

기계학습을 이용한 수출신용보증 사고예측

조재영

한국무역보험공사
(cjy00814@ksure.or.kr)

주지환

KAIST 경영대학
(jhz1012,joo@kaist.ac.kr)

한인구

KAIST 경영대학
(ighan@kaist.ac.kr)

2020년 8월 정부는 한국판 뉴딜을 뒷받침하기 위한 공공기관의 역할 강화방안으로서 각 공공기관별 역량을 바탕으로 5대 분야에 걸쳐 총 20가지 과제를 선정하였다. 빅데이터(Big Data), 인공지능 등을 활용하여 대국민 서비스를 제고하고 공공기관이 보유한 양질의 데이터를 개방하는 등의 다양한 정책을 통해 한국판 뉴딜(New Deal)의 성과를 조기에 창출하고 이를 극대화하기 위한 다양한 노력을 기울이고 있다. 그중에서 한국무역보험공사(KSURE)는 정책금융 공공기관으로 국내 수출기업들을 지원하기 위해 여러 제도를 운영하고 있는데 아직까지는 본 기관이 가지고 있는 빅데이터를 적극적으로 활용하지 못하고 있는 실정이다.

본 연구는 한국무역보험공사의 수출신용보증 사고 발생을 사전에 예측하고자 공사가 보유한 내부 데이터에 기계학습 모델을 적용하였고 해당 모형 간에 예측성능을 비교하였다. 예측 모형으로는 로지스틱(Logit) 회귀모형, 랜덤 포레스트(Random Forest), XGBoost, LightGBM, 심층신경망을 사용하였고, 평가 기준으로는 전체 표본의 예측 정확도 이외에도 표본별 사고 확률을 구간으로 나누어 높은 확률로 예측된 표본과 낮은 확률로 예측된 경우의 정확도를 서로 비교하였다. 각 모형별 전체 표본의 예측 정확도는 70% 내외로 나타났고 개별 표본을 사고 확률 구간별로 세부 분석한 결과 양 극단의 확률구간(0~20%, 80~100%)에서 90~100%의 예측 정확도를 보여 모형의 현실적 활용 가능성을 보여주었다. 제2종 오류의 중요성 및 전체적 예측 정확도를 종합적으로 고려할 경우, XGBoost와 심층신경망이 가장 우수한 모형으로 평가되었다. 랜덤포레스트와 LightGBM은 그 다음으로 우수하며, 로지스틱 회귀모형은 가장 낮은 성과를 보였다. 본 연구는 한국 무역보험공사의 빅데이터를 기계학습모형으로 분석해 업무의 효율성을 높이는 사례로서 향후 기계학습 등을 활용하여 실무 현장에서 빅데이터 분석 및 활용이 활발해지기를 기대한다.

주제어 : 수출신용보증, 보증사고, 부도예측, 한국무역보험공사, 기계학습

논문접수일 : 2021년 1월 5일 논문수정일 : 2021년 2월 13일 게재확정일 : 2021년 3월 8일
원고유형 : 일반논문 교신저자 : 주지환

1. 서론

4차 산업혁명의 핵심 기술로 다양한 산업에서 이용되는 인공지능은 기존 처리방식을 뛰어넘는 성과를 보여주고 있다. 인공지능과 인간의 대결로 전 세계의 주목을 받은 알파고와 이세돌 9단의 바둑 경기에서 모두의 예상을 뒤엎고 알파고가 압도적으로 승리한 것은 인공지능의 우수성을 보여 준 대표적 사례이다. 최근 인공지능 분

야에서 기계학습을 이용한 예측연구가 활발히 이루어지고 있는데 그 배경은 Python과 R과 같은 데이터 분석언어의 API(Application Programming Interface)가 풍부해지고 클라우드 환경에서 컴퓨팅 자원을 사용할 수 있게 되면서 누구나 쉽게 데이터를 분석할 수 있게 되었기 때문이다.

우리나라는 개인정보 보호법이나 의료법 등에 따른 데이터 이용의 규제로 세계적인 4차 산업혁명 흐름을 따라가지 못한다는 비판을 받아 왔

으나 최근 규제가 완화되고 가명 및 익명 정보의 개념이 도입되는 등 데이터 접근 및 활용에 우호적인 여건이 조성되고 있다. 2020년 발표된 한국판 뉴딜 종합계획에는 ‘디지털 뉴딜’을 통해 경제 전반의 디지털 혁신 및 역동성을 촉진하는 내용이 포함되어 있는데 데이터 댐을 통해 데이터의 공유 환경을 구축하고 산업간 데이터 융합으로 기술혁신이 이루어질 수 있도록 정책적 지원이 확대되고 있다. 공공기관들도 고유사업을 디지털 중심으로 전환하고 빅데이터, 인공지능 등 신기술 도입에 적합한 업무를 적극적으로 발굴하여 한국판 뉴딜의 성과를 조기에 창출하고 이를 극대화하기 위한 노력을 기울이고 있다.

한국무역보험공사는 정책금융 공공기관으로 수출기업들을 지원하는 다양한 제도를 운영하고 있다. 최근 금융과 기술의 융합인 핀테크(Fintech)를 전담하는 부서를 신설하고 기술 변화에 능동적으로 대응하고 있으나 빅데이터 분석을 활용한 사업성과는 미비한 실정이다. 공사의 수출신용보증 인수심사는 모두 심사자의 수기심사로 진행되며 심사 완료까지 수일이 소요되고 있는데 최근 여러 금융기관에서 빅데이터 기반의 인공지능 기술을 통해 비대면 온라인 채널에서 즉시 심사 결과를 보여주는 상품을 출시하여 고객의 이용 편의와 심사자의 업무 효율을 높이는 것과는 상반되는 상황이다. 과거 인수된 수출신용보증의 사고정보를 학습시킨 사고 예측 인공지능을 보증 심사에 사용한다면 발생확률이 낮은 소액 보증은 심사자의 개입 없이 즉시 한도를 책정하여 고객에게 빠른 서비스를 제공할 수 있고 심사자는 발생확률이 높은 특이건과 거액 보증에 심사 역량을 집중하여 성과를 창출할 수 있을 것이다. 더불어 유효한 보증기간 중 사고예

측 인공지능을 사용하여 사고 발생 가능성을 사전 점검하고 위험이 높은 기업들을 집중 관리하는 등의 위험관리 체계를 마련할 수 있을 것이다. 이에 본 연구는 한국무역보험공사가 보유한 내부 데이터에 기계학습을 적용함으로써 수출신용보증 사고의 발생을 사전에 예측하는 인공지능 모형을 개발하였고 모형 간 예측성능을 비교하였다. 예측을 위한 모형으로는 로지스틱 회귀 모형, 랜덤포레스트, XGBoost, LightGBM, 심층신경망을 사용하였고 평가 기준으로는 전체 표본의 예측 정확도, 제2종 오류 이외에도 표본별 사고 발생확률을 분석하여 높은 확률로 예측된 표본과 낮은 확률로 예측된 경우의 정확도를 비교하였다. 각각의 모형들은 69~71.1% 수준의 예측 정확도를 보여주었고 사고를 건전으로 예측하는 제2종 오류는 33.2~34.6% 수준으로 나타났다. 개별 표본을 사고 확률 구간별로 세부 분석한 결과 양 극단의 확률구간(0~20%, 80~100%)에서 90~100%의 예측 정확도를 보여 모형의 현실적 활용 가능성을 보여주었다. 본 연구는 한국무역보험공사의 빅데이터를 기계학습모형으로 분석해 업무의 효율성을 높이는 사례로서 기계학습을 이용한 공공기관의 빅데이터 분석과 활용의 계기가 되기를 기대한다.

본 연구의 구성은 다음과 같다. 2장에서 한국무역보험공사의 무역보험 및 수출신용보증을 소개하고 기업의 부실 예측에 관한 선행 연구 및 기계학습의 개념을 설명한다. 3장에서는 연구모형, 변수 선정, 데이터의 처리 및 평가 기준을 제시하고 4장에서는 모형별 예측 결과를 분석한다. 마지막 5장에서는 연구 결과를 요약하고 시사점 및 후속 연구방향에 대하여 논의한다.

