

# 인공지능기법을 이용한 온라인 P2P 대출거래의 채무불이행 예측에 관한 실증연구

## Artificial Intelligence Techniques for Predicting Online Peer-to-Peer(P2P) Loan Default

배재권(Jae Kwon Bae)\*, 이승연(Seung Yeon Lee)\*\*, 서희진(Hee Jin Seo)\*\*\*

### 초 록

온라인 P2P 대출(Online Peer-to-Peer Lending)이란 대출자(차입자)들이 인터넷 및 모바일 P2P 플랫폼을 통해 대출을 신청하면 P2P 플랫폼 기업이 이를 심사하고, 공개하여 불특정 다수가 자금을 빌려주고 이자를 받는 대출중개 서비스를 말한다. 국내외적으로 P2P 대출시장의 성장과 수익률에 대한 관심이 커진 상황에서 현재는 P2P 대출에 대한 안정성 측면에서 문제가 제기되고 있다. P2P 대출시장은 높은 수익률을 제공하지만 P2P 업체의 연체율과 부실률(채무불이행률)도 함께 높아지고 있는 실정이다. P2P 금융시장의 신뢰도를 높이기 위해서는 P2P 대출의 연체율과 채무불이행률을 줄이는 것이 무엇보다 중요하다. 본 연구는 세계적인 P2P 기업인 렌딩클럽(Lending Club)의 P2P 대출거래데이터베이스를 이용하여 인공지능기반의 P2P 채무불이행 예측모형을 구축하고자 한다. 구체적으로 벤치마크(benchmark) 모형으로 통계기법인 판별분석과 로지스틱 회귀분석을 이용하고, 인공지능기법으로는 신경망, CART, 그리고 C5.0을 이용하여 P2P 대출거래의 채무불이행 예측모형을 구축하고자 한다. 연구결과, P2P 대출거래의 채무불이행 예측을 위해 우선 고려해야 할 변수는 대출이자율이며, 중요도 3순위에 가장 많이 언급된 대출금액과 총부채상환비율도 고려해야 할 요인으로 추출되었다. 전통적인 통계기법보다는 인공지능기법의 예측성도가 더 좋은 것으로 나타났으며, 신경망의 경우 모든 데이터 셋에서 오분류율이 가장 낮은 예측모형으로 나타났다.

### ABSTRACT

In this article, an empirical study was conducted by using public dataset from Lending Club Corporation, the largest online peer-to-peer (P2P) lending in the world. We explore significant predictor variables related to P2P lending default that housing situation, length of employment, average current balance, debt-to-income ratio, loan amount, loan purpose, interest rate, public records, number of finance trades, total credit/credit limit, number of delinquent accounts, number of mortgage accounts, and number of bank card accounts are significant factors to loan funded successful on Lending Club platform.

We developed online P2P lending default prediction models using discriminant analysis,

\* First Author, Corresponding Author, Dept. of Management Information Systems, Keimyung University (jkbae99@kmu.ac.kr)

\*\* Dept. of Statistics, Keimyung University(stat\_lsy@naver.com)

\*\*\* Dept. of Management Information Systems, Keimyung University(lion186@naver.com)

Received: 2018-07-25, Review completed: 2018-08-14, Accepted: 2018-08-17

logistic regression, neural networks, and decision trees (i.e., CART and C5.0) in order to predict P2P loan default. To verify the feasibility and effectiveness of P2P lending default prediction models, borrower loan data and credit data used in this study. Empirical results indicated that neural networks outperforms other classifiers such as discriminant analysis, logistic regression, CART, and C5.0. Neural networks always outperforms other classifiers in P2P loan default prediction.

**키워드** : 온라인 P2P 대출, P2P 채무불이행예측, 신경망, 의사결정나무, 렌딩클럽  
Online Peer-to-Peer(P2P) Lending, P2P Loan Default Prediction, Neural Networks, Decision Trees, Lending Club Platform

## 1. 서 론

현재 금융산업은 금융과 첨단 정보기술을 융합한 핀테크(FinTech) 산업으로 진화하고 있으며 그 중에서 P2P 금융(Peer-to-Peer Financing)과 P2P 대출(Peer-to-Peer Lending) 영역은 가장 급성장하고 있는 핀테크 분야이다. P2P 대출은 대출자(차입자)들이 인터넷 및 모바일 P2P 플랫폼을 통해 대출을 신청하면 P2P 플랫폼 기업이 심사하고, 이를 공개하여 불특정 다수가 여유자금을 빌려주고 이자를 받는 대출중개 서비스를 말한다. 온라인 마켓플레이스(online marketplace)를 통해 모든 대출과정을 자동화하여 지점운영비용, 인건비, 대출영업비용 등의 불필요한 경비 지출을 최소화하여 대출자에게는 보다 낮은 금리를, 투자자에게는 보다 높은 수익을 제공하고 있다. P2P 대출은 전통적인 금융기관을 통해서만 가능했던 기존의 금융거래를 인터넷을 통해 이루어지게 함은 물론 대출자와 투자자 모두에게 합리적인 이율을 제공한다는 점이 특징이다. 또한 제2금융권, 캐피탈, 저축은행 등에서 대출거래를 수행한 경우 신용등급에 부정적인 영향을 미칠 수 있으나, P2P 대출이력은 신용등급과 무관하며 성공적으로 대출상환을 수행한 경우 오히려 신용등급

을 높일 수 있다. 현재 P2P 대출서비스는 개인 신용대출, 중소기업대출, 주택건설자금대출, 그리고 학자금대출 등으로 구성되어 있다.

영국의 조파(Zopa Ltd.)라는 기업이 2005년에 세계 최초로 P2P 대출서비스를 시작하였고, 현재 영국 내 P2P 금융산업 중 신용대출분야에서 약 20%의 점유율을 유지하고 있다. 현재 조파의 2017년 말 누적대출금액은 약 2조 3천억원에 달하고 있다. 영국 이외에 미국과 중국, 일본 등의 선진 금융시장에서도 P2P 대출영역은 빠르게 성장하고 있다. 2017년 기준 미국의 전체 개인신용대출 900조 원 중 P2P 대출이 45%(약 40조 원)를 점유하고 있으며, 이로 인해 가계 부채의 질적인 부분도 상당 부분 개선되고 있다. 미국에서 가장 먼저 P2P 대출을 운영한 프로스퍼(Proper Marketplace Inc.)의 2017년 기준 누적 대출금액이 약 2조 7천억 원으로 매년 30%씩 늘어나고 있으며, 세계적인 P2P 대출업체인 렌딩클럽(Lending Club)의 총 누적대출금액도 약 8조 원으로 늘어나면서 미국시장은 세계에서 가장 큰 P2P 시장규모를 형성하고 있다. 2006년부터 P2P 대출서비스를 시작한 중국은 P2P 대출업체가 2010년 10개에서 2017년에는 약 4,000개로 급증하였으며, 2017년 P2P 누적 대출금액이 182조 원을 돌파하였다[6].

이처럼 금융선진국은 P2P 금융을 은행 예적금과 증권시장 수익률을 보완할 대체 투자 상품으로 인식하고 있다. 국외 P2P 기업들은 시중은행과 저축은행, 대부업 사이의 중금리 대출금리(5~20%)를 제공하고 있으며, 투자자 입장에서 시중은행보다 높은 수익을 기대할 수 있다.

국내 P2P 대출시장은 국내 최초의 P2P 대출업체인 머니옥션(Moneyauktion)이 2006년 8월에 설립된 이후로 현재는 렌딧, 테라펀딩, 루프펀딩, 빌리, 에잇퍼센트 등 200여 개 이상의 P2P 대출업체들이 운영 중에 있다. 한국 P2P 금융협회(Korea P2P Finance Association)에 따르면, 2016년 6월 누적대출액은 1,525억 원이었으나 2018년 2월에는 23배 증가한 2조 원을 넘어서면서 P2P 대출분야가 가장 급성장하고 있는 핀테크 분야로 진화하고 있다. 국내 P2P 대출의 대표적인 형태인 신용대출과 담보대출의 수익률 또한 증가함에 따라 P2P 대출의 성장세에 지속적인 관심을 나타내고 있다.

