

고객주소 변경유형에 따른 신용카드 부도대상군 분류에 관한 연구

이 태 훈¹ · 김 형 중^{2*}

¹고려대학교 정보보호대학원 빅데이터응용및보안학과 석사과정

²고려대학교 정보보호대학원 빅데이터응용및보안학과 교수

A Study on the Classification of Credit Card Defaulters according to the Types of Change of Customer Addresses

Tae-Hoon Lee¹ · Hyoung-Joong Kim^{2*}

¹Master's Course, Department of Bigdata Application and Security, Korea University, Seoul 02841, Korea

²Professor, Department of Bigdata Application and Security, Korea University, Seoul 02841, Korea

[요 약]

신용카드사에서 1년간 발생하고 있는 직장이나 자택의 주소변경은 588만건인데, 이러한 주소정보 변경 이벤트가 고객 유형에 따라 부도위험에서 차별화 된다는 점을 발견하였다. 주소변경 건을 직장주소 변경과 자택주소 변경으로 크게 나누고, 주소변경 유형을 1)신규등록, 2)동일 읍면동내에서 변경, 3)읍면동 이상의 이동, 4)주소 삭제의 4가지로 분류하여 규모와 부도위험을 측정하였다. 이중에 직장 주소가 삭제된 경우의 부도위험이 타 유형에 비해 2~3배 높은 수준으로 나타났다. 최근에는 비금융거래에 기반한 대안정보를 활성화하여 신용평가 체계를 정교화 하고자 하는 금융정책이 도입되고 있는데, 이에 착안하여 금융거래정보를 배제하고 주소변경과 신상정보를 간단히 혼합하여 고객 부도위험 분류모형을 구축하고자 시도하였다. 현장에서 Rule-Base기반의 의사결정 분류모형을 이용한다면 활용성이 좋을 것으로 기대한다.

[Abstract]

There are 5.88 million changes in address at work or home that occur in credit card companies over a year. We found these changes in address information events have differentiated customers from risk of default. They were greatly divided into changes of work place address and home address, into 4 types to measure the size and risk of default.; 1) the types of change of address newly registered, 2) the changes within the same town-facing area, 3) then moves beyond township-style, and 4) the deleted addresses. Among them, the risk of default of when workplace addresses are deleted is two to three times higher than other types. Recently, a financial policy was introduced to refine the credit rating system by activating alternative information based on non-financial transactions. Accordingly, we tried to exclude financial transaction information and attempted to establish a classification model for customer default risk by simply mixing address changes and personal information. it is expected that the use of the rule-based default prediction model will be useful.

책임어 : 주소정보 변경, 신용카드, 주소삭제, 부도위험, 대안정보

Key word : address information change, credit card, deleted addresses, risk of default, alternative information

<http://dx.doi.org/10.9728/dcs.2020.21.1.185>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 15 November 2019; **Revised** 11 December 2019

Accepted 23 January 2020

***Corresponding Author; Hyoung-Joong Kim**

Tel: +82-2-3290-4895

E-mail: kjh-@korea.ac.kr

I. 서론

1-1 연구 배경 및 목적

한국의 신용카드 이용은 선진국에 비해 짧은 역사를 가졌지만, IMF시의 경기침체를 극복하기 위한 신용카드 활성화 정부 정책으로 선진국을 능가하는 카드사용율 및 인당이용금액, 높은 카드 이용률, 개인 대출 상품으로 소비주도의 단기 경제 회복을 이루었다. 하지만 길거리 모집, 과도한 현금서비스 한도 부여, 불완전한 자격 심사 및 신용평가 관리 미비로 몇 개 카드사가 유동성 위기에 빠진 '04년에는 카드부실 사태를 경험하였으며, 자본잠식, 공적자금 투입등 이후 카드업계 및 관련 금융업계는 구조조정 되고 많은 국민들이 연체나 파산, 개인회생 상태에 이르러 어렵고 뼈아픈 자구의 시간을 보내야 했던 경험이 있다.

그러나 아이러니컬하게도 이 경험 덕분에 한국의 신용카드 사들은 엄격한 리스크 관리를 위한 보수적인 신용평가 모형을 구축할 수 있게 된 계기가 되었으며, 금융회사들과 학계에서도 연체관리 및 신용평가의 중요성을 깨닫고 많은 연구를 진행하게 되었다. 이러한 연구들의 도움으로 리만브라더스 사태('08년)와 유럽 국가 부채위기('11년)등 전세계적인 몇번의 경제적 충격에도 한국의 금융회사들은 양호한 리스크 관리를 하고 있으며, '13년이후 세계적인 재정완화 및 저금리에 기반한 풍부한 시장 유동성과 안정적인 신용관리 정책에 따라 '19년 현재 역대 최저 수준의 부실율 수준을 유지하고 있는 상황이다.

'16년3월 알파고와 바둑대국에서 이세돌 9단을 압도적으로 승리한 이벤트는 인공지능에 대한 관심으로 이어졌으며 한국에는 빅데이터에 기반한 4차혁명 및 인공지능을 앞다퉈 공부하는 열풍이 불게 되었다. 때마침 해외에서는 고성능 하드웨어 단가의 급격한 하락과 빅데이터에 기반한 인공지능 기술을 알세워 구글, 아마존 등의 온라인 플랫폼 업체가 산업을 전방위적으로 확장하고 있고 온라인 결제, 가상화폐, 블록체인등의 디지털 기술은 저렴하거나 제로 수준의 수수료로 개인들에게 다가와 전세계인의 일상 생활을 변화시키고 있다.

정통적인 금융기관인 신용카드사 또한 단순히 쌓이고 있던 비정형데이터 및 미처 보지 못한 데이터에 기반한 연구와 관심이 생기기에 이르러 본 연구자는 정통적인 신용평가 모형에서 무관심하던 영역을 발견하게 되었으며, 몇가지 신상정보를 추가하여 비금융정보만으로 부도 위험이 높은 고객을 사전에 판단하는 방법을 연구하고자 한다.

1-2 정통적인 연구와의 차별성

앞서 언급한 IMF와 카드사 사태를 통한 신용카드 평가 체계의 유래없는 발전으로 인하여, 현재 한국의 금융사들은 세계 상위 수준의 신용평가 모형을 운영중이다.

2018년 금융위는 신용정보법의 개정 추진과 아울러 빅데이터 및 신용평가에 관한 몇가지 정책을 발표하였는데, 그 중 하

나가 비금융정보를 활용한 신용평가 개선유도이다[1]. 비금융정보란 금융사에서 거래된 신용거래정보가 아니면서 신용도 판단을 위해 사용가능한 정보를 의미하는데 대안정보라는 용어로도 불리고 있다.

한국의 경우도 금융거래가 없거나 새로 금융거래를 시작하는 사회초년생들이나 금융거래가 활발히 일어나지 않는 일부 노년층의 경우 thin-filer라는 명칭으로 소극적이고 보수적인 신용평가를 받고 있는데, 이를 개선하기위해 선진국들은 부동산 정보, 통신정보, 세금 납부 정보, 전자상거래 정보나 심지어 SNS상의 정보와 같은 인성 판단 정보까지 이용하여 도입하고자 연구를 진행하고 있는 초기 단계이다[2].