<Table 1> Acquisition Performance by Insurance System (Unit: 100 Million KRW)

| Year | Short-term Export Credit Program | Medium and Long-term Export Credit Program | Export Credit Guarantee | Foreign Exchange Risk Insurance |
|------|----------------------------------|--|-------------------------|---------------------------------|
| 2014 | 1,694,183 | 127,940 | 64,934 | 15,096 |
| 2015 | 1,494,365 | 118,441 | 56,291 | 12,262 |
| 2016 | 1,394,770 | 112,568 | 47,278 | 11,060 |
| 2017 | 1,321,246 | 90,922 | 30,993 | 3,526 |
| 2018 | 1,351,620 | 99,951 | 26,955 | 7,748 |

2. 문헌검토

2.1. 무역보험제도 및 수출신용보증

1968년 수출보험법이 제정되고 1969년 대한재보험공사(현 코리안리)에서 최초의 수출보험업무가 시작되었다. 이후 1977년에 수출보험업무의 대행이 한국수출입은행으로 이관되었고 1982년 한국수출보험공사가 한국수출입은행에서 분리 설립되면서 수출보험업무 전담기관의 역할을 수행하게 되었다. 2010년 수출보험법은 무역보험법으로 개정되었고 한국수출보험공사의 사명도 한국무역보험공사로 변경되었다.

한국무역보험공사의 무역보험제도는 단기성보험, 중장기성보험, 수출신용보증, 환변동보험 등의 4가지로 분류할 수 있다. 단기성보험인 단기수출보험(선적후)은 결제기간 2년 이하의 수출계약에서 국외기업의 물품 편취 및 수출대금 미지급으로 발행하는 손실을 보상한다. 또한 수출채권 유동화를 이용하면 물품이 수출된 후 환어음 등의 선적서류를 근거로 외국환은행으로부터 자금을 미리 받아 수출대금을 조기에 회수하는 효과를 얻을 수 있다. 중장기성보험은 결제기간 2년 이상의 수출거래를 지원한다. 중장기성보험 중 해외사업금융보험은 수출증진이나 외화

획득의 효과가 있을 것으로 예상되는 해외사업에 필요한 자금을 중장기적으로 공급하여 국내기업의 수출경쟁력 향상을 지원한다. 수출신용보증의 대표 종목인 수출신용보증(선적전)은 담보능력이 부족한 국내기업이 은행으로부터 원재료 구매나 인건비 등의 자금을 대출받기 어려운 경우 한국무역보험공사의 지급보증을 담보로 필요한 자금을 대출받을 수 있다. 환변동보험은 물품의 계약 시점과 대금의 회수 시점의 환율변동에 따른 환차손익을 사전에 확정시키면서 환율변동에 따른 위험을 헷지할 수 있다. 아래 <Table 1>에서는 각 보험제도(단기수출보험, 중장기수출보험, 수출신용보증, 환변동보험)별 인수실적을 보여주고 있다.

신용보증은 담보 능력이 약한 기업의 채무를 보증하여 자금 유통을 원활하게 한다. 특히 공적 신용보증은 기술력 및 성장성이 있음에도 신용이나 담보가 부족한 중소기업들이 자금을 조달할 수 있도록 함으로써 중소기업이 성장할 수 있는 디딤돌 역할을 한다. 단기간의 신용보증은 일반적으로 적은 비용으로 기업의 영업활동을 지원하는 긍정적 효과를 발생시킨다. 그러나 장기간의 신용보증은 무능한 기업을 존속하게 만들 수도 있고 신규기업들의 자금조달의 기회를 축소 시킨다. 그러므로 10년 이상의 장기보증을 지

양하고 일정 기간 이상 보증을 이용하는 기업에 대해 보증료를 큰 폭으로 인상하여 보증을 조기 졸업하게 만드는 유인이 필요하다. 또한 상위 신용등급의 기업보다는 신용보증이 필요한 저 신용등급의 중소기업에 더 많은 지원이 필요하다 (Kim et al., 2014).

우리나라의 신용보증제도는 1961년 중소기업 금융지원 정책으로 신용보증준비금 제도를 도입함으로써 시작되었고 1976년 신용보증기금이 창립되면서 독립기관으로 업무를 시작하게 되었다. 1989년 기업이 가진 기술력을 평가하여 매출이 없더라도 신용보증을 받을 수 있는 기술보증기금이 설립되었고 1992년 수출기업들의 수출이행자금에 대한 신용보증인 수출신용보증제도가 한국무역보험공사에서 시행되었으며 1996년 경기신용보증재단을 시작으로 전국 16개 지역에 지역신용보증재단이 설립되었다. 1997년 IMF사태에 따른 국가적 경제위기에서 신용보증의 공급이 급속하게 확대되었고 이를 통해 빠르게 위기를 극복할 수 있는 원동력이 되었다(Seo and Lee, 2006).

수출신용보증은 수출기업의 수출 관련 채무에 대한 보증을 말하며 선적전, 선적후, 매입, 네고(Nego) 등의 세부 종목으로 구성된다. 본 연구의 데이터로 사용한 수출신용보증(선적전)은 외국 환은행이 수출기업에 대출하여 발생한 상환채무를 한국무역보험공사가 연대보증하는 제도로써 담보력이 부족한 수출기업이 수출물품의 제조, 가공, 조달 및 원자재 구매에 필요한 자금을 금융기관으로부터 원활히 조달할 수 있도록 하여 수출진흥에 기여하는 제도이다. 은행은 한국무역보험공사의 수출신용보증서로 지급보증을 받은 기업이 채무를 불이행하면 한국무역보험공사에 해당 사고를 통지하고 한국무역보험공사는

면책사유가 없으면 보증상대방 은행에 대해 대위변제금을 지급한다.

2.2. 부실예측 선행연구

Beaver(1966)는 단변량 판별분석을 통해 기업의 부실을 예측하는 모형을 개발하였으나 기업의 부도가 단일 요인에 의해 결정되는 것이 아니므로 예측에 한계를 보였다. 이후 Altman(1968)은 다변량 판별분석을 통해 Z-Score를 산출하여 기업의 부실을 예측하는 모형을 개발하였다. Ohlson(1980)은 판별분석을 이용한 기존 예측모형을 보완하여 로지스틱 회귀분석을 이용한 새로운 모형을 제시하였고 Elmer and Borowski(1988)은 인공지능을 이용해 기업의 부실을 예측하였다. 우리나라에서도 1980년대부터 기업부실 예측 연구가 활발히 진행되었다. Kim(1987)는 재무비율의 통계적 분석을 통해 판별함수를 도출하고 부실예측 모델을 개발하였다. Han et al.(1995; 1997)은 인공신경망 등을 이용해 기업의 부실을 예측하고 지능형 신용평가시스템을 개발하였다. Kim and Kim(2001)은 인공신경망을 이용하여 부실기업 예측모형을 연구하였다. 최근에는 랜덤 포레스트(Random Forest)와 같은 앙상블 방식이나 SVM(Support Vector Machine), 인공신경망 등의 기법들을 이용해 부실기업을 사전에 예측하려는 연구가 꾸준히 수행되어왔다. 2016년 알파고와 이세돌의 바둑 대결 이후로 인공신경망의 진화한 모형인 딥러닝에 대한 관심이 높아지면서 딥러닝(Deep Learning)을 이용한 다양한 연구가 이루어졌다. Cha and Kang(2018)은 딥러닝을 이용해 기업부도 예측모형의 유용성을 검증하였다. Yoon(2019)은 딥러닝을 이용해 신용카드 부도위험 예측의 효용성을 검증하였다. Lee(2019)

은 P2P 소셜 대출 채무자의 부도예측을 위해 딥러닝을 사용하였고 로지스틱 회귀분석, 의사결정나무와 예측력을 비교하였다.

2.3. 예측모형

본 연구의 예측모형으로 로지스틱 회귀모형, 랜덤포레스트, XGBoost, LightGBM, 심층신경망 등 총 5가지 모형을 사용하였다. 기존 연구에서 많이 소개된 로지스틱 회귀모형을 제외하고 나머지 모형에 대해 간단히 설명할 것이다.

