

한국 NPL시장 수익률 예측에 관한 연구

A study on the prediction of korean NPL market return

저자 (Authors)	이현수, 정승환, 오경주 Hyeon Su Lee, Seung Hwan Jeong, Kyong Joo Oh
출처 (Source)	지능정보연구 25(2) , 2019.6, 123-139(17 pages) Journal of Intelligence and Information Systems 25(2) , 2019.6, 123-139(17 pages)
발행처 (Publisher)	한국지능정보시스템학회 Korea Intelligent Information Systems Society
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE08746401
APA Style	이현수, 정승환, 오경주 (2019). 한국 NPL시장 수익률 예측에 관한 연구. 지능정보연구, 25(2), 123-139
이용정보 (Accessed)	고려대학교 163.***.133.25 2021/02/21 20:10 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

한국 NPL시장 수익률 예측에 관한 연구

이현수

연세대학교 투자정보공학과
(dm1hs@naver.com)

정승환

연세대학교 산업공학과
(jsh0331@yonsei.ac.kr)

오경주

연세대학교 산업공학과
(johanoh@yonsei.ac.kr)

국내 NPL (Non performing loan) 시장은 1998년에 형성되었지만, 본격적으로 활성화 된 시기는 2009년으로 역사가 짧은 시장이다. 이로 인해 NPL 시장에 대한 연구도 아직까지는 활발히 진행되지 않고 있는 상황이다. 본 연구는 NPL 시장의 각 물건 별 기준 수익률 달성 유무를 예측할 수 있는 모델을 제안한다. 모델 구축에 사용되는 종속변수는 물건 별 최종 수익률이 기준 수익률 수치 도달 여부를 나타내는 이항변수를 사용하였고, 독립변수로는 물건의 특성을 나타내는 11개의 변수를 대상으로 one to one t-test와 logistic regression stepwise, decision tree를 수행하여 의미있는 7개의 독립변수를 선별하였다. 그리고 통상적으로 사용되는 기준 수익률 수치(12%)가 의미있는 기준 수치인지 확인하기 위해 수치 값을 조절해가며 종속변수를 산출하여 예측모델을 구축해보았다. 그 결과 12%의 기준 수익률 수치로 산출한 종속변수를 이용하여 구축한 예측모델의 평균 Hit ratio가 64.60%로 가장 우수하다는 결과를 얻었다. 다음으로 선별된 7개의 독립변수들과 12%를 기준으로한 수익률 달성유무 종속변수를 이용하여 판별분석, 로지스틱 회귀분석, 의사결정나무, 인공신경망, 유전자알고리즘 선형 모델의 5가지 방법론을 적용해 예측모델을 구축해보았다. 5가지 방법론으로 도출한 예측 모델 간 Hit ratio를 비교한 결과 인공신경망을 이용하여 구축한 예측모델의 Hit ratio가 67.4%로 가장 우수한 결과를 도출해내었다. 본 연구를 통해 추후 NPL시장 신규 물건 매매에 있어서 7가지의 독립변수들과 인공신경망 예측 모델을 활용하는 것이 효과적임을 증명하였다. 물건의 12% 수익률 달성 여부를 사전에 예측해봄으로써 유동화회사가 투자 의사결정을 하는 데에 도움을 줄 것으로 예상하며, 나아가 NPL 시장의 거래가 적정한 가격 선에서 진행됨으로 인해 유동성이 더욱 높아질 것이라 기대한다.

주제어 : 로지스틱 회귀분석, 유전자 알고리즘, 의사결정나무, 인공신경망, NPL

논문접수일 : 2019년 3월 9일 논문수정일 : 2019년 3월 26일 게재확정일 : 2019년 4월 29일
원고유형 : 일반논문 교신저자 : 오경주

1. 개요

국내 NPL (Non performing loan) 시장은 1998년 IMF 외환위기 직후 정부와 외국자본에 의하여 형성이 되었고, 이후 금융기관의 대규모 구조조정 및 신용카드사태 등을 거치면서 성장하였다. 이후 2000년대 중반 금융기관의 부실채권 비율이 다소 안정되는 시기를 거치면서 성장이 정체되었고, 소수의 전문투자자들의 간헐적인 투

자만이 이루어졌다. 이러한 태동기를 거쳐 2009년 국제금융위기 이후 실물경기침체로 부실채권이 증가하기 시작하면서 본격적으로 NPL 시장이 활성화되기 시작하였다. 특히 2010년 은행들이 출자하여 설립한 연합자산관리 (UAMCO)의 출현 등으로 은행들의 NPL매각 (공급) 물량이 급격히 늘어나기 시작하였고, 국내 자본시장의 투자자본이 NPL시장에 본격적으로 진출하면서 부터인 최근에서야 NPL이 시장에서 주요 투자

상품으로 자리 잡게 되었다. 따라서 자본시장의 NPL 시장에 대한 투자 역사가 짧고, 이에 대한 연구도 일천한 상황이다.

이러한 가운데 국내 은행의 부실채권은 최근 10년간 연평균 25조 정도 발생하고 있고, 이중 매년 약 5조 원 가량의 매물이 시장에 쏟아져 나오고 있다. 국내 NPL시장의 짧은 역사, 이로 인한 NPL 투자에 대한 Closing 사례 부족, 관련 데이터의 부족 등으로 아직은 NPL 자체적인 통계만으로 NPL Portfolio의 기초자산인 다양한 부동산담보채권 Portfolio의 적정 가격을 정확히 예측하기는 어려운 것이 현실이다. 그동안은 과거의 경험치에 의한 경매시장에서 해당 담보 부동산의 평균 낙찰가율, 주변 시세 등을 참고하여 일정 할인을 통해 매입하는 것이 통상적인 가격 결정 방법이었다. 그러나 최근 NPL 시장의 과열 및 수익성 하락과 부동산 경기변동에 따른 가격 변동 등의 사유로 좀 더 과학적이고 체계적인 분석을 통한 가격 결정이 요구되고 있다.

본 연구에서는 이러한 시장의 요구에 맞추어 NPL 시장 관련 데이터들을 활용하여 기준 수익률의 달성유무를 판별할 수 있는 예측모델을 제안하려고 한다. 이러한 시장분석 및 기준 수익률 달성유무 판별 모델을 통해 NPL 시장의 형성과 유통을 원활하게 하는 데에 본 연구의 목적이 있다.