이처럼 국내외적으로 P2P 대출시장의 성장과 수익률에 대한 관심이 커진 상황에서 현재는 P2P 대출에 대한 안정성 측면에서 문제가 제기되고 있다. 국내 P2P 대출시장은 높은 수익률을 제공하지만 P2P 업체의 연체율과 부실률(채무불이행률)도 함께 높아지고 있는 실정이다. 금융감독원에 따르면 국내 P2P 대출의 연체율(1~3개월 이내)이 2016년 1.24%에서 2017년 7.12%로 급증하고 있고, 부동산 프로젝트파이낸싱(PF) 비중이 높은 P2P 대출업체의 경우 연체율이 20%를 넘어서면서 투자자들의 불만이 커지고 있다. 이로 인해 투자자 1인이 1개의 P2P 기업 채권에 2,000만 원 이상 투자할 수 없는 규제가 생겨났으며, P2P 기업이 부도

(bankruptcy)가 발생하더라도 대출 채권이 부실화되지 않도록 막는 입법 과정과 투자자 보호대책 등의 가이드라인도 준비하고 있다. P2P 대출에 대한 내재된 위험은 대출자에 대해서는 제한적인 반면에 투자자에게는 상대적으로 크게 나타난다. 투자자는 기존 은행보다 높은 수익을 가져다주는 P2P 금융서비스의 이점만을 고려하지 말고, P2P 대출상품마다 부실률이 천차만별이므로 본인이 감수해야 할 위험수준이 다르다는 것을 인식해야 한다.

따라서 본 연구의 목적은 다음과 같다. P2P 금융시장의 신뢰도를 높이고, 시장 활성화를 위해서는 P2P 대출의 연체율과 채무불이행률을 줄이는 것이 무엇보다 중요하다. 또한 기존 금융기관의 대출상환예측모형은 P2P 환경에 적합하지 않으므로 P2P 대출거래에 적합한 채무불이행 예측모형 구축이 필요하다. 본 연구에서는 첫째, P2P 대출거래의 채무불이행에 미치는 요인을 파악하기 위해 기존 금융기관에서 이용하고 있는 대출상환예측모형과 신용평가모형을 분석하고, 글로벌 P2P 업체의 대출거래를 분석하여 이들 차이를 비교하고자 한다. 둘째, 세계적인 P2P 업체인 렌딩클럽(Lending Club)의 P2P 대출거래데이터베이스를 이용하여 인공지능기반의 P2P 채무불이행 예측모형을 구축하고자 한다. 구체적으로 예측모형 성과비교를 위해 벤치마크(benchmark) 모형으로 전통적인 통계기법인 판별분석과 로지스틱 회귀분석(로지스틱 분석)을 이용하고, 인공지능기법으로는 신경망, CART, 그리고 C5.0을 이용하여 P2P 대출거래의 채무불이행 예측모형을 구축하고자 한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 국내외 P2P 대출기업 현황과 채무불이행률을

분석하고, P2P 대출의 상환실패요인 및 채무불이행률에 관한 선행연구를 정리한다. 제 3장에서는 자료 수집 및 변수 선정 과정을 설명하고, 분석 절차 및 분석 방법을 제시한다. 제 4장에서는 P2P 대출거래의 채무불이행 예측모형 설계과정을 설명하고, 채무불이행에 미치는 영향요인 중요도 순위와 채무불이행 예측모형의 오분류율 결과를 비교 분석한다. 마지막 제 5장에서는 결론 및 향후 연구방향에 대해 기술하고자 한다.

## 2. 이론적 배경

### 2.1 국내외 P2P 대출기업 현황과 채무불이행률

본 절에서는 <Table 1>과 같이 국내외 P2P 대출기업 현황과 연체율, 그리고 부실율(채무불이행률)에 대해 분석하고자 한다. 국외 주요 P2P 대출기업(누적대출액 기준)에는 렌딩클럽(Lending Club), 프로스퍼(PROSPER), 소피(SoFi)가 있고, 국내 P2P 상위기업은 (주)렌딧, (주)에잇퍼센트, (주)팝펀딩, (주)어니스트펀드 등이 있으며 이들 기업의 연체율과 채무불이행률을 분석하고자 한다.

P2P 대출기업의 재무건전성 및 수익성, 그리고 비즈니스 모델 평가를 위한 가장 중요한 척도가 바로 연체(delinquency)와 부실(default)이다. 연체란 상환일로부터 30일 이상과 90일 미만 동안 상환이 지연되는 현상을 말하며 연체율은 현재 미상환된 대출잔액 중 연체 중이건의 잔여원금의 비중을 말한다. 부실은 정상 상환일로부터 90일 이상 장기 연체되는 것을 말하며, 채무불이행률은 현재 취급된 총 누적대출취급액 중 90일 이상 연체가 된 건의 잔여원금의 비중을 말한다.

미국은 핀테크 기술을 기반으로 인터넷전문은행, P2P 대출중개업체 등 다양한 온라인 플랫폼 기업들이 중저금리 대출을 통해 가계부채 부담을 줄이는 역할을 수행하고 있다. 렌딩클럽(Lending Club)은 2006년도에 설립되었으며, P2P 업체 최초로 2014년 12월에 뉴욕증권거래소(New York Stock Exchange, NYSE)에 상장된 바 있다. 렌딩클럽은 전체 대출 중 카드대환과 타 금융권 대환 등 대환대출 비중이 약 70%로 나타나고 있다. 렌딩클럽은 P2P 대출투자에 기관투자자(전통적인 금융기관)가 적극 참여하고 있으며, 이들 투자자가 전체 투자의 80%를 넘고 있다. 금융기관들이 보유한 전문적인 리스크 관리팀이 렌딩클럽의 대출채권운영방식을 면밀히 검토한 후 투자를 결정하기 때문에

<Table 1> Cumulative Amount and Default Rate of P2P Loan Decisions

P2P Platform Corporations	Date of Establishment	Cumulative Amount of Loan	Delinquency Rate	Default Rate
Lending Club	Mar, 2007	21,000Million USD	5.23%	6.86%
PROSPER	Feb, 2006	6,000Million USD	6.92%	8.76%
LENDIT	Apr, 2015	93Million USD	0.84%	1.36%
SPERCENT	Dec, 2014	102Million USD	0.60%	1.59%
PopFunding	May, 2007	79Million USD	14.70%	3.84%
HonestFund	Jun, 2015	86Million USD	0.62%	0.98%

금융기관의 투자 참여는 개인투자자를 간접적으로 보호하는 효과로 이어진다. 렌딩클럽의 2018년 2월 기준 P2P 누적대출액이 약 23조 원으로 급성장하면서 동시에 연체율도 2007년부터 지속적으로 증가하여 2018년 2월 기준 약 6.86%의 채무불이행률을 나타내고 있다.

프로스퍼(PROSPER)는 미국 최초의 P2P 업체로 2006년 설립 이후 10여년 간 총 60억 달러(6조 4,590억 원) 규모의 대출을 중개하였다. 프로스퍼는 여신 금액을 웹뱅크(Web Bank)라는 여신회사가 담당하여 대출거래를 수행하고, 투자자를 보호하고 있다. 프로스퍼의 평균 투자 수익률은 미국 3대 P2P 대출기업 중 가장 높은 9.3%를 나타내고 있으며, 평균 대출이자율은 17.05%이다. 2018년 2월 기준 프로스퍼의 연체율은 6.92%이고, 채무불이행률은 8.76%이다.