한국도 CB사 주도로 부동산정보 및 통신정보 위주의 도입을 시작하고자 하나 너무나 강한 개인정보법의 장벽과 초기 고정비가 많이 투입되며 구축되고 있어, 민간기업들이 합리적인 가격 수준으로 비금융정보 서비스를 도입하여 의미있는 활용을 하기에는 시간이 다소 걸릴 것으로 예상된다.

본 연구자는 고객의 주소변경 이벤트 발생시, 기업들이 이미 갖고있는 고객의 간단한 신상정보를 간단히 혼합하여 위험고객군을 미리 예측하고자 하므로, 주로 대출상품 이용, 다중채무, 연체이력에 기반한 정통적인 신용평가 모형과는 차별화되며, 앞서 설명한 비금융정보의 요인(factor)으로써 활용성을 검토하는 연구를 진행하고자 한다.

II. 본론

2-1 선행연구

신용카드의 연체에 관한 연구는 한국보다 신용카드 도입이 빨랐던 해외의 사례를보면 Whately, Windram과 Cox(2004)는 가계소득이 연체의 주요한 원인이라고 설명을 하고 있고[3], Gross와 Souleles(2002)는 개인의 위험속성 보다는 개인의 연체 수요에 의해 영향을 받는다고도 하였다[4]. 한국의 경우는 2004년 카드 사태를 기점으로 활발히 연구되어 오고 있는데 이용금액과 한도 소진율이 클수록, 직업이 불안정 할수록 연체율이 높은 상관 관계를 보인다는 대표적인 연구가 있다[5].

연체 가계는 비연체 가계에 비해 보유 자산 규모와 소득 수준에 있어 열위를 보이고 있다는 연구도 있다[6].

종합해보면 신용 카드 연체의 주요한 요인은 대체로 소득 수준과 카드 및 카드 대출(현금서비스,카드론) 이용 규모라고 판단이 가능하며, 이미 대부분의 금융사의 신용 평가 모형에 주요한 요인으로 도입되어 있다. 그런데 카드사태 이전의 인구통계학적 정보를 이용한 연구에 따르면 여성일수록 미혼일수록 연령이 적고 비수도권일수록 신용카드 발급 갯수가 많을수록 부실화 될 가능성이 높다고 한다[7]. 최근에는 해외사례등을 참조하여 이동통신 빅데이터 정보를 활용한 개인신용평가 모형을 구축시 앞서 thin-filer로 명명된 한국의 금융거래 소외자 계층에 대한 신용우대가 가능하다고 제시되고 있어, 활용이 기대

된다.

본 연구에서는 국내 시장점유율 1위이자 신용관리력이 우수한 A카드사의 빅데이터를 통해 고객의 주소 변경 유형에 따라 신용카드 부도 위험에 있어 차이가 있음을 증명하고, 그중에서도 잠재 위험성이 높은 집단을 발굴할 수 있는 모형을 구축하고자 하므로 대체로 신용 능력 정보인 소득이나 금융 거래 이력을 이용하는 정통적인 선행 연구들과는 차별화하고자 한다.

2-2 분석데이터 및 사전정의

A카드사의 2017년7월부터 2018년6월까지 12개월간 직장주소 또는 자택주소를 변경한 고객의 변경건 588만건을 대상으로 변경이후 12개월간 추적분석을 하였다. 추적분석은 주소변경이후 12개월 부도율을 측정하기 위함이고, 부도는 비부도 상태에서 12개월이내에 연체90일 이상을 경험하거나 사전에 규제로 정의된 사유의 신용불량이 신용정보원등에 등재된 경험이 있는 경우를 의미 하는데, 현재 위험이 발생되지 않았으나 미래에 위험이 발생할 잠재위험을 측정하기 위한 주요한 리스크관리 모니터링 Tool로 쓰이고 있는 지표이다. 본 연구에서는 주소 변경 고객의 개인 신용카드의 부도율을 종속변수로 활용하며, 주소 변경을 직장 주소 변경과 자택 주소 변경 으로 나누고, 주소 변경 유형을 4가지로 분류하였다. 첫번째는 신규등록이다. 기존에 주소가 등재되지 않았다가 등재가 되는 경우이며, 대체로 신규입회되는 경우 또는 직장을 처음 구하였을 경우에 발생된다. 두번째는 읍면동 단위내 변경이다. 최근에 한국의 주소체계가 도로명 주소 체계로 변경이 되었으며 주소 상세명 및 우편번호 자릿수등의 변경이 신용카드사에 다소 지연되며 반영되는 경우가 대다수이다. 세번째는 읍면동 단위 이상의 이동이다. 도로명 주소체계의 변경과는 구분이 되며, 의미있는 생활지역의 변동 상황이 부도 위험에 영향을 더 크게 미칠 것이라 판단하였다. 네번째는 주소의 삭제이다. 주소 정보가 있던 고객의 정보가 삭제된 경우이며, 고객이 공개하기를 원치 않거나 실업등의 사유가 대부분일 것으로 판단된다. 어찌되었던 주소 정보는 고객의 중요한 식별정보이므로 전화 상담원의 실수로 인한 삭제 등의 가능성은 희박하며, 고객의 의지에 의한 결정임은 분명해보인다. 그리고 추가적인 분석들을 위해 직장명 변경여부, 직장 전화번호 변경여부, 변경전후 지역의 부동산 시세, 성별, 연령대, 직업등의 변수를 이용하고자 한다.

2-3 연구가설

본 연구에서는 카드사에서 주소 변경을 한 고객이 미래에 더 부실해지지 않은가 라는 첫번째 가설을 세워보고자 한다. 왜냐하면 앞선 선행연구에서 보듯 신용카드의 연체는 소득 수준과 상관 관계가 크므로 자택 주소의 변경의 경우 거주 수준의 변동을 의미하며, 서울 및 대도시의 거주 밀도가 높은 한국은 부동산 거주지로 소득을 유추하기에 용이하기 때문이다. 직장 주소 변경의 경우에도 지역에 따라 취업 환경을 유추하기에 그리 어

렵지 않다. 그리고 앞서 사전 정의된 4가지 유형별 고객의 부도 위험은 서로 차이가 있는가 라는 두번째 가설을 세워보고, 특히 고객의 의지가 반영된 삭제된 경우에 부도 위험이 높지 않은가에 초점을 맞추고자 한다. 주소가 변경된 경우에도 부동산 시세가 낮은 지역으로 이동시에 부도 위험이 커지는 지도 파악하고자 한다.

2-4 연구분석1: 주소변경 유형별 부도위험 분석

A카드사의 주소 변경 건수는 한해동안 총합 588만건이 발생하고 있으며, 그림1에서 보듯이 자택 주소 변경이 362만건, 직장 주소 변경이 226만건 발생하고 있다. 대체로 이사 성수기인 3~5월이 높은 편이고, 자택 주소 변경과 직장 주소 변경간의 발생 빈도 트렌드는 아주 유사한 편이다.