본 논문의 구성은 다음을 따른다. 2장 연구배경에서는 NPL 시장이 무엇인지, 그리고 과거에 어떠한 관련 연구들이 진행되었는지 서술하고, 3장 연구방법에서는 NPL 시장을 분석하기 위한 방법론들을 서술하였다. 4장 실증분석에서는 NPL 시장 데이터와 분석 방법론을 활용하여 분석한 내용을 서술하였고, 5장 결론 및 제언에서는 본 연구 실험의 결과 정리 및 결론을 제시한다.

2. 연구배경

NPL이란 은행 등 금융기관에서 대출이 이루어진 후 채무자의 경제력 상실이나 부도 등으로 인하여 대출채무 원리금 상환이 불이행상태로 되어 상환 가능성이 낮은 대출채권을 의미한다. 은행업 감독규정에서 금융기관의 자산 건전성을 정상, 주의, 고정, 회수의문, 추정손실의 5개 단계로 구분하였을 때, 고정, 회수의문, 추정손실의 여신을 NPL로 간주하며, 일반적으로 시장에서 부실채권, 무수익여신으로 통칭되고 있다. NPL 종류는 대출채권의 담보 존재 여부와 회수 방식 (경매, 회생 등)에 따라 구분되며, 일반적으로 경매절차에 따라 진행되는 부동산담보 NPL (Regular NPL)이 시장의 대부분을 차지하고 있다. 국내 NPL 시장은 1998년 IMF 금융위기 직후 금융기관의 대규모 구조정이 필요했던 시기에 정부 (한국자산관리공사)와 론스타와 같은 외국인 투자자들이 시장을 주도하였다. 급격히 늘어난 부실채권의 효율적 정리를 위하여 ‘자산 유통화에 관한 법률 (SPC)’을 중심으로 ‘산업발전법 (CRC)’, ‘기업구조조정투자회사법 (CRV)’, ‘금융산업의 구조개선에 관한 법률’ 등이 제정되어 법률적 환경의 조성과 함께 시장의 토대가 형성되었다. 우리나라 NPL 시장은 초창기 정부와 외국 자본에 의하여 시장이 형성되기는 하였으나 약 10여 년간 금융시장의 안정화와 성숙기를 거쳤으며, 실제로 NPL 시장이 활성화되고 국내 자본이 본격적으로 진출한 것은 국제금융위기 발생 후인 2010년 이후로 그리 오래되지 않았다.

NPL 시장과 관련된 기존의 국내 연구를 살펴보면, 서울지역 경매 아파트물건을 중심으로 부실채권 회수 관행에 대한 개선방안을 제안하는 연구 (Woo and Kim, 2009) 와 투자자들이 NPL

에 투자할 때 고려하는 중요 요인에 대한 연구 (Choi and Lee, 2011) 가 있었다. 또한, 부실채권 정리 기구와 방안을 연구하여 부실채권 정리와 회생의 기회를 마련할 수 있는 정책적 대안을 제시하는 연구 (Kim, 2014) 와 NPL 투자의사결정에서 중요하게 고려할만한 변수들과 그 중요도 및 특성을 분석한 연구 (Sung and Park, 2014) 도 있었다. 이외에도 NPL의 투자만족도에 영향을 미치는 요인을 분석한 연구 (Kim and Lee, 2014) 와 NPL의 실태를 조사하여 문제점을 파악하여 개선점을 제안하는 연구 (Shin, 2015), NPL 시장이 공정하고 건전한 투자시장으로 발전하기 위한 문제점 및 개선점을 제안한 연구 (No, 2015), 다중회귀분석을 사용하여 NPL 비즈니스 의사결정 능력을 제시한 연구 (Jeong and Kim, 2016) 등이 진행되었다.

이에 그동안 NPL시장 최종구매자에 대한 분석과 NPL과 관계된 부동산경매시장의 여러 변수의 상관관계에 대한 연구 자료는 다소 있었지만, NPL 공급자인 은행 등 금융기관과 수요자인 자본시장의 NPL전문투자자 (유동화회사)의 투자에 대한 이론적 의사결정에 대한 연구는 전무한 상황이다.

NPL 시장과 관련된 해외 연구를 살펴보면, 미국 NPL 시장에 영향을 주는 요인들을 분석한 연구 (Saba et al., 2012)와 거시 경제 지표들과 은행 관련 요소들 중에서 터키 NPL 시장에 직접적인 영향을 주는 요인들을 분석한 연구 (Vatansever and Hepsen, 2013)가 있었으며, 유럽 NPL 시장에 있어서 직접적으로 관련있는 요인들을 분석한 연구 (Škarica, 2014) 등이 진행되었다. 이와 같이 해외에서도 NPL 투자의 수익률을 예측한 연구보다는 NPL 시장에 영향을 주는 요인을 분석한 연구 위주로 진행되어왔다.

3. 연구방법

NPL 시장을 분석하기 위하여 판별분석, 의사결정나무, 로지스틱 회귀분석, 유전자알고리즘, 인공지능망의 5가지 방법론을 활용하였다. 선행 연구들은 Table 1과 같다.

3.1 판별분석 (Discriminant analysis)

판별분석은 데이터들이 2개 이상의 집단 중에서 어떤 집단에 속할지를 분별해주는 방법론이다 (Kim, 2006). 이 방법론은 데이터들을 잘 분류해주는 판별변인을 찾고, 이를 이용하여 수식 $Z = W_1X_1 + W_2X_2 + \dots + W_nX_n$ 과 같은 선형 판별함수를 도출한다. Z 는 판별 값, W 는 판별 계수, X 는 분류변수를 각각 의미한다. 이 함수를 이용하여 데이터들을 특정 집단으로 분류하였을 때 분류 오차를 최소로 하는 집단을 선정해 준다.

본 연구에서는 입력한 독립변수를 모두 사용하는 옵션을 적용하였으며, 사전확률은 모든 집단이 동일하다는 가정하에 집단 내 공분산 행렬을 사용하는 옵션으로 예측모델을 도출하였다.