2011년 미국 샌프란시스코에서 설립된 소피(Social Finance, SoFi)는 학자금 대출서비스에 주력하고 있는 P2P 기업이다. 소피는 대학생들 대상으로 고금리로 조달한 학자금 대출을 저금리로 대환해주고 있다. 학자금 대출신청부터 승인까지 15분 안에 완료되는 신속한 대출서비스를 제공하고 있다. 최근에는 자산 규모를 늘려 생존력을 강화하는 방안으로 부동산 자산에 투자규모를 확대하고 있다. 소피의 평균 대출이자율은 19.52%이고, 미국 P2P 기업 중 대출가능금액(65,000달러, 6,945만 원)이 가장 많은 곳이다.

국내 개인신용대출시장은 제1금융권인 은행에서 대출을 받을 수 없는 개인이 20% 이상의 고금리 시장으로 이동하고 있는 등 양극화되고 있다. 이로 인해 다양한 P2P 대출기업이 설립되면서 현재는 약 200여 개의 P2P 기업이 대출중개서비스를 제공하고 있다. 저축은행 등 제2금융권의 고금리(평균 20%)로 부채 부담을 안

고 있는 저신용자 차주들에게 P2P 대출은 평균 연 10%의 금리로 대출서비스를 제공하면서 부채 부담을 줄여주고 있다. 개인신용 및 법인신용 누적대출액 기준으로 국내 P2P 상위기업은 (주)렌딧, (주)에잇퍼센트, (주)팝펀딩, (주)어니스트펀드 순이다.

2015년 4월에 설립된 (주)렌딧(LENDIT)은 주로 개인신용대출을 취급하고 있으며, 창업 3년 만에 누적대출액 1,000억 원을 기록하는 등 개인신용대출 분야 시장점유율 1위 기업이다. 2018년 2월 기준 렌딧의 고객당 평균 대출액은 1,371만 원이고, 평균 대출이자율은 10.9%이다. 렌딧은 1차적으로 나이스신용평가(NICE신용평가)의 300여 가지 금융정보를 바탕으로 1차 대출심사를 진행하며, 2차 심사에서는 렌딧 웹사이트에서 대출신청자가 보여주는 행동유형과 페이스북(Facebook) 정보 등의 비금융 데이터를 수집, 분석하여 최종적으로 렌딧신용등급(Credit Scoring System, CSS)를 산출하여 대출심사를 진행하고 있다. 현재 렌딧은 빅데이터 기반과 인공지능기법을 활용하여 대출심사 평가모형을 정교화해 나가고 있다. 2018년 2월 기준 렌딧의 연체율은 0.84%이고, 채무불이행률은 1.36%이다.

(주)에잇퍼센트(8PERCENT)는 국내에서 최초로 중금리 신용대출 플랫폼을 만든 신생 벤처기업이다. 에잇퍼센트는 국내 최초로 최저금리 보상제를 시작하였으며, IT기술을 기반으로 투자채권에 균등하게 분산투자가 가능한 자동분산투자 추천서비스를 제공하고 있다. 에잇퍼센트의 2018년 2월 기준 누적대출액은 1,141억 원이며, 이는 1년 전에 비해 2배 증가한 수치이다. 에잇퍼센트의 2018년 2월 기준 연체율은 0.60%이며, 채무불이행률은 1.59%를 기록하고 있다.

2007년 5월에 설립된 팝핀딩(주)은 재고자산 담보대출상품인 동산담보 기업투자 서비스에 핵심역량을 가지고 있으며 개인신용대출, 기업 신용대출, 영화제작펀딩 등 다양한 영역에서 P2P 대출서비스를 제공하고 있다. 팝핀딩은 개인신용보다 법인신용대출서비스에 집중하고 있으며, 현재 법인신용대출 시장점유율 1위 기업이다. 팝핀딩의 2018년 2월 기준 누적대출액은 891억 원이며, 연체율은 14.70%, 채무불이행률은 3.84%를 기록하고 있다. 팝핀딩의 연체율이 주요 P2P 업체보다 높은 이유는 법인신용대출과 기타 담보대출 비중이 높기 때문이다.

2015년 6월에 설립된 (주)어니스트펀드는 개인 신용, 법인신용, 그리고 부동산담보 등의 P2P 대출서비스를 제공하고 있으며, 2017년보다 약 5배 성장한 P2P 기업이다. 어니스트펀드의 2018년 2월 기준 누적대출액은 973억이며, 연체율은 0.62%, 그리고 채무불이행률은 0.98%를 기록하고 있다.

한국 P2P금융협회 회원사(64개 기업)의 2018년 2월말 기준 누적대출액은 1년 전보다 232% 급증한 2조 822억 원을 나타내고 있으며, 평균 연체율은 1.90%, 채무불이행률은 3.71%을 기록하고 있다. 국내 P2P 대출거래 연체율의 경우 매년 지속적으로 높아지고 있으며, 국내 P2P 금융시장의 활성화를 위해서는 보다 정확한 대출자들의 채무불이행률 여부를 파악하고 분석하는 것이 필요한 시점이다.

## 2.2 P2P 대출의 상환실패요인 및 채무 불이행률에 관한 연구

P2P 대출서비스와 관련된 연구는 P2P 대출과 기존 은행 금융거래의 비교연구[2], 차입자

의 신용정보 및 대출정보를 고려한 대출상환성공 및 실패요인에 관한 연구[3], 그리고 차용인과 투자자의 온라인 플랫폼 상에서의 사회적 활동변수를 도출한 연구[9] 등의 집단행동 연구가 주로 수행되었다.

재무학 및 금융학에서 오랫동안 수행되었던 금융기관(카드사, 은행 등)의 대출상환 성공요인 및 실패요인과 신용평가모형에 관한 연구는 공통적으로 인구통계학적인 변수와 재무적 변수에 집중하여 연구가 수행되어왔다. 인구통계학적인 변수에는 연령, 교육수준, 결혼유무, 성별, 직업 등이 있으며 재무적 변수에는 가계소득, 주택소유형태, 신용등급, 총부채상환비율, 금융서비스사용액, 사금융(고리대금) 사용유무, 신용카드 연체율 등이 있다.

Lee and Huh[10]는 차입자의 인구통계학적 변수(나이, 교육수준, 결혼유무, 직업), 경제적 변수(가계소득, 주택소유형태), 그리고 재무적 특성 변수(금전관리능력, 부채비율, 금융서비스사용액) 중에서 재무적 변수가 차입자의 연체율 및 채무불이행에 가장 유의미한 영향을 미친다고 주장하였다.

Herzenstein et al.[4]는 P2P 대출거래의 상환 성공요인을 제시하면서 차입자의 재정상태와 신용등급이 대출금액, 대출이자율, 그리고 대출기간 등에 유의한 영향을 미친다고 주장하였다. Weiss et al.[15]는 프로스퍼(Prosper)의 P2P 대출거래 데이터를 이용하여 대출상환성공요인으로 차입자의 대출금액과 대출이자율을 제시하였다. 이들은 차입자의 대출금액과 대출이자율이 높을수록 대출상환성공 가능성이 낮고, 신용도 높은 차입자가 온라인 커뮤니티(online community) 활동이 활발한 경우 대출상환가능성이 높아진다고 주장하였다. Duarte et al.[2]도

프로스퍼의 대출거래 데이터를 기반으로 P2P 대출거래에서 차입자의 외모가 투자자의 대출 의사결정에 미치는 영향도를 측정하였다. 이들은 차입자의 외모가 신뢰를 준다면 대출이자율이 낮아지고, 대출경매 성공가능성도 높아진다고 주장한 바 있다.

Shin and Choi[14]는 온라인 P2P 대출거래의 상환실패요인을 분석하기 위해 차용인의 인구통계학적특성, 신용정보, 그리고 차용인 및 투자자의 행동특성요인을 제시하였다. 이들 연구에서는 특수기록(파산, 면책, 워크아웃, 회생 등의 기록)이 없을수록 상환실패 비율이 높고 차용인의 낙찰 전 온라인상에서의 질의응답횟수가 많을수록 상환실패 비율이 높다고 주장하였다. 또한 신용등급과 월소득이 낮을수록 상환실패 비율이 높다고 주장한 바 있다.

