

전통적인 통계와 기계학습 기반 중국 문화산업 기업의 재무적 곤경 예측모형 연구

Research on Financial Distress Prediction Model of Chinese Cultural Industry Enterprises Based on Machine Learning and Traditional Statistical

원도*, 왕콘**, 란희***, 배기형***

세종대학교 경제학과/서남대학 상업무역학원*, 세종대학교 경제학과/허난경제무역직업대**,
세종대학교 경제학과***

Tao Yuan(nikeny90@naver.com)*, Kun Wang(kung-777@naver.com)**,
Luan Xi(luanxi0720@naver.com)***, Ki-Hyung Bae(baekh@sejong.ac.kr)***

요약

본 연구의 목적은 전통적인 통계와 기계학습(Machine Learning)을 통해 중국 문화산업 기업의 재무적 곤경을 정확하게 예측하는 분석 모형을 탐색하는 데 있다. 예측모형을 구축하기 위하여 중국 128개 문화산업 상장 기업의 데이터를 수집하였다. 25개 설명변수로 이뤄진 데이터베이스를 토대로 판별분석과 로지스틱 회귀(Logistic) 등 전통적인 통계 방법과 서포트 벡터 기계(SVM), 결정 트리(Decision Tree)와 랜덤 포레스트(Random Forest) 등 기계학습을 이용한 예측모형을 구축하고 각 모형의 성능 평가를 위해 Python 소프트웨어를 사용한다. 분석 결과, 예측 성능이 가장 좋은 모형은 랜덤 포레스트(Random Forest) 모형으로 95%의 정확도를 보였다. 그 다음은 서포트 벡터 기계(SVM) 모형으로 93%의 정확도를 보였다. 그 다음은 결정 트리(Decision Tree) 모형으로 92%의 정확도를 보였다. 그 다음은 판정분석 모형으로 89%의 정확도를 보였다. 예측 효과가 가장 낮은 모형은 로지스틱 회귀(Logistic) 모형으로 88%의 정확도를 보였다. 이는 중국 문화산업 기업의 재무적 곤경을 예측하면서 기계학습 모형이 전통적인 통계 모형보다 더 좋은 예측 효과를 얻을 수 있음을 설명한다.

■ 중심어 : | 기계학습 | 전통적인 통계 | 재무적 곤경 | 문화산업 |

Abstract

The purpose of this study is to explore a prediction model for accurately predicting Financial Difficulties of Chinese Cultural Industry Enterprises through Traditional Statistics and Machine Learning. To construct the prediction model, the data of 128 listed Cultural Industry Enterprises in China are used. On the basis of data groups composed of 25 explanatory variables, prediction models using Traditional Statistical such as Discriminant Analysis and logistic as well as Machine Learning such as SVM, Decision Tree and Random Forest were constructed, and Python software was used to evaluate the performance of each model. The results show that the Random Forest model has the best prediction performance, with an accuracy of 95%. The SVM model was followed with 93% accuracy. The Decision Tree model was followed with 92% accuracy. The Discriminant Analysis model was followed with 89% accuracy. The model with the lowest prediction effect was the Logistic model with an accuracy of 88%. This shows that Machine Learning model can achieve better prediction effect than Traditional Statistical model when predicting financial distress of Chinese cultural industry enterprises.

■ keyword : | Machine Learning | Traditional Statistical | Financial Difficulties | The Cultural Industry |

I. 서론

최근 중국의 문화산업은 지속적으로 발전하고 있다. 중국 국가통계국(中國國家統計局) 데이터에 따르면, 2010년~2019년 중국 문화산업의 증가치는 2010년 1조1,052억 위안에서 2019년 4조 4,363억 위안으로 300% 증가하였다. 중국 GDP에서 문화산업의 부가치 비중도 2010년 2.75%에서 2019년 4.5%로 향상하였다[1].

따라서 중국 국민경제에서 문화산업의 지위는 부단히 제고되고 중국경제에 대한 문화산업의 발전의 견인 작용이 꾸준히 증가하고 있음을 알 수 있다. 하지만 코로나 19 및 중국 국가산업정책조정의 영향으로 최근 2년간 중국 문화산업의 증가치에는 성장 속도가 느리고 발전 저력이 부족한 등 문제가 나타났다. 또한 문화산업은 다른 산업에 비해 초기자금 투입이 많고 원금 회수 주기가 길고 기업자금 유통이 많아 부채비율이 높기 때문에 실제 이익은 있지만, 위험부담도 높다. 이에 따라 충분한 연구가 없는 상황에서 문화산업 기업의 투자가 불합리하거나 부실하게 운영될 경우 자금 부족이나 자금줄이 끊어지는 등 재무적 곤경과 심각한 재무위기를 초래할 수 있다. 중국의 현재 상장기업의 상장 퇴출 위험 기준에 따르면 2년 연속 적자를 내면 상장 퇴출 위험 조기경보(ST로 표시)를 실시한다[2]. 따라서 과학적이고 효과적인 재무적 곤경 조기경보 모형의 구축은 문화산업기업의 재무 위험 식별능력을 제고 시켜 문화산업 기업의 발전을 촉진할 수 있다.

