

개인사업자 부도율 예측 모델에서 신용정보 특성 선택 방법

홍동숙[†] · 백한중 · 신현준

The Credit Information Feature Selection Method in Default Rate Prediction Model for Individual Businesses

Dongsuk Hong[†] · Hanjong Back · Hyunjoon Shin

ABSTRACT

In this paper, we present a deep neural network-based prediction model that processes and analyzes the corporate credit and personal credit information of individual business owners as a new method to predict the default rate of individual business more accurately. In modeling research in various fields, feature selection techniques have been actively studied as a method for improving performance, especially in predictive models including many features. In this paper, after statistical verification of macroeconomic indicators (macro variables) and credit information (micro variables), which are input variables used in the default rate prediction model, additionally, through the credit information feature selection method, the final feature set that improves prediction performance was identified. The proposed credit information feature selection method as an iterative & hybrid method that combines the filter-based and wrapper-based method builds submodels, constructs subsets by extracting important variables of the maximum performance submodels, and determines the final feature set through prediction performance analysis of the subset and the subset combined set.

Key words : credit information, individual business default rate, feature selection, deep learning, prediction performance

요약

본 논문에서는 개인사업자 부도율을 보다 정확하게 예측하기 위한 새로운 방법으로 개인사업자의 기업 신용 및 개인 신용 정보를 가공, 분석하여 입력 특성으로 활용하는 심층 신경망기반 예측 모델을 제시한다. 다양한 분야의 모델링 연구에서 특성 선택 기법은 특히 많은 특성을 포함하는 예측 모델에서 성능 개선을 위한 방법으로 활발히 연구되어 왔다. 본 논문에서는 부도율 예측 모델에 이용된 입력 변수인 거시경제지표(거시변수)와 신용정보(미시변수)에 대한 통계적 검증 이후 추가적으로 신용정보 특성 선택 방법을 통해 예측 성능을 개선하는 특성 집합을 확인할 수 있다. 제안하는 신용정보 특성 선택 방법은 통계적 검증을 수행하는 필터방법과 다수 래퍼를 결합 사용하는 반복적·하이브리드 방법으로, 서브 모델들을 구축하고 최대 성능 모델의 중요 변수를 추출하여 부분집합을 구성 한 후 부분집합과 그 결합셋에 대한 예측 성능 분석을 통해 최종 특성 집합을 결정한다.

주요어 : 신용정보, 개인사업자 부도율, 특성 선택, 딥러닝, 예측 성능

1. 서론

금융 분야는 금융 위험, 사기 활동 등을 식별하기 위해

다양한 빅데이터를 활용해야하는 영역 중 하나이며, 이러한 금융 빅데이터에 대한 예측 모델 설계는 국가 경제의 건전성을 유지하는 데 필수적이다. 빅데이터 관점에서 금융 빅데이터는 금융 거래 데이터, 기업, 개인 등 고객 정보, 시장 데이터 등 다양한 데이터 소스를 포함하고 있으며 엄청난 양의 빅데이터에 대한 효율적인 처리를 필요로 한다(Girija 등, 2019).

기업 신용 리스크로 인한 금융 파산은 사회적으로 상

Received: 26 December 2020, Revised: 10 February 2021,
Accepted: 15 February 2021

[†] Corresponding Author: Dongsuk Hong

E-mail: dshong@kcredit.or.kr

Korea Credit Information Services, Republic of Korea

당한 경제적 비용을 초래할 수 있으며 극단적인 경우 기업 부도뿐만 아니라 경제 침체로 이어질 수 있기 때문에 (박호연 등, 2018) 기업의 부도가능성을 예측하는 것은 금융 분야에서 중요한 예측 문제 중 하나이다.

이러한 부도가능성을 정확히 예측하는 것은 신용 리스크 측정을 위해 매우 중요하므로 기업 부도 예측의 정확도를 향상시키기 위한 다양한 연구가 진행되어 왔다. 이러한 연구는 예측 정확도와 효율성을 향상시키기 위한 실험 데이터의 전처리, 학습 모델 구축, 검증 등을 다루고 있으며, 본 논문은 특히 데이터의 전처리 단계에서 핵심 이슈인 특성 선택에 초점을 둔다.

기업 신용 리스크 관리에 관한 지금까지의 많은 연구는 기업 재무 정보를 중심으로 한 부도 예측에 집중되었으나, 본 논문은 첫째, 일반 기업(상장 법인)이 아니라 경기에 매우 민감한 취약 차주인 개인사업자를 대상으로 한다. 둘째, 개별 개인사업자 차주에 대한 부도 여부를 예측하는 모형이 아닌 개인사업자의 주요 업종별 부도율을 예측하는 모형을 개발하는 것을 목적으로 한다. 셋째, 연단위 부도율이 아닌 월단위 부도율을 산출한다. 넷째, 개인사업자 업종별 부도율을 예측하기 위한 특성으로 기업 재무 정보가 아닌 신용정보를 새로운 특성으로 제시한다.

본 논문에서 제안하는 신용정보 특성 선택 방법은 “AI 기반 개인사업자 업종별 부도율 예측에 관한 연구”(홍동숙 2020)를 기초로 예측 성능을 개선한 연구로, 최근 다양한 분야에서 우수성이 실증되고 있는 하이브리드 특성 선택 방법을 적용함으로써 기존 예측 성능을 개선한 신용정보 특성 집합을 제시한다.

2. 관련 연구

2.1 부도부도율 예측 연구

기업 부도 예측 모델의 첫 번째 모델은 기업의 재무 정보를 활용하여 통계 기법에 기반한 Z-score를 산출하는 Altman의 모델(Altman, 1968)이며, 이후로 판별분석, 프로빗, 생존분석 등과 같은 통계 기반 모델들이 소개되었다. 머신러닝과 딥러닝은 비교적 최신의 방법론으로 금융 및 재무 분야에서는 전통적인 통계 방법에 의한 예측연구에 비하여 연구의 양과 질 모두 부족하지만 최근 연구가 증가하는 추세이다(Hong 등, 2020).

Altman 회계 모형은 운전자본/총자산, 이익잉여금/총자산, 영업이익/총자산, 자기자본 시장가치/총부채의 장부가치, 매출액/총자산 5개 재무비율을 이용하여 부도위험 예측 판별함수를 도출하였고, Ohlson(1980)은 판별분

석 예측 모델을 개선하기 위한 로지스틱 회귀 모델을 제안하였으며, 분석 결과 기업규모, 총부채/총자산, 운전자본/총자산 등이 중요한 예측 변수로 확인되었다. Merton(1974)은 재무 정보를 이용한 회계 모형에서 벗어나 시장에서 거래되는 가격정보를 예측에 이용하는 시장기반 모형인 EDF(Expected Default Frequency model)를 제안하였으며, 이후 재무 정보에 시장 정보를 통합 분석하는 연구(이인로 등, 2015)가 소개되었다.