Kim et al.[7]는 P2P 대출의 상환성공요인에 관한 연구에서 파산, 면책, 워크아웃, 그리고 회생 등의 파산경험이 있는 차입자와 주거비용, 생활비용, 의료비용 등을 목적으로 대출받은 차입자는 대출상환가능성이 높다고 주장하였다. 또한 사회적 교류활동 변수로 제시한 질의응답수와 대출상환투표율도 대출상환성공에 유의한 영향을 미친다는 연구결과를 제시한 바 있다.

Lim et al.[11]는 P2P 대출이용자 측면의 가치탐색 연구에서 텍스트마이닝(textmining) 기법을 이용하여 P2P 대출 관련 기사의 단어별 빈도분석을 수행하였다. 이들 연구에서는 리스크, 위험, 손실, 우려, 대부업 등의 P2P 대출에 대한 부정적 인식에 대한 키워드가 높은 빈도로 도출되어 P2P 대출시장의 신뢰도를 높이는 방안과 정교화된 리스크 관리시스템 구축이 필요하다고 주장하였다.

Lin et al.[12]는 P2P 대출서비스의 사회적 교류활동에 관한 연구를 통해 온라인 커뮤니티 상에서 차입자와 투자자가 친분관계를 쌓으면 대출경매 성공률은 높아지고, 대출이자율은 낮아진다고 주장하였다. Lin et al.[13]는 중국의 P2P 플랫폼 기업의 대출거래 데이터를 이용하여 채무불이행률에 영향을 미치는 주요 변수를 제시하였다. 이들은 인구통계학적 변수(나이, 성별, 결혼유무, 학력수준, 근무경력)와 재정변수(급여, 대출금액, 총부채상환비율, 연체기록)가 채무불이행 여부에 영향을 주는 변수라고 언급하였다. Zhang et al.[17]은 중국 P2P 대출서비스의 대출상환에 관한 연구에서 이자율, 신용등급, 이전 대출성공 횟수 등이 대출상환가능성을 높여준다고 주장하였다. 높은 대출이자율과 신용등급은 대출상환 성공률을 높여주며 상환기간이 길면 대출상환 성공률은 낮아진다고 주장하였다.

Yang and Lee[16]는 국내 주요 P2P 업체(에잇퍼센트, 빌리, 펀듀, 펀다)의 대출거래 자료를 바탕으로 중소기업과 소상공인 P2P 대출의 투자결정요인을 제시하였다. 이들은 투자수익률, 대출기간, 차입자 속성(신용등급, 재무상황), 그리고 P2P 대출플랫폼의 정보제공이 투자결정에 중요한 요인이라고 주장한 바 있다.

### 3. 연구 방법

#### 3.1 자료 수집 및 변수 선정

본 연구의 목적은 P2P 대출거래에서 채무불이행 요인을 추출하여 이를 사전에 경고할 수 있는 요인을 투자자에게 알려주어 채무불이행

율을 낮추는 것이다. 이를 위해 세계적인 P2P 대출업체인 렌딩클럽(Lending Club)의 공개용 데이터베이스(<http://www.lendingclub.com>)를 활용하고 통계기법과 인공지능기반의 P2P 대출거래 채무불이행 예측모형을 구축하고자 한다. 렌딩클럽은 2007년부터 채무상태와 채무정보, 발행된 모든 대출거래에 대한 데이터, 그리고 렌딩클럽의 신용정책에 충족하지 못한 거절된 대출신청들의 세부사항에 대한 데이터를 공개하고 있다. 발행된 P2P 대출데이터에는 총 151개의 변수가 있고, 거절된 대출신청 데이터는 총 9개의 변수가 있다.

본 연구는 2016년 1월부터 2017년 12월까지 2년 간 렌딩클럽의 P2P 대출거래 데이터를 수집하여 총 759,298개의 대출거래 데이터를 획득하였다. <Table 2>는 본 연구에서 종속변수로 사용하게 될 채무상태를 나타내고 있다. 렌딩클럽에서는 채무상태를 연체 정도와 대출상환만료 여부에 따라 (1) Current(상환 진행 중인 대출), (2) In Grace Period(연체일 15일 이하인 대출), (3) 연체일(16~30일 이하), (4) 연체일(31~120일 이하), (5) Fully Paid(상환완료), (6) Default(채무불이행)의 6가지 속성으로 구분한다.

본 연구는 P2P 대출거래의 채무불이행 예측 모형 구축이 목적이므로 현재 상황이 진행 중이거나 연체 중이어서 상환성공과 채무불이행 여부를 알 수 없는 데이터는 제외하고, 상환완료(Fully Paid)와 채무불이행(Default) 속성을 종속변수로 활용하고자 한다. <Table 2>에서 보는 바와 같이 상환완료된 대출은 120,434건이며, 채무불이행은 32,816건이다. 채무불이행은 연체일이 150일을 초과할 경우 그 대출은 채무불이행으로 분류하여 더 이상 채무를 상환할 의지가 없다고 간주한다.

<Table 2> Loan Application Status of Lending Club(Jan 2016 to Dec 2017)

Application Status	Number of Loans
1) Current	577,331
2) In Grace Period	9,818
3) Delinquency(16 to 30 days)	3,242
4) Delinquency(31 to 120 days)	15,657
5) Fully Paid	120,434
6) Default	32,816

본 연구에서는 상환 완료된 대출과 채무불이행 대출의 특징을 설명하는 151개의 변수 중에서 채무불이행 예측에 사용할 수 없는 속성을 제외한 73개 속성의 데이터에 대해 전처리를 진행하였다. 다음으로 신용평가모형과 관련된 선행연구와 렌딩클럽의 P2P 대출거래시스템에서 제공하는 핵심변수를 고려하여 차입자의 인구통계학적 변수를 포함한 차입자정보 관련 변수, 대출정보변수, 그리고 재무 및 신용변수로 이루어진 13개의 독립변수를 선정하였다. <Table 3>에서 보는 바와 같이 차입자정보 관련 변수에는 주택소유형태, 차입자의 근무경력, 6개월 평균 잔고액, 총부채상환비율(debt to income, DTI) 등으로 구성된다. 대출정보에는 대출신청금액을 로그(log)화한 대출금액과 대출목적, 대출이자율 등으로 구성된다. 대출목적은 생계필수비용(주거비, 생활비, 의료비), 사업용, 부채상환용 등으로 구분된다. 대출이자율은 P2P 차입자의 대출이자율을 말한다. 신용정보 변수에는 차입자의 신용과산기록, 신용조회건수, 신용한도대비 신용사용금액, 연체계좌수, 모기지론 계좌 수, 차입자의 발급받은 신용카드 계좌 수 등으로 구성된다. 신용과산기록은 차입자의 과산, 면책, 워크아웃, 회생 등의

기록유무를 말하고, 신용조회건수는 투자자들이 대출자에 대해 신용을 조회하는 횟수를 말한다. 마지막으로 대출상환성공더비는 대출상환 성공 시 '1'의 값, 채무불이행시 '0'의 값을 갖는 더미변수를 말하며, P2P 대출시스템에서 가장 중요한 측정지표인 채무불이행률(default rate)을 구성하는 변수이다.

