

전자결제서비스 이용 사업자 폐업 예측에서 비재무정보 활용을 통한 머신러닝 모델의 정확도 향상에 관한 연구

공현정

한국과학기술원 경영공학부
(hj.gong@alumni.kaist.ac.kr)

황유진

한국과학기술원 경영공학부
(hegene3686@kaist.ac.kr)

박성혁

한국과학기술원 경영공학부
(sunghyuk.park@kaist.ac.kr)

기업 부도 예측에 관한 연구는 재무정보를 중심으로 연구되어 왔다. 기업의 재무정보는 분기별로 갱신되기 때문에 실시간으로 기업의 폐업 가능성을 예측하는 데 있어 적시성이 부족하게 되는 문제가 발생한다. 이를 개선하고자 하는 평가 기업에서는 대상 기업의 건전성을 판단하기 위한 재무정보 외의 정보를 활용한 기업의 건전성을 판단하는 방법이 필요하다. 이를 위해 정보 기술의 발달로 기업에 대한 비재무정보 수집이 용이해지면서 기업 부도 예측에 재무정보 외의 추가적인 변수와 여러 가지 방법론을 적용하는 연구가 진행되어 왔으며, 이 중에서도 어떤 변수들이 기업의 부도를 예측하는데 영향을 주는지를 밝히는 것이 중요한 연구 과제가 되었다. 본 연구에서는 전자결제서비스를 이용하는 사업자의 폐업을 예측할 때 비재무정보를 구성하는 전자결제 정보들이 얼마나 영향을 미치는지를 살펴보고, 재무정보와 비재무정보 결합에 따른 폐업 예측 정확도 차이를 살펴보았다. 구체적으로, 재무정보 모형과 비재무정보 모형, 그리고 이를 결합한 모형으로 구성된 세 가지 연구 모형을 설계하였으며 Multi Layer Perceptron(MLP) 알고리즘을 포함한 여섯 가지 알고리즘으로 폐업 예측 정확도를 확인하였다. 재무정보와 비재무정보를 결합한 모형이 가장 높은 예측 정확도를 보였으며, 그 다음으로는 비재무정보 모형, 재무정보 모형의 순서로 예측 정확도가 확인되었다. 알고리즘별 폐업 예측 정확도는 여섯 가지의 알고리즘 중 XGBoost가 가장 높은 예측 정확도를 보였다. 사업자의 폐업 예측에 활용된 전체 87개의 변수를 대상으로 상대적 중요도를 살펴본 결과 폐업 예측에 중요하게 영향을 미친 변수는 상위 20개 중 70% 이상이 비재무정보인 것으로 확인되었다. 이를 통해 비재무정보의 전자결제 정보가 사업자의 폐업을 예측하는 중요한 변수임을 확인하였으며, 비재무정보가 재무정보의 대안적 정보로서 활용할 수 있는 가능성 역시 살펴볼 수 있었다. 본 연구를 기반으로 사업자의 폐업을 예측할 수 있는 정보로서 비재무정보의 수집과 활용에 대한 중요성을 인식하고 기업의 의사결정에 활용할 수 있는 방안에 대해서도 다루었다.

주제어 : 폐업, 예측 정확도, 전자상거래, 전자결제서비스, 전자결제대행사

논문접수일 : 2023년 6월 8일

논문수정일 : 2023년 9월 17일

게재확정일 : 2023년 9월 18일

원고유형 : Fast Track

교신저자 : 박성혁

1. 서론

최근 금리 상승의 영향으로 기업의 예상 부도 확률은 높아지고 기업의 신용 위험에 대한 경계감은 커져 기업에 대한 평가가 더 중요해지고 있다. 기업의 부도는 이해관계자들에게 연쇄적인 피해를 줄 수 있어 기업의 건전성을 판단하는 일은

매우 중요하다. 기업의 부도를 예측하기 위해 과거에서부터 기업의 경영 성과를 반기, 연도별로 나타낼 수 있는 객관적 지표로서 재무정보를 활용하였다. 하지만 이와 같은 정보는 기업의 부도를 예측하는 적시성이 부족하여 기업의 현재 상태를 판단하는 데에는 한계가 있다. 이에 따라 기업에서는 이해관계에 있는 상대 기업의 건전성을 적시에

판단하기 위한 재무정보 외의 정보를 활용하여 기업의 건전성을 판단하는 방법이 필요하다.

기업의 부도 예측 연구는 두 가지 형태로 발전되었다. 기업의 부도 예측에 있어 기본 변수로 활용되었던 재무정보 외에 어떠한 변수가 기업의 현 상태를 나타낼 수 있는 정보인지를 중심으로 거시 지표, 시장 정보, 뉴스 기사 등의 정보 원천을 활용한 측면과 부도 예측 모형 구축에 적용할 수 있는 통계적, 기계적 데이터 분석 방법에 대한 연구로 발전되고 있다. 특히 최근에는 정보 기술의 발달로 과거 활용이 어려웠던 비정형 데이터의 수집이 가능해지면서 비재무정보가 기업 부도 예측에 영향을 주는 요소인지에 대한 유용성 판단이 중요한 연구 과제가 되었다.

본 연구에서는 전자결제대행사(Payment Gateway)와 전자결제서비스를 이용하는 사업자 간에 발생할 수 있는 채권과 채무의 위험에 기인하여 전자결제 정보들이 전자결제서비스 이용 사업자의 폐업 예측에 어떠한 영향을 미치는지 확인하였다. 전자결제대행사는 전자결제서비스를 이용하는 사업자에게 전자결제를 통해 발생한 매출액을 지급한 후에 거래 취소가 발생하는 경우 해당 금액을 회수해야 하기 때문에 전자결제서비스를 이용하는 사업자의 건전성을 판단하는 일은 매우 중요하다. 전자상거래에서 전자결제서비스를 이용하는 사업자는 재무정보의 신뢰도가 낮은 개인 사업자의 비중이 전체 이용 사업자의 약 69%를 차지(전자결제대행사 A사 기준)하기 때문에 본 연구는 재무정보 외의 어떤 정보들이 전자상거래 사업자의 폐업 예측에 유용한 정보인지를 확인하는 데에 의의가 있다. 본 연구에서는 전자결제 정보가 전자결제서비스 이용 사업자의 예측 정확도 향상에 미치는 영향을 살펴보기 위해 재무정보와 재무정보 외의 정보로 전자상거래의 실제

거래 데이터를 활용하여 재무정보와 비재무정보, 이를 결합한 세 가지 모형을 설계하였으며 전자결제 정보들이 재무정보의 대안적 정보로서의 활용 가능성을 확인하고자 한다.

2. 선행연구

기업의 부도 예측 연구는 부도를 예측할 수 있는 변수로 재무정보 외의 정보를 활용하여 기업 부도 예측 정확도를 높이는 방향의 연구가 꾸준히 진행되고 있다. 장윤신(2013)은 기업 부도에 영향을 주는 재무비율 외의 변수로 거시경제 지표를 활용하여 금융 위기 기간과 금융 위기 이전으로 어떠한 영향이 있는지 검증하였다. 설명변수는 거시경제 지표와 재무비율로 종속변수는 EDF(Expected Default Frequency)로 설정하여 정상 기업과 부도기업의 유의한 차이를 검증하였다. 재무변수와 거시변수를 각각 고려한 경우보다 재무변수와 재무변수 외의 정보로서 거시변수를 함께 고려한 경우가 EDF와의 상관관계가 높았다. 재무정보만 고려한 경우보다 거시변수를 함께 고려하였을 때 예상 부도 확률을 설명하는 영향력이 커지며 이는 거시적 환경을 나타내는 거시변수가 기업의 부도율에 영향을 미치는 변수로서 기업 부도 예측 모형에 거시변수를 포함해야 하는 필요성과 활용성을 제시하였다. 권누리, 김영민, 최광신(2019)는 2000년~2016년 상장기업 중 금융사를 제외한 기업을 대상으로 건설기업과 부도 기업에 시장 자료와 거시경제 자료를 기업 부도 예측 설명변수로 사용하였다. 재무정보의 변수로 감가상각비, 매출액, 부채총계, 영업이익, 이익잉여금, 자본총계, 총자산 등을 적용하였고 재무정보 외의 정보로서 시장변수와 거시경제 변수를 활용

하여 기업 부도 예측 모형의 성능을 최적화할 수 있는 변수를 선별하였으며 국제 경제 상황에 민감하게 반응하는 우리나라의 경우에는 기업의 부도 예측을 위해 재무정보를 주요 변수로 고려하는 것보다 거시 경제 상황을 주요 변수로 고려하는 것을 제안하였다.