2.3.1. 랜덤포레스트

랜덤포레스트는 다수의 결정트리(Decision Tree)들을 학습하게 되는데 랜덤 노드 최적화 및 배깅(Bagging)을 결합하는 CART(Classification And Regression Tree)를 이용하여 상관관계가 없는 트리들로 숲(Forest)을 만들게 된다. CART는 설명변수의 비선형성(Nonlinearity) 및 상호작용(Interactions)을 고려함으로써 종속변수에 대한 영향을 판단한다. 설명변수를 중요도 기준에 근거하여 줄기(Branch)를 만들어 가면서 마지막 노드에서 종속변수에 대한 판단을 내리게 된다.

CART는 하나의 결정트리만 생성하는 반면 랜덤포레스트는 여러 개의 결정트리로 숲(Forest)을 만들게 된다. 여러 개의 결정트리를 만들어 내기 위하여 예측인자와 관측값에 대하여 랜덤하게 표본을 추출하는 과정을 반복한다. 많은 수의 결정트리로부터 예측 범주를 얻은 뒤 다수결 투표(Majority Voting)를 거쳐 최종적으로 범주를 예측하게 된다. 결정트리를 형성하는 과정을 랜덤하게 함으로써 독립된 결정트리들을 반복적으로 생성할 수 있으며 이는 예측오차를 감소시킨다. 예측인자 및 관측값의 랜덤한 추출을 위하여

붓스트랩(Bootstrap) 기법이 활용된다. CART에서는 하위 노드가 많아지면 예측오차의 편향이 감소하는 반면에 분산은 증가하는 문제점이 존재한다. 그러나 랜덤포레스트의 경우에는 동일하게 분포되어 있는 결정트리를 반복하여 생성하므로 예측오차의 분산을 감소시킬 수 있게 된다. 랜덤포레스트는 종속변수가 이항변수, 다항변수, 또는 연속형인 경우에도 적용이 가능하고 설명변수가 연속형 및 범주형인 경우에도 적용할 수 있다.

2.3.2. XGBoost

XGBoost 모형은 트리 기반 앙상블 알고리즘으로서 결정트리를 차례대로 학습하며 각각의 트리는 선행하는 트리에 존재하는 오류를 개선하는 방식으로 생성된다. XGBoost는 GBM(Gradient Boosting Machine)을 발전시킨 모형으로 Chen and Guestrin(2016)에 의해 개발되었다. 병렬화와 가지치기를 이용하여 시스템을 최적화하며 병렬화로 인하여 GPU(Graphics Processing Unit)를 사용할 수 있는 것이 장점이다. XGBoost모형은 기존의 GBM 모형에 비해서 속도가 빠르고 일반화된 모형으로 주목을 받게 되었다. XGBoost는 훈련 시 손실함수와 규제항으로 구성된 목적함수를 최소화하게 된다. 여기에서 규제항은 모형의 과적합(Overfitting)을 제한하기 위하여 추가된 항이다. 따라서, 예측값과 목적함수는 아래의 식 (1)과 같이 표현된다.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i),$$

$$\text{Min} \left[\sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \right] \quad (1)$$

여기에서 f_k 는 k 번째 의사결정트리로, 총 K 개 가진다고 가정하였고, \hat{y}_i 와 y_i 는 각각 i 번째 예측값과 실제값이며, $l(\hat{y}_i, y_i)$ 는 손실함수, 그리고 $\Omega(f_k)$ 은 규제항이다. 결정트리는 기존의 리프노드에 두 가지의 새로운 리프노드를 추가하여 목적함수를 최소화하는 변수의 값에서 분리되어 만들어지게 된다. 이는 결정트리가 연구자에 의해 정해진 최대 깊이까지 도달하도록 반복하여 수행하게 된다. 여기에서 각 노드는 목적함수에서 테일러급수 전개와 1차 및 2차 미분을 통하여 유도되는 식에서 최적의 가중치를 얻게 된다. 각 결정트리는 이전 트리의 잔차를 기반으로 학습을 한다.

2.3.3. LightGBM

LightGBM 모형은 XGBoost과 마찬가지로 GBM (Gradient Boosting Machine)기반의 모형으로서 Ke et al.(2017)에 의하여 제시되었다. 기존의 GBM 계열에서 사용하던 것과는 달리 Leaf-wise 방법으로 트리를 분할함으로써 정확도를 향상시켰다. 또한 GOSS(Gradient-based One-Side Sampling) 및 EFB(Exclusive Feature Bundling)를 이용하여 메모리 사용량을 감소시키고 훈련 속도를 빠르게 만들었다. 그러나 과적합이 되기 쉬운 모형으로 10,000개 이상의 충분한 데이터 수가 확보되어야 활용하는데 문제가 없다.

2.3.4. 심층신경망(Deep Neural Network, DNN)

인공신경망(Artificial Neural Network, ANN)은 인간의 두뇌에서의 생물학적 문제해결 방식에 근거해 개발한 인공지능의 모형으로 입력층(Input Layer), 은닉층(Hidden Layer), 그리고 출력

층(Output Layer)을 포함하는 신경망 구조로 이루어진다. 인공신경망의 각 층은 하나 또는 여러 개의 노드로 구성되어 있고, 해당 노드들을 연결할 때마다 가중치를 부여하고 학습시킴으로써 최적화가 이루어진다. 은닉층 내에 존재하는 노드들은 가중치를 곱하여 계산된 값을 입력으로 받게 되고, 활성화함수(Activation Function)를 통과하고 출력되어 그 다음 노드로 전달된다. 인공신경망은 관계가 복잡하고 방대한 양을 가지는 비선형 데이터를 분석하는데 유용하며 예측능력이 뛰어나서 다방면에 활용된다(Ahn et al., 2005).

심층신경망은 인공신경망을 깊은 구조로 쌓은 형태인데 최근 자연어처리 등의 여러 분야에서 패턴 분류, 인식, 또는 예측 시 우수한 성능을 보여주면서 새롭게 주목을 받고 있다. 심층신경망은 입력층과 출력층 사이에 2개 이상의 은닉층들로 구성된 신경망으로 단 하나의 입력층과 출력층으로만 구성되는 단층신경망 분석에 비해 고도화된 알고리즘이다(Schmidhuber, 2015). 심층신경망은 보통의 인공신경망과 유사하게 복잡한 비선형 관계를 고려한 모형이나 더욱 많은 층이 계층들을 활용함으로써 하위 계층들이 가지는 특성을 보다 잘 파악할 수 있게 해준다. 수학적으로는 아래 식 (2)와 같이 표현할 수 있는데, v_i 는 입력층 또는 이전 은닉층 신호고 b_j 와 b_k 는 은닉층과 출력층의 편향(Bias)을 의미한다. 또한 w_{ij} 와 w_{jk} 는 은닉층과 출력층의 계수를 나타낸다. 함수 f 는 활성화함수로 시그모이드(Sigmoid)나 ReLU(Rectified Linear Unit) 등의 비선형 함수를 활용하여 복잡한 관계를 보다 효과적으로 표현하게 된다. 다음 식 (2)의 경로를 거치면 출력층의 결과 값(Y_k)을 얻게 된다. 물론 은닉층의 수가 많으므로 연산 비용이 증가하게 되나, 최근

컴퓨팅 능력이 향상되고 관련된 API가 풍부해지면서 별다른 어려움 없이 분석을 수행할 수 있게 되었다.

$$Y_k = \sum_{j=1}^m f\left(\sum_{i=1}^n v_i w_{ij} + b_j\right) w_{jk} + b_k \quad (2)$$

심층신경망은 신경망을 훈련시킬 때 오류 역전파(Back Propagation) 알고리즘을 활용하여 마지막 출력의 오류를 사용해 은닉층의 가중치까지 학습한다. 이는 출력층에서 아래 층으로 정보를 역전파하여 아래 층에서는 그 값을 기준으로 가중치를 업데이트한다. 입력값들은 은닉층에 전달되고 또한 은닉층에서 출력층으로 전달되므로, 현재 가중치에 대한 실제값과 예측값 사이의 오차를 계산해 이를 최소화하는 방향으로 가중치를 학습시키는 것이다. 가중치를 학습할 때 델타 규칙(Delta Rule)을 이용하여 출력층에서의 각 뉴런들의 오차를 활용하여 가중치를 업데이트하게 된다. 해당 알고리즘은 가중치를 업데이트할 때 경사 하강법(Gradient Descent Algorithm)을 이용해 오차를 최소화하는데, 한 번 오차가 계산되고 나면 출력층에서 이전 층으로 역으로 정보가 업데이트되기 때문에 역전파라고 한다.