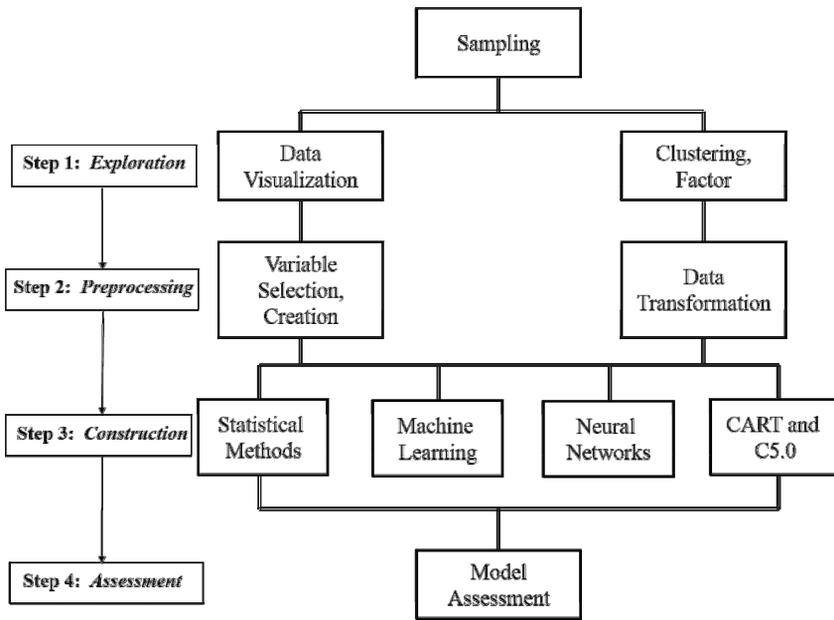
### 3.2 분석 절차 및 분석 방법

본 연구는 <Figure 1>과 같은 분석 절차에 따라 진행된다. 단계 1은 데이터에 대한 이해와 분석이다. P2P 대출거래의 구조와 특징을 이해하고, P2P 대출거래에서 사용 가능한 데이터 파악, 데이터 원천들에 대한 위치와 구조, 데이

터 테이블의 필드와 그들의 코드 분석, 그리고 데이터의 신뢰성, 정확성, 유용성을 검토하였다 [8]. 단계 2는 데이터 준비단계로 데이터 사전처리와 연구변수의 전처리 과정을 실시하였다. 채무불이행 예측에 사용할 수 없는 속성을 제외한 데이터에 대해 전처리 과정을 진행하였다. 상환 완료된 대출과 채무불이행 대출 건인 153,250개의 대출거래에서 전처리 과정을 거쳐 총 143,639건의 대출거래를 분석에 활용하고, 모든 분석은 학습용과 검증용의 두 가지 데이터 셋으로 구성하였다. 즉, 학습용 데이터 셋 (training dataset)으로 60%(86,183/143,639)를 사용하고, 나머지 40%(57,456/143,639)는 검증용 데이터 셋(validation dataset)으로 사용하였다. 또한 보다 일반화된 연구결과를 얻기 위해

<Table 3> Variable Used in the Study

Characteristics	Variables	Descriptions
Borrower Characteristics	Housing Situation(X1)	Our values are: rent, own, mortgage, other
	Length of Employment(X2)	Employment length in years
	Average Current Balance(X3)	Average current balance of all accounts
	Debt to Income Ratio(X4)	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations
Loan Characteristics	Loan Amount(X5)	The listed amount of the loan applied for by the borrower
	Loan Purpose(X6)	14 loan purposes: debt consolidation, credit card, home improvement, major purchase, medical, car loan, moving, vacation, house, and other
	Interest Rate(X7)	Interest rate on the loan
Credit Information	Public Records(X8)	Number of derogatory public records
	Number of Finance Trades(X9)	Number of finance trades
	Total Credit/Credit Limit(X10)	Total high credit/credit limit
	Number of Delinquent Accounts(X11)	The number of accounts on which the borrower is now delinquent
	Number of Mortgage Accounts(X12)	Number of mortgage accounts
	Number of Bankcard Accounts(X13)	The number of bankcard accounts in the borrower's credit file
Default Rate	Loan Funded Successful(Y)	Dependent variable



〈Figure 1〉 Analysis Procedures

5회의 상호검증방법(5-fold cross-validation method)을 사용하였다. 단계 3은 데이터 분석 및 모형화 단계로 분석용 데이터 마트(data mart)를 이용하여 P2P 대출거래 채무불이행 예측모형을 구축하는 단계이다. 본 연구에서는 벤치마크 모델로 통계기법인 판별분석과 로지스틱 회귀분석을 이용하였고, 인공지능기법으로는 인공신경망과 의사결정나무로 가장 널리 활용되고 있는 CART와 C5.0 알고리즘을 이용하여 P2P 대출거래 채무불이행 예측모형을 구축하였다. 인공지능기법의 설계과정에서는 R-studio 프로그램을 이용하였다. 단계 4에서는 P2P 대출거래의 검증용 데이터를 예측모형에 적용시킨 후 다양한 평가도구들을 이용하여 이들 모형의 예측정확도를 비교하여 그 유용성을 검증하고 최종적인 예측모형을 제시하고자 한다.

#### 4. 연구모형 설계 및 연구 결과

##### 4.1 통계기법과 인공지능기법의 P2P 대출거래 채무불이행 예측모형 설계

종속변수가 명목척도로 측정된 범주형 질적 변수인 경우에 모형개발을 위해서 사용할 수 있는 대표적인 통계분석방법이 판별분석과 로지스틱 회귀분석(로지스틱)이다. 채무불이행 여부를 종속변수로 설정하고, 차입자정보, 대출정보, 그리고 재무 및 신용정보로 이루어진 13개의 독립변수로 정의된 판별분석과 로지스틱 분석을 수행하였다. 판별분석은 특정한 관심대상이 어느 집단에 속하는지를 예측하는 모형을 개발하는데 사용되는 대표적인 분석방법이다. 차입자정보, 대출정보 그리고 신용정보 등의 특성을 나타내는 독립변수 값들을 판별함수식에 대입

하여 얻은 결과 값을 이용해서 정상적으로 상환 완료된 대출집단과 채무불이행 집단을 예측하였다. 로짓분석은 분석대상들이 두 집단 혹은 그 이상의 집단으로 나누어진 경우에 개별 관측치들이 어느 집단으로 분류될 수 있는가를 분석하고 이를 예측하는 모형을 개발하는데 이용된다. 로짓분석에서는 연구변수의 다중공선성(multicollinearity) 문제를 해결하기 위해서 변수선택법을 고려하였다. 변수선택법 중에서는 가장 안정적인 방법인 단계별선택법(stepwise selection method)을 이용하여 회귀모형을 구축하였다. 단계별선택법은 다른 독립변수가 회귀식에 존재할 때 종속변수에 영향력이 있는 변수들만을 회귀식에 포함시키기 때문에 예측함수를 개발하는데 효과적이다[1].

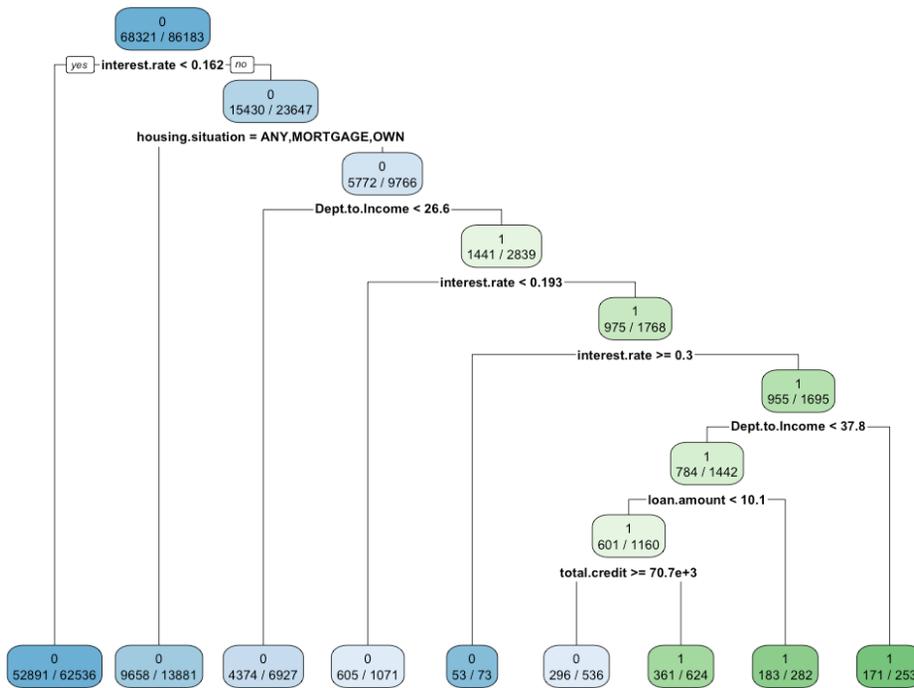
신경망(neural networks)은 인간의 뇌가 수많은 신경들로부터 입력과 출력의 신호를 전달하는 과정을 착안하여 모델화한 방법이다. 신경망은 최근까지 부도예측, 신용등급예측, 배당금예측, 외환시장거래, 신용카드 부정행위 발견 등 금융분야에서 성공적으로 적용되었다. 신경망의 구조는 다양한 형태가 있으나 일반적으로 많이 쓰이는 형태는 관리학습에 적합한 다층전방향 신경망(multilayer feedforward networks)이다. 이는 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)의 삼층구조를 이루며 각 층마다 다수의 뉴런(neuron) 또는 노드(node) 등의 처리단위를 소유하고 있다. 서로 다른 층에 존재하는 처리단위는 서로 연결되어 있으며 연결강도는 연결가중치(interconnection weight)로 계산된다. 신경망의 구조에서 입력계층의 입력노드 수는 학습자료 내의 독립변수 수인 '13개'로 설정하였다. 출력계층의 출력노드 수도 학습자료 내의 종속변수