최근 재무정보 외의 정보로서 비정형데이터를 활용한 기업 부도 예측 연구도 진행되고 있다. 김찬송과 신민수(2019)는 기업 관련 뉴스의 감정 분석을 하여 부도 예측에 미치는 영향과 이에 따른 효과를 연구하였다. 부도 시점의 이전과 이후의 뉴스 데이터를 활용하여 뉴스의 단어 감성을 점수화하여 기업의 점수를 산출하고 재무변수와 결합하여 부도 예측성을 평가하였다. 부도 예측에 가장 큰 효과가 있는 뉴스 분류는 부도 상황을 전달하는 스트레이트 뉴스보다 해설 및 논평, 칼럼이 부도 나기 전 상황에 대한 감성을 잘 반영하여 부도 예측에 더 효과적인 변수임을 밝혔으며, 재무변수만을 적용했을 때의 부도 예측 성능 보다 비정형데이터인 뉴스 감정 변수를 결합하였을 때의 예측 성능이 더 높음을 확인하였다. 모델의 성능 측정은 로지스틱 회귀분석, SVM, 인공신경망 알고리즘을 사용하였으며 인공신경망 알고리즘에서 가장 높은 예측력을 보였다. 또한, 최정원, 오세경, 장재원(2017)도 기업 부도 예측에서 비재무정보 중 비정형 데이터의 뉴스 정보를 활용하여 새로운 정보 원천으로서의 활용 방법을 제안하였다. 기본 정보의 원천인 재무정보와 재무정보 외의 변수들인 거시경제 정보, 증권 시장 정보를 점진적으로 추가한 모형을 설계하여 기업 부도 예측 정확도를 비교하였을 때 비정형 데이터인 뉴스 정보를 추가함으로써 기업 부도 예측 성능이 향상되는 것을 확인하였으며, 뉴스 정보를 기반으로 한 예측 모형과 시장 정보 기반의 예측 모형인

KMV 모형 간 유사한 결론을 도출하면서 기업 부도 관련 뉴스가 실제 부도가 발생하는 시점 이전에 부도 가능성을 알려줄 수 있는 지표로서의 중요성을 언급하였다.

3. 연구의 이론적 방법론

3.1. 이진분류 방법론

Logistic Regression은 기업의 부도 여부를 판단하는 기본적인 예측 방법론으로 활용된다. 독립변수와 종속변수를 입력하여 예측값이 실제값과 가깝도록 학습하여 새롭게 입력되는 데이터가 어떤 범주에 속할 확률을 로짓 변환과 최대우도추정법을 통해 0에서 1 사이의 값으로 예측하고 그 확률에 따라 가능성이 더 높은 범주에 속하는 것으로 분류해주는 지도 학습 알고리즘이다. Logistic Regression은 적은 데이터로도 계수를 이용하여 각 변수의 중요성을 쉽게 판단할 수 있으며 독립변수와 종속변수 간의 인과관계를 설명하기에 용이하여 기업 부도 예측의 방법론 간 예측 정확도를 비교하는 기준으로 활용된다.

Decision Tree Classifier 모형은 예측값에 대한 의사결정 규칙을 나무 형태로 나타내 주는 모형이다. 수집된 데이터들을 분석하여 데이터들 사이에 존재하는 패턴들의 속성을 조합으로 나타내는 분류 모형으로 새로운 데이터에 대한 분류나 해당 범주 값을 예측하는 목적으로 사용한다. 데이터의 속성으로부터 분류기준을 판별하고 연속적으로 발생하는 의사결정 문제를 나무 모양으로 시각화할 수 있기 때문에 독립변수와 종속변수 간 해석이 용이하여 전자결제서비스를 이용하는 사업자의 정보들이 사업자의 폐업여부를

예측하는 데 어떠한 정보가 중요하게 작용하는지 확인할 수 있다.

3.2. 머신러닝 기반의 방법론

전자결제서비스를 이용하는 사업자의 폐업 전 조기 경보로서 활용하기 위해서는 높은 성능의 알고리즘을 활용해야 한다. 이에 따라 폐업 예측 정확도를 높일 수 있는 정보들과 해당 정보들도 가장 높은 예측 정확도의 알고리즘을 확인하는 것이 중요하다. 본 연구에서는 비재무정보의 전자결제 정보가 사업자의 폐업 예측 성능에 어떠한 영향을 미치는지 확인하고 이 정보들을 활용하여 최적의 예측 정확도를 보이는 알고리즘을 확인하고자 여러 가지 방법론에 적용하여 예측 정확도를 비교하였다.

Random Forest는 무작위로 Decision Tree를 랜덤하게 형성한 여러 개의 모델 결과를 결합하여 데이터를 분류하거나 예측하는 앙상블 기법으로 Bagging 방식과 Boosting 방식이 있다. Random Forest는 각 의사결정 나무를 훈련할 때 독립변수를 추출하는 과정의 무작위성을 주기 위해 부트스트랩(Bootstrap) 방법을 활용하여 데이터를 결합하는 Bagging 방식으로 예측한다. 이는 각 데이터를 샘플링하여 학습한 뒤 모든 분류기가 voting을 통해 예측하기 때문에 데이터의 변동에 따라 다른 결과가 나올 수 있는 불안정성과 데이터가 적은 경우에 발생하는 과대 적합의 문제를 보완한다. Breiman(2001)은 Decision tree보다 높은 예측 방법론으로 Random forests 방법론을 제안하였으며, 김성진과 안현철(2016)은 Random forests가 불균형한 이항 분류 예측에 높은 예측 정확도를 보인다고 하였다. 또한 해당 알고리즘은 독립변수 선정에 비교적 제약이 적어 다양한 전자결제 정보

들을 독립변수로 활용하는 본 연구에 적합한 방법론이 될 수 있다.

eXtreme Gradient Boosting(XGBoost) 알고리즘은 기본 학습기를 의사결정나무로 하며 학습된 모형의 약점을 잔차로 보완하면서 새로운 모형을 순차적으로 결합하여 학습하는 앙상블 기법의 Boosting 방식의 모형이다. 잔차를 이용하여 이전 학습된 모형에서 실제 값을 정확하게 예측하지 못한 차이를 줄여주는 최적의 함수를 찾아 최종 모형을 생성하는 Gradient Boosting 알고리즘의 과적합을 보완하는 파라미터가 추가된 모형이다. 사업자의 폐업을 예측함에 있어 폐업 여부에 영향을 미치는 변수 간의 설명보다 사업자의 건전성 유무에 대한 판단 결과가 중요할 때 tree 기반 알고리즘의 Decision Tree Classifier와 Random Forest, XGBoost 간의 예측 정확도를 비교하여 활용할 수 있다.

이재식과 한재홍(1995)은 재무정보만을 활용한 기업 부도 예측에 한계가 있음을 언급하고 비재무정보를 활용하여 인공지능경망 기반의 부도 예측 모형을 제시하였고, 재무정보가 부족한 기업일수록 예측 정확도가 더 높음을 제시하였다. 전자결제서비스 이용 사업자의 경우에도 재무정보의 신뢰성이 부족한 사업자의 비중이 높기 때문에 비재무정보 활용에 따른 최적의 예측 정확도를 확인하기 위해 머신러닝 기반의 Multi Layer Perceptron을 활용하였다. 전자결제 사업자의 폐업예측에 활용할 변수들의 다양성을 고려할 수 있는 알고리즘으로 과적합 발생이 상대적으로 낮고 선형 분류와 비선형 분류에서도 사용될 수 있는 Support Vector Machine(SVM)을 활용하였다(박정민, 김경재, 한인규, 2002). SVM은 분류에 사용되는 머신러닝 기법으로 두 집단을 구분하는 판정 경계선에서 두 집단의 관측치 데이터 간의 거리인 마진(margin)을 최대화하여 마진의 넓이를 최대화하는 서포트

벡터(Support Vector)를 찾는 기법으로 비선형 분류에 커널 함수를 사용하여 예측을 수행한다.

4. 연구방법 및 연구설계

4.1. 데이터 수집

비재무정보의 데이터는 '16년~ '21년 동안 전자결제대행사 A기업의 전자결제서비스를 신청한 사업자의 전자결제 이용 정보 및 거래 정보를 수집하였으며 해당 사업자의 재무정보는 기업신용평가사인 A사로부터 수집 가능한 최근 회계연도를 기준으로 1년 전(t-1)과 2년 전(t-2)의 재무정보를 수집하였다. 사업자가 폐업한 경우에는 폐업한 연도의 직전 회계연도(1월~12월)를 기준하였다. 수집한 재무정보는 재무상태표와 손익

계산서 상의 재무 계정 31개이며<표 1>, 수집한 비재무정보는 거래와 관련된 지표를 기준으로 33개이다<표 2>.