거시경제 변수들이 개별 기업의 신용 위험에 미치는 영향의 중요성에 대해 많은 학자들이 언급해 왔으며, 기업 부도 확률 예측 모형에 거시경제변수를 포함시켜 예측력을 높이려는 시도가 이어졌다(Chava 등, 2004). 권누리 등(2019)은 개방경제 시스템에 속한 대한민국은 국제경제 상황에 민감하게 반응하므로 개별 상장 기업들의 부도 확률 예측을 위해 거시경제 상황을 주요 변수로 고려해야 함을 제안하였고, 원재환 등(2012)은 기업 내부의 미시정보와 기업 외부의 거시정보를 통합하여 이용하는 방법을 국내 기업들에 적용하여 확인하였으며, 미시요인과 거시요인을 모두 고려할 때 정확한 기업 부도 예측이 가능하다는 것을 실증하였다.

개별 기업의 신용 위험은 해당 기업의 경영활동에 의해서만이 아니라 그 기업이 속한 산업의 상황에도 영향을 받으므로 기업 부도 예측을 위해 산업의 신용위험 분석이 필요하며(남기정 등, 2008), 금융 분야에서의 리스크 관리 강화를 위한 취약 업종 선정, 관리의 중요성 또한 커지고 있다. 그러나 산업별 신용 위험 분석에 대한 연구는 부족한 실정인데, 자료 수집의 어려움, 소비자 선호변화, 새로운 경쟁자 출현, 대체상품 등장 등으로 인해 산업 위험에 대한 계량적 접근이 쉽지 않고, 또한 산업별 신용 위험이 재무요인, 비재무요인, 거시경제환경 요인, 산업간 신용위험 전이 등 다양한 요인들에 의해 결정되기 때문에 정교한 분석이 어렵기 때문으로 알려져 있다.

이치송(2005)은 제조업, 건설업, 도매 및 소매업, 서비스업 등 4개 산업별 부도율을 구하고 VAR 모형을 이용하여 거시경제변수의 관계를 분석한 결과 제조업에서 설명력이 가장 높은 변수는 실업률이고 서비스업에서는 경기종합지수임을 제시하였다. 김한균(2000)의 논문은 기업이 속한 산업의 평균 부실정도를 예측하고 산업간에 상대적 비교를 제시하였다. 김창배 등(2008) 및 남주하 등(2008)은 PCA를 통해 예측력이 향상될 수 있다는 점에 착안해 18개 거시경제변수의 주성분과 총 21개 산업별 부도율 간 회귀분석을 한 결과, 실업률, KOSPI, 주택가격, 원화가치, 신용스프레드 등이 산업별 부도율에 영향

을 주는 것을 확인하였다. 박근영 등(2013)은 자동차부품 제조업을 대상으로 재무 데이터 없이 선행적으로 산업의 부도위험을 예측하기 위한 방법론으로 전후방산업의 위험 영향을 적용하였다. 그 외 Tong(2014)은 고수의 기업에 대한 월별 부도율을 예측하기 위한 회귀 모델을 제안하였다.

앞서 언급한 기존 연구를 포함하여 지금까지의 부도 예측 연구는 대부분 상장 기업을 대상으로 하고 있으며, 본 논문에서 예측하고자 하는 대상인 개인사업자에 대한 연구는 드물다. 개인사업자는 부가치세와 소득세 납부 의무가 있는 사업자로, 경기 변동에 민감하고, 기업이면서 동시에 개인이며, 여러 가지 측면에서 일반 상장 기업과는 상이한 특성을 가진다. 따라서 이를 고려하여 종합적이고 정교한 예측 분석이 필요하지만 일반 법인과 달리 기업 정보 등 관련 정보를 확보하기 어렵다. 본 논문은 재무 정보 등 부도 예측을 위해 필요한 기업 정보를 확보하기 어려운 영세 취약 차주(개인사업자)의 부도율 예측을 위해 신용정보를 가공한 새로운 특성을 제시하고자 한다.

개인사업자와 유사한 특성을 갖는 취약 차주로 소상공인, 창업기업, 자영업자 등에 대한 연구에서 실증 분석된 특성을 살펴본다. 먼저, 윤상용 등(2016)은 소상공인의 자금조달 환경이 중·대기업에 비해 상대적으로 열악하여 경영자 개인의 자금여력에 상당부분 의존하고 있다는 점, 소상공인의 재무적 정보와 비재무적 정보의 획득이 쉽지 않은 점, 매출 장부에 대한 신뢰성을 보장하기 어려운 점 등을 언급하였고, 개인CB등급이 소상공인 신용위험에 유용함을 실증하였다. 박주완 등(2017)은 소상공인 신용평가모형에서 로지스틱회귀모형, 의사결정나무모형, 신경망모형 등 기계학습 기법 적용 가능성을 확인하였고, 대출 총 기관수, 채무불이행 등록 총건수, 1년 내 연체 총 건수, 미상환대출총금액 중 캐피탈&저축은행업권 비중 등이 모형 구축에 이용되었다. 윤경영 등(2013)은 거시경제변수가 소상공인 신용위험에 미치는 영향을 살펴보고, 환율, 종합주가지수가 높으면 신용보증사고율은 감소하고 실업률, 신용스프레드가 높으면 신용보증사고율이 증가하는 것을 실증하였다.

그 외 연구로 남기정 등(2019)은 설립 5년 미만의 창업기업을 대상으로 부실에 영향을 미치는 비재무정보를 분석하였다. 비재무정보란 재무제표에 나타나지 않는 기업 내부의 계량 정보와 시장의 외부정보, 산업정보, 창업자에 대한 정보를 포함한 기업과 관련된 모든 정성적(텍스트) 정보를 포함하는 개념으로, 저자는 회계정보의 수

집이 어렵거나 정보의 신뢰성이 낮은 창업기업의 부실 예측에서 비재무정보 활용의 중요성을 언급하였다.

윤병우 등(2020)은 자영업 가구의 채무불이행을 설명하기 위한 요인으로, 신용활동 특성과 함께 인구·사회학적 특성, 고용·소득·주거 특성 등을 고려하였으며, 분석 결과 유용한 신용활동 특성으로 부채/자산 비율, 신용대출 및 제2금융권 대출 비중을 제시하였다.

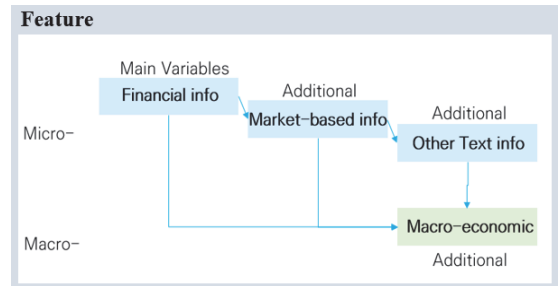


Fig. 1. The feature categorization of existing research

본 논문에서는 기업(상장기업, 소상공인, 창업기업, 자영업 등 포함)과 산업의 신용위험 연구에서 살펴본 다양한 거시경제지표를 거시변수 후보로 정의하여 이러한 거시경제지표가 개인사업자 부도율 예측에도 유용한 특성인지 살펴본다. 한편, 본 논문의 예측 모형은 한국신용정보원(Korea Credit Information Services)이 집중, 관리, 제공하고 있는 신용정보(기업 신용정보 및 개인 신용정보 포함)를 기존의 재무정보를 대체할 주요 변수로 정의하고 어떠한 신용정보 특성이 예측에 유용한지 식별하고자 한다.