수로 설정하였다. 은닉계층의 적절한 은닉노드 수는 신경망 초기에 미리 알 수는 없으며, 실험을 통해 적절한 은닉노드 수를 결정한다. 일반적으로 은닉노드 수가 많아지면 학습자료에 과도학습(over-fitting)되어 일반화 오류가 커지는 경향이 있으며, 은닉노드수가 너무 적으면 학습자료에 대한 예측력이 떨어지게 된다. 본 연구에서는 시행착오법과 Hornik[5]의 선행연구를 토대로 다음과 같이 신경망 아키텍처(architecture)를 결정하였다. 다층 전방향 신경망의 구조에서 은닉층의 수는 1과 2로 설정(3층과 4층 신경망)하였고, 은닉노드의 수는 3개, 7개( $n/2$ ), 10개, 13개( $n$ ), 20개, 26개( $2n$ ), 39개( $3n$ )개로 설정하였다. 다음으로 학습률(learning rate)은 학습초기에 학습률을 '1'로 설정하여 빠른 학습을 유도하고, 그 다음에는 가중치의 조정이 반복(iteration)됨에 따라서 천천히 학습률을 감소시켰다.

본 연구는 보다 효과적인 P2P 대출거래의 채무불이행 여부를 사전에 구분하기 위한 방법으로 규칙(rule)이라는 형태를 이용하고자 의사결정나무 방법을 채택하였다. 의사결정나무 알고리즘에는 CHAID, CART, QUEST, C5.0 등이 있으며 본 연구에서는 이산형(discrete) 목표변수 분류를 위해 널리 쓰이는 CART와 C5.0 알고리즘을 사용하였다. 의사결정나무 모형의 설계과정에서 고려해야 할 요인으로는 분리기준(splitting criterion)과 정지규칙(stopping rule)이 있다(<Table 4> 참고). 본 연구는 이산형 목표변수 분류를 위한 것으로 CART는 지니지수(Gini index), C5.0는 엔트로피 지수(Entropy index)를 분리기준으로 각각 사용한다. 정지규칙은 더 이상 분리가 일어나지 않고 현재의 마디가 얹히 되도록 하는 규칙을 말한다. 정지

<Table 4> Maximum Tree Depth and Splitting Criterion of Decision Trees

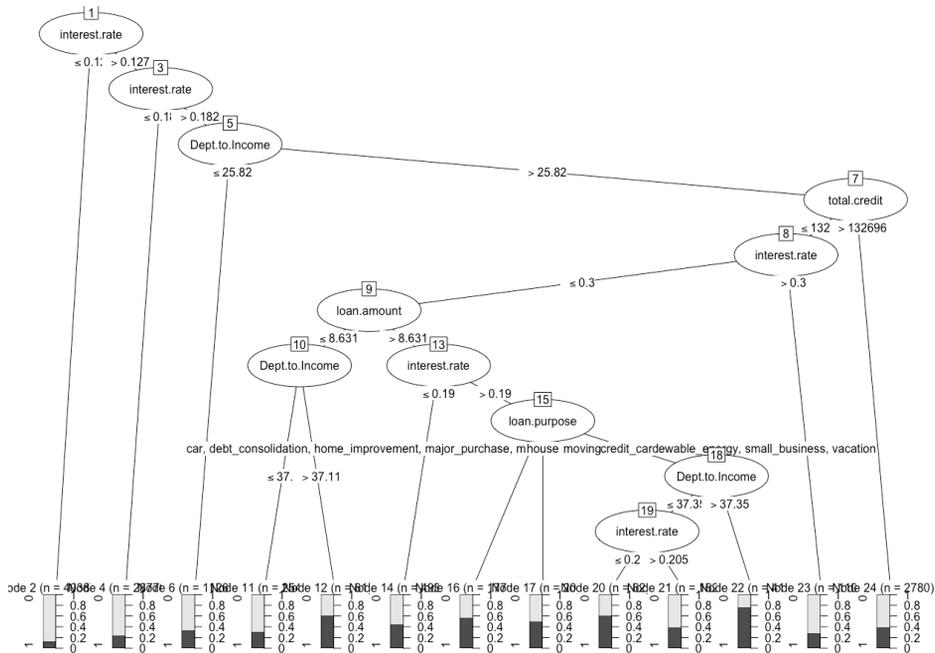
Algorithms	Splitting Criterion	Maximum Tree Depth	Minimum Number of Cases
CART	Gini Index	5	Parent Node: 5 Child Node: 1
C5.0	Entropy Index	5	Parent Node: 5 Child Node: 1



<Figure 2> An Example of CART Rules from Training Dataset 1

규칙에서 최대 나무구조의 깊이(maximum depth of tree)를 5로 지정한 이유는 최대 나무구조의 깊이 수치가 4일 때까지 부모마디(parent node)의 순수도(purity)에 비해서 자식마디(child node)들의 순수도가 증가하였으나 5 이상의 수치에서는 부모마디의 순수도에 비해 자식마디들의 순수도가 증가하지 못하였다. <Figure 2>와 <Figure 3>은 CART와 C5.0로 추출된 분류나무(classification tree)의 규칙을 나타낸 것이다. 첫 번째 데이터 셋(dataset 1) 기준으로 CART

에서는 가장 영향력 있는 변수로 추출된 뿌리마디(root node)의 위치에 있는 대출이자율이 분류되었고, 다음으로 주택소유형태, 총부채상환비율, 총신용사용금액/신용한도, 대출금액 순으로 분류되었다. C5.0의 경우에도 대출이자율이 가장 영향력 있는 변수로 선정되었고, 다음으로 대출금액, 총부채상환비율, 대출목적, 총신용사용금액/신용한도 순으로 중요도가 분류되었다. 다른 데이터 셋에서의 변수중요도 연구 결과는 <Table 5>에 정리하였다.