3. 연구설계

3.1. 연구자료

본 연구에서 사용할 데이터는 2014년~ 2018년까지 한국무역보험공사의 수출신용보증(선적전)을 이용한 기업이다. 해당 데이터의 특성상 하나의 관측값이 기업 단위가 아닌 보증서 단위이고,

특정 기업이 보증서를 2개 이상 보유하거나 보증기간(1년)이 경과할 때마다 새로운 보증서를 발급 받으므로 초기 데이터에 대한 조정 과정을 아래 <Table 2>와 같이 진행하였다. 수출신용보증(선적전)에 의한 보증부 대출 기간은 1년이나 일반적으로 신용보증은 여러 해에 걸쳐 연장된다. 즉, 대부분의 기업들은 대출의 상환보다는 만기 연장을 하고 있다. 따라서 만기 연장의 결격사유가 없다면 추가 보증료를 납부하고 별도의 대환처리 없이 계속 사용할 수 있으므로 1년 이상 보증서를 이용 중인 기업은 가장 최근 만기 연장을 심사한 시점을 기준으로 하였다. 보증서 단위로 하면 건전 사례는 20,793건, 부실 사례는 1,896건이고 이를 기업 단위로 정리하면 건전 기업은 4,684곳, 사고 기업은 1,605곳이다.

본 데이터의 특성 상 건전 대 사고 기업의 비율이 9:1 정도이다. 이러한 불균형 데이터에 대해 모형을 훈련시킬 경우 건전 또는 사고에 대한 예측에서 모두 건전기업으로 예측하려는 경향이 관찰된다. 클래스 불균형을 조정하기 위한 여러 방법이 존재하는데 적은 수의 표본 개수에 맞추어 표본을 삭제하는 언더샘플링과 많은 수의 표본에 개수에 맞추어 임의의 표본을 생성하는 오버샘플링으로 구분된다. 기본적으로 언더샘플링은 삭제되는 표본에 담긴 유용한 정보를 잃어버리는 단점이 존재하고 표본 수가 충분하지 않으면 방법을 적용하기 어려워진다. 반면 오버샘플링은 표본에 담긴 모든 정보를 활용할 수 있으나 생성된 표본이 실제 표본을 적절히 대체하지 못하여 과적합되는 현상이 발생하기도 한다. 본 연구에서는 선행 연구를 참조하여 다양한 샘플링 방법들을 적용해 보았으나 표본 수가 충분하고 방법 간 예측 정확도 차이가 유의하지 않았으므로 CNN(Condensed Nearest Neighbour) 언더샘플

〈Table 2〉 Adjustment of Initial Data by Year

| Year | Initial Data (Unit: Guarantee) | | Adjusted Data (Unit: Firm) | | Final Data after CNN Under-sampling (Unit: Firm) | |
|-------|-----------------------------------|---------|-------------------------------|---------|--|---------|
| | Normal | Default | Normal | Default | Normal | Default |
| 2014 | 4,719 | 411 | 4,684 | 1,605 | 1,581 | 1,374 |
| 2015 | 4,397 | 426 | | | | |
| 2016 | 4,118 | 344 | | | | |
| 2017 | 3,888 | 379 | | | | |
| 2018 | 3,671 | 336 | | | | |
| Total | 20,793 | 1,896 | | | | |

〈Table 3〉 Selected Variables

| Classification | Variable | Classification | Variable |
|----------------------|--|-------------------------|--|
| Profitability (5) | Rate of Sales Cost | Cash Flow (4) | Debt Service Coverage Ratio |
| | Rate of Selling and Administrative Expenses | | Cash Operating Profit to Interest Expenses |
| | Return on Invested Capital(ROIC) | | Cash Operating Profit to Short-Term Borrowings |
| | Return on Equity(ROE) | | Free Cash Flow to Sales |
| | EBITDA to Sales Ratio | | Net Working Capital Turnover Ratio |
| Stability (9) | EBITDA to Interest Expenses | Activity (2) | Total Assets Turnover Ratio |
| | EBITDA to Total Borrowings | Growth (5) | Sales Growth Rate |
| | Total Borrowings and Bonds Payable to Total Assets | | Operating Income Growth Rate |
| | Short-Term Borrowings to Total Borrowings | | Net Income Growth Rate |
| | Fixed Assets to Net Worth and Fixed Liabilities | | Total Asset Growth Rate |
| | Interest Expenses to Sales | | Stockholders Equity Growth Rate |
| | Debt to Equity Ratio | Firm Overview (5) | Sector |
| | Total Borrowings to Sales | | Firm Size |
| | Current Ratio | | Firm Type |
| | | | Initial Public Offering |
| | | Firm Age | |

링을 사용하여 사고기업과 건전기업의 클래스 불균형을 조정하였다. CNN 언더샘플링은 Hart(1968)가 제안한 방법으로 데이터가 많은 클래스에 밀집되어 있는 데이터가 없을 때까지 제거해 주어 해당 분포 상 대표성을 갖는 데이터만 남기게 된다. 이 방법의 경우 이미 선택된 데이터에 근접하면서 동일한 클래스에 속한 다른 데

이터는 선택되지 않도록 함으로써 불균형을 효과적으로 조정해 준다. 따라서, 최종적으로 분석에 활용된 건전 기업은 총 1,581곳, 사고 기업은 1,374곳이다.

분석에 활용하기 위한 변수선택 절차는 다음과 같다. 우선 기업 기본정보로서 업종(중공업, 도소매, 경공업, 건설업, 기타), 기업규모(중소기

업, 중견기업, 기타), 기업유형(법인, 개인), 기업 공개(외감, 비외감, 기타), 업력 등 5가지의 비재무변수를 사용하였다. 재무변수는 수익성, 안정성, 현금흐름, 활동성, 성장성의 5개 그룹으로 분류된 34가지 재무비율 정보를 활용하였다. 재무비율은 인수심사 시점을 기준으로 가장 최근 재무제표상에 있는 값으로 활용하였다. 다만 상관관계가 높은 변수들을 모형에 학습시킬 경우 다중공선성의 문제가 있으므로 재무변수 간 상관계수를 검토하여 0.7보다 높은 경우 변수의 중요도를 고려하여 중요하지 않은 9가지 변수를 제거하였다. 결과적으로 <Table 3>과 같이 재무변수 25개 및 비재무변수 5개를 포함하여 총 30개 변수를 최종 선택하였다.

범주형 변수인 업종은 4개의 더미변수, 기업 규모 및 기업공개는 2개의 더미변수, 기업유형은 1개의 더미변수를 이용하여 변환하였다. 종속변수는 0과 1의 값을 가지는 더미변수이며 보증사고가 발생한 경우는 '1'(사고기업), 보증사고가 발생하지 않은 경우를 '0'(건전기업)으로 처리하였다.

3.2. 예측모형 및 평가기준

본 연구의 예측모형으로 전통적으로 많이 활용되는 로지스틱 회귀모형, 앙상블 방식 중 하이퍼 파라미터의 조정 없이도 좋은 성능을 보여주는 랜덤포레스트, 최근 기계학습 경진대회에서 가장 좋은 성능을 보여주고 있는 그래디언트 부스팅 방식인 XGBoost 및 LightGBM, 심층신경망(DNN) 등 5가지 모형을 적용하였다.