종속변수로 활용된 사업자 휴·폐업 정보는 국세청의 공공 데이터 포털에서 오픈 API를 통해 수집하였다. 사업자의 폐업 여부를 관찰한 기간은 연구 분석에 활용한 표본 5,725개를 대상으로 '22년 11월까지로 한정하였다

4.2. 변수의 정의

4.2.1. 종속변수

본 연구에서의 종속변수는 전자결제대행사를 기준으로 전자결제서비스 이용 사업자의 폐업 여부를 판단하는 상태로 정의하였다. 폐업자는 국세청에서 정보를 보유하고 있는 사업자 중 휴·

<표 1> 재무정보의 유형

구분	분류	수집한 재무 데이터
재무정보	손익계산서 (13개)	매출액, 영업이익, 당기순이익, 자산총계, 매출원가, 이자비용, 매출충이익, 판매비와관리비, 영업외수익, 감가상각비, 이자수익, 영업외비용, 법인세 비용 차감 전 영업이익
	대차대조표 (18개)	자본총계, 유동자산, 유형자산, 재고자산, 단기차입금, 비유동자산, 부채총계, 비유동부채, 현금및현금성자산, 매입채무, 유동부채, 자본금, 유동성장기차입금, 기타당좌자산, 단기금융상품, 매출채권, 납입자본금, 장기차입금

<표 2> 비재무정보의 유형

구분	분류	수집한 비재무 데이터
비재무정보	사업자 기본 정보 (10개)	업종구분, 사업자유형, 서비스가입경로, 보증보험가입여부, 주요판매상품, 실물/비실물, 정산한도, 서비스신청ID개수, 정산주기I, 정산주기II
	전자결제 이용 정보 (6개)	전자결제이용기간, 서비스사용ID개수, 전자결제이용결제수단, 간편결제수단사용여부, 소비자불만접수건, 소비자불만접수월
	거래 정보 (17개)	전체결제건, 전체결제금액, 순결제건, 순결제금액, 취소건, 취소금액, 한달정산횟수, 한달정산금액, 미수금발생건, 미수금발생누계, 미수금상계기간, 직전3개월거래건, 직전3개월거래금액, 직전3개월순결제건, 직전3개월순결제금액, 직전3개월취소건, 직전3개월취소금액

폐업 신고(부가가치세법 제8조 및 동법 시행령 제13조)에 따라 사업자가 실질적으로 사업장의 사업을 중단하여 폐업 신고를 한 사업자이다. 계속사업자는 국세청에서 사업자 정보를 보유하고 있는 사업자 중 폐업 신고가 되지 않은 사업자로 휴업자를 포함하여 정의하였다. 사업자의 폐업 유형은 고려하지 않았으며, 휴업자의 경우 사업을 재개할 가능성이 있으므로 계속사업자로 분류하였다. 계속사업자는 폐업 신고를 하지 않은 사업자에 해당하며 휴업자를 포함하였다.

4.2.2. 독립변수

독립변수로 활용한 재무정보는 일정 기간 동안 일어난 거래에 대한 정보를 재무제표에 작성하는 정보로 대차대조표와 손익계산서에 기록된 재무 계정으로 정의하였으며 비재무정보는 재무 정보 외의 정보로 전자결제서비스 이용 사업자의 특성 및 전자결제에 관한 거래 정보로 정의하였다. 수집한 재무정보의 재무계정 31개를 재무비율 50개로 구성하여 재무변수로 활용하였으며 <표 3> 수집한 비재무정보 33개를 37개의 변수

로 구성하여 비재무변수로 활용하였다<표 4>.

4.3. 연구분석 모형 및 연구방법

재무변수 50개와 비재무변수 37개로 구성된 독립변수 87개를 세 가지 유형으로 나누어 연구분석 모형을 설계하였다. 재무변수를 적용한 모형과 비재무변수를 적용한 모형, 재무변수와 비재무변수를 결합한 모형을 구축하여 여섯 가지의 알고리즘을 적용하였다. 알고리즘별로 연구분석 모형의 예측 정확도 순위를 비교하고 연구분석 모형별로 예측 정확도가 가장 높은 알고리즘을 확인하였다. 연구분석 모형 A의 독립변수는 재무정보이며 연구분석 모형 B는 비재무정보, 연구분석 모형 C는 재무정보와 비재무정보의 독립변수를 결합한 정보가 독립변수이다<그림 1>.

4.4. 검증 및 평가 방법

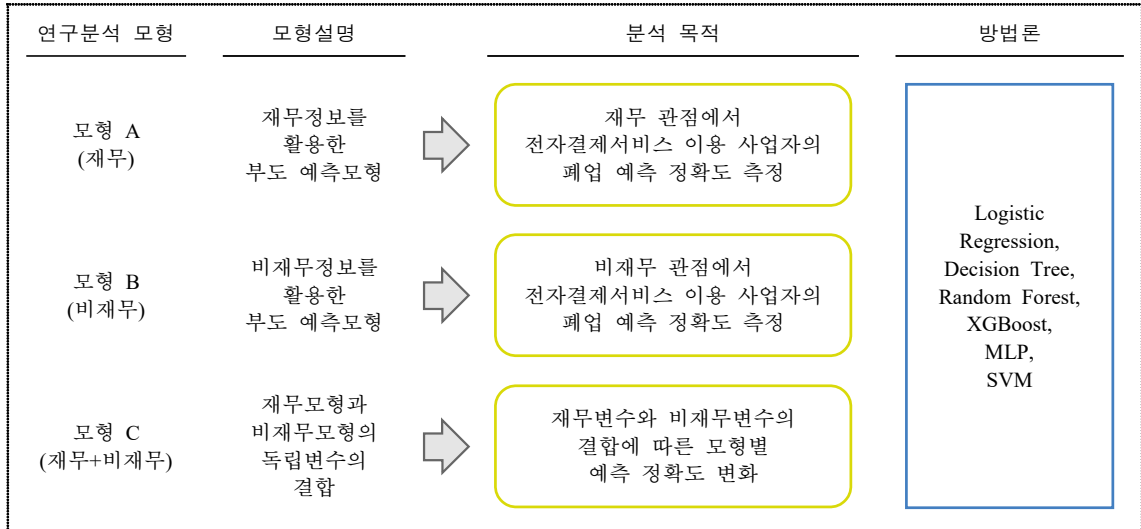
MLP를 포함한 여섯 가지 알고리즘 성능 평가는 Confusion Matrix의 Accuracy, F1-score 값과 AUC(Area Under the ROC Curve)로 판단하였다.

<표 3> 재무정보의 재무비율 변수

구분	분류	재무비율
재무정보	성장성 (9개)	매출액증가율, 영업이익증가율, 순이익증가율, 총자산증가율, 자기자본증가율, 법인세차감전순이익증가율, 유동자산증가율, 유형자산증가율, 재고자산증가율
	수익성 (14개)	매출원가율, 판매비율, 영업이익률, 법인세차감전순이익률, 당기순이익률, EBITDA마진율, 순금융비용부담률, 금융비용대총비용비율, 매출총이익률, EBITDA, 영업외수지율, 수지비율, 총자산순이익률(ROA), 총자산영업이익률,
	안정성 (20개)	이자보상배수(배), 차입금의존도, 고정장기적합률, 부채비율, 자기자본비율, 유동비율, EBITDA/단기차입금, 차입금/자기자본비율, 비유동부채대순운전자본비율, 비유동부채비율, 비유동비율, 매입채무대재고자산비율, 순운전자본대총자본비율, 유동부채비율, 유동자산구성비율, 재고자산대순운전자본비율, 재고자산대유동자산비율, 비유동자산비율, 총자산부채비율, 현금비율
	활동성 (7개)	재고자산회전율(회), 순운전자본회전율(회), 총자본회전율(회), 비유동자산회전율(회), 유동자산회전율(회), 유형자산회전율(회), 자기자본회전율(회)