3.2 의사결정나무 (Decision Tree)

의사결정나무는 기존 사례들을 기반으로 일정한 의사결정 기준을 도출하는 기법으로, 두 개 이상의 변수들을 결합하여 종속 변수 값을 결정하는 데에 사용되는 방법론이다 (Kim and Oh, 2012; Kim and Cho, 2012). 의사결정나무는 결과에 대한 근거를 나뉘어 가지 형태로 추적하기 때문에 많은 변수들을 대상으로 종속변수에 영향이 높은 변수를 선택할 수 있다는 장점이 있다. 이

로 인해 다른 통계적 방법 및 인공지능 방법을 사용하기에 앞서 변수선택 단계에서 사용되기도 한다. 하지만 종속변수가 연속형일 때에는 쓸 수 없고, 독립변수가 연속형일 때 낮은 예측능력을 보인다는 취약점이 있다. 또한 자료의 추가에 의하여 나무구조가 바뀔 수 있다는 단점을 가지고 있다.

본 연구에서는 성장방법으로 CHAID 방법을 사용하였으며, 확장 단계는 자동으로, 최소 케이스수는 부모 노드의 경우 100, 자식 노드의 경우 50으로 설정하였다. 그리고 CHAID 방법에서 유의수준은 노드 분할과 범주 합치기 모두 0.05로 설정하였으며, Bonferroni 방법을 사용하여 유의성 값을 조정하는 옵션을 적용하였다.

3.3 로지스틱 회귀분석 (Logistic regression)

로지스틱 회귀분석은 각 데이터가 어느 집단에 속하는지를 판단하기 위해 사용하는 회귀분석방법론이다 (Shim et al., 2012). 일반적인 선형 회귀분석의 경우 독립변수 및 종속변수에 대한 제약조건은 없는 편이지만, 로지스틱 회귀분석은 종속변수가 0 또는 1과 같은 이항 형태에 특화된 방법론이다. 예측 정확도도 상당히 높은 편이므로 일반적으로 유/무 결정과 같은 결정 값이 이항 데이터일 때 예측모델로 주로 사용된다.

본 연구에서는 예측 모델 도출 시 분류 분리점은 0.5로, 최대반복계산은 20으로 설정하였고, 모형에 상수를 포함하는 옵션을 적용하였다.

3.4 유전자 알고리즘 (Genetic algorithm)

유전자 알고리즘은 유전자의 진화에서 아이디어를 얻어 만들어진 인공지능 방법론이다. 확률적 탐색과 학습 및 최적화를 이용한 기법으로서,

보통 이진수 형태의 유전자와 적합도로 이루어진 개체들의 집합을 이용하여 최적화를 수행하는 알고리즘이다 (Jo et al., 2018; Hong and Shin, 2003). 일반적으로 선택 (Selection)과 교차 (Crossover), 돌연변이 (Mutation)의 세 가지 단계의 유전자 조작 (Genetic operations)을 사용한다. 선택은 교차를 하기위한 유전자를 선택하는 단계이며, 교차는 선택된 유전자의 일부분을 생물학적 교차 방법과 동일하게 일부분을 교차시키는 단계이다. 마지막으로 돌연변이는 교차를 거친 유전자의 순서를 바꾸어 새로운 유전자를 만드는 단계이다. 이렇게 다양한 유전자를 만들어 가면서 기존에는 미처 생각하지 못하였던 범주의 데이터까지도 만들어 주기 때문에, 기존에 주로 사용되던 통계적 방법론과는 다양성 측면에서 차별점이 있다고 할 수 있다.

유전자 알고리즘을 이용하여 예측 값을 도출할 경우에는 $Y = A_0 + A_1X_1 + A_2X_2 + \dots + A_nX_n$ 의 수식과 같이 일반적인 회귀모형과 같은 선형 모델의 계수에 회귀계수 대신 유전자 알고리즘 값을 넣어 사용하는 유전자 알고리즘 선형 모델의 형태로 많이 사용되어진다. (Kim and Han, 2000; Chung and Oh, 2014; Jo et al., 2018) Y 는 종속변수이고, A 는 GA로 최적화 한 계수(비중) 값, X 는 입력변수를 각각 의미한다.

본 연구에서 유전자 알고리즘의 population size는 50, generation 은 20, crossover rate는 0.5, mutation rate는 0.1로 설정하였으며, progress는 0.01%의 maximum change가 20000의 trials 동안 나타나지 않으면 값을 도출하도록 설정하였다.

3.5 인공신경망 (Artificial neural network)

인공신경망은 큰 범주에서 본다면 회귀분석과

형태는 비슷하겠지만, 가중치 값을 구하는 방법이 숨겨져 있다는 특징을 가지는 인공지능 방법론이다 (Oh et al., 2011; Jo et al., 2015; Lee and Park, 2016). 어떠한 확률 분포나 변수들 간의 선형적 관계도 가정하고 있지 않기 때문에 전통적인 통계기법들보다 적용 범위가 넓다. 입력 패턴들을 그 특성에 따라 분류하는 Pattern classification이나, 불완전한 패턴으로부터 완전한 패턴을 생성하는 Pattern completion, 입력 패턴에 대응하는 출력 패턴을 생성하는 Function approximation 등의 문제들에 적합한 기법이다.

본 연구에서는 은닉층의 경우 자동 신경망 설계의 옵션을 적용하였으며, 은닉층의 최소 노드 수는 1, 최대 노드 수는 50으로 설정하였다. 인공 신경망 모델의 학습 유형은 배치 방법을 적용하였으며, 최적화 알고리즘은 척도화된 켈레기울기 알고리즘을 사용하였다. 학습 옵션으로는 초기 람다는 0.0000005, 초기 시그마는 0.00005, 구간 중심은 0, 구간 변위는 ± 0.5 로 설정하였다.

4. 실증 분석

실증 분석을 위해 2013년 12월부터 2017년 12월까지의 약 4년 동안의 국내 NPL 데이터를 활용했다. 데이터의 물건 수는 총 2291개이며, 물건별 최종 수익률이 특정 기준치를 넘었으면 1, 그렇지 않았다면 0으로 설정된 종속변수를 사용했다. 종속변수를 연속형 변수 자체로 사용하지 않고 이항 변수로 변형하여 사용한 이유는, 연속형 변수를 예측하는 모델보다 이항 변수를 예측하는 모델의 정확도가 높고, 이러한 예측 모델의 정확도는 모델의 실효성 측면과 직결되기 때문이다. 또한 실질적으로 유동화 전문 회사의 경우

수익을 얻기 위해 해당 물건을 매입할 것인지 하지 않을 것인지가 최대 관심사인데, 이는 최종적으로 특정 기준치의 수익률을 달성했는지의 여부만으로도 의사결정에 충분하기 때문이다.