평가기준으로 전체 표본의 예측의 정확도뿐만 아니라 개별 표본들의 사고 확률을 구간으로 나누어 예측 정확도와 그 비중을 분석하였다. 또

한 사고기업을 건전기업이라고 예측하는 제2종 오류에 초점을 맞추어 모형을 검토하였다. 건전기업을 사고기업이라고 예측하는 제1종 오류의 손실은 영업기회의 상실 및 신용보증에 따른 보증료 수익의 감소인 반면 사고기업을 건전기업이라고 예측하는 제2종 오류의 경우의 손실은 보증계약자의 채무 불이행에 따른 신용보증금액 및 부실채권 회수 및 관리를 위한 비용이 될 것이다. 따라서 제2종 오류의 손실이 제1종 오류의 손실보다 압도적으로 크기 때문에 전체 표본의 예측 정확도가 높게 나와도 높은 제2종 오류를 보인다면 좋은 모형이라고 할 수 없다.

4. 실험 결과

4.1. 로지스틱 회귀모형

로지스틱 회귀모형과 같은 선형모형은 독립변수의 분포에 따라 예측 성능이 달라질 수 있기에 평균 0, 분산 1을 가지는 정규분포로 변환해서 학습을 진행하였고 예측 결과는 아래의 <Table 4>와 같다. 로지스틱 회귀모형이 건전기업으로 예측한 사고 확률 0~50% 사이 구간에서는 상대적으로 높은 예측 정확도를 보여주고 있다. 건전기업을 강하게 확신한 0~10% 구간에서는 100%의 정확도로 예측을 했고 사고기업과 건전기업의 경계구간인 50% 확률 구간에 가까워질수록 예측 정확도가 낮아지는 합리적 결과를 보였다. 반면 사고기업으로 예측하는 사고 확률 50~100% 사이 구간은 앞선 건전기업 분류보다 상대적으로 낮은 예측 정확도를 보여주었다. 80~90% 구간부터 사고기업과 건전기업의 경계구간인 50% 확률 구간에 가까워질수록 예측 정

확도가 낮아지는 결과를 보였지만 사고기업을 가장 강하게 확신한 사고 확률 90~100% 구간에서 예측 정확도가 61.5%로 저하되는 이상 현상을 보였다. 표본을 변경하여 교차 검증을 진행했으나 모두 동일한 결과를 보였다.

<Table 4> Results from Logistic Model

| | | Probability (%) | Actual | | Accuracy (%) |
|------------|---------|-----------------|---------|--------|--------------|
| | | | Default | Normal | |
| Prediction | Default | 90-100 | 8 | 5 | 61.5 |
| | | 80-90 | 18 | 2 | 90 |
| | | 70-80 | 45 | 17 | 72.6 |
| | | 60-70 | 81 | 40 | 66.9 |
| | | 50-60 | 97 | 68 | 58.8 |
| | Normal | 40-50 | 70 | 99 | 58.6 |
| | | 30-40 | 40 | 133 | 76.9 |
| | | 20-30 | 29 | 84 | 74.3 |
| | | 10-20 | 4 | 35 | 89.7 |
| | | 0-10 | 0 | 12 | 100 |
| Total | | | 392 | 495 | 69 |

4.2. 랜덤포레스트

랜덤포레스트는 하이퍼 파라미터의 조정 없이 사용해도 비교적 좋은 결과를 보여준다고 알려져 있다. 조정 가능한 하이퍼 파라미터로는 생성할 의사결정나무의 수, 각각의 의사결정나무가 가질 독립변수의 수, 의사결정나무의 최대 깊이(Max Depth) 등이 있다. 적절한 하이퍼 파라미터의 선정을 위해 격자탐색(Grid Search)을 이용할 수 있는데 이는 하이퍼 파라미터에 들어갈 수 있는 여러 값들을 시뮬레이션하고 가장 높은 성능을 보이는 값을 찾는 방법이다. 격자탐색을 통해 시뮬레이션을 수행한 결과 의사결정나무의 최대 깊이가 3보다 커지면 훈련 데이터를 100%의 정확도로 분류하는 과적합 현상을 보여 깊이를 제한하였으나 훈련데이터의 분류 정확도가 낮아지

면 평가데이터의 예측 정확도가 동시에 낮아지는 현상을 보였다. 이에 따라 훈련데이터의 과적합을 고려하지 않고 평가 데이터의 예측 정확도가 가장 높게 나타나도록 모형을 구성했다. 최종적으로 의사결정나무 500개, 각각의 의사결정나무가 가지는 독립변수 6개, 최대 깊이는 제한하지 않도록 구성하였다.

랜덤포레스트 모형에 의한 예측결과는 아래의 <Table 5>와 같다. 예측한 사고 확률은 대부분 20~80% 구간에 속하고 강하게 확신한 0~20% 구간과 80~100% 구간에 속하는 값들은 많지 않았다. 다만 강하게 확신한 구간에서는 숫자는 적지만 높은 예측 정확도를 보여주었고 사고기업과 건전기업의 경계 구간인 50%에 가까워질수록 예측 정확도가 낮아지는 바람직한 결과를 보였다. 최종적으로 훈련 표본을 변경하여 교차 검증을 실시하였고 모두 동일한 결과를 보임을 확인했다.

<Table 5> Results from Random Forest Model

| | | Probability (%) | Actual | | Accuracy (%) |
|------------|---------|-----------------|---------|--------|--------------|
| | | | Default | Normal | |
| Prediction | Default | 90-100 | 0 | 0 | - |
| | | 80-90 | 9 | 1 | 90 |
| | | 70-80 | 43 | 12 | 78.2 |
| | | 60-70 | 88 | 28 | 75.9 |
| | | 50-60 | 119 | 91 | 56.7 |
| | Normal | 40-50 | 84 | 138 | 62.2 |
| | | 30-40 | 32 | 111 | 77.6 |
| | | 20-30 | 16 | 71 | 81.6 |
| | | 10-20 | 1 | 40 | 97.6 |
| | | 0-10 | 0 | 3 | 100 |
| Total | | | 392 | 495 | 70.1 |

4.3. XGBoost, LightGBM

캐글(www.kaggle.com)은 구글이 운영하는 기계학습 대회 플랫폼으로 전 세계 기계학습 전문

가들이 모여 더 높은 예측 정확도를 달성하기 위해 다양한 모형들을 사용하고 있다. XGBoost와 LightGBM은 케글의 기계학습 대회에서 가장 많이 우승한 모형으로 다양한 분야에서 우수한 예측 정확도를 보여주었다. 두 모형 모두 의사결정나무 기반의 앙상블 모형이고 반복해서 예측을 진행하면서 에러를 줄여나가는 그래디언트 부스팅 방식이다. XGBoost는 트리의 깊이보다는 좌우 균형을 맞추어 수평 성장하는 Level-wise tree growth 방식을 사용하고 LightGBM 모형은 리프노드가 분할하면서 트리의 깊이가 깊어지는 Leaf-wise tree growth 방식을 사용한다. LightGBM의 수직성장 방식은 불균형한 트리를 만들지만, 손실(Loss)의 변화를 기준으로 분할하므로 수평 성장 방식에 비해 계산속도가 빠르다. 하이퍼 파라미터 선정 시에도 이와 같은 모형의 특성을 고려하여 선정이 필요하다. 두 방식 모두 하이퍼 파라미터에 따라 많은 성능 차이를 보여주기 때문에 격자탐색을 통해 시뮬레이션하고 하이퍼 파라미터를 선정하였다.

XGBoost 모형의 분류기(Estimators)는 500개로 설정하였고, 훈련데이터의 학습 오류가 감소하지 않을 경우 조기 종결횟수(Early Stopping Round)는 100회로 설정하였다. 의사결정나무의 깊이는 4로 설정하고 오차에 따른 업데이트가 중치인 학습률(Learning Rate)은 0.1로 설정하였다. LightGBM 모형도 비슷한 하이퍼 파라미터를 사용했을 때 좋은 결과를 보였는데 이는 사고 기업을 분류하는데 특정 변수들이 높은 영향을 주기 때문에 의사결정나무가 깊게 형성되지 않아도 충분히 예측 가능한 모형이 구축되기 때문이다. XGBoost 및 LightGBM 모형에 의한 예측 결과는 각각 <Table 6>, <Table 7>과 같다. 두 모형

모두 강하게 확신한 90~100%와 0~10% 구간에서 90% 이상의 높은 예측 정확도를 보였다. XGBoost모형은 60~70%구간에서 53%로 가장 낮은 정확도를 보였고 위 아래로 정확도가 증가하는 합리적인 패턴을 보였다. LightGBM은 사고 기업과 건전기업의 경계 구간인 50%에 가까워질수록 예측 정확도가 낮아지는 추세이지만 70~80% 구간 보다 60~70% 구간이 정확도가 높아지는 역전 현상도 나타났다.