자택 주소 변경의 유형은 그림2와 같이 동일 읍면동 내의 변경이 46.2%로 제일 많고, 읍면동 단위 이상의 이동이 44.3%, 신규 등록이 9.4%순이며 정보 삭제의 경우는 0.1%로 거의 미미하였다. 직장 주소 변경의 경우 읍면동 단위이상 이동이 42.8%로 제일 많고, 동일 읍면동 내의 변경이 29.7%, 신규 등록이 21.0%, 정보삭제는 6.5%로 자택 주소 변경에 비해서 비중이 많은 편이다.

그림3을 살펴보면 자택 주소 변경자의 부도율은 1.4%, 직장 주소 변경자의 경우 1.5%로 관찰되었는데, 이는 알려진 카드 업계의 수준인 1%~1.5% 수준과 부합하므로, 주소 변경을 하였다는 것 만으로 전체 건의 잠재위험이 높다고 생각하긴 어렵다.

주소 변경 유형별로는 동일 읍면동내 변경이 제일 낮으나, 신규 등록 및 읍면동 이상 변경과 의미있는 차이를 보이지 않는다. 다만, 정보를 삭제한 경우는 다른 유형에 비해 2~3배 높은 수준으로 나타나고 있다.

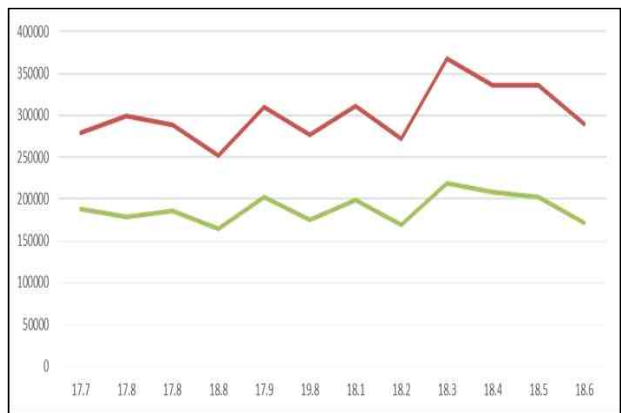


그림 1. 직장주소 변경(上)과 자택주소 변경(下)의 월별 발생 건수
Fig. 1. Monthly number of workplace(U) and home(D) address changes Trend

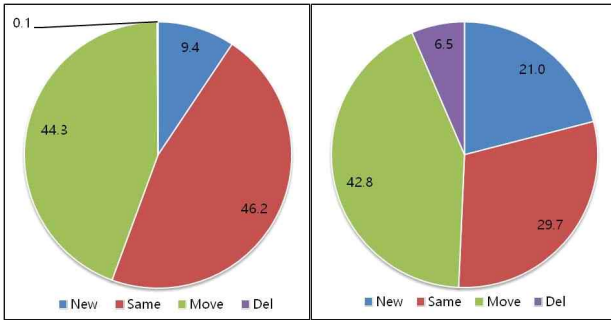


그림 2. 변경 유형별 비중 (左:주택주소, 右:직장주소)
 Fig. 2. Percentage by type of change (left:home address, right: work address)

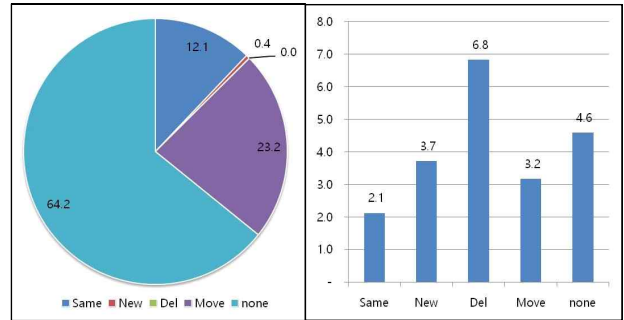


그림 4. 직장주소 삭제자의 주택주소 상세 변경 유형별 비중(左)과 부도율(右)
 Fig. 4. Percentage(left) and default rate(right) of workplace address deleters by type of home address change

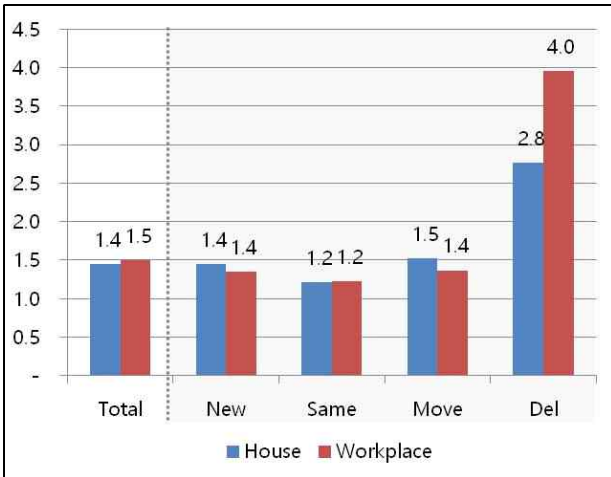


그림 3. 주택과 직장주소 상세 변경 유형별 부도율
 Fig. 3. Rate of default by type of home/work address change

이는 고객이 의지를 가지고 주소정보나 직장 정보를 삭제했으며 고객은 주거 수준의 하향 이동, 실업과 같은 상황을 겪고 있어 한도의 하향, 신규 대출 거절등의 불이익이 발생됨을 염려하여 스스로 자의적 행위를 취하였을 가능성이 높아 보인다. 더 나아가 연체가 예상되고 상환할 의지가 없을 경우 카드 이용 대금의 회수를 회피할 목적을 가졌을 개연성도 있어 보인다. 다만, 부도율은 해당 집단이 부실화될 가능성이 높다는 확률 수치 일뿐 선량한 대부분은 프라이버시의 보호를 위해 단순 삭제를 하였을 지도 모른다. 직장 주소를 삭제한 경우중 주택 주소 변경 여부를 살펴보면 그림4에서 보듯 주택 주소를 동시에 변경하지 않은 경우가 64.2%로 대부분을 차지하고 있어 대체로 취업 상태의 변화만 반영되는 경우가 많은 것으로 추정된다. 그 다음으로는 읍면동 이상의 주택주소 변경이 23.2%를 차지하고 있고, 동일 읍면동내의 변경이 12.1%, 신규 등록이 0.4%이며, 주택 주소까지 삭제한 경우는 0% 수준이다.

이 경우의 부도율을 살펴보면, 직장 주소와 주택 주소 모두 삭제한 경우가 제일 높지만 유형 발생이 거의 안되는 수준으로 현장에서 관리 효율성을 감안할 때 의미는 없어 보인다. 특이한 점은 직장 주소를 삭제했다라도 주택 주소가 동일 읍면동내 변

경된 경우는 부도율이 낮아 오히려 리스크 관리가 가능한 우량한 대상군으로 판단된다. 이 대상군은 세분화하여 변별할 경우에 의미있는 대상군으로 분류되어야 할 것으로 보인다.

주택 주소가 읍면동 이상 이동하며 변경된 경우, 거주지 주거비용의 상하향에 따라 부도위험에 있어 차이가 있는지 살펴보기위해 주소변경 이전과 이후의 지역 부동산 시세차이를 살펴 보았다. '18년6월의 K은행의 부동산 시세 정보를 활용한 동 단위 1m²당 아파트 매매가 평균의 Gap의 차이는 그림5와 같다.