2.2 특성 선택

금융 빅데이터를 포함하여 빅데이터를 활용하는 대부분의 경우에서 예측과 연관성이 없거나 적은 데이터 혹은 중복적인 데이터를 제거하는 이슈는 모델의 성능과 신뢰도에 직접적인 영향을 미친다. 일반적으로 특성의 개수를 차원의 크기라고 하는데 차원이 증가하면 그것을 표현하기 위한 데이터 양이 기하급수적으로 증가하며 지나친 고차원의 데이터들은 의미를 제대로 표현하기 어렵다. 따라서, 특성이 너무 많은 경우 더 좋은 특성만 가지고 학습에 사용하기 위해 데이터의 의미를 제대로 표현하는 특성을 추려내야 하는데 이를 차원 축소라 한다. 차원 축소를 하는 방법은 기존 특성의 부분 집합을 유지하는 특성 선택(feature selection)과 기존 특성으로부터 새로운 특성을 생성해 내는 특성 추출(feature extraction)이 있다.

특성 선택은 모델의 정확도를 향상시키기 위해, 데이터에서 가장 좋은 성능을 보여 줄 수 있는 특성의 부분집합을 찾는 것을 뜻하며(김경륜 등, 2020), 이를 통해 계산 시간을 줄일 수 있고, 유효하지 않은 특성들에 의한 기계 학습 성능 저하를 방지할 수 있으므로 모델의 성능과 신뢰도를 높일 수 있다(이상근 등, 2004; Li 등, 2018).

특성 선택은 패턴인식, 기계학습, 데이터 마이닝, 텍스트 분류, 이미지 검색, 침입 탐지 및 계능 분석과 같은 다양한 분야에서 널리 적용되고 있다(서재현 2018). 이러한 특성 선택은 데이터 확보 비용을 절감하고, 변수 간 다중 공선성을 제거하며 과적합 문제를 해소하여 일반화 성능을 개선할 뿐만 아니라 학습시간과 예측 시간을 단축하기 위한 목적을 가진다(고우석 등, 2019; 김경륜 등, 2020; 이대범 등, 2019; Li 등, 2018). 특성 선택에 중요한 것은 특성 부분집합 생성 알고리즘과 그에 대한 평가이며, 특성 선택 프로세스는 일반적으로 부분집합의 생성, 평가, (조건에 의한) 종료, 결과 검증의 4가지 단계를 포함한다(Girija 등, 2019).

특성 선택 방법은 크게 필터 기반(filter-based), 래퍼 기반(wrapper-based) 방법이 있고, 이들을 결합한 하이브리드(hybrid) 방법과 반복적(iterative) 방법 등이 있다.

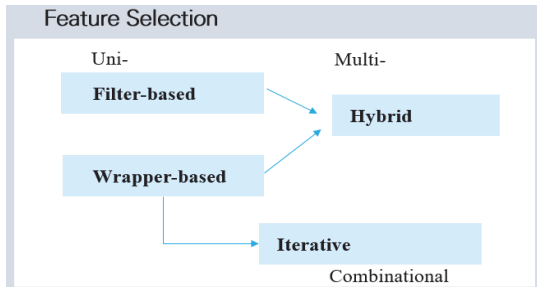


Fig. 2. The feature selection methods of existing research

필터 기반 방법은 특성의 부분집합에 대한 평가 기준이 독립적인 방법으로 어떤 특성을 어떠한 방법으로 평가하는지에 따라 실제 성능이 상이해질 수 있으며 평가 시간은 빠르나 실제 모델에서의 성능에서는 큰 차이를 보일 수 있는 단점이 있다. 래퍼 기반 방법은 특성의 부분집합에 대한 평가에 종속적인 방법으로, 예측 성능을 기반으로 해당 부분집합의 상대적 유용성을 평가한다. 즉, 특성 선택 알고리즘을 호출하여 각 특성의 부분집합을 평가하는 서브 루틴(서브 모델)을 사용한다. 평가에 걸리는 시간이 길고 특히 특성의 개수가 많아질수록 유용성을 계산하는데 필요한 시간이 늘어나는 단점이 있다.

나 실제 모델에서의 성능이 우수한 특성집합을 찾는 데 유리하다(서재현 2018; 어균선 등, 2019; Kohavi 등, 1997).

하이브리드 방법은 필터 방법과 래퍼 방법을 결합하여 사용한다. 일반적으로 통계적 유의성 검증을 하는 필터 기반으로 일차 선택을 한 후, 학습 모델을 이용하는 래퍼 기반 선택을 이차적으로 수행한다. 반복적 방법은 단일 모델 기반의 래퍼 기반 방식을 확장하여 다수 모델을 반복적으로 이용하여 특성의 부분집합을 선택한다. 계산 비용이 큰 단점을 가진다.

Girija 등(2019)은 IG(Information Gain)를 기반으로 하는 서브모듈 최적화 함수를 제안하였고, 고우석 등(2019)은 여러 후보 변수 중 상관성(correlation-based)을 기반으로 예측 모델에 사용될 대표적인 변수를 추출하였다. 홍승현 등(2003)은 부도예측 모형의 최적 입력변수 선정을 위해 인공지능망을 적용하였고 이를 필터 기반 선택 방법의 성능과 비교하였다.

이재식 등(2008)은 FeaSUA(래퍼)와 LVF(필터)를 결합하여 새로운 하이브리드 특성 선택 방법을 제안하였고, 박호연 등(2018)에서는 분산분석을 통해 일차적으로 선정한 피쳐 서브셋 중 시뮬레이티드 어니일링 기법을 통해 최종 특성을 결정한다. 이대범 등(2019)의 연구에서는 다양한 알고리즘을 이용하여 특성 부분집합을 추출하고 가장 우수한 부분집합을 추출한 알고리즘을 제시하였는데, 이를 평가하는 알고리즘으로 SVM, k-NN, 의사결정 트리를 사용하였다. 김경륜 등(2020)은 IG 필터와 래퍼로서 MLP 모델을 결합하여 활용한 하이브리드 방법을 채택하였다. 그 밖에 객체 검출 분야에서 효율적인 특성 선택을 위하여 정병우 등(2013)은 특성의 형태와 특성간의 거리로부터 특성의 유사도를 계산하고 이미 선택된 특성과 유사도가 큰 특성들을 제거하는 개선된 방법을 제안하였고, Nguyen 등(2010)은 침입탐지시스템 분야에서 효율적인 특성 선택을 위해 CFS(Correlation Feature Selection) 방법을 제안하였다.