<Table 6> Results from XGBoost Model

| | Probability (%) | Actual | | Accuracy (%) | |
|------------|-----------------|---------|--------|--------------|------|
| | | Default | Normal | | |
| Prediction | Default | 90-100 | 11 | 1 | 91.7 |
| | | 80-90 | 57 | 11 | 83.8 |
| | | 70-80 | 80 | 22 | 78.4 |
| | | 60-70 | 48 | 43 | 52.7 |
| | | 50-60 | 64 | 52 | 55.2 |
| | Normal | 40-50 | 49 | 71 | 59.2 |
| | | 30-40 | 38 | 83 | 68.6 |
| | | 20-30 | 29 | 86 | 74.8 |
| | | 10-20 | 14 | 91 | 86.7 |
| | | 0-10 | 2 | 35 | 94.6 |
| Total | | 392 | 495 | 70.6 | |

<Table 7> Results from LightGBM Model

| | Probability (%) | Actual | | Accuracy (%) | |
|------------|-----------------|---------|--------|--------------|------|
| | | Default | Normal | | |
| Prediction | Default | 90-100 | 18 | 2 | 90 |
| | | 80-90 | 79 | 17 | 82.3 |
| | | 70-80 | 59 | 43 | 57.8 |
| | | 60-70 | 75 | 34 | 68.8 |
| | | 50-60 | 50 | 49 | 50.5 |
| | Normal | 40-50 | 29 | 65 | 69.1 |
| | | 30-40 | 32 | 66 | 67.3 |
| | | 20-30 | 24 | 82 | 77.4 |
| | | 10-20 | 23 | 92 | 80 |
| | | 0-10 | 3 | 45 | 93.8 |
| Total | | 392 | 495 | 71.1 | |

4.4. 심층신경망

인공신경망 기반의 모형은 설계 방식에 따라 심층신경망(Deep Neural Network, DNN), 합성곱신경망(Convolutional Neural Network, CNN), 순환신경망(Recurrent Neural Network, RNN), 장단기 기억 신경망(Long Short-Term Memory Recurrent Neural Network, LSTM) 등이 존재한다. 은닉층의 개수나 노드의 개수를 자유롭게 구성할 수 있고 Drop-out을 사용해 특정 노드를 확률적으로 비활성화시킬 수 있으며 활성화 함수에 따라 학습방식을 변경할 수 있다. 본 연구에서는 최적의 모형을 찾기 위해 하이퍼 파라미터를 변경하며 반복 실험을 진행하였고 다음과 같은 모형을 구성하였다. 입력계층과 출력계층 사이에 위치한 은닉계층은 2개로 구성하였고 각각 128개의 노드를 가지도록 하였다. 최종 계층의 출력 함수는 0과 1사이의 값을 출력하는 시그모이드(Sigmoid) 함수를 사용하고 최종 결과 값을 사고 확률로 사용했다. 은닉층의 활성화 함수는 tanh, 최적화 도구(Optimizer)는 adam, 오차에 따른 업데이트 가중치인 학습률은 0.1, 손실함수는 binary cross-entropy, 과적합 방지를 위한 드롭아웃은 0.3, 미니 배치는 10개를 사용하여 150번 반복 학습을 진행했고 <Table 8>과 같은 결과를 얻었다. 예측 확률 80~100%, 0~10% 구간에서 높은 분류 정확도를 보여주고 경계 구간인 50%에 가까워질수록 분류 정확도가 낮아지는 합리적 결과를 보였다.

<Table 8> Results from DNN Model

| | | Probability (%) | Actual | | Accuracy (%) |
|------------|---------|-----------------|---------|--------|--------------|
| | | | Default | Normal | |
| Prediction | Default | 90-100 | 7 | 1 | 87.5 |
| | | 80-90 | 37 | 5 | 88.1 |
| | | 70-80 | 64 | 29 | 68.8 |
| | | 60-70 | 89 | 44 | 66.9 |
| | | 50-60 | 71 | 57 | 55.5 |
| | Normal | 40-50 | 55 | 83 | 60.1 |
| | | 30-40 | 30 | 92 | 75.4 |
| | | 20-30 | 25 | 101 | 80.2 |
| | | 10-20 | 14 | 64 | 82.1 |
| | | 0-10 | 0 | 19 | 100 |
| Total | | | 392 | 495 | 70.7 |

4.5. 실험 결과 종합

아래 <Table 9>와 같이 전체 표본을 총 5가지 모형별로 산출한 분류 정확도는 70% 전후로 나타났다. 로지스틱 회귀 모형이 69%로 가장 낮고 LightGBM 모형이 71.1%로 가장 높은 정확도를 보였다. 표본별 사고 발생 확률을 구간으로 나누어 분류 정확도를 살펴보고 사고기업을 건전기업으로 예측하는 제2종 오류에 대해서도 살펴보고자 한다.

사고 발생 확률 구간별 분류 정확도를 살펴보면 로지스틱 회귀모형은 건전기업이라고 가장 강하게 예측하는 확률 구간(0~10%)에서 가장 높은 분류 정확도(100%)를 보이고 사고와 건전의 경계 구간인 50%에 가까워질수록 분류 정확도가 낮아지는 합리적인 결과를 보였으나 사고기업이라고 강하게 예측한 확률 구간(90~100%)에

<Table 9> Overall Prediction Accuracy and Type 2 Error

| | Logit | Random Forest | XGBoost | LightGBM | DNN |
|-------------------------|-------|---------------|---------|----------|------|
| Prediction Accuracy (%) | 69 | 70.1 | 70.6 | 71.1 | 70.7 |
| Type 2 Error (%) | 34.6 | 33.8 | 33.2 | 34 | 33.7 |

<Table 10> Type 2 Error for High(80-100%) Prob. of Occurrence of a Default

| | Logit | Random Forest | XGBoost | LightGBM | DNN |
|-----------------------|-------|---------------|---------|----------|-----|
| Type 2 Error Rate (%) | 21.21 | 10 | 15.00 | 16.38 | 12 |
| The # of Cases | 33 | 10 | 80 | 116 | 50 |

서 상대적으로 낮은 분류 정확도(61.5%)를 보여 주는 이상 현상이 나타났다. LightGBM도 사고기업과 건전기업을 경계 구간인 50%에 가까워질수록 예측 정확도가 낮아지는 추세이지만 70~80% 구간 보다 60~70% 구간이 정확도가 높아지는 역전 현상을 보였다. 반면 랜덤포레스트, XGBoost, 심층신경망 모형은 사고기업 및 건전기업을 확신하는 정도가 높은 확률 구간에서 분류 정확도가 높게 나타나고 사고와 건전의 경계 구간인 50%에 가까워질수록 분류 정확도가 낮아지는 합리적 추세를 보였다.

다음으로 사고 발생 확률구간별 표본들의 분포를 위의 <Table 10>에서 살펴보면 LightGBM과 XGBoost모형은 사고기업과 건전기업 분류 모두에서 다른 모형들에 비해 사고 및 건전을 높게 확신하는 구간에 많은 표본들이 분포하고 있다(각각 80건 및 116건). 사고기업 예측에서 랜덤포레스트 모형은 80~100% 사고 확률 구간으로 단 10건의 표본만 분류했지만 LightGBM 모형은 116개 표본을 분류했다. 랜덤포레스트 모형이 적은 수의 표본을 조금 더 정확하게 분류하는 보수적인 결과를 보여주었다면, LightGBM과 XGBoost와 같은 부스팅 기반 알고리즘은 예측 정확도는 조금 떨어지지만 많은 양의 표본을 양극단의 범주로 분류했다.