〈Figure 3〉 An Example of C5.0 Rules from Training Dataset 1

#### 4.2 P2P 대출거래 채무불이행 예측모형의 오분류율 비교분석

〈Table 5〉는 판별분석, 로짓분석, 신경망, 의사결정나무 알고리즘을 이용하여 P2P 채무 불이행에 미치는 독립변수들의 중요도 순위를 나타낸 것이다. 변수중요도 순위는 5회의 상호 검증방법(5-fold cross-validation method)을 이용한 검증용 데이터의 결과를 정리한 것이다. 판별분석에서는 대출이자율, 대출금액, 신용과 산기록, 모기지대출계좌수, 총부채상환비율 순으로 중요도가 분류되었다. 로짓분석의 경우 대출이자율, 총부채상환비율, 대출금액, 총신용 사용금액/신용한도, 신용조회건수의 순서로 영향력 있는 변수를 판별하였다. 즉, 전통적인 통계기법에서는 공통적으로 대출이자율, 대출금액, 총부채상환비율이 P2P 대출거래의 채무불

이행을 예측하는데 가장 영향력있는 변수로 추출되었다. 신경망에서는 신용카드계좌수를 가장 영향력 있는 변수로 판별하였고, 다음으로 대출목적, 근무경력, 대출금액, 총신용사용금액/신용한도 순으로 중요도가 분류되었다. CART에서는 대출이자율이 가장 영향력 있는 변수로 선정되었고, 다음으로 주택소유형태, 총부채상환비율, 총신용사용금액/신용한도, 대출금액 순으로 분류되었다. C5.0의 경우에도 대출이자율이 가장 영향력 있는 변수로 선정되었고, 다음으로 대출금액, 총부채상환비율, 대출목적, 총신용사용금액/신용한도 순으로 선정되었다. 전통적인 통계기법과 인공지능기법을 비교하면, 대출목적은 통계기법에서는 중요한 변수로 선별되지 않았으나 인공지능기법에서는 대출목적이 중요한 변수로 추출되었다. 종합하면, P2P 대출거래의 채무불이행 예측을 위해 우선 고려

〈Table 5〉 The Determinant Factors for Predicting P2P Lending Default

Ranking	MDA	LR	MLP	CART	C5.0
Top 1	Interest Rate	Interest Rate	Number of Bankcard Accounts	Interest Rate	Interest Rate
Top 2	Loan Amount	Debt to Income Ratio	Loan Purpose	Housing Situation	Loan Amount
Top 3	Public Records	Loan Amount	Length of Employment	Debt to Income Ratio	Debt to Income Ratio
Top 4	Number of Mortgage Accounts	Total Credit/Credit Limit	Loan Amount	Total Credit/Credit Limit	Loan Purpose
Top 5	Debt to Income Ratio	Number of Finance Trades	Total Credit/Credit Limit	Loan Amount	Total Credit/Credit Limit

Note: MDA(Multiple Discriminant Analysis), LR(Logistic Regression), MLP(Multi-Layer Perceptron), CART(Classification and Regression Tree), C5.0(Commercial version 5.0).

〈Table 6〉 Misclassification Rate of P2P Lending Default Prediction Models

Data Set no.	Types	MDA	LR	MLP	CART	C5.0
Dataset 1	Training	0.2136	0.3630	0.1766	0.2008	0.2276
	Validation	0.2300	0.3427	0.1802	0.2301	0.2064
Dataset 2	Training	0.2013	0.3557	0.1705	0.2116	0.2054
	Validation	0.1994	0.3866	0.1892	0.2378	0.2189
Dataset 3	Training	0.2296	0.3398	0.1634	0.2366	0.2343
	Validation	0.1923	0.3722	0.1911	0.2210	0.1965
Dataset 4	Training	0.2322	0.3688	0.1723	0.2111	0.2312
	Validation	0.2145	0.3955	0.1899	0.2499	0.2207
Dataset 5	Training	0.1953	0.3492	0.1632	0.2375	0.2188
	Validation	0.2023	0.3740	0.1902	0.2177	0.2094
Average	Training	0.2144	0.3553	0.1692	0.2192	0.2275
	<b>Validation</b>	<b>0.2077</b>	<b>0.3742</b>	<b>0.1881</b>	<b>0.2313</b>	<b>0.2100</b>

Note: MDA(Multiple Discriminant Analysis), LR(Logistic Regression), MLP(Multi-Layer Perceptron), CART(Classification and Regression Tree), C5.0(Commercial version 5.0).

해야 할 변수는 대출이자율이며, 중요도 3순위에 가장 많이 언급된 대출금액과 총부채상환비율도 고려해야 할 요인으로 추출되었다.

다음으로 판별분석, 로짓분석, 신경망, CART, 그리고 C5.0 알고리즘을 이용하여 구축한 P2P 채무불이행 모형의 예측성과를 비교 분석하기 위해 교차검증을 통한 오분류율(misclassification rate)을 제시하였다. <Table 6>에서 보는 바와 같이 전통적인 통계기법인 판별분석의 평균 오분류율(검증용 데이터 기준)은 0.2077이며, 로짓분석의 평균 오분류율은 0.3742로 나타나

판별분석이 로짓분석에 비해 예측성과가 더 좋은 결과를 보여주고 있다. 인공지능기법인 신경망의 평균 오분류율은 0.1881, CART의 평균 오분류율은 0.2313, 그리고 C5.0의 평균 오분류율은 0.2100으로 나타나 인공지능기법 중에서는 신경망의 예측성과(은닉노드의 수: 10개)가 더 좋은 것으로 나타났다. 전체적으로 전통적인 통계기법(평균 오분류율: 0.2009)보다는 인공지능기법(평균 오분류율: 0.2008)의 예측성과가 더 좋은 것으로 나타났으며, 신경망의 경우 모든 데이터 셋에서 오분류율이 가장 낮은 예측모형으로 나타났다.

## 5. 결 론

핀테크 산업의 발전으로 등장한 P2P 대출(금융)은 제도권 은행차입이 제한적인 저신용자와 저소득자의 새로운 차입수단과 자금조달수단으로 활용되고 있다. P2P 대출의 연체와 채무불이행은 전통적인 금융기관 대출에 비해 높고, 부실위험을 전적으로 불특정 다수의 비전문가인 개인 투자자들이 부담해야 한다는 점에서 차입자의 상환 실패에 따른 투자자의 위험부담이 중요한 관심사항이 되고 있다. P2P 대출의 핵심은 투명한 자금중개를 위한 정교한 심사평가모형의 개발에 있으며, 이를 위해 금융빅데이터 분석과 다양한 인공지능기법의 활용이 필요한 시점이다.

따라서 본 연구의 목적은 P2P 대출거래의 신뢰를 높이기 위해 투자자에게 P2P 대출거래의 채무불이행을 사전에 고지할 수 있는 P2P 대출거래의 채무불이행 예측모형을 구축하여 연체율과 채무불이행을 낮추는 것이다. 연구데이터는 차입자의 인구통계학적 변수를 포함한 차입자정보, 대출정보, 그리고 채무 및 신용정보로 이루어진 변수를 이용하여 통계적 기법 및 인공지능기반의 P2P 대출거래 채무불이행 예측모형을 구축하였다. 분석용 데이터는 온라인 P2P 대출업체인 렌딩클럽의 2016년부터 2년간 총 143,639건의 P2P 대출거래를 획득하였고, 일반화된 연구결과를 얻기 위해 상호교차검증 방법론을 이용하였다.

본 연구의 결과는 다음과 같이 요약할 수 있다.

첫째, 판별분석에서는 P2P 채무불이행 여부에 가장 영향력 있는 변수로 대출이자율이 선정되었고, 다음으로 대출이자율, 대출금액, 신용과산기록, 모기지대출계좌수, 총부채상환비율

순으로 중요도가 분류되었다. 로짓분석의 경우에도 대출이자율이 채무불이행 예측에 가장 영향력 있는 변수로 선정되었고, 다음으로 총부채상환비율, 대출금액, 총신용사용금액/신용한도, 신용조회건수 등이 영향력 있는 변수로 판별되었다.

둘째, 인공지능기법인 신경망에서는 신용카드계좌수를 가장 영향력 있는 변수로 판별하였고, 다음으로 대출목적, 근무경력, 대출금액, 총신용사용금액/신용한도 순으로 중요도가 선정되었다. CART에서는 통계기법과 마찬가지로 대출이자율이 가장 영향력 있는 변수로 선정되었고, 다음으로 주택소유형태, 총부채상환비율, 총신용사용금액/신용한도, 대출금액 순으로 선정되었다. C5.0의 경우에도 대출이자율이 가장 영향력 있는 변수로 선정되었고, 다음으로 대출금액, 총부채상환비율, 대출목적, 총신용사용금액/신용한도 순으로 선정되었다.