독립변수로는 매입 시점을 구분하는 SPC (Special Purpose Company)와 매입년도, 시도와 시군구, 용도 대분류와 용도 소분류, 감정가, 매입원가, OPB (Outstanding Principle Balance), HP (Holding Period), TP (Target Price)를 사용했다. SPC는 대출채권이나 외상 매출채권 등의 자산을 통합하고, 이를 증권형태로 매각하는 유동화 전문회사를 가리킨다. 이 변수를 통해 한 유동화 전문회사가 어떤 물건들을 매입하였는지 확인할 수 있고, 나아가 어떤 시기에 어떤 물건들이 동시에 매입되었는지도 확인할 수 있다. 매입년도는 물건이 어느 해에 매입이 되었는지를 나타내는 변수이고, 시도와 시군구는 물건의 지역을 나타낸 변수이다. 용도 대분류와 용도 소분류는 물건의 용도를 구분하여 나타내는 변수이고, 감정가는 물건에 대한 감정 가격, 매입원가는 물건의 매입 시 가격을 나타내는 변수이다. OPB는 매각 대상자산에 있어서 현재까지 갚지 못하고 남아 있는 채권 원금(미상환 원금잔액)을 의미한다.

HP와 TP는 유동화 전문 회사가 자체적으로 물건의 여러 조건들을 고려하여 결정하는 값들인데, HP는 수익을 얻기 위해 물건을 어느 정도 보유할 예정인지를 나타내는 예상보유기간이고, TP는 예상보유기간을 고려하였을 때 얼마만큼의 수익을 목표를 할 예정인지를 나타내는 예상 목표 수익률 값이다. 이 두 변수의 경우 물건이 선정되는 동시에 결정되는 값들이므로 독립변수에 포함시켰다.

독립변수 중에서 SPC와 시도, 시군구, 용도 대분류, 용도 소분류의 경우 범주형 데이터이므로,

〈Table 1〉 Methodological theorem of prior research

Author	Method	Data	Contents
Chung and Oh (2014)	GA RS	KOSPI200 Futures	Proposal of Rough set optimal strategy for KOSPI200 Futures market using GA
Hong and Shin (2003)	ANN GA	Default Corporate	Using the genetic algorithm to select optimal variables and using the artificial neural network model to predict the default company.
Jo et al. (2018)	GA	KOSPI200	Scoring model to determine trade timing based on genetic algorithm
Kim (2006)	ANN LR MDA	Hotels bankruptcy	Factors affecting hotel bankruptcy
Kim and Cho (2012)	DT	Business management analysis index	An important indicator for predicting rights offering
Kim and Han (2000)	ANN GA	KOSPI200	Proposal of hydride GA and ANN solution for weight optimization
Kim and Oh (2012)	DT RS	KOSPI200 Futures	Proposal of futures market trading strategy using rough set and decision tree
Lee and Park (2016)	ANN SE	Exchange rate related data	An economic crisis prediction model using ANN and structural equation model
Oh et al. (2011)	ANN AR	KOSPI200	Proposal of financial market instability index using artificial neural network and AR model
Shim et al. (2012)	LR PCA	Foreign exchange	Proposal of foreign exchange market investment strategy using principal component analysis and logistic regression analysis

- * ANN : Artificial Neural Network
- * AR : Auto-Regressive Model
- * DT : Decision Tree
- * GA : Genetic Algorithm
- * LR : Logistic Regression
- * MDA : Multi-Discriminant Analysis
- * PCA : Principal Component Analysis
- * RS : Rough Set
- * SE : Structural Equation Model

추가적으로 데이터로 변환을 시켜주어 실험을 진행하였다. 데이터 변환은 동일한 범주 데이터를 동일한 숫자 데이터로 변환을 하였으며, 문자 데이터의 오름차순으로 낮은 숫자를 배정하였다. 이는 SPSS 통계 프로그램에서 분석 시 변수의 속성을 범주형 변수로 지정할 경우, 자체적으로 더미변수 처리를 해주기에 이와 같은 방법으

로만 데이터를 변환해주었다.

4.1 변수 선정

종속변수에 의미 있는 변수를 선별하기 위해 one to one t-test와 logistic regression stepwise, decision tree를 활용하였다. One to one t-test는

각각의 독립변수를 종속변수와 일대일 대응시켜 독립표본 t-test를 수행하는 방법으로써, 독립변수들이 개별적으로 종속변수에게 유의미한 영향을 주는지 판단할 때 사용될 수 있다. Logistic regression stepwise는 변수 단계선택법을 수행하여 최적의 로지스틱 회귀모형을 구성하는 변수들을 선정해주는 방법이고, decision tree는 데이터들이 종속변수를 판단할 때 의미있는 독립변수를 기준으로 트리가 구성된다는 개념을 기반으로 변수를 선정하는 방법이다. One to one