제2종 오류의 중요성 및 전체적 예측 정확도를 종합적으로 고려하면 XGBoost와 심층신경망 모형이 전체 5가지 모형 중에서 가장 우수한 것으로 평가된다. 다음으로 랜덤포레스트와

LightGBM이 그 다음으로 우수하며 로지스틱 모형이 가장 낮은 성과를 보였다. 각 모형들이 가지는 장점도 분명히 존재하는데 랜덤포레스트 모형이 높은 사고 확률로 예측한 표본은 거의 100%의 정확도로 사고가 발생했으므로 모형별 예측 결과에 가중치를 부여하여 투표를 통해 사고 확률을 추정하는 것은 좋은 대안이 될 수 있다.

5. 결론

5.1. 연구 요약

본 연구에서는 한국무역보험공사의 수출신용보증(선적전) 이용기업의 사고 발생 가능성을 예측하고자 총 5개의 기계학습모형(로지스틱 회귀 모형, 랜덤포레스트, XGBoost, LightGBM, 심층신경망)을 사용하였다. 이를 위해 2014년부터 2018년 동안 수출신용보증(선적전) 데이터 2만건을 수집하였고 CNN 언더샘플링을 통해 사고기업과 건전기업을의 비율을 맞추어 3천건으로 축소하였다. 입력변수는 상관관계를 고려한 25개 재무변수와 5개 비재무 변수를 사용하였다. 모형의 성능 판단을 위해 개별 표본들의 사고 확률을 모두 구한 후 10% 단위의 10개 구간으로 나누어 구간별 예측 정확도 및 그 비중을 확인하고 사고기업을 건전기업으로 예측하는 제2종 오류 구간인 사고확률 50~100% 구간을 중점적으

로 분석하였다.

예측모형들은 69~71.1% 수준의 예측 정확도를 보여주었고 제2종 오류는 33.2~34.6% 수준으로 나타났다. 부스팅 기반의 XGBoost와 LightGBM 모형과 인공신경망 기반의 심층신경망 모형은 많은 표본들을 사고 확률이 0~20%, 80~100%의 양극단구간으로 예측하여 80%이상의 높은 예측 정확도를 보였고, 랜덤포레스트 모형은 적은 수의 표본을 양극단의 사고 확률 구간에서 90%이상 정확도로 예측하여 모형의 현실적 활용 가능성을 높였다. 반면 로지스틱 회귀모형은 제2종 오류 구간에서 가장 극단 구간의 예측 정확도가 61.5%로 낮게 측정되는 이상 현상이 존재했다. 구간별 예측 정확도 및 제2종 오류를 종합적으로 고려하면 XGBoost와 심층신경망이 5개의 모형 중 가장 우수한 것으로 평가된다.

5.2. 시사점 및 향후 연구

선행 연구들은 사고 확률이 50%인 경우를 기준으로 사고기업과 건전기업을 분류하고 ROC Curve 등의 기법을 이용하여 하나의 Score를 가지고 모형의 성능을 측정했으나 본 연구는 개별 표본들의 사고 확률을 구간으로 나누어 예측 정확도와 그 비중을 분석하였다. 이를 통해 각각의 모형은 전체 표본의 예측 정확도는 유사하지만 사고 확률 구간별 예측 정확도와 그 비중은 상당히 달라짐을 확인하였다.

본 연구는 한국무역보험공사가 보유한 수출신용보증 데이터에 기계학습 기법을 적용한 최초의 연구이며 다음과 같은 실무적인 시사점을 갖는다. 첫째, 낮은 사고 확률로 예측된 소액 보증건은 심사자의 개입 없이 즉시 한도를 책정하여 고객에게 빠른 서비스를 제공할 수 있다. 이를

통해 사고 확률이 높은 특이 건과 거액 심사 건에 역량을 집중하여 인적자원의 관리를 효율화할 수 있다. 둘째, 기발급된 보증서의 사고 확률을 주기적으로 점검하고 보증기간 동안 사고 위험이 높은 기업들을 집중적으로 감시하는 사후 위험관리 체계를 구축할 수 있다.

본 연구에서 사용한 학습변수가 일반적인 기업부도 예측에서 사용되는 재무변수들로 한정되어 있다는 점에서 한계를 보였으나 향후 연구로 공사의 다른 보험종목들의 사고 예측이 진행되는다면 추가적인 학습변수를 확보하여 예측 모형의 정확도를 높일 수 있을 것이다. 또한 수출 활동을 하는 기업들은 환율 변동(Krugman, 1986)이나 해외에 자회사를 소유하였는지 여부(Gumpert et al., 2020) 등의 영향을 많이 받는 것으로 알려져 있는데, 특정 시점의 거시경제 지표 등을 추가변수로 고려한다면 보다 우수한 예측 정확도를 가질 수 있고 설득력을 향상시킬 수 있을 것으로 기대된다.

본 연구를 계기로 삼아 실무에서 공사 등이 보유한 양질의 데이터에 다양한 기계학습 및 빅데이터 분석이 활발하게 적용되기를 바란다. 예를 들어, 물품을 수출한 후 수입자로부터 수출대금을 받을 수 없게 되어 발생하는 손실을 보상하는 단기수출보험의 경우 대금 미지급 수입자의 데이터를 기계학습모형으로 학습시키고 물품 수출 이전에 대금 미회수 확률을 예측하면 위험관리를 효과적으로 수행할 수 있을 것이다. 또한 공사가 관리하는 국내 및 국외 부실채권의 회수 데이터를 이용해 기계학습으로 학습시키고 회수 가능 금액을 추정하는데 활용할 수도 있다. 회수율이 낮은 채권은 적극적인 아웃소싱을 통해 관리의 부담을 경감시킬 수 있고 회수 가능성이 높은 채권에 적극적인 회수 역량을 투입하여 회수

율을 극대화할 수 있을 것이다.

참고문헌(References)

- Ahn, H., K. Kim, and I. Han, "Purchase Prediction Model using the Support Vector Machine," *Journal of Intelligence and Information Systems*, Vol.11, No.2 (2005), 69~81.
- Ahn, S. M., "Deep Learning Architectures and Applications," *Journal of Intelligence and Information Systems*, Vol.22, No.2 (2016), 127~142.
- Altman, E. I., "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *Journal of Finance*, Vol.23, No.4 (1968), 589~609.
- Beaver, W. H., "Financial ratios as predictors of failure," *Journal of Accounting Research*, Vol.4 (1966), 71~111.
- Cha, S. J. and J. S. Kang, "Corporate Default Prediction Model using Deep Learning Time Series Algorithm, RNN and LSTM," *Journal of Intelligence and Information Systems*, Vol.24, No.4 (2018), 1~32.
- Chen, T. and C. Gustrin, "XGBoost: A Scalable Tree Boosting System," 22nd ACM KDD Conference on Knowledge Discovery and Data Mining (2016), 785~794.
- Elmer, P. J. and D. M. Borowski, "An Expert System Approach to Financial Analysis: The Case of S&L Bankruptcy," *Financial Management*, Vol.17, No.3 (1988), 66~76.
- Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *In Advances in Neural Information Processing Systems*, (2014), 2672~2680.
- Gumpert, A., H. Li, A. Moxnes, N. Ramondo, and F. Tintelnot, "The Life-Cycle Dynamics of Exporters and Multinational Firms," *Journal of International Economics*, Vol.126 (2020), 1~38.
- Han, I. G., Y. S. Kwon, and K. C. Lee, "Development of Intelligent Corporate Credit Evaluation System," *Korean Management Review*, Vol.24, No.4 (1995), 91~118.
- Han, I. G., H. K. Cho, and K. S. Shin, "The Hybrid System for Credit Rating," *Journal of the Korean Operations Research and Management Science Society*, Vol.22, No.3 (1997), 163~173.
- Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," 31st Conference on Neural Information Processing Systems, (2017), CA, USA.
- Kim, K. W., "Symptoms and Predictions of Business Failure based on Financial Ratios," *Korean Management Review*, Vol.16, No.2(1987), 263~316.
- Kim, J. R., S. B. Kim, and J. H. Nam, "The Performance Measurement of Credit Guarantee and Methods of Improvement of Its System," *Korea Review of Applied Economics*, Vol.16, No.2 (2014), 33~64.
- Kim, Y. T. and M. H. Kim, "An Artificial Neural Network Model for Business Failure Prediction," *Korean Journal of Accounting Research*, Vol.6, No.1 (2001), 275~294.
- Krugman, P., *Pricing to Market when the Exchange Rate Changes*, MIT Press, Massachusetts, 1986.