3. 제안 방법

본 논문에서는 개인사업자 부도율을 보다 정확하게 예측하기 위한 새로운 방법으로 거시경제지표(거시변수)와 개인사업자의 기업 신용 및 대표자 개인 신용정보(미시변수)를 가공, 분석하여 입력 특성으로 활용하는 DNN 기반 예측 모델을 제시한다. 특히 미시변수인 최종 신용정보 특성을 선택하기 위하여 통계적 검증을 수행하는

필터방법과 최대 성능 특성집합을 선택하기 위한 다수 래퍼(서브 모델)를 결합 사용하는 반복적·하이브리드 방법의 새로운 특성 선택 방법을 제안한다.

3.1 실험 데이터

본 실험에서 사용되는 데이터셋은 국내 제조 분야 개인사업자(약 45만 차주)의 대출, 연체 현황 등과 같은 신용정보, 거시경제 정보, 부도율을 포함한다. 2014년 1월부터 2019년 6월까지의 데이터를 분석하였고 신용정보는 약 170종, 거시경제지표는 33종을 활용하였다. 본 논문에서 부도는 신용리스크의 예측 관리 목적에 따라 어음교환업무규약 상 부도 뿐만 아니라 금융감독원(신BIS 협약 기준) 내부등급법의 부도를 포함하는 개념으로, 가계수표/당좌수표/약속어음 최종부도, 채무불이행, 파산면책, 대출금 연체, 지급보증대지급 등을 포괄한다. 최종 모델의 종속변수인 부도율은 신용보유 전체 차주 중 부도 차주의 월단위 비율이며, 특성 선택 알고리즘에서 설명될 서브모델의 종속변수는 개별 차주의 부도 여부이다.

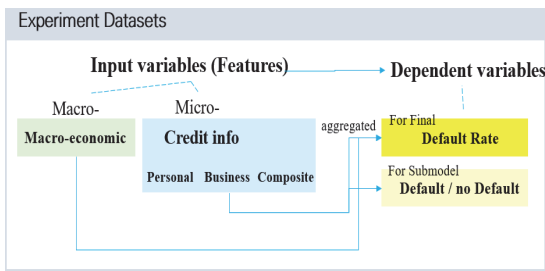


Fig. 3. Experiment Datasets

3.2 초기 연구 : AI기반 개인사업자 업종별 부도율 예측

본 논문에서 제안하는 신용정보 특성 선택 방법은 초기 분석 결과를 공개한 “AI기반 개인사업자 업종별 부도율 예측에 관한 연구”(홍동숙 2020, 이하 초기 연구)를 기반으로 하므로 본 절에서 대략적으로 소개한다.

본 연구는 개인사업자의 부도율 예측 시 신용정보의 유용성을 검증하고, 머신러닝과 딥러닝의 예측 성능을 비교함으로써 개인사업자 부도율 예측 모형에서 신용정보와 AI 기법의 적용 가능성을 살펴보기 위한 것으로 전반적인 연구 절차의 흐름도는 Fig.4와 같다.

(1) 전문가 선정과 문헌 조사 등을 통해 도출 한 초기 후보 특성집합으로부터 (2) 첫 번째 단계(1st Step)의 통계적 검증을 거쳐 일부 특성을 선택하고, (3-4) 이차(2nd Step) 특성 선택을 통해 개인사업자 부도율 예측에

유용한 최종 특성을 선정하여 (5) 이를 입력 변수로 하는 AI 기반 모형의 성능 평가를 통해 연구 결과를 검증한다. 여기서 특성 선택은 신용정보 특성을 개인신용, 기업신용, 복합신용, 그리고 이들의 조합인 개인+기업신용, 개인+복합신용, 기업+복합신용으로 총 6종의 서브셋으로 분류하여 이들 중 최대 성능을 나타내는 중요변수를 추출하는 것이 특징이다. 이는 부도율 예측 문제에서 신용정보 변수의 적용 가능성을 실험한 초기 연구로, (4)단계에서 최종 선택된 입력 특성을 고정한 후 Ridge 기초 모형을 생성하여 기초 모형을 여타 알고리즘(Random Forest, SVR(Support Vector Regression), DNN(Deep Neural Network; 심층신경망), Lasso, Dropout 등)과 비교함에 따라 알고리즘 변화에 따른 성능 변화를 확인하기 위해 기초 모형에서의 Ridge를 특성 선택 시 래퍼로 채택하였다.

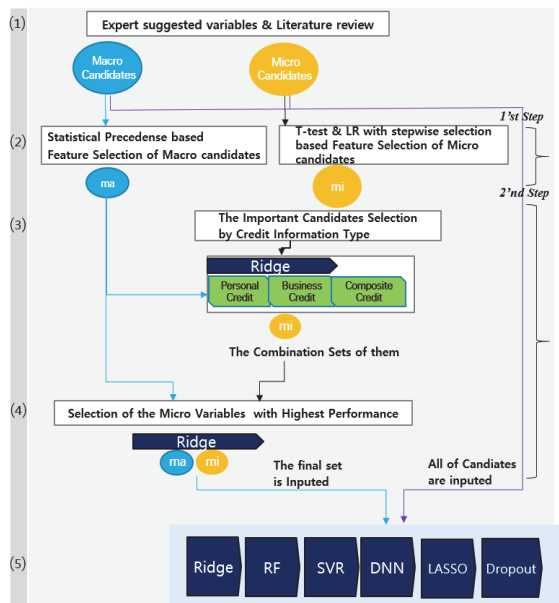


Fig. 4. Flow diagram of the earlier work

주요 분석 결과는 첫째, 신용정보 특성을 거시 변수와 함께 사용할 때 예측 성능이 향상되었고, 신용정보 유형 별로 볼 때 개인이나 기업신용정보 보다는 이를 가공한 복합신용정보가 가장 유용하며, 다음으로 개인신용, 기업신용 순으로 유용성이 검증되었다. 둘째, 대표적 AI 기법들 중 DNN, SVR, Lasso 순으로 예측 성능이 우수한 것으로 실증되어 개인사업자 부도율 예측 문제에 있어서도 여타 머신러닝 기법보다 DNN이 효과적임을 알 수 있었다. 셋째, 특성 선택과정을 생략하여 전체 후보 특성집합

을 대상으로 모형별로 실험한 결과 최종 선택된 특성 부분집합에 비해 전반적으로 성능이 저하됨을 확인한 바, AI 예측 연구에 있어서 정제된 양질의 특성을 선택하는 것이 성능에 영향을 준다는 것을 반증하였다.

3.3 특성 선택 알고리즘

개인사업자 부도율 예측 성능을 개선하기 위하여 초기 연구에서의 특성 선택 알고리즘(이차 단계)을 개정하고, 실험을 위한 최종 모델은 초기 연구에서 가장 우수한 성능을 보였던 DNN을 적용하였다. Fig.5는 개정된 특성 선택 알고리즘에 따라 추가, 수정된 사항(서브모델과 최종 변수선택 방법)을 전체 연구절차 상에 나타낸다(보라색 점선 박스 부분에 해당).