셋째, 신경망 모형을 제외한 모든 예측모형에서 대출이자율이 P2P 대출거래의 채무불이행 예측에 가장 영향력 있는 변수로 선정되었고, 중요도 3순위 안의 변수로 대출금액과 총부채상환비율이 영향력 있는 변수로 선정되었다. 이는 대출정보와 재무적 변수가 채무불이행에 미치는 영향이 크다는 선행연구 결과와 일치한다. 특히 총부채상환비율이 높은 차입자의 경우 연체율 및 채무불이행률이 대체적으로 높은 특징을 나타내고 있다.

넷째, 차입자의 대출신청금액이 많을수록 채무불이행 확률이 높은 것으로 나타났고, 대출목적 또한 P2P 대출거래의 채무불이행에 유의한 영향을 미치는 것으로 나타났다. 주택구입비용, 생활비, 의료비 등의 생계필수비용 목적으로 대출받은 차입자의 경우 채무불이행률이 낮은 것으로 나타났다. 위의 연구결과는 P2P

대출서비스가 서민금융과 금융소외계층에게 필요 자금조달수단이 되는 동시에 신용과산기록이 있는 차입자의 신용회복을 위한 수단으로 활용될 수 있다는 것을 의미한다.

마지막으로, 검증용 데이터의 상호교차검증을 통한 P2P 채무불이행 예측모형의 성과로 5개의 데이터 셋에서 신경망이 가장 낮은 평균 오분류율(0.1881)을 나타내었고, 다음으로 판별분석(0.2077), C5.0(0.2100) 순으로 나타났다. 가장 우수한 예측모형 성과를 보인 신경망에서는 다른 예측모형과는 다르게 ‘신용카드계좌수’ 변수를 가장 영향력 있는 변수로 판별하여, 위 변수도 채무불이행 예측에 우선적으로 고려해야 할 필요성이 있다.

현재 국외 P2P 업체들은 금융 빅데이터(financial bigdata) 분석 기반의 대출심사평가모형을 도입하여 대출자 입장에서는 세분화된 대출금리 서비스를 제공받고 있다. 대출신청자의 객관적 신용정보 이외에 주관적 정보를 반영하여 차별화된 대출금리를 산출하고 있다. 따라서 향후 연구에서는 P2P 대출거래에서 차용인의 재무적 특성과 같은 객관적 신용정보와 더불어 주관적 정보 또는 행동특성요인과 집단행동이론(herding behavior theory) 관련 변수를 고려한 채무불이행 예측모형을 구축하고자 한다. 대출신청자의 금융거래실적과 재무상황이 주 분석 대상이었던 전통적 금융기관의 대출심사평가와 달리 기존 신용정보 이외에 심리적 행동데이터, 사회관계망 데이터, 대출신청자가 사용한 로그(log) 및 키워드 분석, 그리고 대출신청자가 명시한 자금사용목적 및 상환계획 등의 주관적이고 비정형적인 데이터를 반영하여 P2P 대출거래의 채무불이행 예측모형을 구축할 필요성이 있겠다.

---

## References

---

- [1] Choi, C., Kim, J., Kim, J., Kim, H., Lee, W., and Kim, H., “Development of Heavy Rain Damage Prediction Function Using Statistical Methodology,” *The Korean Society of Hazard Mitigation*, Vol. 17, No. 3, pp. 331-338, 2017.
- [2] Duarte, J., Siegel, S., and Young, L., “Trust and Credit: The Role of Appearance in Peer to Peer Lending,” *Review of Financial Studies*, Vol. 25, No. 8, pp. 2455-2484, 2012.
- [3] Herrero-Lopez, S., “Social Interactions in P2P Lending,” *Proceedings of the 3<sup>rd</sup> Workshop on Social Network Mining and Analysis*, Paris, France, June 28, 2009, ACM Press, New York, 2009.
- [4] Herzenstein, M., Andrews, R., and Dholakia, U., “The Democratization of Personal Consumer Loans? Determinants of Success in Online Peer-to-Peer Lending Communities,” *Working Paper*, Available at SSRN [www.prosper.com](http://www.prosper.com), 2008.
- [5] Hornik, K., “Approximation Capabilities of Multilayer Feedforward Networks,” *Neural Networks*, Vol. 4, No. 2, pp. 251-257, 1991.
- [6] IResearch(艾瑞諮詢), “Trends and Long Term Outlook for the China’s Online Peer-to-Peer(P2P) Lending Market,” *The Strategic Research Report*, Vol. 1, pp. 1-18, 2017.

- [7] Kim, H. K., Park, G. W., Lee, B. T., and Choi, E. H., "A Study on Determinants of Loan Repayment in Peer-to-Peer Lending," *Asian Review of Financial Research*, Vol. 26, No. 3, pp. 381-415, 2013.
- [8] Kim, J. H., Bae, J. K., and Jeon, H. C., "A Study on the Information Cascades Effects of the Offline WOM and Online Review," *The Journal of Society for e-Business Studies*, Vol. 15, No. 1, pp. 39-60, 2010.
- [9] Lee, E. and Lee, B., "Herding Behavior in Online P2P Lending: An Empirical Study Investigation," *Electronic Commerce Research and Applications*, Vol. 11, No. 5, pp. 495-503, 2012.
- [10] Lee, E. Y. and Huh, E. J., "Korean Households' Delinquent Behavior and the Determinants of Debt," *Journal of Consumer Studies*, Vol. 16, No. 1, pp. 179-194, 2005.
- [11] Lim, E. J., Lee, H. J., and Jeong, S., "A Study on Consumers' Core Value on P2P Loan Service," *Journal of Consumer Studies*, Vol. 26, No. 6, pp. 267-291, 2015.
- [12] Lin, M., Prabhala, N., and Viswanathan, S., "Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending," *Management Science*, Vol. 59, pp. 17-35, 2013.
- [13] Lin, X., Li, X., and Zheng, Z., "Evaluating Borrower's Default Risk in Peer-to-Peer Lending: Evidence from a Lending Platform in China," *Applied Economics*, Vol. 49, No. 35, pp. 3538-3545, 2017.
- [14] Shin, D. H. and Choi, M. S., "An Empirical Study on the Default Factor in Online P2P Lending," *Korea Journal of Business Administration*, Vol. 25, No. 5, pp. 2233-2254, 2012.
- [15] Weiss, G., Pelger, K., and Horsch, A., "Mitigating Adverse Selection in P2P Lending Empirical Evidence from Prosper.com," Working Paper, University of Bochum, 2010.
- [16] Yang, Q. and Lee, Y. C., "Influencing Factors on the Lending Intention of Online Peer-to-Peer Lending: Lessons from Renrendai.com," *Journal of Information Systems*, Vol. 25, No. 2, pp. 79-110, 2016.
- [17] Zhang, Y., Li, H., Hai, M., Li, J., and Li, A., "Determinants of Loan Funded Successful in Online P2P Lending," *Procedia Computer Science*, Vol. 22, pp. 896-901, 2017.

## 저 자 소 개



배재권  
2009년  
2010년~2012년  
2012년~현재  
관심분야

(E-mail: jkbae99@kmu.ac.kr)  
서강대학교 경영학과 MIS (박사)  
동양대학교 철도경영학과 전임강사  
계명대학교 경영정보학과 부교수  
금융빅데이터 분석, 핀테크 서비스, 지능형 시스템



이승연  
2015년~현재  
관심분야

(E-mail: stat\_lsy@naver.com)  
계명대학교 통계학과 (학사과정)  
금융빅데이터 분석, P2P금융



서희진  
2013년~현재  
관심분야

(E-mail: lion186@naver.com)  
계명대학교 경영정보학과 (학사과정)  
금융빅데이터 분석, P2P금융