- Kwon, H. K., D. K. Lee, and M. S. Shin, "Dynamic Forecasts of Bankruptcy with Recurrent Neural Network Model," *Journal of Intelligence and Information Systems*, Vol.23, No.3 (2017), 139~153.
- Lee, H. J., "A Study on Prediction Model of Peer-to-Peer (P2P) Social Lending Debtor using Deep Learning Technique," *Journal of Digital Contents Society*, Vol.20, No.7 (2019), 1409~1416.
- Ohlson, J. A., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, Vol.18, No.1 (1980), 109~131.
- Schmidhuber, J., "Deep Learning in Neural Networks: An Overview," *Neural Networks*, Vol.61 (2015), 85~117.
- Seo, C. S. and B. K. Lee, "A Study on the Optimal Credit Guarantee Fund Operation Model: Focused on Local Credit Guarantee Foundations," *Korea Trade Review*, Vol.31, No.5 (2006), 197~217.
- Shin, S. H., H. J. Lee, and J. J. Ahn, "A Study on Initial Price Change Prediction of IPO Shares using Non-financial Information", *Journal of the Korean Data And Information Science Society*, Vol.29, No.2 (2018), 425~439.
- Yoon, J. M., "Effectiveness Analysis of Credit Card Default Risk with Deep Learning Neural Network," *Journal of Money and Finance*, Vol.33, No.1 (2019), 151~183.

Abstract

The Prediction of Export Credit Guarantee Accident using Machine Learning

Cho, Jaeyoung* · Joo, Jihwan** · Han, Ingoo***

The government recently announced various policies for developing big-data and artificial intelligence fields to provide a great opportunity to the public with respect to disclosure of high-quality data within public institutions. KSURE(Korea Trade Insurance Corporation) is a major public institution for financial policy in Korea, and thus the company is strongly committed to backing export companies with various systems. Nevertheless, there are still fewer cases of realized business model based on big-data analyses.

In this situation, this paper aims to develop a new business model which can be applied to an ex-ante prediction for the likelihood of the insurance accident of credit guarantee. We utilize internal data from KSURE which supports export companies in Korea and apply machine learning models. Then, we conduct performance comparison among the predictive models including Logistic Regression, Random Forest, XGBoost, LightGBM, and DNN(Deep Neural Network).

For decades, many researchers have tried to find better models which can help to predict bankruptcy since the ex-ante prediction is crucial for corporate managers, investors, creditors, and other stakeholders. The development of the prediction for financial distress or bankruptcy was originated from Smith(1930), Fitzpatrick(1932), or Merwin(1942). One of the most famous models is the Altman's Z-score model(Altman, 1968) which was based on the multiple discriminant analysis. This model is widely used in both research and practice by this time. The author suggests the score model that utilizes five key financial ratios to predict the probability of bankruptcy in the next two years. Ohlson(1980) introduces logit model to complement some limitations of previous models. Furthermore, Elmer and Borowski(1988) develop and examine a rule-based, automated system which conducts the financial analysis of savings and loans.

Since the 1980s, researchers in Korea have started to examine analyses on the prediction of financial distress or bankruptcy. Kim(1987) analyzes financial ratios and develops the prediction model. Also, Han

* Korea Trade Insurance Corporation

** Corresponding author: Jihwan Joo

School of Management Engineering, College of Business, KAIST

85 Hoegiro, Dongdaemun-gu, Seoul, Korea 02455

Tel: +82-2-958-3604, Fax: +82-2-958-3620, E-mail: jh1012.joo@kaist.ac.kr

*** College of Business, KAIST

et al.(1995, 1996, 1997, 2003, 2005, 2006) construct the prediction model using various techniques including artificial neural network. Yang(1996) introduces multiple discriminant analysis and logit model. Besides, Kim and Kim(2001) utilize artificial neural network techniques for ex-ante prediction of insolvent enterprises. After that, many scholars have been trying to predict financial distress or bankruptcy more precisely based on diverse models such as Random Forest or SVM. One major distinction of our research from the previous research is that we focus on examining the predicted probability of default for each sample case, not only on investigating the classification accuracy of each model for the entire sample.

Most predictive models in this paper show that the level of the accuracy of classification is about 70% based on the entire sample. To be specific, LightGBM model shows the highest accuracy of 71.1% and Logit model indicates the lowest accuracy of 69%. However, we confirm that there are open to multiple interpretations. In the context of the business, we have to put more emphasis on efforts to minimize type 2 error which causes more harmful operating losses for the guaranty company. Thus, we also compare the classification accuracy by splitting predicted probability of the default into ten equal intervals.

When we examine the classification accuracy for each interval, Logit model has the highest accuracy of 100% for 0~10% of the predicted probability of the default, however, Logit model has a relatively lower accuracy of 61.5% for 90~100% of the predicted probability of the default. On the other hand, Random Forest, XGBoost, LightGBM, and DNN indicate more desirable results since they indicate a higher level of accuracy for both 0~10% and 90~100% of the predicted probability of the default but have a lower level of accuracy around 50% of the predicted probability of the default.

When it comes to the distribution of samples for each predicted probability of the default, both LightGBM and XGBoost models have a relatively large number of samples for both 0~10% and 90~100% of the predicted probability of the default. Although Random Forest model has an advantage with regard to the perspective of classification accuracy with small number of cases, LightGBM or XGBoost could become a more desirable model since they classify large number of cases into the two extreme intervals of the predicted probability of the default, even allowing for their relatively low classification accuracy.

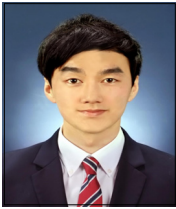
Considering the importance of type 2 error and total prediction accuracy, XGBoost and DNN show superior performance. Next, Random Forest and LightGBM show good results, but logistic regression shows the worst performance. However, each predictive model has a comparative advantage in terms of various evaluation standards. For instance, Random Forest model shows almost 100% accuracy for samples which are expected to have a high level of the probability of default. Collectively, we can construct more comprehensive ensemble models which contain multiple classification machine learning models and conduct majority voting for maximizing its overall performance.

Key Words : Credit Guarantee, Guarantee Accident, Bankruptcy Prediction, KSURE, Machine Learning

Received : January 5, 2021 Revised : February 13, 2021 Accepted : March 8, 2021

Corresponding Author : Jihwan Joo

저 자 소개



조재영

현재 한국무역보험공사 정보화사업부에서 과장으로 재직 중이다. 서울과학기술대학교 컴퓨터공학과에서 학사학위를 한국과학기술원 정보미디어경영대학원에서 석사학위를 취득하였다. 주요 관심분야는 핀테크 기술의 비즈니스 활용, 기계학습 기법을 이용한 빅데이터 분석 등이다.



주지환

현재 KAIST 경영공학부 박사과정에 재학 중이다. 중앙대학교 경영학부를 졸업하였다. 주요 관심분야는 기업공시자료분석, 기업가치평가, 재무제표분석 등이다.



한인구

현재 KAIST 경영대학 교수로 재직 중이다. 서울대학교에서 국제경제학을 전공하고 KAIST에서 경영과학 전공으로 석사학위를 받고 University of Illinois at Urbana-Champaign에서 회계정보시스템 전공으로 경영학박사학위를 받았다. 주요 관심분야는 인공지능을 이용한 재무분석, 신용평가시스템 및 가치평가 등이다. 국내외 학술지에 150여편의 논문을 발표하였다. 한국지능정보시스템학회장, 한국지식경영학회장, 한국경영정보학회장, 한국경영학회장을 역임하였다.