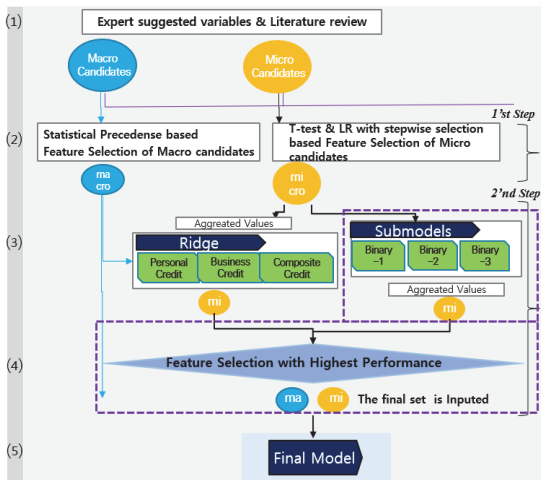


Fig. 5. Flow diagram of proposed method

제안하는 특성 선택 방법은 첫째, 최대 성능을 나타내는 특성을 결정하기 위하여 초기 연구에서 활용한 Ridge 외의 서브모델들을 래퍼로 추가 채택하였고, 둘째, 업종 단위 부도율 예측을 위한 특성을 찾기 위해 개별 차주의 부도 예측에 유용한 특성을 추가적으로 분석하여 최종 변수에 결합시킨다. Fig.6에서 세부 내용을 살펴보고자 한다.

먼저 거시변수의 경우 초기 후보 특성집합 중 부도율에 대한 교차상관과 시차상관의 유의한 통계적 관계가 있는 특성을 선택한다. 선택 기준은 시차상관계수는 +/- 0.3 이상이며 시차상관계수, 교차상관계수의 상관 방향성이 경제적 이론을 근거로 추정된 가설과 일치하는가이다. 각 거시변수에 대한 시차상관 분석을 통해 통계적으로 선행성이 확인된 거시변수의 시차를 예측 시 적용하였으

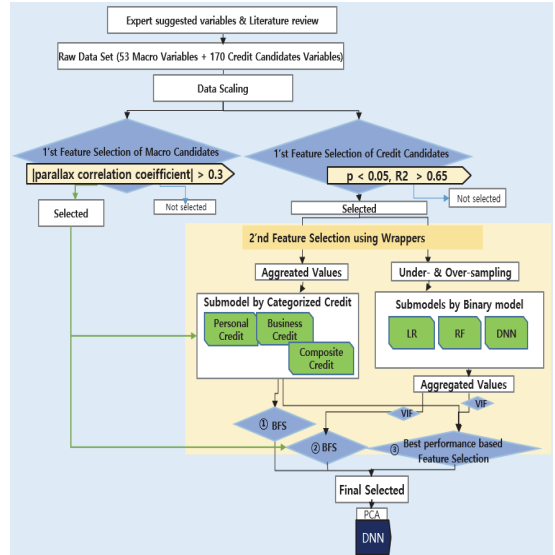


Fig. 6. The proposed algorithm

며 미시변수는 종속변수 대비 3개월 전 변수로 일괄 적용하였다. 즉, 신용정보 변수로 부터 3개월 후 부도율 예측을 가정한다.

미시변수는 독립표본 T-test와 단계적 선택에 의한 Logistic Regression(이하 LR)에 의해 1차 선택을 한 후, 2단계에서 4개의 서브모델을 기반으로 최대성능 특성 부분집합을 최종 특성으로 선택한다. 1차 선택 기준은 p값 0.05 이하, R2값 0.65 이상이며 2단계에서는 1단계에서 선택된 신용정보의 월별 집계값(평균 및 총계)을 개인-기업-복합 조합 서브셋 6종으로 분류하고 이들의 중요변수를 선택한다(첫 번째 서브모델). 중요변수 선정을 위해 학습된 모델에서 데이터셋의 특정 변수 값의 순서를 임의로 뒤섞을 때 모델의 예측 결과에 미치는 영향력을 변수 중요도 수치로 산출하여 변수의 랭킹 순위를 계산한다. 첫 번째 서브모델의 6종 서브셋 중 최대 성능 중요변수를 ① 특성집합이라 한다.

다음은 개별 차주 부도 예측 시 유용한 특성을 추출하기 위한 LR(두 번째 서브모델), Random Forest(이하 RF, 세 번째 서브모델), DNN(네 번째 서브모델) 래퍼를 생성한다. 이 때 데이터 불균형 문제를 해결하기 위해 언더샘플링 및 오버샘플링 전처리를 수행한다. 각 서브모델로부터 중요변수 3종을 월별 집계값으로 가공 후 다중공선성이 일정 수준보다 높음(VIF 값 5이상) 특성을 삭제한 부분 집합을 거시변수와 통합하여 ②-1 특성집합으로 한다. 이들 3종의 ②-1 특성집합 중 최대 성능 특성을 ② 특

성집합이라 한다.

① 특성집합 즉, 신용정보 유형별 최대 성능 특성집합에 ②-1 특성집합 즉, 개별 차주 부도 예측 서브모델별 중요변수를 결합한다. 결합 조건은 ① 특성집합을 기준으로 ②-1 특성집합의 특성들을 변수 중요도 순위에 따라 하나씩 추가해가며 성능 개선에 기여하는 특성을 추가하는 것이다. 따라서 ① 특성집합과 ②-1 특성집합의 결합을 통해 최대성능 특성집합(③ 특성집합)을 선택할 수 있다. 마지막으로 3종의 특성집합 중 성능이 가장 우수한 최종 특성집합을 선택한다.

4. 예측 성능 분석 결과

본 연구에서는 기존 연구와 달리 개인사업자의 고유한 특성을 반영하여 업종별·월별 부도율을 보다 정확하게 예측하기 위한 새로운 신용정보 특성 선택 방법을 제안하며 제조업 약 13.5만 개인사업자 데이터를 대상으로 실증 분석함으로써 제안방법과 기존 방법을 비교하고 신용정보 미시변수를 포함하는 최대 성능 특성 집합을 제시한다.

4.1 실험 환경

본 실험을 위해 Anaconda 2020.02, Scikit learn 0.22.1, Python 3.7.6, Tensorflow 2.3.0, Keras 2.4.3 등의 오픈 소스 패키지를 활용하였다. 개인사업자 데이터는 일반 기업 데이터와 마찬가지로 특정 범주의 빈도가 다른 범주에 비해 과도하게 높은 데이터 불균형 특성을 지닌다. 부도 기업은 정상 기업에 비해 상당히 적은 빈도(전체 데이터셋의 1% 미만)를 나타내므로 이로 인한 왜곡과 과적합을 피하기 위해 부도와 정상 차주의 비율이 1:1이 될 수 있도록 언더샘플링과 오버샘플링을 혼합 적용하는 것이 필요하였다. 다수 클래스인 정상 차주에 대해서는 Random Under Sampling(RUS)을 적용하고 소수 클래스인 부도 차주에 대해서는 ADASYN(ADaptive SYNthetic sampling approach) 오버샘플링을 적용함으로써 부도와 정상 차주의 비율을 1:1로 조정하였으며 결과적으로 실험 데이터는 45만 전체 차주의 약 30%에 해당되는 13.5만 차주를 포함한다.