그림5의 가로축의 1단위는 주소변경이후 시세-이전 시세의 m² 당 백만원 단위를 의미하며 32평형이라 불리는 약 109 m²의 경우 약 1억원 수준이다. 대부분의 경우가 -1~+1사이에 모여있어 대부분의 경우는 생활 수준이 유사한 지역으로 이동을 하고 있다고 판단되며, 가로축의 음의 방향인 시세가 낮은 지역으로 이동하더라도 양의 방향과 비교해볼 때 특이구간(-20)을 제외 하고는 부도율과 상관 관계가 보이지 않는다.

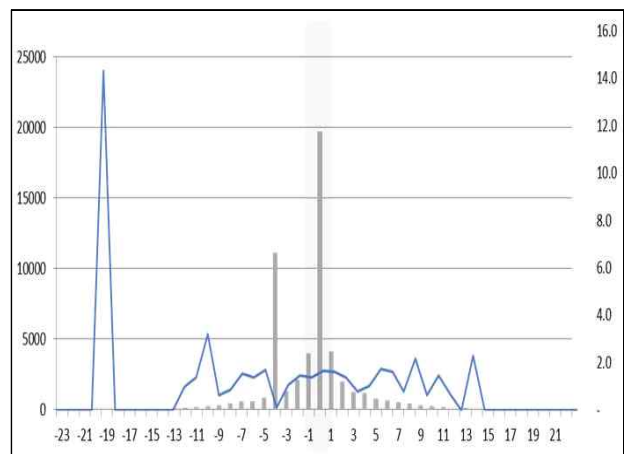


그림 5. 주택주소 읍면동이상 이동 변경자의 부동산 시세 Gap(이후-이전) 구간대별 비중(막대)과 부도율(선)
 Fig. 5. Percentage(bar) and default rate(line) by Gap (after-before) section of real estate price changeers moving from their home address to their home address

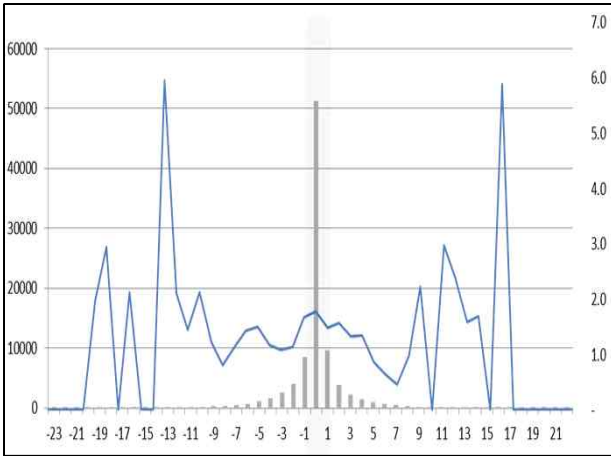


그림 6. 직장주소 읍면동이상 이동 변경자의 부동산 시세 Gap(이후-이전) 구간대별 비중(막대)과 부도율(선)
Fig. 6. Percentage(bar) and default rate(line) by Gap (after-before) section of real estate price changers moving from their home address to their office address

그림6은 동일한 방법으로 직장 주소가 읍면동 이상 이동시 시세 변화를 살펴본 것으로, 부도율과 상관 관계가 보이지 않는다는 점에서 자택 주소의 경우와 같다.

주소변경 유형에 따른 잠재 리스크는 부도율 뿐 아니라 신용등급의 이동을 통해서도 유추가 가능한데, 주소변경 6개월후 A 신용평가사의 신용 등급의 이동을 살펴보면, 그림7과 그림8에서 보듯 자택과 직장 모두 주소가 삭제된 경우의 신용 등급 하락 비중이 타유형보다 높은 편이다. 동기간 A 신용평가사가 파악하고 있는 카드 업권 전체의 하락 이동 수준인 16.4%와 비교 시에도 다소 높은 수준으로 신용 등급의 하락이 부도 위험과 상관성이 높음을 감안시 부도 위험이 높은 집단임을 알 수 있다.

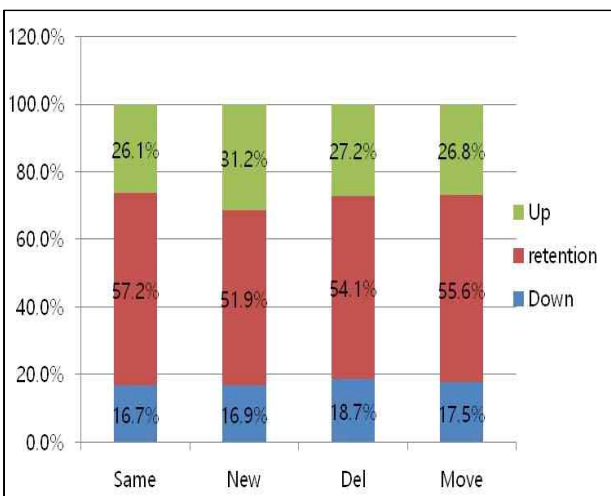


그림 7. 자택주소 변경유형별 신용등급의 이동 변화
Fig. 7. Changes in credit rating by type of home address change

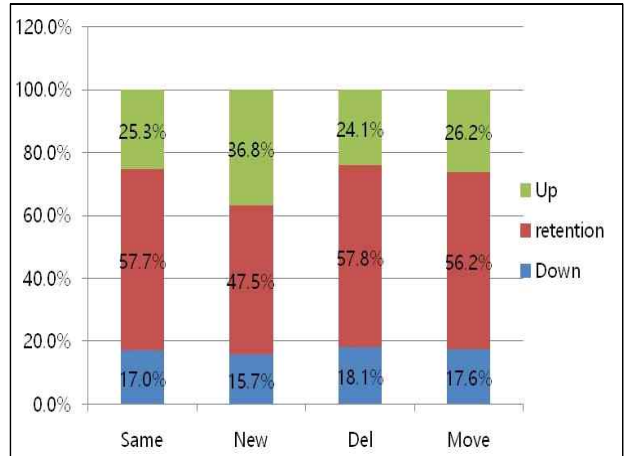


그림 8. 직장주소 변경유형별 신용등급의 이동 변화
Fig. 8. Changes in credit rating by type of work address change

2-5 연구분석2: 신상정보를 추가 활용한 고객 부도위험 분류

앞서 연구 배경 및 목적에서 설명한 바와 같이 금융거래 이력에 기반한 정통적인 신용평가 모형과 차별화 하기 위해, 금융거래 이력 및 외부 신용 평가 등급의 활용을 배제하고 간단한 rule로 부도 위험에 따라 고객군을 분류하고자 하므로, 주소 변경 정보와 신상 정보를 활용하여 살펴보고자 한다.

성별 정보를 결합하여 살펴보면, 그림9과 그림10에서 보듯 전 유형에 걸쳐 남성이 여성보다 부도율이 높은 편이고, 주소 삭제시에는 특히 높은편이다. 여기에서 한가지 특이한 점은 직장 주소를 삭제한 경우는 여성도 높지만, 자택 주소를 삭제한 경우는 여성이 다른 유형과 비슷하게 낮아 진다는 점이다.

이번에는 연령대 정보와 결합하여 살펴보면, 그림 11과 그림 12에서 보듯 삭제의 경우를 제외한 유형들에서는 의미있는 수준의 차이가 보이지는 않는 편이다.