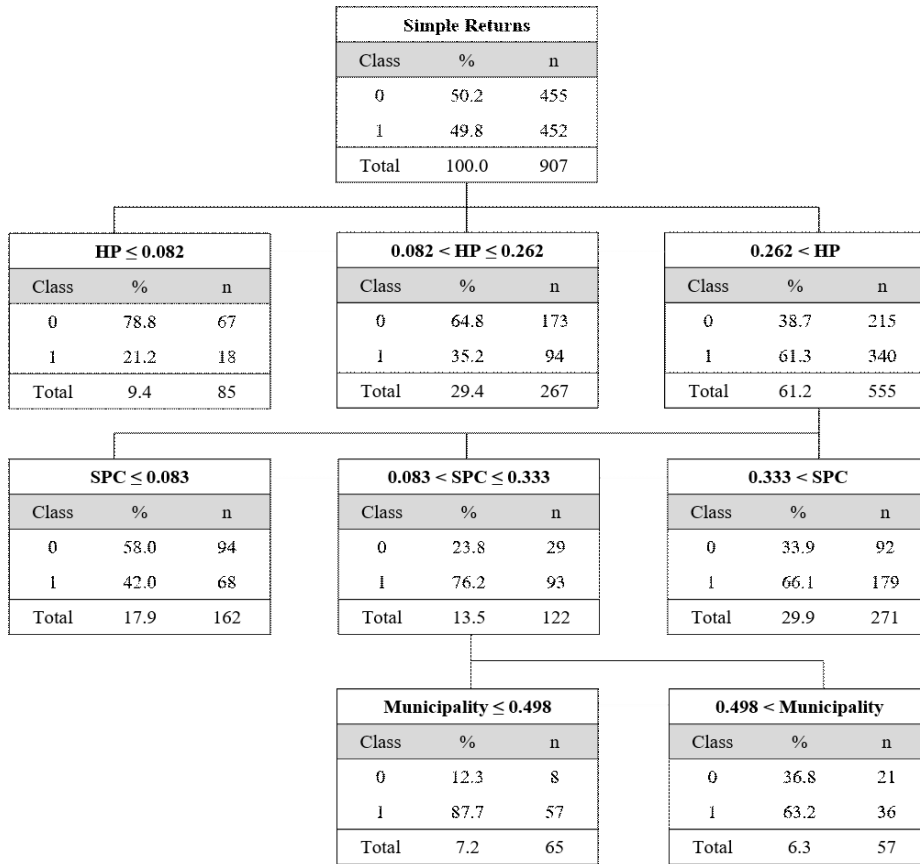
t-test를 수행하였을 때, 유의수준 5% 하에서 Table 2과 같이 SPC와 OPB, HP가 유의미한 변수라고 선정되었고, logistic regression stepwise를 통해서는 Table 3과 같이 SPC, 매입 년도, 시군구, 감정가, 매입원가, OPB, HP가 선정되었다. 마지막으로 decision tree를 통해서는 Figure 1과 같이 SPC와 HP가 의미 있는 변수로 선정되었다. 이러한 3가지 방법론을 통해 최종적으로 Table 4과 같은 7가지 변수를 선정하였다.

〈Table 2〉 Input variable selection for one to one t-test

Variable	t	Degree of freedom	p-value	Selection
Purchase Year	-1.172	905.000	0.242	X
SPC	-2.115	902.064	0.035	O
City	-0.472	905.000	0.637	X
Municipality	0.713	905.000	0.476	X
Major classification of use	0.885	885.924	0.376	X
Sub-classification of use	-0.333	885.112	0.739	X
Appraised value	-1.581	468.200	0.115	X
Purchase cost	-1.532	586.536	0.126	X
OPB	-2.090	642.512	0.037	O
Holding Period	-6.601	872.332	0.000	O
Target Price	-1.694	593.147	0.091	X

〈Table 3〉 Input variable selection for logistic regression stepwise

Variable	B	S.E.	Wald	p-value
Purchase Year	-2.039	0.821	6.170	0.013
SPC	2.644	0.799	10.943	0.001
City	-0.413	0.240	2.951	0.086
Appraised value	47.936	23.807	4.054	0.044
Purchase cost	-35.930	12.218	8.648	0.003
OPB	23.082	9.246	6.233	0.013
Holding Period	3.104	0.511	36.971	0.000
Constant	-0.875	0.237	13.592	0.000



〈Figure 1〉 Input variable selection for decision tree

〈Table 4〉 Input variable selection

Variable	One to one t-test	LR stepwise	Decision tree	Selection
Purchase Year	X	O	X	O
SPC	O	O	O	O
City	X	X	X	X
Municipality	X	O	O	O
Major classification of use	X	X	X	X
Sub-classification of use	X	X	X	X
Appraised value	X	O	X	O
Purchase cost	X	O	X	O
OPB	O	O	X	O
Holding Period	O	O	O	O
Target Price	X	X	X	X

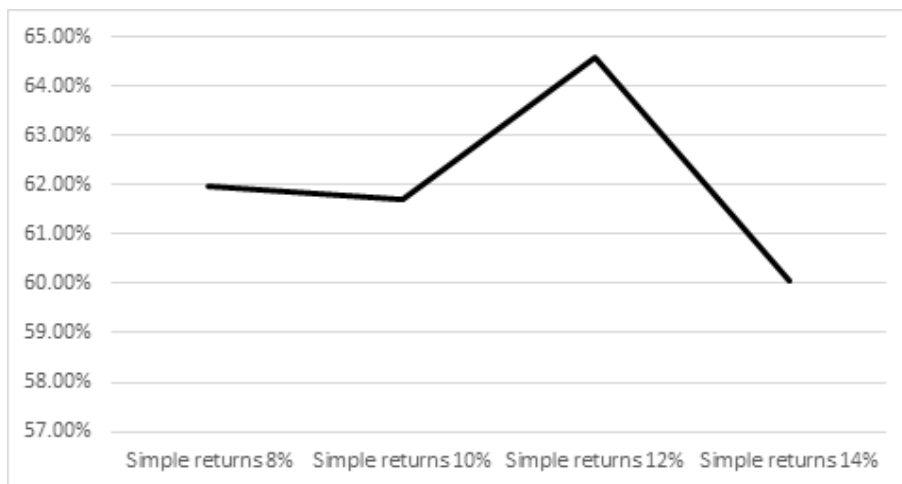
4.2 최적의 기준 수익률 설정

유의미한 예측모델을 만들기 위한 최적의 기준 수익률을 선정하기 위해 단순수익률을 2%씩 바꾸어가며 Hit ratio average가 최고점인 값을 찾았다. 이는 현업에서 사용되어지는 기준수익률이 적절한 지에 대한 분석이 필요하였기에 진행하였다. 실험은 판별분석 (DA), 로지스틱 회귀분석 (LR), 의사결정나무 (DT), 인공신경망 (ANN), 유전자알고리즘 선형모델 (GA linear model)의 5가지 방법론을 활용하여 진행하였고, 이에 대한 hit ratio를 확인하였다. 각 방법론 별 설정 값은 3절에 제시한 옵션과 동일하게 적용하였으며, 유

전자 알고리즘 선형모델의 경우는 최종 예측 값이 0.5 이상일 경우 1로, 0.5 미만일 경우 0으로 처리해주었다. 실험결과 Table 5와 같이 단순수익률 12%를 기준으로 한 종속변수가 가장 높은 Hit ratio average를 보여주었고, 단순수익률 별 Hit ratio average 추이는 Figure 2와 같다. 단순수익률 12%는 현재 현업에서 추구하는 기준 수익률과 동일한 값이다.