4.2 실험 방법

다양한 특성집합에 대한 예측 성능 평가를 위하여 데이터를 여러 번 반복해서 나누어 모델을 학습하여 각 모델을 평가하여 나온 결과를 측정한다. 먼저 훈련 데이터와 시험 데이터를 분리(75:25)하고 교차 검증 시 훈련테

이터를 다시 학습 데이터와 검증 데이터로 분리한다. 학습을 위한 데이터는 훈련 데이터의 80%로 전체 데이터의 60%, 검증 데이터는 교차 검증을 통해 최적의 파라미터를 찾기 위한 데이터로 전체 데이터의 15%에 해당된다. 테스트 데이터는 새로운 데이터에 대한 예측력을 평가하기 위한 시험 데이터로 전체데이터의 25%이다.

최대 성능 특성집합을 결정하고 예측 성능을 평가하기 위해 싸이킷런에서 제공하는 GridSearchCV를 활용하여 서브 모델 및 최종 모델의 교차검증을 수행한다. 교차 검증에 의해 반복적으로 측정된 성능의 평균이 최대성능을 갖는데 활용된다. 동 분석에서는 분할되는 폴드 수 설정(cv=5)에 따라 5번씩 반복 검증을 수행하였다.

또한 일반화 성능을 최대로 높여주는 하이퍼파라미터 값을 찾는 과정으로 모델 별로 가능한 하이퍼파라미터 조합을 시도한다. Table1은 실험에서 설정한 하이퍼파라미터의 튜닝 범위를 제시한다. Ridge 모델의 알파값, LR의 C값과 반복 횟수, RF의 결정트리 개수, 최대 깊이, DNN의 은닉층 노드구성, 활성화 함수, 알파값, 학습률, 솔버, Lasso의 알파값 등을 설정하였으며, 공통적으로는 PCA 차원 축소를 위한 차원수를 설정하였다.

Table 1. Hyper-parameters for models (in python)

Model	Hyper-parameters
Ridge	‘pca_n_components’: range(1, 36), ‘model_alpha’: [0.001, 0.01, 0.1, 1.0, 10, 100]
LR	‘pca_n_components’: range(1, 10), ‘model_C’: [0.01, 0.1, 1.0, 10], ‘model_max_iter’: [100, 10000]
RF	‘pca_n_components’: range(1, 10), ‘model_n_estimators’: [100, 200], ‘model_max_depth’: [6, 8, 10, 12]
DNN (Submodel)	‘pca_n_components’: range(1, 16), ‘model_hidden_layer_sizes’: [(5,5)], ‘model_activation’: [‘relu’], ‘model_alpha’: [0.0001, 0.05], ‘model_learning_rate’: [‘constant’, ‘adaptive’], ‘model_solver’: [‘adam’], ‘model_max_iter’: [10000,100000]
DNN	‘pca_n_components’: range(1, 16), ‘model_hidden_layer_sizes’: [(5,5,5)], ‘model_activation’: [‘relu’, ‘tanh’, ‘logistic’], ‘model_alpha’: [0.0001, 0.05], ‘model_learning_rate’: [‘constant’, ‘adaptive’], ‘model_solver’: [‘adam’], ‘model_max_iter’: [100, 1000, 10000, 100000]
Lasso	‘pca_n_components’: range(1, 9), ‘model_alpha’: [0.01, 0.1, 1.0, 10., 100], ‘model_max_iter’: [10000, 100000]

Table 2. Input feature type and model for experiments

No.	Input Feature Type	Model	
		Feature selection	Final
1	All Candidates	-	DNN
2		Lasso	Lasso
3	Macro + Micro(correlation-based selected)	Ridge	DNN
4	Macro + Micro(wrapper-based selected)	Ridge	Ridge
5		DNN	DNN
6	Macro + Micro(Step1-based selected)	Ridge	DNN
7		DNN	DNN
8	Macro + Micro(① feature-sets)	Ridge	DNN
9		Lasso	Lasso
10	Macro + Micro(② feature-sets)	Ridge	DNN
11		Lasso	Lasso
12	Macro + Micro(③ feature-sets)	Ridge	DNN
13		Lasso	Lasso

Table2의 실험 조건은 입력 변수의 종류에 따라, 그리고 특성 선택 시 활용된 알고리즘에 따라 분류, 정의한다. 특성집합의 종류에 따라 전체 후보 특성집합을 대상으로 하는 경우(1~2)와 1차 선택된 거시변수에 미시변수를 추가한 경우로 분류하고, 특성 부분집합의 성능 평가를 위한 알고리즘(8~13)은 Ridge와 Lasso로 분류한다.

(8~13)의 실험은 2단계 특성 선택의 ① 특성집합, ② 특성집합, ③ 특성집합에 따른 예측 성능을 평가하는 것이며, 그 밖에 필터 방법(3), 단순 래퍼 방법(4-5), 1단계 특성 선택(6~7)에 따른 비교 실험을 실시한다. 한편, 최종 모델의 검증은 초기 연구에서 가장 우수한 성능을 보였던 DNN으로 고정하였으나, 기존 방법 중 (2, 9, 11, 13) Lasso에 의한 특성 선택, (4) 래퍼 방식을 제안 방법과 비교하기 위해 각각의 경우에는 학습 모델에서 활용된 Lasso, Ridge를 최종 모델로 추가 실험하였다.

4.3 실험 결과 및 최종 특성집합

제안 방법을 검증하기 위하여 훈련 데이터셋에서의 R-sqaure(결정계수), 시험 데이터셋에서의 R-sqaure와 MSE(평균오차)를 성능 지표로 산출하였고 Table3에서 시험 데이터셋에서의 R-sqaure와 MSE값을 기준으로 상위 3개 순위를 기술하였다. 세부적으로는 전체 후보 특성 집합을 모두 입력한 경우 혹은 전체 후보 특성집합에서 Lasso에 의해 선택된 특성집합을 입력한 경우 보다 본 연

구에서 제안한 1차 선택된 거시변수에 미시변수를 추가 입력한 경우 성능에 개선이 있는 것을 알 수 있다. 본 연구에서 제안한 신용정보 특성 선택방법에 따라 생성한 ③ 특성집합의 성능이 가장 우수하였고 최대 성능 특성 집합으로 선택되었다. 한편 DNN 기반 래퍼는 MSE 평균 오차가 크게 나타났고 Lasso-Lasso 모델은 Ridge-DNN 모델과 비교해 다소 낮은 성능을 나타냈다.