〈표 4〉 비재무정보의 변수

구분	분류	변수 설명
전자결제 서비스 신청정보 (10개)	업태구분	소비자에게 제공하는 제품 또는 서비스의 판매 방법
	사업자유형	사업자 번호에 따른 사업자 유형 : 개인,법인
	주요판매재화	판매하는 주요 상품 분류
	실물/비실물	판매 물품 또는 서비스에 따라 실물과 비실물로 분류
	서비스가입경로	사업자가 전자결제서비스를 신청하는 경로 : 호스팅, 영업사원, 일반 가입 등
	보증보험가입여부	이행보증보험 가입 여부
	서비스신청ID 개수	결제서비스 이용을 위해 사업자가 신청한 고유ID 개수
	정산한도	사업자가 한 달에 정산 받을 수 있는 금액의 한도
	정산주기I	사업자가 매출액을 정산 받는 일 단위 주기
	정산주기II	사업자가 매출액을 정산 받는 주 단위 주기
전자결제 서비스 이용정보 (7개)	서비스이용 결제수단개수	결제 수단 : 신용카드,계좌이체,무통장입금,휴대폰,상품권
	간편결제수단 사용여부	간편결제수단 사용 여부
	서비스이용기간	결제서비스를 이용한 기간
	서비스신청ID 사용비율	(서비스사용ID개수)/(서비스신청ID개수)
	소비자불만접수월	상품 및 서비스에 대한 소비자 불만이 접수된 월
	평균소비자 불만건I	(소비자 불만 접수월)/(거래 발생월)
	평균소비자 불만건II	(소비자 불만 접수건)/(소비자 불만 발생월)
거래정보 (20개)	평균거래발생월	(거래발생월)/(서비스이용기간월)
	평균순결제건I	(누계 순결제 건)/(거래 발생월)
	평균순결제금액I	(누계 순결제 금액)/(거래 발생월)
	평균취소건I	(누계 취소건)/(거래 발생월)
	평균취소금액I	(누계 취소금액)/(거래 발생월)
	취소건비중I	(누계 취소건)/(누계 순결제건)*100
	취소금액비중I	(누계 취소금액)/(누계 결제금액)*100
	미수금발생여부	사업자에게 지급할 매출액의 정산금액보다 취소금액이 큰 경우 발생
	미수금발생비중	(미수금발생누계)/(정산한도)*100
	미수금규모	미수금이 발생한 시점부터 다시 매출 금액이 (+)가 되기까지의 소요 시간과 규모에 따라 구분함
	미수금상계기간	미수금 발생 이후 미수금이 상계 되는 기간의 평균
	미수금발생누계	발생한 미수금의 누계
	평균순결제건II	(직전 3개월 순결제건)/(직전 3개월 거래발생월)
	평균순결제금액II	(직전 3개월 순결제금액)/(직전 3개월 거래발생월)
	평균취소건II	(직전 3개월 취소건)/(직전 3개월 거래발생월)
	평균취소금액II	(직전 3개월 취소금액)/(직전 3개월 거래발생월)
	취소건비중II	(직전 3개월 취소건)/(직전 3개월 평균 순결제건)*100
	취소금액비중II	(직전 3개월 취소금액)/(직전 3개월 평균 결제금액)*100
	순결제건비중	(직전 3개월 순결제건)/(직전 3개월 전체 결제건)*100
	순결제금액비중	(직전 3개월 순결제금액)/(직전 3개월 전체 결제금액)*100



〈그림 1〉 연구분석 모형

5. 실증연구

5.1. 표본의 구성

재무정보 수집이 가능한 사업자 중 전자결제 서비스를 이용하는 5,725개의 사업자를 표본으로 구성하였다. 계속사업자가 5,621개 폐업자가 104개로 구성된 표본은 계속사업자와 폐업자의 비중이 98:2로 편중이 심한 표본으로 예측 모형의 정확도를 과대하게 평가할 수 있는 우려가 존재하였다. 박정민 등(2005)은 연구에 사용되는 건전 기업과 부도기업은 일반적으로 불균형한 데이터를 가지고 있어 이러한 불균형 데이터를 처리하는 방법이 중요하며, 부도 관련 데이터는 채무의 불이행에 관한 데이터를 수집해야 하는 경우도 있기 때문에 대안적인 방법으로 상장폐지 데이터 등을 사용하는 경우도 있다고 하였다. 이를 토대로 본 연구에서는 계속사업자와 폐업자의 데이터 비중의 불균형을 해소하기 위해 계속사업자의

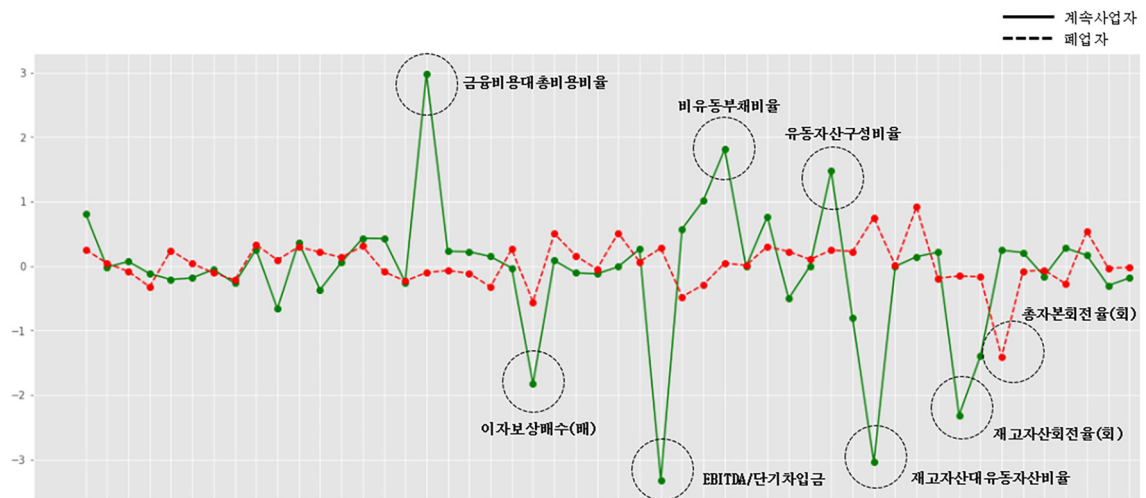
이상치만 제거한 후 오버샘플링(SMOTE) 기법을 사용하여 최종 2,558개를 연구에 활용하였다. 표본으로 활용된 데이터의 계속사업자와 폐업자의 수의 비중은 5:5이며 학습데이터와 검증 데이터는 8:2의 비율로 사용하였다.

5.2. 데이터 처리 및 정제

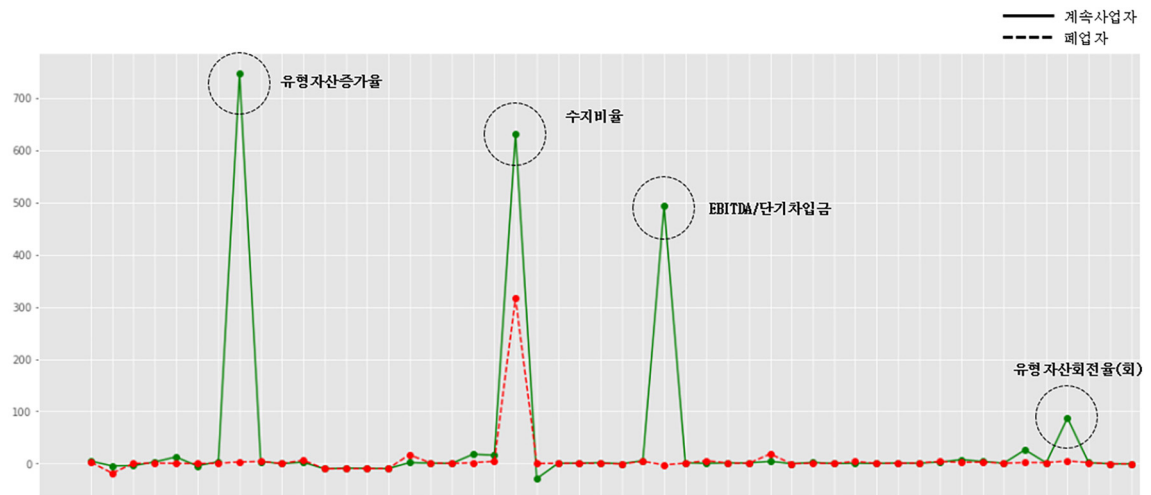
연구 분석에 사용한 데이터의 독립변수 87개는 범주형 변수 12개와 수치형 변수 75개로 구성되어 있다. 12개의 범주형 변수 중 범주가 3개 이상인 8개의 변수에 대해 해당 범주가 순서의 의미가 있는 지급주기 I, 지급주기 II, 미수금규모, 미수금상계기간의 4개의 변수에는 label encoding을 하였고 순서의 의미가 없는 업태구분, 주요판매채화, 가입경로, 사업자유형의 변수에 대해서는 one-hot encoding을 하였다. 범주형 변수 데이터 값의 분포는 전체적으로 다양하게 이루어져 있었으며 수치형 변수 75개 중 이상치는 재무변수

에서 1개, 비재무변수에서 4개에서 확인되었고 이상치는 데이터 손실을 최소화하기 위해 quantile 0.1 이하, 0.9 이상의 값만 이상치로 처리하였다. 또한, 재무변수와 비재무변수는 폐업 여부와의 상관관계가 거의 존재하지 않았지만 계속사업자