4.3 학습 (train) 및 검정 (test) 데이터의 구성

먼저 종속변수의 0 또는 1의 값 개수를 맞춰주었다. 이 과정에서 2291개의 데이터 중에서 종속



〈Figure 2〉 Simple returns value comparison

〈Table 5〉 Selection of optimal simple returns

Based on yield	DA	LR	DT	ANN	GA linear model	Hit ratio average
Simple returns 8%	58.83%	59.69%	63.64%	66.90%	60.89%	61.99%
Simple returns 10%	60.97%	61.60%	60.13%	64.35%	61.39%	61.69%
Simple returns 12%	63.57%	65.40%	65.10%	69.25%	59.69%	64.60%
Simple returns 14%	61.06%	61.68%	61.37%	57.32%	58.88%	60.06%

변수가 0인 데이터 647개와 1인 데이터 647개를 임의로 추출하여 사용하였다. 다음으로 반복적인 실험을 위해 10 fold validation method를 활용하여 데이터를 10가지 set으로 만들었다.

〈Table 6〉 Training data set and testing data set

DA, DT, LR, GA-linear model		ANN	
Training	70%	Training	40%
		Validation	30%
Testing	30%	Testing	30%

그리고 Table 6와 같이 각 set 별로 training set과 testing set의 비율은 70%와 30%로 고정하였으며, 인공신경망의 경우는 추가적으로 hyper parameter를 설정해주는 validation set이 필요하여 training set과 validation set, testing set을 각각 40%, 30%, 30%의 비율로 고정하여 랜덤으로 나누어주었다.

4.4 최적 단순 수익률에 대한 예측모델 실험

4.1절에서 선정된 매입년도, SPC, 시군구, 감정가, 매입원가, OPB, HP의 7가지 독립변수들과 4.2절에서 선정된 단순수익률 12%를 기준으로 한 종속변수를 사용하여 실험을 진행했다.

단순수익률 12% 달성 유무를 기준으로 다섯 가지 방법론(판별분석 (DA), 로지스틱 회귀분석 (LR), 의사결정나무 (DT), 인공신경망 (ANN), 유전자알고리즘 선형모델 (GA linear model))을 이용하여 모델을 만들어 보았을 때 Table 7와 같은 결과가 나타났다.

Set1의 경우 GA linear model의 정확도는 약 60%, Discriminant Analysis 모델의 정확도는 약 63%, Logistic regression과 Decision tree의 경우 약 65%의 정확도를 보였습니다. 반면에 Artificial neural network 모델의 경우 69%로 다른 모델보다는 상당히 우수한 정확도를 얻을 수 있었다.

그리고 10개의 데이터 셋으로 모델을 형성

〈Table 7〉 Empirical study result

Set	DA	LR	DT	ANN	GA linear model
Set 1	63.57%	65.37%	65.12%	69.25%	59.69%
Set 2	61.64%	62.15%	62.15%	66.50%	60.87%
Set 3	66.67%	67.44%	63.82%	66.15%	66.93%
Set 4	62.95%	64.25%	64.77%	65.13%	60.26%
Set 5	66.23%	66.75%	65.97%	71.17%	58.70%
Set 6	68.23%	68.49%	62.50%	68.49%	61.20%
Set 7	62.08%	62.08%	63.64%	64.16%	60.26%
Set 8	64.60%	64.86%	62.02%	65.63%	58.91%
Set 9	64.27%	64.78%	58.61%	70.44%	58.87%
Set 10	63.75%	65.04%	66.84%	67.10%	59.38%
Average	64.40%	65.12%	63.54%	67.40%	60.51%

〈Table 8〉 Empirical study McNemar test result

	ANN-DA	ANN-LR	ANN-DT
N	1294	1294	1294
Chi-square	12.540	9.552	5.568
p-value	0.000	0.002	0.018

한 후 결과의 평균을 내보니 위의 표와 같이 Artificial neural network 모델이 67.40%의 Hit ratio를 보이며 우수한 성능을 보였다.

ANN 모델의 성능이 다른 모델과의 비교에서도 유의미한 차이를 나타내는지 확인하기 위해 Mc Nemar test를 수행하였다. 총 10개의 데이터셋을 통해 도출된 결과 값을 종합한 후 McNemar test를 수행하였으며, ANN모델을 제외한 Hit ratio 상위 3개의 모델(DA, LR, DT)에 대해 분석을 진행하였다. 분석 결과 Table 8과 같이 ANN 모델은 다른 3개의 모델과 유의수준 5%하에서 유의미한 차이가 있음을 확인할 수 있었다.

5. 결론

국내 NPL 시장에 있어서 투자 의사 결정을 제 안해주는 연구는 많이 부족한 상황이다. 본 연구를 통해서 NPL 시장의 물건 데이터들은 단순수익률 달성유무를 판단할 때 매입년도와 SPC (Special Purpose Company), 시군구 (Municipality), 감정가 (Appraised value), 매입원가, OPB (Outstanding Principle Balance), HP (Holding Period)의 변수들이 의미가 있음을 제안하며, 단순수익률 12%를 기준으로 하였을 때 비교적 높은 예측력을 가지는 모델을 만들 수 있음을 확인하였다. 독립 변수들 중에서 매입년도 변수가 선

택된 것을 보아 단순 수익률 12% 달성 유무(종속변수)를 결정하는 데에 있어 시기가 중요함을 확인할 수 있었다. 그리고 물건 별 SPC가 동일하다는 것은 해당 물건들이 동시에 매입이 되었음을 의미하므로 이 변수 또한 종속변수 결정에 있어서 시기가 중요함을 뒷받침해준다. 다음으로 시도(City) 변수는 선택이 되지 못한 것에 반해 시군구(Municipality) 변수는 선택된 것을 보아, 시도(City)의 경우 한 범주 내에 다양한 물건들이 속해 있기에 종속변수를 결정하는 데에 큰 영향을 주지 못하였지만 시군구(Municipality)의 경우 보다 세분화 되어있기에 종속변수 결정에 의미 있는 영향을 준다는 것을 확인하였다. 그리고 은행에서 결정해주는 감정가나 OPB의 경우도 최종적인 단순수익률 달성 유무를 결정하는 데에 있어서 유의미한 역할을 한다는 것을 볼 수 있다. 마지막으로 유동화 회사에서 자체적인 산출 방식으로 도출해왔던 매입원가와 HP, TP 중에서 매입원가와 HP만 의미있는 변수로 선택되었는데, 이를 통해 특정 수익률 달성에는 물건별 매입 설정 가격과 보유기간은 중요했지만, 할인율을 적용하여 산출된 목표가격의 크기는 실제 단순 수익률 달성 유무에 영향을 미치지 않음을 확인할 수 있었다.