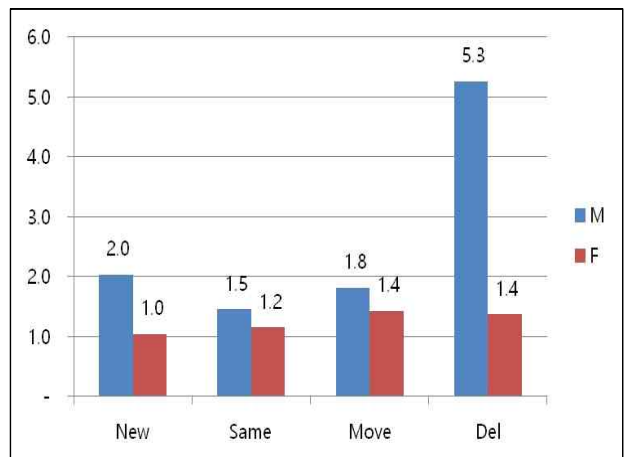


그림 9. 자택주소 변경유형과 성별 부도율
Fig. 9. Rate of default by gender group of change of home address

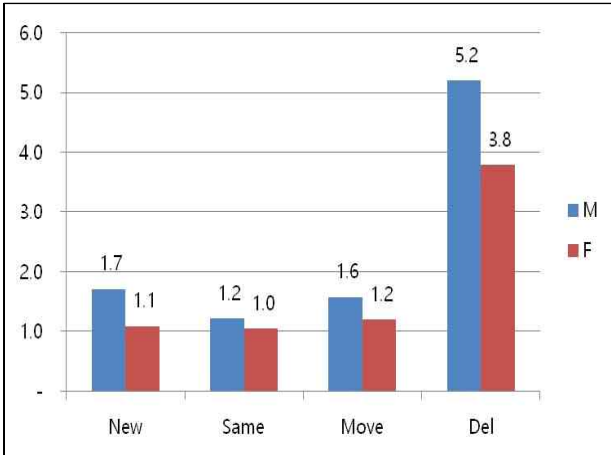


그림 10. 직장주소 변경유형과 성별 부도율
 Fig. 10. Rate of default by gender group of change of work address

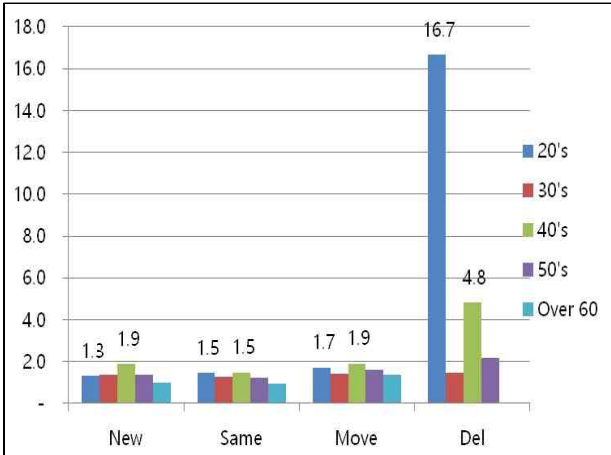


그림 11. 자택주소 변경유형과 연령대별 부도율
 Fig. 11. Rate of default by age group of change of home address

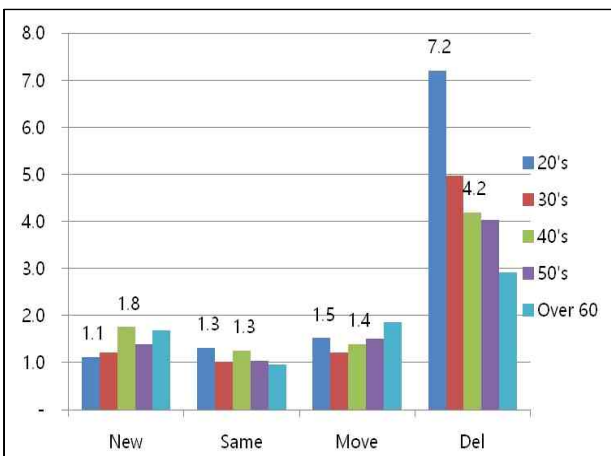


그림 12. 직장주소 변경유형과 연령대별 부도율
 Fig. 12. Rate of default by age group of change of work address

직장이든 자택이든 주소의 삭제의 경우에는 20대의 부도율이 상당히 높고, 직장 주소의 삭제는 30,40대가 자택 주소의 삭제의 경우는 40대가 그 다음 순으로 높다.

그 다음으로는 직업 정보를 결합하여 살펴보고자 한다. 고객의 직업을 8가지 그룹군으로 분류를 하였는데, 그림13와 그림 14에서 보듯 리스크 수준과 통용되는 유사한 직업 분류를 감안하여, 그룹군 A는 30대 기업, B는 자본금 3억이상의 기업, C는 자본금 3억미만 기업, D는 금융권, E는 공무원/교육/종합병원, F는 전문직/군인/학원/주부, G는 금융소독자, H는 자영업/판매/기능직/운전 이다. 그림12와 그림13을 살펴보면 주소 변경 유형들이 직업군별로 유사한 트렌드를 보이고 있는바 직업별 부도율의 차이가 주소 변경 유형보다 영향력이 큰 것으로 보인다.

직장 주소가 삭제된 경우는 타 유형보다 부도율이 높은 편이고, 특정 직업군과 결합시 부도율이 더 커지거나 부도율이 감소하여 변별력 확보에 도움이 될 것으로 판단된다.

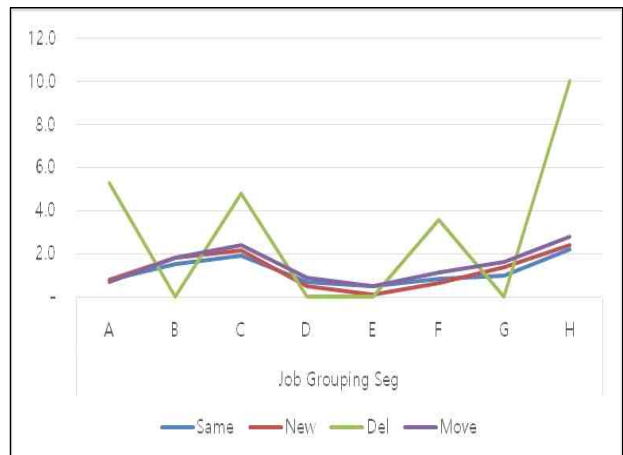


그림 13. 자택주소 변경유형과 직업군별 부도율
 Fig. 13. Rate of default by job group of change of home address

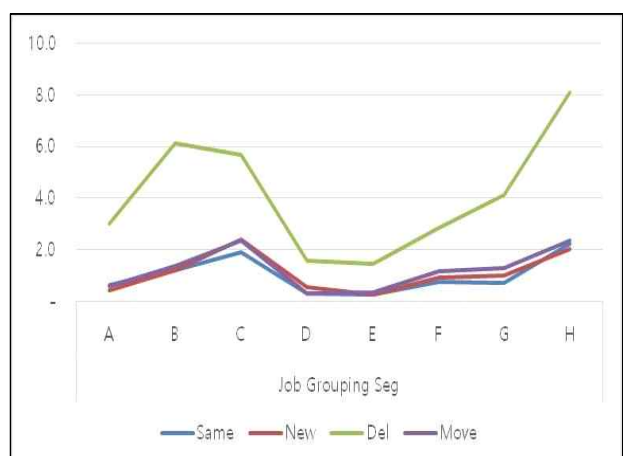


그림 14. 직장주소 변경유형과 직업군별 부도율
 Fig. 14. Rate of default by job group of change of work address

앞서 살펴본 3가지의 신상 정보를 활용하여 부도위험에 유의한 고객군을 분류하는 간단한 모형을 만들어 보고자 하였다. 모형은 카드사의 주요 분석 Tool인 SAS의 E-Miner를 이용하여 구축하였다.