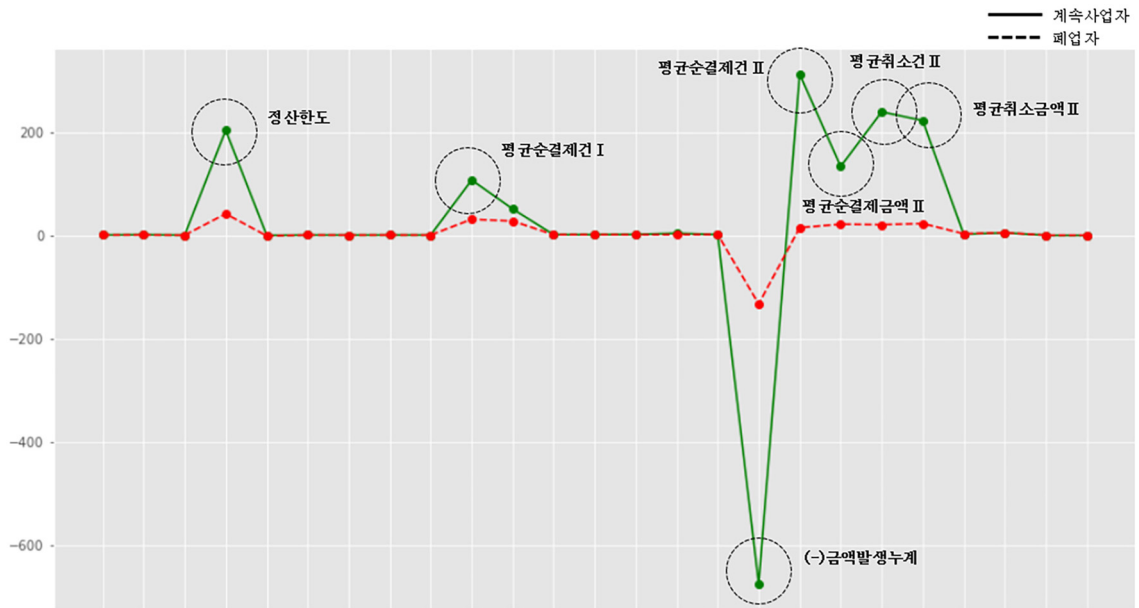
와 폐업자의 두 집단 간 변수들의 차이를 확인할 수 있었다. 계속사업자와 폐업자의 변수들 간 차이는 재무정보 데이터를 정제하기 전과 후<그림 2>, <그림 3>와, 비재무정보 데이터 정제 전과 후 <그림 4>, <그림 5>로 각각 비교하였다.



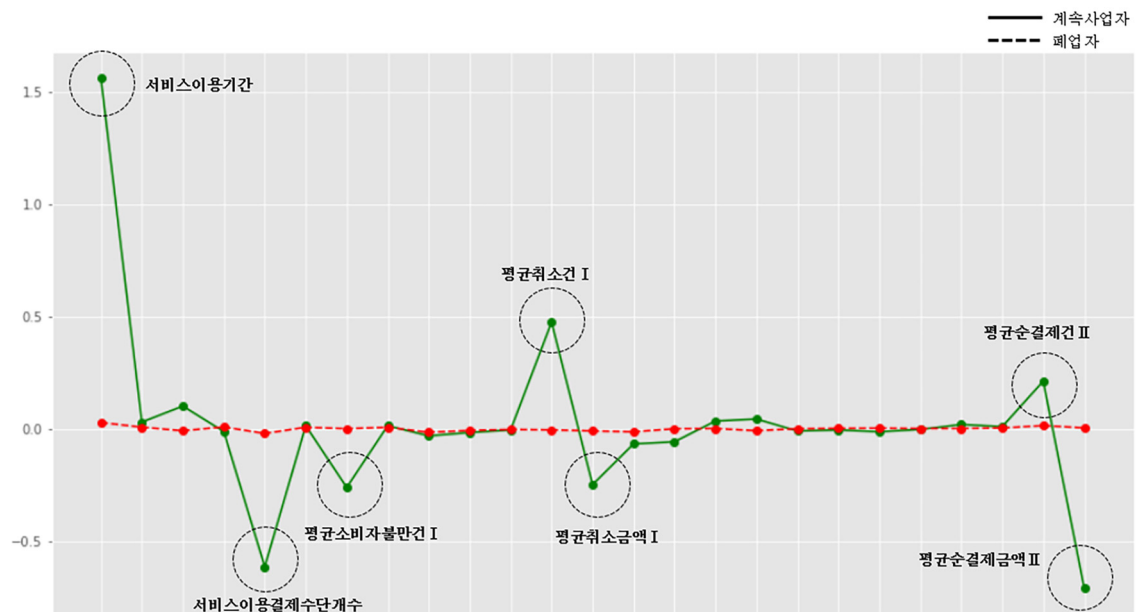
<그림 2> 재무정보 변수의 평균 비교 (데이터 정제 전)



<그림 3> 재무정보 변수의 평균 비교 (데이터 정제 후)



〈그림 4〉 비재무정보 변수의 평균 비교 (데이터 정제 전)



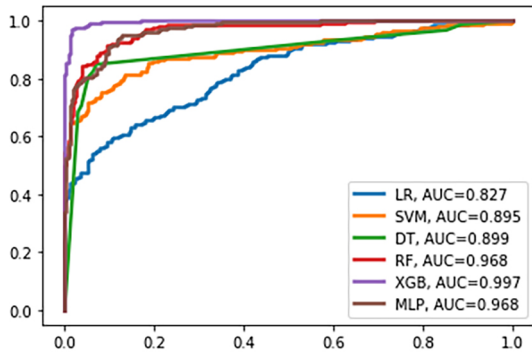
〈그림 5〉 비재무정보 변수의 평균 비교 (데이터 정제 후)

5.3. 연구분석 모형

5.3.1. 재무정보 모형 A

재무정보를 적용한 모형 A에서는 XGBoost, Decision Tree, Random Forest, MLP, SVM, Logistic 알고리즘 순서로 정확도가 높았다. tree 기반의 XGBoost의 정확도가 0.977, F1-score는 0.976로 가장 높았고, 그 다음 Decision Tree의 정확도는 0.891, F1-score 0.881로 XGBoost의 정확도와는 약 0.086의 차이가 있었다. F1-score 값은 XGBoost, Decision Tree, MLP, Random Forest, SVM, Logistic 순서로 높았다<표 5>.

AUC는 XGBoost 0.997, Random Forest와 MLP 0.968, Decision Tree 0.899, SVM 0.895, Logistic 0.827 순서로 높았다<그림 6>.

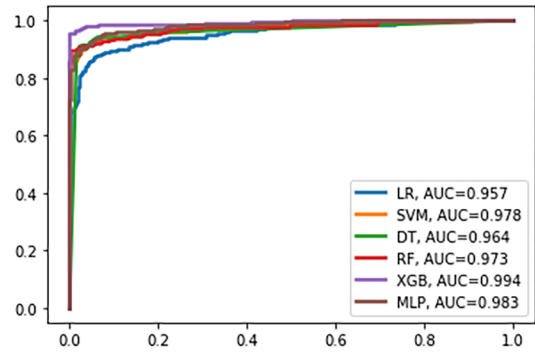


<그림 6> 모형 A의 알고리즘별 AUC

5.3.2. 비재무정보 모형 B

전자결제 정보로 구성된 비재무정보를 적용한 모형 B는 XGBoost의 정확도가 가장 높았으며 Decision Tree와 Random Forest, SVM, MLP, Logistic 알고리즘 순서로 정확도가 높았다. 또한, 모든 알고리즘에서 비재무정보 모형 B의 정확도가 재무정보의 모형 A에서보다 높았으며 Logistic 알고리즘에서 가장 큰 차이를 보였다. F1-score 값과 AUC 값 또한 재무정보를 적용한 모형 A에서보다 비재무정보 모형 B에서 높았다. 모형 B의 알고리즘에서 F1-score 값은 높은 순서는 XGBoost, Decision Tree, Random Forest, MLP, SVM, Logistic이다<표 6>.

AUC를 살펴보면 XGBoost 0.994, MLP 0.983, SVM 0.978, Decision Tree 0.964, Logistic 0.957이다<그림 7>.