최종적으로 이러한 7가지 독립변수들과 종속변수를 활용하였을 때 인공지능망을 이용한 예측모델이 67.40%의 비교적 높은 예측 정확도를

보였다. 즉, 이항 데이터를 예측함에 있어서 인공신경망 모델이 판별분석이나 의사결정나무, 로지스틱회귀분석보다는 높은 성능을 보였고 성능간의 차이도 유의미함을 확인하였다. 반면에 GA linear model의 경우 종속변수를 이항 변수로 변환해주는 과정이 추가적으로 필요하였기 때문에 비교적 정확도가 떨어졌다.

본 연구는 NPL 시장에 투자를 하는 유동화회사가 어떠한 물건에 투자를 하는 데에 있어서 과연 매입원가나 HP 등이 적절한지를 판단해줄 수 있다. 예를 들어 추후 특정 지역에 새로운 NPL 물건이 등록 되었다면, 해당 물건의 매입년도, SPC, 시군구, 감정가, OPB 등과 유동화 회사가 기존에 의사결정을 위해 설정한 매입원가 및 HP를 기반으로 인공신경망 예측모델에 적용시켜볼 수 있다. 이때에 최종 12% 단순수익률 달성이 가능할 것으로 예측이 된다면 해당 물건을 제시한 매입원가와 HP 기준으로 매입을 진행하면 되지만, 만약 단순수익률이 달성하지 못한다고 예측이 된다면 매입원가나 HP 등을 조절하여 적정 매입원가 혹은 HP를 찾아볼 수 있도록 한다.

본 연구를 통해 투자자는 합리적인 의사결정을 할 수 있기에 적절한 보유기간이나 적정 투자 금액을 결정하는 데에도 상당한 기여를 할 것으로 예상된다. 투자자에게 있어서 효율적인 의사결정을 제안함은 NPL 시장의 거래가 적정한 가격 선에서 진행될 수 있음을 의미하기에 시장 유동성 또한 더욱 높아질 것으로 기대한다.

하지만, 본 연구의 제안모델은 결정된 매입원가나 HP가 단순수익률 12%를 달성하기에 적정한지를 판단해줄 뿐이며, 어떠한 변수가 어떻게 조절되면 단순수익률 12%를 달성할 수 있는지를 구체적으로 알려주지 못한다는 한계점이 존재한다. 즉, 특정 물건이 단순수익률 12%를 달성

하지 못한다고 나왔을 때, 매입원가를 줄여야할지 혹은 HP를 늘려야할 지에 대한 의사결정을 제시해주지 못한다.

따라서 본 연구를 기반으로 하여 단순수익률 기준의 달성유무로 예측만을 해주는 모델에서 더 나아가 물건별 적정 매입원가나 Holding period, Target price를 예측해주는 모델에 대한 연구로 발전시킬 수 있다. 이러한 발전된 연구는 유동화회사뿐만 아니라 NPL 시장 자체가 더욱 활성화되는 데에 기여할 것이라 기대한다.

참고문헌(References)

- Choi, S. H. and S. Y. Lee, "Analysis of the Importance Level of Decision Making Criteria for the Investment of Asset-Backed Non-Performing-Loans," *Korean Association For Housing Policy Studies*, Vol.19, No.2 (2011), 47~71.
- Chung, S. H. and K. J. Oh, "Using genetic algorithm to optimize rough set strategy in KOSPI200 futures market," *Journal of the Korean Society of Cadastre*, Vol.25, No.2 (2014), 281~292.
- Hong, S. H. and K. J. Shin, "Using GA based Input Selection Method for Artificial Neural Network Modeling: Application to Bankruptcy Prediction," *Korea Intelligent Information Systems Society*, Vol.9, No.1(2003), 227~249.
- Jeong, J. H. and G. C. Kim, "A Study on Profitability Determinants of Non-Performing Real Estate Loan - Focused on NPL Investment Company's Loan-Level Data in Seoul, Korea -, " *Korea Real Estate Society*, Vol.33, No.1(2016), 5~24.

- Jo, K. H., S. H. Jeong, K. S. Kim and K. J. Oh, "Scoring model to determine trade timing based on genetic algorithm," *Journal of the Korean Data & Information Science Society*, Vol.29, No.3(2018), 1~11.
- Jo, N. O., Kim, H. J. and K. S. Shin, "Bankruptcy type prediction using a hybrid artificial neural networks model," *Korea Intelligent Information Systems Society*, Vol.21, No.3(2015), 79~99.
- Kim, H. H. and K. J. Oh, "Using rough set to develop the optimization strategy of evolving time-division trading tin the futures market," *Journal of the Korean Data & Information Science Society*, Vol.23, No.5(2012), 881~893.
- Kim, J., "NPL study on the current situation and investment management survey - Overseas case centred on a comparison -," *Korea Real Estate Society*, Vol.32, No.1(2014), 209~220.
- Kim, K. J. and I. G. Han, "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index," *Expert systems with Applications*, Vol.19, No.2(2000), 125~132.
- Kim, M. K. and Y. H. Cho, "Development of predictive model for rights issues using financial analysis indices and decision tree technique," *Korea Intelligent Information Systems Society*, Vol.18, No.4(2012), 59~77.
- Kim, S. G. and C. S. Lee, "A Empirical Study on the Investment Satisfaction of Asset-backed Non Performing Loans(NPL)," *Korea Real Estate Academy*, Vol.59, (2014), 87~98.
- Kim, S. Y., "Prediction of hotel bankruptcy using multivariate discriminant analysis, logistic regression and artificial neural network," *Journal of Tourism Sciences*, Vol.30, (2006), 53~75.
- Lee, H. Y. and J. M. Park, "The prediction of currency crises through artificial neural networks," *Korea Intelligent Information Systems Society*, Vol.22, No.4(2016), 19~43.
- No, H. J., "Study on Fairness Consolidation of Real Estate Auctions Secured for Bank NPLs," *The Korea Contents Association*, Vol.15, No.11(2015), 397~409.
- Oh, K. J., T. Y. Kim, K. W. Jung and C. H. Kim, "Stock market stability index via linear and neural network autoregressive model," *Journal of the Korean Data & Information Science Society*, Vol.22, No.2(2011), 335~351.
- Saba, I., R. Kouser and M. Azeem, "Determinants of Non Performing Loans: Case of US Banking Sector," *The Romanian Economic Journal*, Vol.44, No.6(2012), 125~136.
- Shim, K. S., J. J. Ahn and K. J. Oh, "Multi-currencies portfolio strategy using principal component analysis and Logistic regression," *Journal of the Korean Data And Information Science Sociaty*, Vol.23, No.1 (2012), 151~159.
- Shin, S. D., "The Present Situation and Problems of non-performing loans," *Legal Theory & Practice Review*, Vol.3, No.2(2015), 281~299.
- Škarica, B. "Determinants of non-performing loans in Central and Eastern European countries," *Financial theory and practice*, Vol.38, No.1 (2014), 37~59.
- Sung, S. K. and C. G. Park, "A Study on the Investment Decision making for the Real Estate Mortgage NPL," *Journal of The Residential Environment Institute of Korea*, Vol.12, No.4(2014), 151~166.
- Vatansever, M, and A. Hepsen, "Determining impacts on non-performing loan ratio in

