 중이다. KAIST에서 경영정보시스템을 전공하여공학석사와 공학박사를 취득하였다. 주요 관심분야는 비즈니스 애널리틱스, 딥러닝, 오피니언 마이닝, CRM 등이다. 주요 논문을 Expert Systems, Expert Systems with Applications, Information Processing&Management, Asia Pacific Journal ofInformation Systems, 정보시스템연구 등에 게재하였다.

배 경 일 (Bae, Kyoung Il)



1994년 2월: 한국과학기술원 수학과 학사
1996년 2월: 한국과학기술원 산업경영학과 석사
2002년 2월: 한국과학기술원 경영공학과 박사
2002년 4월~2005년 7월, 2011년 11월~2014년 2월: 한국IBM
2014년 2월~2016년 6월: 포스코기술투자
2018년 10월~현재: (주)아토피스랩
<관심분야> 블록체인, 암호학, 금융공학

<Abstract>

A Personal Credit Rating Using Convolutional Neural Networks with Transformation of Credit Data to Imaged Data and eXplainable Artificial Intelligence(XAI)

Won, Jong Gwan · Hong, Tae Ho · Bae, Kyoung Il

Purpose

The purpose of this study is to enhance the accuracy score of personal credit scoring using the convolutional neural networks and secure the transparency of the deep learning model using eXplainable Artificial Intelligence(XAI) technique.

Design/methodology/approach

This study built a classification model by using the convolutional neural networks(CNN) and applied a methodology that is transformation of numerical data to imaged data to apply CNN on personal credit data. Then layer-wise relevance propagation(LRP) was applied to model we constructed to find what variables are more influenced to the output value.

Findings

According to the empirical analysis result, this study confirmed that accuracy score by model using CNN is highest among other models using logistic regression, neural networks, and support vector machines. In addition, With the LRP that is one of the technique of XAI, variables that have a great influence on calculating the output value for each observation could be found.

Keyword: Convolutional Neural Networks, eXplainable Artificial Intelligence, Deep learning, Imaged data, Personal credit rating

* 이 논문은 2021년 11월 10일 접수, 2021년 12월 10일 1차 심사, 2021년 12월 20일 게재 확정되었습니다.