표1은 모형을 구축하기위한 데이터에 대한 레이어아웃이다. A 카드사내에 존재하는 Raw형태의 변수들을 기반으로 변수 변환등을 통해 유형을 정의하고, 결측값이 없는 여부의 형태로 가공한후 모형 개발을 위한 데이터 셋을 구축하였다.

Pudo_TF를 종속 변수(Target)로 하는 변수들의 기초 통계량은 그림15와 같다.

먼저 전체 355만건의 데이터를 50:50의 비율로 분석용(Train) 데이터셋과 평가용(Test) 데이터셋으로 나눴다. 그리고 부도율은 희귀사건(Rare Event)이며 Pudo_TF=1에 해당하는 데이터의 수가 매우 적기 때문에 Pudo_TF=0인 경우와 분리하여 샘플링 하여야 하므로, 층화추출의 방법으로 그중에서도 레벨 기반의 Oversampling기법을 이용하였다. 원래 부도의 비율은 1%대에 불과하지만 본 연구자는 Pudo_TF=1 : Pudo_TF=0 이 10 : 90의 비율로 추출하도록 하였다. 즉 샘플 데이터에서 이벤트의 비율이 10%가 되게끔 하였고 분석용과 평가용에 동일하게 적용하였다.

첫번째로 분류모형에 주로 쓰이는 의사결정트리(Decision Tree)모형을 만들어 부도 가능성이 높은 고객을 판별해 보고자 한다. 의사결정트리 모델은 시인성이 좋고 변수의 개수가 적을 때 유용하게 활용 가능하며, Rule-Base기반의 간단한 모형을 구축할때 활용 가치가 높다.

그림16는 의사결정트리(Decision Tree)모형의 서브트리 평가도표이며, 오분류 최소화 관점에서는 잎의 개수가 7개일 때 최적적이고 초반에는 오분류율이 급격히 감소하지만 그뒤로는 변화가 없으므로 단순모델로도 충분하다는 것을 알 수 있다.

표 1. 모형 구축을 위한 변수 구성

Table 1. Configure Variables for Model Deployment

Variable label	Variable Description	properties
AGE_CCD	age group code	Class
HOU_cd	change of home address types	Class
OFF_cd	change of workplace address types	Class
OFF_juso_TF	change of workplace address TF	Class
OFF_name_TF	change of workplace name TF	Class
OFF_phn_TF	change of workplace phn number TF	Class
Pudo_TF	default TF	Class
HOU_af_rlstate	property prices after change home address	Numeric
OFF_af_rlstate	property prices after change workplace address	Numeric
SEX_CCD	sex group code	Class
Job_cd	jog group code	Class

Data Role	Variable Name	Role	of Levels	Missing	Mode	Mode percentage	Mode2	Mode2 percentage
TRAIN	AGE_CCD	INPUT	3	0	5	56.91	6	25.46
TRAIN	HOU_cd	INPUT	5	0	2.CON	37.08	3.MOV	33.1
TRAIN	OFF_cd	INPUT	5	0	9.N/A	50.13	3.MOV	21.93
TRAIN	OFF_juso_TF	INPUT	2	0	0	52.95	1	47.05
TRAIN	OFF_name_TF	INPUT	2	0	0	62.87	1	32.13
TRAIN	OFF_phn_TF	INPUT	2	0	0	71.81	1	28.19
TRAIN	SEX_CCD	INPUT	2	0	F	51.53	M	48.47
TRAIN	Job_cd	INPUT	9	0	C	36.16	F	15.27
TRAIN	Pudo_TF	TARGET	2	0		98.81	1	1.19

Data Role	Variable Name	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
TRAIN	HOU_af_rlstate	INPUT	373.0117	287.9192	68259	31741	0	279	2410	2.577462	9.463178
TRAIN	HOU_bf_rlstate	INPUT	381.0219	307.7165	59670	40330	0	280	2410	2.57055	9.110774
TRAIN	OFF_af_rlstate	INPUT	403.3091	335.1792	34758	65242	0	277	2410	1.979652	4.489446
TRAIN	OFF_bf_rlstate	INPUT	419.6887	353.8181	29712	70288	0	280	2410	1.822718	3.516332

그림 15. 모형 구축을 위한 변수들의 기초 통계량

Fig. 15. Basic Statistics of Variables for Model Deployment

그림17은 의사결정트리(Decision Tree)모형 결과로써, 변수 별로 나뉘어갈라진 결과를 해석해보면, 직업군 코드의 영향이 제일 크고 직장 주소의 삭제 여부, 직장 이름의 변경 여부, 성별 정도가 유의한 판별 기준이 될 것이다. 가장 변별력이 높은 rule은 직업이 30대기업이 아닌 기업체이거나 자영업/판매/기능직/운전인 경우이면서 직장의 주소와 직장명을 삭제한 경우이며 자택 주소 변경이 없거나 1m² 당 138.5백만원미만, 남자가 조금 더 나쁘다는 판단 Rule이다. 거꾸로 우량한 고객군은 직업이 상기 직업군이 아닐 경우만 분리해 내어도 전체 평균보다 낮은 부도율로 필터링 가능하다. 다만 의사결정트리(Decision Tree)모형의 가지구조의 특성상 Target에 접근하는 방향으로 가지치기가 흘러가므로, 반대의 경우(우량한 고객군 판별)는 별도의 모형화 작업이 필요할 것으로 보인다.