<그림 7> 모형 B의 알고리즘별 AUC

<표 5> 재무정보 모형 A의 예측 정확도

	Logistic Regression	Decision Tree	Random Forest	XGBoost	MLP	SVM
Precision	0.850	0.916	0.956	0.976	0.806	0.908
Recall	0.580	0.849	0.804	0.976	0.967	0.722
F1-score	0.689	0.881	0.874	0.976	0.879	0.805
AUC	0.827	0.899	0.968	0.997	0.968	0.895
Accuracy	0.750	0.891	0.889	0.977	0.873	0.832

〈표 6〉 비재무정보 모형 B의 예측 정확도

	Logistic Regression	Decision Tree	Random Forest	XGBoost	MLP	SVM
Precision	0.954	0.950	0.961	0.988	0.931	0.965
Recall	0.845	0.922	0.910	0.967	0.935	0.902
F1-score	0.896	0.936	0.935	0.977	0.933	0.932
AUC	0.957	0.964	0.973	0.994	0.983	0.978
Accuracy	0.906	0.939	0.939	0.979	0.936	0.938

〈표 7〉 결합 모형 C의 예측 정확도

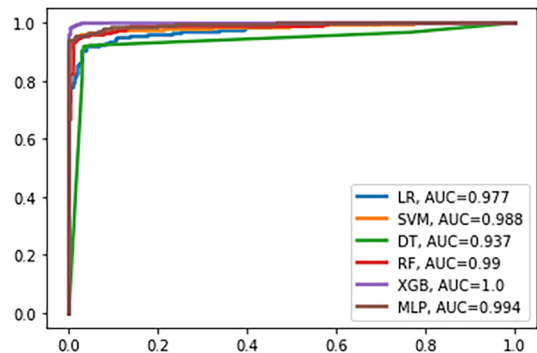
	Logistic Regression	Decision Tree	Random Forest	XGBoost	MLP	SVM
Precision	0.964	0.957	0.991	0.996	0.963	0.991
Recall	0.869	0.918	0.935	0.984	0.955	0.943
F1-score	0.914	0.938	0.962	0.990	0.959	0.967
AUC	0.977	0.937	0.990	1.000	0.994	0.988
Accuracy	0.922	0.941	0.965	0.990	0.961	0.969

5.3.3. 재무정보와 비재무정보를 결합한 모형 C

재무정보와 비재무정보를 결합한 모형에서 XGBoost의 정확도가 0.990로 가장 높았으며 SVM 0.969, Random Forest 0.965, MLP 0.961, Decision Tree 0.941, Logistic 0.922 순서로 정확도를 보였다. F1-score 값이 높은 순서도 정확도가 높은 알고리즘 순서와 동일하였다. 모형 C는 재무정보의 모형 A와 비재무정보의 모형 B보다는 모든 알고리즘에서 전반적으로 정확도가 높았으며 F1-score 또한 세 가지 모형 중 가장 높았다 <표 7>.

재무정보와 비재무정보를 결합한 모형 C의 각 알고리즘별 AUC는 XGBosst 1.000, MLP 0.994,

Random Forest 0.990, SVM 0.988, Logistic 0.977, Decision Tree는 0.937이다<그림 8>.



〈그림 8〉 모형 C의 알고리즘별 AUC

5.4. 분석결과

5.4.1. 연구분석 모형별 예측 성능

모형별로 예측 성능을 정확도와 AUC 측면에서 비교하였을 때, 연구분석 모형 C, 모형 B, 모형 A 순으로 높은 성능을 기록했다. 이는 재무정보만을 활용한 예측 모델 보다 비재무정보만을 활용하거나 재무정보와 비재무정보를 결합한 모형의 예측 성능이 높다는 것을 의미한다<표 8>. 특히, 비재무정보의 모형 B의 알고리즘의 평균 정확도는 재무정보의 모형 A와 비교하여 7.1%p 높았으며 평균 AUC는 4.9%p 높았다. 재무정보와 비재무정보가 결합된 모형 C의 알고리즘 평균 정확도는 재무정보 모형 A와 비교하여 10.2% 향상되었고 평균 AUC는 5.9% 향상되었다. 이때, Logistic 알고리즘의 정확도 향상폭이 22.9%로 가장 컸으며 다음으로는 SVM 16.5%이다. 이러한 결과를 통해 재무정보와 비재무정보만을 각각 적용하는 경우보다 재무정보와 비재무정보를 결합하여 적용하는 경우가 전자결제서비스 이용 사업자의 폐업 예측의 성능을 높일 수 있음을 확인하였다. 또한, 비재무정보에 재무정보를 결합하였을 때 정확도가 1.9%,

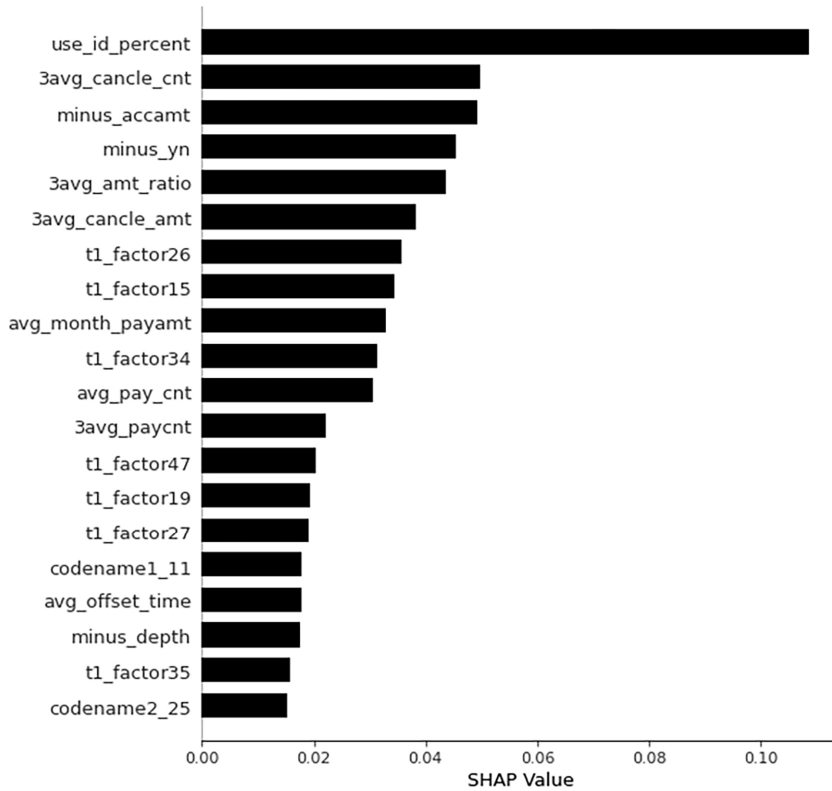
AUC가 0.6% 향상된 것 보다 재무정보에 비재무정보를 결합하였을 때 정확도 10.2%, AUC 5.9%로 더 크게 향상되는 것을 확인하였다.

5.4.2. 예측 중요 변수

각 모형에서 폐업 예측 정확도에 영향을 미친 변수들의 중요도를 학습한 모델과 독립적인 해석이 가능한 SHapley Additive exPlanations(SHAP) 기법으로 살펴보았다. SHAP는 개별 예측 값에 대한 각 변수들의 영향력을 모형 클래스에 상관없이 실제로 존재하는 변수만의 영향력의 합으로 나타내어 Shapley Value로 영향력을 측정한다. SHAP에서 확인된 전체 87개의 변수의 중요도는 중요도가 높은 상위 20개 중 비재무정보가 약 70%를 차지하였다. 상위 20개 중 13개의 비재무정보는 전자결제 정보들 중 거래 유동성을 나타내는 변수들로 업태구분, 주요판매채화, 서비스신청ID사용비율, 평균거래발생월, 평균순결제금액 I, 미수금발생여부, 미수금규모, 미수금상계기간, 미수금발생누계, 평균순결제금액II, 평균취소건II, 평균취소금액II, 순결제금액비중이 확인되었다<그림 9>. 이 중에서 서비스신청ID사용비율, 미수금 규모는

<표 8> 연구분석 모형별 정확도 산출 결과

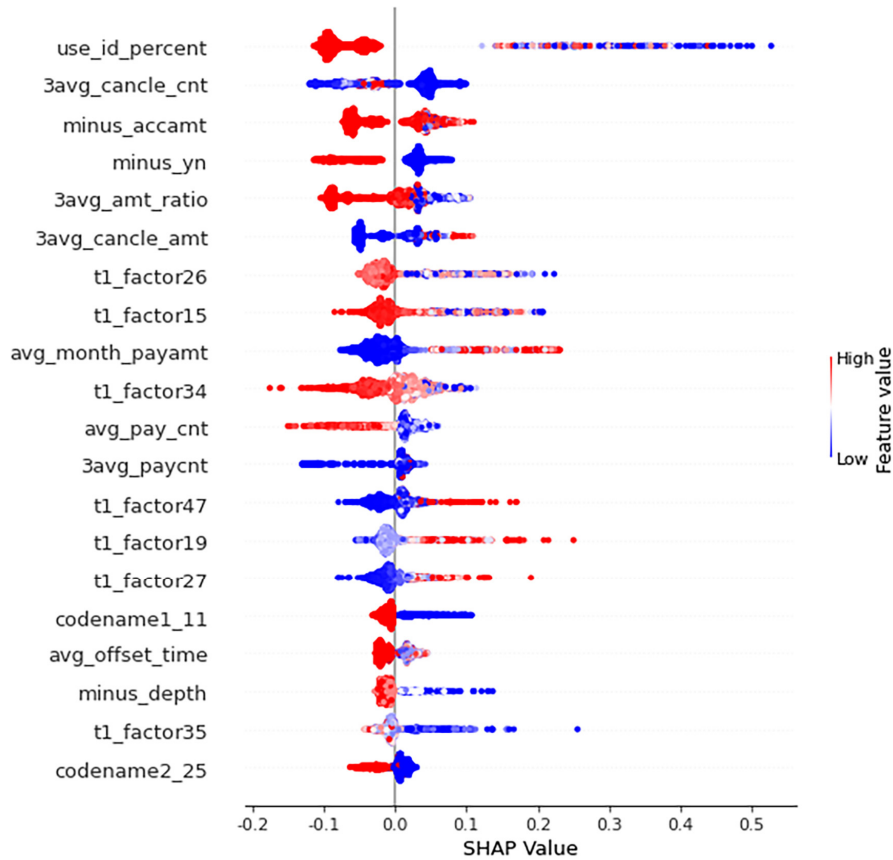
	모형 A		모형 B		모형 C	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Logistic Regression	0.750	0.827	0.906	0.957	0.922	0.977
Decision Tree	0.891	0.899	0.939	0.964	0.941	0.937
Random Forest	0.889	0.968	0.939	0.973	0.965	0.990
XGBoost	0.977	0.997	0.979	0.994	0.990	1.000
MLP	0.873	0.968	0.936	0.983	0.961	0.994
SVM	0.832	0.895	0.938	0.978	0.969	0.988
평균	0.869	0.926	0.940	0.975	0.958	0.981