Table 3. Evaluation of the model performance

No.	Performance			Top Rank (testsets)	
	R-sqaure (for trainingsets)	R-sqaure (for testsets)	MSE (for testsets)	by R-sqaure	by MSE
1	0.957	0.964	3.409	-	-
2	0.997	0.970	2.808	-	-
3	0.988	0.971	2.696	-	-
4	0.990	0.987	1.187	3	-
5	0.606	0.585	38.892	-	-
6	0.989	0.966	3.231	-	-
7	0.998	0.208	74.216	-	-
8	0.997	0.993	0.619	2	2
9	0.983	0.982	0.935	-	-
10	0.998	0.976	2.231	-	-
11	0.965	0.969	1.453	-	-
12	0.998	0.994	0.601	1	1
13	0.989	0.987	1.021	3	3

Table 4는 성능 평가 결과에 따라 최종 선택된 특성집합이다. 최종 특성집합은 5종의 거시변수(콜금리, 기준금리, 고용지수, 국고채금리, KOSPI)와 평균 개인신용카드 발급건수, 평균 미해제 개인평균연체기간 등 7종의 개인 신용정보 가공 변수, 개인 및 기업 총대출 중 기업 대출 잔액 비중의 총계, 개인대출 잔액 구간의 총계 등 5종의 복합신용정보 가공 변수, 기업카드연체건수의 총계, 평균 운전자금대출기관수 등 3종의 기업신용정보 가공 변수 등이며, 총계 및 평균은 월 단위 집계값을 뜻한다.

연체는 기준년월 해제여부에 따라 해제건과 미해제건으로 구분될 수 있다. 본 연구의 신용정보 특성(후보)은 해제건과 미해제건, 그리고 두가지를 포괄한 경우를 구분, 분석하였다. 복합신용정보 변수에서 기업대출잔액구간과 개인대출잔액구간은 각각 5분위 분석을 통해 각 구간을 수치화하여 활용하였으며, 결과적으로 기업대출잔액구간의 평균과 개인대출잔액구간의 총계값이 예측을 위한 최종 특성에 포함되었다.

Table 4. Final features

Type (FID)		Contents
Macro (t-2~ t-6)	E2_2	Call rate (t-2),
	E2_4	Base rate (t-3),
	E4_3	Number of employees (t-6),
	E8_1	Treasury bond interest rate (t-6),
	E9_3	KOSPI (t-3)
TOT P1_5	Sum of total number of private non-bank loan institutions,	
TOT P2_4	Sum of total number of personal overdue experiences,	
AVG P2_8	Mean of personal average overdue period (non-cancellation),	
TOT P2_8	Sum of personal average overdue period (non-cancellation),	
AVG P2_12	Mean of total number of personal card overdue,	
AVG P3_4	Mean of personal credit card issuance cases,	
TOT P3_4	Sum of personal credit card issuance cases	
Micro (t-3)	AVG C7_8	Mean of number of default codes for self-employment business (cancellation),
	TOT C9_2	Sum of ratio of business loan balance among self-employment loans,
	AVG C9_4	Mean of ratio of business loan balance among self-employment loans in non-banking sectors,
	AVG C10_2	Mean of business loan balance section for self-employment business,
	TOT C10_1	Sum of personal loan balance section for self-employment business
	AVG C1_9	Mean of total number of institutions for corporate working capital loan,
	TOT C3_12	Sum of total number of corporate card overdue,
	AVG C6_9	Sum of corporate credit card issuance cases

(* t : base month of the default rate of the dependent variable)

5. 결론

본 연구에서는 기존 연구와 달리 개인사업자의 신용정보 가공 데이터를 활용함으로써 개인사업자의 고유한 신용 특성을 반영하여 업종별-월별 부도율을 예측하기 위한 새로운 신용정보 특성 선택 방법을 제안하며 방대한 양의 초기 신용정보 특성집합으로부터 다양한 부분집합을 생성, 성능을 비교함으로써 최대 성능 특성 집합을 제시한다.

많은 신용정보 특성(후보)을 포함하는 개인사업자 업종별 부도율 예측 모델에서 유용한 특성 선택을 위한 제안 방법을 적용하여 기존 연구단순 Filter 방법, Wrapper 방법, 초기 연구 방법 대비 예측 성능을 개선(평균오차를 줄이고 R2 결정계수를 향상) 시킴으로써 유용성을 입증하였다.

본 연구는 개인사업자 취약 업종 리스크 관리를 위해 개별 기업의 부실이 아닌 산업의 신용위험을 예측하는 모형에서 이용될 수 있는 특성 선택 방법이며, 기존 기업 부도 예측 연구에서 주로 사용되었던 분기 단위 재무정보가 아닌 월단위 신용정보를 활용함으로써 분기 혹은 연단위 부도율이 아닌 월단위 부도율을 산출한다.

제안 방법을 통해 식별한 최종 거시 및 신용정보 특성 집합을 이용하여 개인사업자 취약 업종에 대한 보다 정교하고 정확한 신용 리스크 관리가 가능하게 될 것을 기대하며, 개인사업자 부도율 예측 뿐만 아니라 개인사업자와 유사한 특성을 가지는 자영업자, 창업기업, 비상장기업 등의 부도율 예측에도 폭넓게 활용될 것을 기대한다.

References

고우석, 윤춘경, 이한필, 황순진, 이상우 (2019), “머신러닝 기반 CFS(Correlation-based Feature Selection) 기법과 Random Forest 모델을 활용한 BMI(Benthic Macroinvertebrate Index) 예측에 관한 연구”, *한국물환경학회지*, 제35권, 제5호, pp.425-431.

권누리, 김영민, 최광신 (2019), “거시경제 변수를 고려한 한국기업부도 모형 구축 방법 연구”, *한국데이터정보과학회지*, 제30권, 제5호, pp.1037-1050.

김경륜, 김재권, 이종식 (2020), “특성집합 IG-MLP 평가 기반의 최적화된 특성선택 방법을 이용한 질환 예측 머신러닝 모델”, *한국사물레이선학회논문지*, 제29권, 제1호, pp.11-21.

김창배, 남주하 (2008), “산업별 신용위험 결정요인 분석: 거시변수와 산업간 부실 전이효과를 중심으로”, *국제지역연구*, 제12권, 제1호, pp.95-116.

김한균 (2000), “기업부실예측모델의 산업에의 적용”, *산은조사월보*, pp.1-24.

남주하, 김창배 (2008), “산업별 신용위험에 대한 연구”, *응용경제*, 제10권, 제3호, pp.157-181.

남기정, 이동명, 진로 (2019), “비재무정보를 이용한 창업기업의 부실요인에 관한 실증연구”, *벤처창업연구*, 제14권, 제1호, pp.139-149.

박근영, 한현수 (2013), “자동차부품제조업의 부도 위험 수준 예측 연구”, *한국데이터베이스학회*, 제20권, 제4호, pp.221-234.

박주완, 송창길, 배진성 (2017), “기계학습 기법을 이용한 소상공인 신용평가모형 구축에 관한 연구”, *한국비즈니스리뷰*, 제10권, 제3호, pp.1-16.