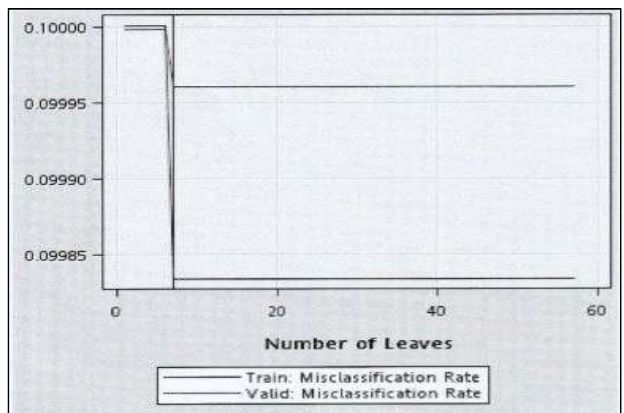


그림 16. Decision Tree 모형의 서브트리 평가도표

Fig. 16. Subtree Evaluation Chart of Decision Tree

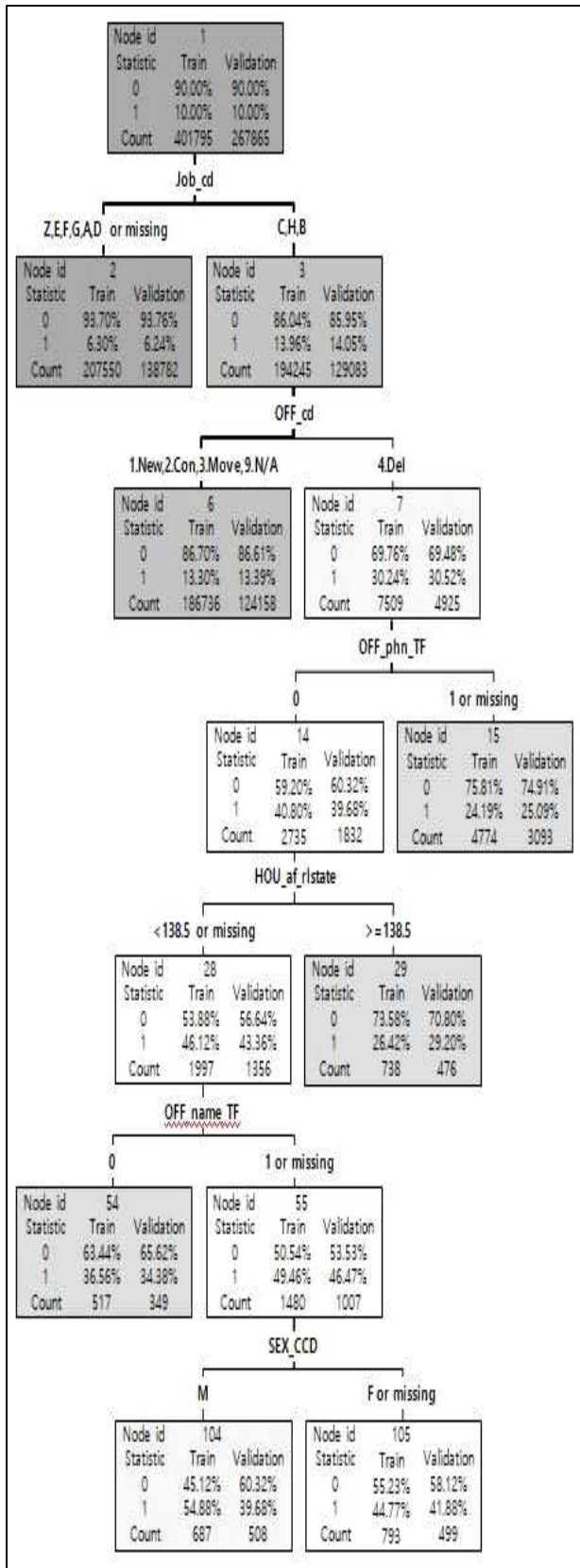


그림 17. Decision Tree 모형 결과
 Fig. 17. Decision Tree Model Result

두번째로 회귀분석 모형을 적용해 보았다. 단계적 선택법(stepwise)을 통하여 SAS의 E-miner 틀을 통하여 간편하게 산출할 수 있었으며, 그림18에서 보듯 연령대(age_ccd),직업군(job_cd),주택주소 변경유형(HOU_cd),직장주소 변경유형(OFF_cd), 직장전화 변경여부(OFF_phn_TF),주택이동 변경전후 부동산시세(hou_bf_rlstate, hou_af_rlstate), 성별(sex_ccd)의 변수로 회귀모형을 구축할 수 있었다.

마지막으로는 신경망 모형을 적용해 보았다. 역시 SAS의 E-miner를 통해 구축되었으며, 단계적 선택법에 의한 회귀분석 결과 변수들을 넣어서 모형을 정교화 하였다. 이렇게 산출한 3가지 모형을 그림20의 ROC 그래프 결과로 확인해보면 신경망모형과 회귀모형이 의사결정트리(Decision Tree)모형보다 다소 높은 위치에 있고, 그림19의 ROC index값으로도 신경망모형 0.68, 회귀모형이 0.68로 의사결정트리(Decision Tree)모형 0.62보다 다소 나은 모형으로 판단된다.

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept	1	-1.6576	0.0350	2239.16	<.0001		0.191
AGE_CCD	01	1.4842	0.0828	321.38	<.0001		4.411
AGE_CCD	02	0.0551	0.0207	7.09	0.0078		1.057
AGE_CCD	03	-0.2408	0.0192	157.08	<.0001		0.786
AGE_CCD	04	-0.2309	0.0192	144.77	<.0001		0.794
AGE_CCD	05	-0.4022	0.0201	402.05	<.0001		0.669
HOU_cd	1,NEW	-0.0942	0.0303	9.24	0.0041		0.910
HOU_cd	2,CON	-0.2772	0.0294	89.98	<.0001		0.758
HOU_cd	3,MOV	-0.0854	0.0293	8.48	0.0036		0.918
HOU_cd	4,DEL	0.5007	0.1115	20.42	<.0001		1.655
IMP_HOU_af_rlstate	1	-0.00059	0.00041	209.91	<.0001	-0.00077	0.999
IMP_HOU_bf_rlstate	1	-0.00033	0.00040	69.49	<.0001	-0.00059	1.000
OFF_cd	1,NEW	-0.2793	0.0166	294.64	<.0001		0.756
OFF_cd	2,CON	-0.2382	0.0126	369.77	<.0001		0.788
OFF_cd	3,MOV	-0.2012	0.0119	286.26	<.0001		0.818
OFF_cd	4,DEL	0.9111	0.0178	2617.25	<.0001		2.487
OFF_phn_TF	0	0.0979	0.00790	148.27	<.0001		1.102
SEX_CCD	F	-0.1473	0.00549	719.50	<.0001		0.863
job_cd	A	-0.2064	0.0227	158.89	<.0001		0.751
job_cd	B	0.4201	0.0160	685.29	<.0001		1.522
job_cd	C	0.7328	0.0119	3816.53	<.0001		2.081
job_cd	D	-0.6596	0.0500	173.98	<.0001		0.517
job_cd	E	-1.0434	0.0278	1404.75	<.0001		0.352
job_cd	F	-0.0695	0.0160	17.19	<.0001		0.930
job_cd	G	0.1614	0.0181	79.44	<.0001		1.175
job_cd	H	0.7676	0.0153	2533.28	<.0001		2.155

그림 18. 회귀분석 모형 결과
 Fig. 18. Regression model results

Statistics	Neural	Reg	Tree
Test: Kolmogorov-Smirnov Statistics	0.26	0.25	0.22
Valid : Average Squared Error	0.09	0.09	0.09
Test: ROC Index	0.68	0.68	0.62
Valid : Average Error Function	0.30	0.31	.
Test: Bin-Based Two-Way Kolmogorov-Smirnov Statistics	0.11	0.11	0.10
Test: Cumulative Reaction Detection Rate	23.30	22.45	16.54
Test: Reaction Detection Rate	9.57	9.31	6.70