〈그림 9〉 SHAP 기반 변수 중요도

위계적 회귀분석을 통해 유의미한 변수로 확인된 비재무변수들과 일치하는 변수로 확인되었다. 변수의 증감에 따라 폐업의 영향도에 대해 SHAP Value를 살펴보면<그림 10> 서비스신청ID사용비율, 미수금 발생여부, 평균순결제금액Ⅱ, 평균거래발생월, 미수금 상계기간, 미수금규모는 값이 클수록 사업자의 폐업과의 관련성이 낮게 나왔다. 서비스신청ID사용비율과 평균순결제금액Ⅱ, 평균거래발생월, 미수금발생여부의 값이 크다는 것은 지속적인 매출이 발생하는 것을 의미하며, 미수금상계기간과 미수금 규모의 변수는 해당 값이 클수록 매출액으로 미수금을 상계하는 속도가 빠르며 미수금이 적은 것을 의미하여 폐업이 아닌 것과의 관련성이 높게 작용

한 변수임을 확인할 수 있다. 평균취소금액Ⅱ, 평균순결제금액 I의 두 가지 변수는 특정 기간 동안 실제 거래가 발생한 월을 기준으로 확인한 변수로 거래가 꾸준할수록 값이 작아 사업자의 폐업과의 관련성이 낮게 나온 것으로 확인된다. 특히, 이 중에서 가장 높은 중요도를 가진 서비스신청ID 사용비율의 변수는 사업자가 전자상거래에서 재화를 판매함에 있어 사업자의 필요에 따라 판매 상품별로 ID를 신청하고 실제 매출이 발생하는 ID의 비율을 나타내는 변수로 사업자의 판매 유동성을 판단할 수 있는 지표이다. 서비스신청ID 개수가 많고 매출이 발생하는 ID 개수가 많을수록 해당 사업자가 판매하는 상품 또는 서비스가 많고



〈그림 10〉 SHAP Value

다양하다고 볼 수 있다. 평균취소건Ⅱ과 평균순결제건Ⅱ, 미수금발생누계는 폐업간 상관관계의 방향성은 없지만 평균취소건Ⅱ과 평균순결제건Ⅱ이 낮을수록, 미수금발생누계 값은 높을수록 다른 변수 값들에 따라 폐업 예측에 중요하게 작용하였고 업태구분과 주요제품 변수에서는 소매업, 유통업이 사업자의 폐업과 관련성이 높음을 확인하였다.

SHAP 기반의 중요 변수 상위 20개 중 확인된 재무정보는 자기자본비율과 유동비율, 유동부채비율, 순운전자본대총자본비율, 유형자산회전율(회), EBITDA, EBITDA마진율이 확인되었으며

기업의 안정성을 나타내는 지표가 높은 비중으로 확인되었다.

6. 결론 및 시사점

금융 산업에서 기업의 부도를 사전에 판단하는 일은 매우 중요하다. 기존 연구에서 밝힌 바와 같이 기업 부도 예측에 있어 재무정보는 유의한 정보이며 기업의 상태를 객관적으로 판단할 수 있는 가장 기본적인 척도일 것이다. 하지만, 금융

산업에서 기업을 판단하는 지표로서의 재무정보의 활용은 재무정보가 기업의 상태를 나타내는 시점과 기업을 판단하는 시점 간의 차이로 인해 판단의 적시성이 부족한 한계가 있다. 더욱이 금융 기관 외 기업들은 거래 당사자들의 기업 신용 정보를 확인할 수 없기 때문에 이를 대신할 정보가 필요하며 재무정보가 없는 초기 사업자들의 건전성을 판단할 수 있는 정보가 매우 중요한 실정이다. 따라서 본 연구는 기업의 건전성을 판단함에 있어 이러한 한계점을 보완할 수 있는 정보로서 비재무정보의 유용성에 관해 실증 연구하여 다음과 같은 의의를 지닌다.

첫째, 본 연구에서는 재무정보의 신뢰성이 낮은 사업자가 다수인 전자결제서비스 이용 사업자의 폐업 예측에 있어 비재무정보 중 전자결제 정보가 폐업 예측에 영향을 미치는 유용한 정보로서의 가치를 밝혔다. 재무정보와 전자결제와 관련한 비재무정보로 연구분석 모형 세 가지를 설계하여 그 차이를 비교한 결과, 적용하는 알고리즘에 따라 정확도와 AUC의 차이가 있었으나 재무정보나 비재무정보만을 활용하였을 때보다 재무정보에 비재무정보를 결합함으로써 예측 성능이 전반적으로 높아지는 것을 확인하였다. 이는 비재무정보가 폐업 예측 정확도 향상에 영향을 미치는 정보이며, 재무정보 모형보다 비재무정보 모형의 예측 정확도가 높은 결과를 통해 본 연구에서 비재무정보로 활용한 전자결제 정보가 재무정보의 대안적 지표로 활용될 수 있는 가능성을 확인하였다.

둘째, 본 연구는 재무정보가 반기, 년 단위 정보로서 부도나 폐업을 예측하는 데에 있어 시의적절하지 못하다는 부도 예측 연구의 한계를 보완할 수 있는 데이터로서 전자결제 정보의 가치를 밝혔다는 것에 의의가 있다. 현재까지 부도 예측에 관한 연구는 재무정보 외에 시장정보, 뉴스

기사와 같은 외부 정보를 활용한 연구는 많지만, 정보들의 편중을 해소하거나 해당 기업의 현 상태를 적시에 가장 잘 반영할 수 있는 정보를 활용한 연구는 부족한 실정이다. 본 연구에 활용한 데이터는 시장의 흐름에 따른 거래의 변화를 일 단위, 월 단위로 확인할 수 있는 실시간 거래 데이터로 실증 연구하여 해당 정보들이 전자상거래 사업자의 건전성을 판단하는 시점의 적시성을 갖는 정보로서의 활용성을 확인하였다. 특히, SHAP에서 확인한 상위 20개 중 13개의 비재무변수는 전자결제서비스를 이용하는 사업자 폐업 예측에 중요한 변수로서 기업에서는 사업자 건전성 판단에 해당 변수들을 주요 지표로 활용할 수 있을 것이다.

더 나아가 금융 산업의 기업들은 본 연구를 통해 비재무정보에 대한 중요성을 인식하고 각 기업에서 필요로 하는 기업의 건전성 판단 지표로 다양한 비재무정보들을 수집하고 활용해야 하는 당위성이 생겼을 것으로 생각한다. 이를 통해 금융 산업의 기업들은 이해관계자인 기업의 건전성 판단에 있어 기존 재무정보만으로 판단하기 어렵거나 정보의 편중으로 판단하지 못했던 기업들까지 재무정보의 대안적 정보로 비재무정보를 적극 활용하여 적시에 기업의 건전성을 판단하고 판단의 정확성까지 높일 수 있을 것으로 기대한다.

끝으로 본 연구에서는 비재무정보의 유용성을 밝히는 데에 있어 머신러닝 기반으로 사업자 폐업 예측 연구를 진행하였다. 비재무정보의 수집이 가능해짐에 따라 더 다양한 정보들을 수집, 활용 가능하게 되면서 기업 부도 예측 성과를 향상할 수 있는 더 많은 변수들이 활용될 수 있을 것으로 기대한다. 이에 따라 머신러닝 기반의 폐업 예측 연구에서 딥러닝 기반의 모형으로 비재무정보의 유용성에 관하여 예측 성능을 고도화하는 시도가 필요할 것이다.