Turkey,” *Journal of Finance and Investment Analysis*, Vol.2, No.4(2013), 119~129.

Woo, H. D. and H. J. Kim, “A Study on the Realities of Payback of Bad Credits in Real

Estate – Focused on Apartment Houses for Sale at Auction in Seoul,” *Journal of the Korean Society of Cadastre*, Vol.25, No.2 (2009), 1~15.

Abstract

A study on the prediction of korean NPL market return

Hyeon Su Lee* · Seung Hwan Jeong** · Kyong Joo Oh***

The Korean NPL market was formed by the government and foreign capital shortly after the 1997 IMF crisis. However, this market is short-lived, as the bad debt has started to increase after the global financial crisis in 2009 due to the real economic recession. NPL has become a major investment in the market in recent years when the domestic capital market's investment capital began to enter the NPL market in earnest. Although the domestic NPL market has received considerable attention due to the overheating of the NPL market in recent years, research on the NPL market has been abrupt since the history of capital market investment in the domestic NPL market is short. In addition, decision-making through more scientific and systematic analysis is required due to the decline in profitability and the price fluctuation due to the fluctuation of the real estate business.

In this study, we propose a prediction model that can determine the achievement of the benchmark yield by using the NPL market related data in accordance with the market demand. In order to build the model, we used Korean NPL data from December 2013 to December 2017 for about 4 years. The total number of things data was 2291. As independent variables, only the variables related to the dependent variable were selected for the 11 variables that indicate the characteristics of the real estate. In order to select the variables, one to one t-test and logistic regression stepwise and decision tree were performed. Seven independent variables (purchase year, SPC (Special Purpose Company), municipality, appraisal value, purchase cost, OPB (Outstanding Principle Balance), HP (Holding Period)).

The dependent variable is a bivariate variable that indicates whether the benchmark rate is reached. This is because the accuracy of the model predicting the binomial variables is higher than the model predicting the continuous variables, and the accuracy of these models is directly related to the effectiveness of the model. In addition, in the case of a special purpose company, whether or not to purchase the property is the main concern. Therefore, whether or not to achieve a certain level of return is enough to make a

* Department of Investment Information Engineering, Yonsei University

** Department of Industrial Engineering, Yonsei University

*** Corresponding Author: Kyong Joo Oh

Department of Industrial Engineering, Yonsei University

50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea

Tel: +82-2-2123-5720, Fax: +82-2-364-7807, E-mail: johanoh@yonsei.ac.kr

decision. For the dependent variable, we constructed and compared the predictive model by calculating the dependent variable by adjusting the numerical value to ascertain whether 12%, which is the standard rate of return used in the industry, is a meaningful reference value. As a result, it was found that the hit ratio average of the predictive model constructed using the dependent variable calculated by the 12% standard rate of return was the best at 64.60%.

In order to propose an optimal prediction model based on the determined dependent variables and 7 independent variables, we construct a prediction model by applying the five methodologies of discriminant analysis, logistic regression analysis, decision tree, artificial neural network, and genetic algorithm linear model we tried to compare them. To do this, 10 sets of training data and testing data were extracted using 10 fold validation method. After building the model using this data, the hit ratio of each set was averaged and the performance was compared. As a result, the hit ratio average of prediction models constructed by using discriminant analysis, logistic regression model, decision tree, artificial neural network, and genetic algorithm linear model were 64.40%, 65.12%, 63.54%, 67.40%, and 60.51%, respectively. It was confirmed that the model using the artificial neural network is the best.

Through this study, it is proved that it is effective to utilize 7 independent variables and artificial neural network prediction model in the future NPL market. The proposed model predicts that the 12% return of new things will be achieved beforehand, which will help the special purpose companies make investment decisions. Furthermore, we anticipate that the NPL market will be liquidated as the transaction proceeds at an appropriate price.

Key Words : Artificial Neural Network, Decision Tree, Genetic Algorithm, Logistic Regression, NPL

Received : March 9, 2019 Revised : March 26, 2019 Accepted : April 29, 2019

Publication Type : Regular Paper Corresponding Author : Kyong Joo Oh

저 자 소 개



이 현 수

코레이트 자산운용 대표이사

연세대학교 투자정보공학 협동과정 통합과정 재학 중

관심 연구 분야는 NPL 시장분석, 인공지능기법을 활용한 예측모델 개발 등



정 승 환

연세대학교 산업공학과 통합과정 재학 중

관심 연구 분야는 인공지능기법, 통계적 분석방법, 머신러닝/딥러닝 기법을 활용한 금융 데이터 분석 등



오 경 주

연세대학교 산업공학과 정교수

연세대학교 투자정보공학 협동과정 주임교수

금융시장 및 데이터 분석 관련하여 약 40여 편의 해외저널 논문 게재

주요 연구 분야는 시스템트레이딩, 포트폴리오 최적화, 조기경보 시스템, 통계적 분석 모델링, 고객 관계 관리, 로보어드바이저, 핀테크 전략과 스마트 금융기술 등