그림 19. 모형들의 ROC 평가 결과
 Fig. 19. ROC Evaluation Results of Models

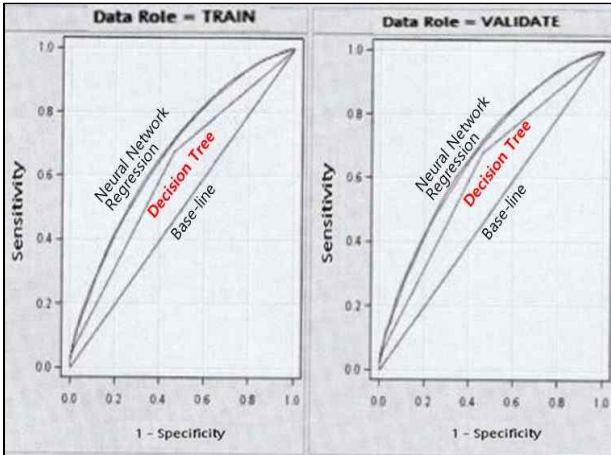


그림 20. 모형들의 ROC 곡선
 Fig. 20. ROC Evaluation Curves of Models

III. 결 론

3-1 요약 및 시사점

A카드사의 연간 주소가 변경된 588만건을 자택 주소 변경과 직장 주소 변경으로 나누고 변경유형을 신규 등록, 행정동내 변경, 이동, 삭제로 나눠서 부도율을 살펴보았다. 직장 주소를 삭제한 경우는 변경건의 6.5%이며 부도율이 타 유형에 비해 2.6 배이상 높았으며, 이중에 자택 주소가 행정동내 변경한 경우는 리스크가 낮은 우량 집단 이었다. 읍면동 이상 이동자의 경우 변경 전후 부동산 시세와 부도율의 연관성은 없었으며 주소 삭제자의 6개월후 신용 등급 하락폭이 다소 높아 리스크가 높은 집단임을 추가로 확인할 수 있었다. 이러한 주소변경 이벤트와 연령/성별/직업등의 간단한 신상정보로 부도 위험 예측 모형을 만들어 보았는데, 직업군과 주소 삭제 여부가 영향이 큰 요인이었으며 6단계 수준의 Decision Tree로 만든 간단한 Rule이 회귀 모형이나 신경망 모형에 못지않은 변별력이 있었다. 다만 적중률은 62%수준으로 통계적 수준의 모형 구축 기준으로는 다소 부족하지만 1%대에 불과한 부도 고객(Target)을 예상하고 관리하는 Pool을 구축하고, 고객을 잠재 위험 수준별로 분류하고 모니터링 및 관리하는데는 큰 무리가 없어 보인다.

3-2 한계점 인식 및 향후 활용방안

본 연구자는 기업들이 이미 갖고있는 고객의 신상 정보와 주소 변경 정보를 간단히 혼합하여 비금융정보로 고객의 부실화 위험을 미리 예측하고자 하므로 아래와 같은 잇점이 있으리라 판단된다. 첫째, 서론에서 언급한 바와 같이 한국의 신용평가 모형은 세계적인 수준으로 사설 신용평가 기관(CB사) 및 대형 금융기관이 금융거래 DB기반으로 구축한 신용평가 모형과 본 연구의 정밀도는 비교 대상이 될수는 없으나, 기존의 신용평가

Tool에 주소 변경 및 삭제에 관한 정보를 추가 활용시에 정밀도 향상에 도움이 될 것이라고 예상한다. 오히려 정교한 신용평가 모형에 비해 작동이 가볍고 간단한 rule기반이므로 기업내 활용도가 높을 것으로 판단된다. 예를들면 추가 영업을 하는 우선 순위에서 고위험 예상 고객을 후순위로 배치하거나, 주소 변경이나 삭제를 원하는 고객과 통화하는 Inbound채널의 전화상담원이 추가 확인을 통해 연체 개연성을 추가 확인하는 활동들을 할 수 있다. 둘째, 금융거래업이 아닌 업종의 산업업에 활용이 가능하다. 폭발적인 매출 신장이 일어나고 있는 렌탈업의 경우 렌탈 고객들의 렌탈료 채무 불이행여부를 사전 감지할 수 있고, 핀테크 기반의 P2P 업체들이 간편 소액 대출 거래에서 대출 연체 위험을 사전 감지하는 목적으로 활용이 가능할 것이다. 더 나아가 CB사가 없거나 신용평가 체계가 한국보다 뒤떨어진 해외 시장에 한국 금융기관이 진출시, 진출국의 현지 사정을 고려하여 기반을 만들 때 리스크 관리 요인으로써 손쉽게 활용할 것이라고 예상한다. 마지막으로 비금융정보 기반의 대안정보로써 통신정보, 공과금정보, 전자상거래정보, SNS 정보 등과 함께 정부의 금융 소외 계층에 대한 신용평가 체계 개선 정책에 기여할 것으로 기대한다.

참고문헌

- [1] Financial Service Commission, A Study on the Advancement of the Credit Information Industry for the Data Economy Available : http://www.fsc.go.kr/info/ntc_news_list.jsp.
- [2] J. Y. Kim, Development of a Personal Credit Scoring model (TELCO Score) Using Big Data in Telecommunications, pp. 1-5, Ph.D. Dissertation, Soongsil University, Seoul, 2019.
- [3] John Whitely, Richard Windram, and Prudence Cox, An empirical Model of Household Arrears, *Bank of England Working Paper*, No. 214, 2004.
- [4] D. B. Gross, and N. S. Souleles, "An empirical analysis of personal bankruptcy and delinquency", *The Review of Financial Studies*, Vol 15, No 1, pp. 319-347, 2002.
- [5] K. R. Min, H. J. Ko, J. H. Lee, and K. W. Wee, "An empirical analysis of the Korean credit card users' delinquency", *The Korean Academic Association of Business Administration*, Vol 20, No 4, pp. 1965-1966, 2007
- [6] E. Y. Lee, and E. J. Huh, "Korean households' delinquent behavior and the determinants of debt repayments", *Journal of Consumer Studies*, Vol 16, No 1, pp. 186-189, 2005
- [7] K. H. Moon, "The model for predicting credit card insolvent", *Journal of Industrial Economics and Business*, Vol 10, No 1, pp. 431-435, 1997



이태훈(Tae-Hoon Lee)

2001년 : 고려대학교 통계학과 (학사)

2019년 : 고려대학교 정보보호대학원 빅데이터응용및보안학과(석사)

2001년~2014년 : 신한카드 CRM분석팀,리스크관리팀,VIP마케팅팀

2015년~ 현재 : 신한카드 리스크관리팀

※관심분야 : 빅데이터, 신용리스크 측정 및 모니터링, 신용평가 모형,
신용정보를 활용한 리스크관리



김형중(Hyoung-Joong Kim)

1978년 : 서울대학교 전기공학과 (학사)

1986년 : 서울대학교 제어계측공학과 (공학석사)

1989년 : 서울대학교 제어계측공학과 (공학박사)

1989년~2006년 : 강원대학교 교수

2006년~ 현재 : 고려대학교 정보보호대학원 빅데이터응용및보안학과 교수

※관심분야 : 암호화폐, 블록체인, 컴퓨터보안, 패턴인식,
가역정보은익, 머신러닝, 빅데이터 분석등