참고문헌(References)

[국내 문헌]

- 권누리, 김영민, 최광신. (2019). 거시경제 변수를 고려한 한국기업부도 모형 구축 방법 연구. *한국데이터정보과학회지*, 30(5), 1037-1050.
- 권재영. (2015). Support Vector Machine 및 Random Forest 기법을 이용한 기업 부도 예측. 이화여자대학교.
- 권혁건, 이동규, 신민수. (2017). RNN(Recurrent Neural Network)을 이용한 기업부도 예측모형에서 회계정보의 동적 변화 연구. *지능정보연구*, 23(3), 139-153.
- 김경원. (2018). 산업별 평가를 위한 뉴스 기사 기반 산업 위험 지표 예측. 건국대학교.
- 김상보, Philip Ji, 조경준. (2011). 부도 예측모형을 이용한 기업부실화의 원인분석. *시장경제연구*, 40(1), 85-106.
- 김성규. (2010). 경기변동을 반영한 부도 예측모형에 관한 실증연구. 한양대학교.
- 김성진, 안현철. (2016). 기업신용등급 예측을 위한 랜덤 포레스트의 응용. *산업혁신연구*, 32(1), 187-211.
- 김수영. (2006). 다변량 판별분석과 로지스틱 회귀 분석, 인공지능망 분석을 이용한 호텔 도산 예측. *관광학연구*, 30(2), 53-75.
- 김찬송, 신민수. (2019). 부도 예측 모형에서 뉴스 분류를 통한 효과적인 감성분석에 관한 연구. *한국IT서비스학회지*, 18(1), 187-200.
- 김형준, 류두진, 조훈. (2019). 기업부도 예측과 기계 학습. *금융공학연구*, 18(3), 131-152.
- 남재건. (2019). 국내 기업신용평가를 위한 기계 학습 알고리즘 성능 비교. 국민대학교.
- 민재형, 이영찬. Support Vector Machine을 이용한 부도 예측모형의 개발-격자탐색을 이용한 커널 함수의 최적 모수 값 선정과 기존 부도 예측모형과의 성과 비교. *한국경영과학회지*, 30(1), 55-74.
- 박유성, 최보승, 이상운. (2009). 중소기업을 위한 효율적인 부도 예측모형 구축을 위한 연구. *Journal of The Korean Data Analysis Society*, 11(3), 1363-1375.
- 박정민, 김경재, 한인구. (2005). Support Vector Machine을 이용한 기업부도 예측. *경영정보학연구*, 15(2), 51-63.
- 송현준, 박도준, 이준기. (2021). 머신러닝을 이용한 외감기업 및 비외감기업의 부도 예측에 관한 연구. *한국IT정책경영학회 논문지*, 13(3), 2521-2517.
- 엄하늘, 김재성, 최상욱. (2020). 머신러닝 기반 기업부도위험 예측모델 검증 및 정책적 제언: 스택킹 앙상블 모델을 통한 개선을 중심으로. *지능정보연구*, 26(2), 105-129.
- 오우석, 김진화. (2017). 인공지능기법을 이용한 기업부도 예측. *대한산업경영학회지*, 15(1), 17-32.
- 이재식, 한재홍. (1995). 인공지능망을 이용한 중소기업 도산 예측에 있어서의 비재무정보의 유용성 검증. *한국전문가시스템학회지*, 1(1), 123-134.
- 이재식. (2010). *데이터 애널리틱스 경기*: 책출판 위키북스.
- 임형준. (2016). 기업 특성별 기술평가정보의 부도 예측력 검증: 관계형성 여부, 업력, 규모를 중심으로. *한국경제의 분석*, 22(1), 81-142.
- 장운신. (2013). 기업 특성변수 및 거시경제 변수와 예상부도 확률간의 상관관계. 성균관대학교.
- 최소운, 안현철. (2015). 퍼지이론과 SVM 결합을 통한 기업부도 예측 최적화. *디지털융복합연구*, 13(3), 155-165.
- 최영수, 장욱. (2007). 재무변수와 시장변수를 결합한

- 은행에 대한 부도 예측 모형. *한국경영학회 융합학술대회*, 1 - 52.
- 최정원, 오세경, 장재원. (2017). 빅데이터와 인공지능 기법을 이용한 기업 부도 예측 연구. *한국재무학회 학술대회*, 396-435.
- CRETOP. 기업재무. 2022년 11월 5일 검색. <https://www.cretop.com/>.
- DATA 공공 데이터 포털. 국세청 사업자등록정보 진위확인 및 상태조회 서비스. 2022년 11월 1일, 11월 30일 검색. <https://www.data.go.kr/data/15081808/openapi.do>.
- [국외 문헌]**
- Addal, S. (2016). Financial forecasting using machine learning. *African Institute for Mathematical, Science(AIMS)*.
- Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589-609.
- Beaver, W.H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 4, 65-100.
- Breiman, L. (2001). Random forests, *Machine Learning*, 45(1), 5-32.
- Cox, D. (1972). Regression Models and Life Tables. *Journal of Royal statistical society, Series B*, 18(2), 187-220.
- Campbell, J. Y., J. Hilscher, J. Szilagyi. (2008). In search of distress risk. *The Journal of Finance*, 63(6), 2899-2939.
- Louma M., E. Laitinen. (1991). Survival Analysis as a Tool for Company Failure Prediction. *Omega*, 19(6), 673-678.
- Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109-131.
- Zmijewski, M. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22, Supplement, 59-82.

Abstract

A study on improving the accuracy of machine learning models through the use of non-financial information in predicting the Closure of operator using electronic payment service

Hyunjeong Gong* · Eugene Hwang* · Sunghyuk Park**

Research on corporate bankruptcy prediction has been focused on financial information. Since the company's financial information is updated quarterly, there is a problem that timeliness is insufficient in predicting the possibility of a company's business closure in real time. Evaluated companies that want to improve this need a method of judging the soundness of a company that uses information other than financial information to judge the soundness of a target company. To this end, as information technology has made it easier to collect non-financial information about companies, research has been conducted to apply additional variables and various methodologies other than financial information to predict corporate bankruptcy. It has become an important research task to determine whether it has an effect. In this study, we examined the impact of electronic payment-related information, which constitutes non-financial information, when predicting the closure of business operators using electronic payment service and examined the difference in closure prediction accuracy according to the combination of financial and non-financial information. Specifically, three research models consisting of a financial information model, a non-financial information model, and a combined model were designed, and the closure prediction accuracy was confirmed with six algorithms including the Multi Layer Perceptron (MLP) algorithm. The model combining financial and non-financial information showed the highest prediction accuracy, followed by the non-financial information model and the financial information model in order. As for the prediction accuracy of business closure by algorithm, XGBoost showed the highest prediction accuracy among the six algorithms. As a result of examining the relative importance of a total of 87 variables used to predict business closure, it was confirmed that more than 70% of the top 20 variables that had a significant impact

* School of Management Engineering, KAIST

** Corresponding author: Sunghyuk Park

School of Management Engineering, KAIST

85 Hoegi-ro, Dongdaemun-gu, Seoul, Korea

Tel: +82-2-958-3433, E-mail: sunghyuk.park@kaist.ac.kr

on the prediction of business closure were non-financial information. Through this, it was confirmed that electronic payment-related information of non-financial information is an important variable in predicting business closure, and the possibility of using non-financial information as an alternative to financial information was also examined. Based on this study, the importance of collecting and utilizing non-financial information as information that can predict business closure is recognized, and a plan to utilize it for corporate decision-making is also proposed.

Key Words : closure, prediction accuracy, e-commerce, payment service, payment gateway

Received : June 8, 2023 Revised : September 17, 2023 Accepted : September 18, 2023

Corresponding Author : Sunghyuk Park

저자 소개



공현정

한국과학기술원 정보미디어경영전문대학에서 정보경영 석사학위를 취득하고, 현재 금융플랫폼사에 재직 중이다. 주요 관심 분야는 비즈니스 애널리틱스, 블록체인, 인공지능 등이다.



황유진

한국과학기술원에서 생명화학공학을 전공하여 학사학위를 취득하고, 현재 한국과학기술원 경영대학에서 경영공학부 석박통합과정에 재학 중이다. 주요 관심 분야는 딥러닝, 시계열 예측, 비즈니스 애널리틱스 등이다. 주요 논문을 IEEE Transactions on Neural Networks and Learning Systems, Artificial Intelligence in Medicine 등에 게재하였다.



박성혁

현재 한국과학기술원 경영대학 조교수로 재직 중이다. 한국과학기술원에서 수학을 전공하여 학사학위를 취득하고, 한국과학기술원 경영대학에서 경영공학을 전공하여 박사학위를 취득하였다. 주요 관심 분야는 비즈니스 애널리틱스, 인공지능, 추천시스템 등이다. 주요 논문을 Information Systems Research, MIS Quarterly 등에 게재하였다.