- 박호연, 김정재 (2018), “시물레이티드 어니얼링 기반의 랜덤 포레스트를 이용한 기업부도예측”, *지능정보연구* 제24권, 제4호, pp.155-170.
- 서재현 (2018), “유전 알고리즘 기반의 비정상 행위 탐지를 위한 특성선택”, *한국융합학회논문지*, 제9권, 제7호, pp.1-7.
- 어균선, 이진창 (2019), “효과적인 오피니언 마이닝을 위한 최적속성선택에 관한 연구”, *한국컴퓨터정보학회 논문지*, 제24권, 제2호, pp.171-177.
- 윤경영, 김석진 (2013), “거시경제변수가 소상공인 신용위험에 미치는 영향”, *경영학연구*, 제42권, 제4호, pp.959-985.
- 윤병우, 권태구 (2020), “자영업가구의 채무불이행 결정요인에 관한 실증연구”, *한국자료분석학회*, 제22권, 제1호, pp.279-292.
- 윤상용, 강만수, 이형탁 (2016), “소상공인 신용평가에서 비재무적 정보는 중요한가”, *경영컨설팅연구*, 제16권, 제2호, pp.37-46.
- 원재환, 반주희 (2012), “부도확률 예측에서 미시정보와 거시정보의 역할”, *금융안정연구*, 제13권, 제2호, pp.25-50.
- 이대범, 서재현 (2019), “특성선택 기법에 기반한 UNSW-NB15 데이터셋의 분류 성능 개선”, *한국융합학회 논문지*, 제10권, 제5호, pp.35-42.
- 이상근, 장병탁 (2004), “베이지안 신경망을 이용한 유전자 발현 데이터에서의 특성 선택 기법”, *한국정보과학회 2004 학술발표논문집*, pp.235-237.
- 이인로, 김동철 (2015), “회계정보와 시장정보를 이용한 부도예측모형의 평가 연구”, *재무연구*, 제28권, 제4호, pp.626-666.
- 이재식, 정미경 (2008), “단변량 분석과 LVF 알고리즘을 결합한 하이브리드 속성선정 방법”, *지능정보연구*, 제14권, 제4호, pp.179-200.
- 이치송 (2005), “거시경제변수와 산업별 신용위험에 관한 연구”, *산업경제연구*, 제18권, 제1호, pp.79-99.
- 정병우, 박기영, 황선영 (2013), “객체검출을 위한 빠르고 효율적인 Haar-Like 피쳐선택 알고리즘”, *한국통신학회논문지*, 제38권, 제6호, pp.486-491.
- 한국신용정보원(Korea Credit Information Services), <http://www.kcredit.or.kr>.
- 홍동숙 (2020), “AI기반 개인사업자 업종별 부도율 예측에 관한 연구 : 신용정보 입력변수의 유용성 검증과 AI 기법 성능 평가 중심으로”, *CIS 이슈리포트 2020-6* 호, pp.1-12.
- 홍승현, 신경식 (2003), “유전자 알고리즘을 활용한 인공 신경망 모형 최적입력변수의 선정: 부도예측 모형을 중심으로”, *한국지능정보시스템학회논문지*, 제9권, 제1호, pp.227-249.
- Altman, E. I. (1968), *Financial Ratios, Discriminant Analysis and the Predication of Corporate Bankruptcy*, *Journal of Finance*, Vol. 23, pp 589-609.
- Chava, S., Jarrow, R. A. (2004), *Bankruptcy Prediction with Industry Effects*, *Review of Finance*, Vol 8, Issue 4, pp.537-569.
- Girija, A., Manohara P. M. M., Radhika, M. P. (2019), *Feature Selection Using Submodular Approach for Financial Big Data*, *Journal of Information Processing Systems*, Vol 15, Issue 6. pp.1306-1325.
- Hong, D. S., Baek, H. J. (2020), *Default Rate Prediction Models for Self-employment in Korea using Ridge, Random Forest and Deep Neural Network*, *Lattice : The Machine Learning Journal*, Vol 1, Issue 3. pp.21-26.
- Kohavi, R., John, G. H. (1997), *Wrappers for Feature Subset Selection*, *Artificial intelligence*, Vol 97, Issue 1-2. pp.273-324.
- Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., Liu, H. (2018), *Feature selection: a Data Perspective*, *ACM Computing Surveys*, Vol 50, no 6, pp.1-45.
- Merton. R. (1974), *On the Pricing of Corporate Debt: The Risk Structure of Interest Rates*, *Journal of Finance* Vol 29, pp.449-470.
- Nguyen, H., Franke, K., Petrovic, S. (2010), *Improving Effectiveness of Intrusion Detection by Correlation Feature Selection in Availability, Reliability, and Security*, *ARES'10 International Conference on*, pp.17-24.
- Ohlson, J. A. (1980), *Financial ratios and the probabilistic prediction of bankruptcy*, *Journal of Accounting Research*, Vol 18, Issue 1, pp.109-131.
- Tong, X. (2014), *Modeling Corporate Default Rates*, *International Journal of Economics and Finance*, Vol 6, No 8, pp.1-14.

홍 동 숙 (ORCID : <https://orcid.org/0000-0003-0236-9357> / dshong@kcredit.or.kr)



1999 건국대학교 컴퓨터공학과 공학사
 2001 건국대학교 정보통신대학원 공학석사
 2008 건국대학교 정보통신대학원 공학박사
 2001~ 2003 쌍용정보통신 모바일GIS기술팀
 2008~ 2009 연세대학교 TMS정보기술사업단 Post Doctoral Researcher
 2009~ 2014 한국과학기술정보연구원 기술사업화센터
 2014~ 2015 전국은행연합회 기술정보부
 2016~ 現 한국신용정보원 빅데이터센터

관심분야 : 빅데이터 전략분석, CreDB (기업신용DB), AI (딥러닝) 예측, 성능분석 등

백 한 중 (ORCID : <https://orcid.org/0000-0001-6188-2906> / click101@kcredit.or.kr)



2001 동국대학교 컴퓨터정보통신공학부 공학사
 2013 서울시립대학교 경영대학원 경영학석사
 2020 숭실대학교 대학원 IT정책경영학과 공학박사
 2003~ 2015 전국은행연합회 전산부
 2016~ 現 한국신용정보원 IT본부 / 빅데이터센터

관심분야 : 빅데이터, AI, 클라우드, IT 기획, IT 아키텍처 설계, 프로젝트 관리 등

신 현 준 (ORCID : <https://orcid.org/0000-0003-0127-3468> / hjs183@kcredit.or.kr)



1991 서울대학교 경영학과 경영학사
 2000 University of Missouri, Columbia, Ph.D in Economics
 1992~ 2004 재무부 국제관세과 / 재정경제부 은행과 · 보험과
 2004~ 2007 대통령비서실 / 美 World Bank
 2008~ 2013 금융위원회 글로벌금융과, 자산운용과, 보험과, 기획재정담당관
 2013~ 2017 주 OECD대표부
 2017~ 2019 우정사업본부 보험사업단
 2019~ 現 한국신용정보원

관심분야 : 빅데이터 전략, 데이터기반 거시경제 분석, 통화정책 등