그러나 현재의 재무적 곤경 예측모형은 지난 세기 30년대에 기원을 두고 있으며 연구 방법은 주로 판별분석, 주성분 분석법 등 전통적인 통계 방법을 이용하여 기업의 재무적 곤경을 예측하고있다. 또한 전통적인 통계 모형에는 복잡한 가설 조건과 엄격한 데이터 요구가 있다. 이것은 재무적 곤경 예측모형을 세우는 것을 통해서 재무 위험을 방지하고자 하는 기업에는 불리하다. 따라서 기존의 연구를 바탕으로 본 연구에서는 전통적인 통계 방법과 기계 학습 알고리즘을 선택하여 기업의 재무적 곤경을 예측하였으며, 두 가지 서로 다른 방법에 대하여 비교 분석을 진행하였다. 또한 문화산업기업의 재무적 곤경 예측모형의 연구에 대한 문헌이 비교적

적이므로 본 연구는 문화산업 기업을 연구표본으로 사용한다. 이는 문화산업 기업의 재무적 곤경 예측모형에 관한 연구의 방법론을 넓힐 수 있다. 이것은 본 연구와 관련된 기존 연구의 차별성이다.

따라서 본 연구의 목적은 전통적인 통계모형과 기계 학습(Machine Learning) 모형을 통해 중국 문화산업 기업의 재무적 곤경을 정확하게 예측하는 분석 모형을 탐색하는 데 있다. 예측모형을 구축하기 위하여 중국 128개 문화산업 상장 기업의 데이터를 수집하였다. 25개 설명변수로 이뤄진 데이터베이스를 토대로 판별분석과 로지스틱 회귀(Logistic) 등 전통적인 통계 방법과 서포트 벡터 기계(SVM), 결정 트리(Decision Tree)와 랜덤 포레스트(Random Forest) 등 기계학습을 이용한 예측모형을 구축하고 각 모형의 성능 평가를 위해 Python 소프트웨어를 사용한다.

본 연구는 효과적인 재무적 곤경 예측모형의 구축을 통하여 중국 문화산업 기업의 재무적 곤경에 대한 조기경보 신호를 얻을 수 있을 것으로 기대한다. 이는 중국 문화산업 기업의 경영자가 재무적 곤경을 예측하고 기업의 재무 위험을 줄이는 데 도움을 준다. 중국 문화산업 기업의 투자자에게 재무적 곤경 예측모형은 중국 문화산업 기업의 내재적 가치를 평가하는데 도움을 주어 투자자들의 투자 결정에 지원을 제공할 수 있다. 또 정부는 재무적 곤경 예측모형을 사용하여 중국 문화산업 기업에 대한 감독 관리를 진행함으로써 사회의 경제 안정을 보장할 수 있다. 따라서, 경영자, 투자자와 정부가 재무 위험을 식별해 대응할 수 있도록 중국 문화산업 기업의 재무적 곤경을 예측할 수 있는 모형을 구축하는 것이 필요하다. 이것도 본 연구의 활용 분야이다.

II. 이론적 배경

1. 기계학습

기계 학습(Machine Learning)은 다양한 분야를 다루는 교차 학과다. 주로 계산기가 어떻게 인류의 학습 행위를 모방하거나 실현한 후에 기존의 지식구조를 재조직하여 자체의 성능을 부단히 개선할 것인가를 연구한다. 기계 학습은 인공지능의 핵심이며 계산기가 지혜

를 구비할 수 있는 근본적인 경로이다.

Wiederhold G and Mccarthy J(2010)는 기계 학습이 컴퓨터가 특정한 코딩이나 명령어에 의존하지 않고 스스로 배우고 일할 수 있도록 하는 연구 분야라고 주장하였다[3]. Chen E(2011)은 기계 학습을 경험(Experience), 작업(Task), 작업 수행의 효과(Performance)를 도입함으로써 기계 학습에 대해 더 자세히 설명하였는데, 그는 기계 학습이란 기계가 경험(Experience)을 학습한 후 작업(Task)을 진행하는 효과(Performance)가 더욱 좋아진 것을 말한다[4]. 빅데이터의 발전은 기계학습을 생산, 생활과 학술 연구에서의 응용을 추진했다. 빅데이터란 정해진 시간에 통상적인 소프트웨어로 수집, 관리, 처리가 불가능한 데이터 집합을 말한다. 빅데이터는 데이터 규모가 크고, 데이터 분류가 많으며, 데이터 가치가 높은 정보 자산이다. 빅데이터의 핵심은 데이터의 가치를 활용하는 것이고, 기계학습은 데이터 가치를 활용하는 핵심 기술이다. 기계 학습에서는 데이터가 많을수록 예측 모델의 정확도를 높일 수 있다. 현대 기업에 대해, 거대한 상업 가치를 포함하고 있는 이런 데이터 정보는 기업이 과학적인 관리와 의사결정을 분석하는 기초이다. 하지만 현재 기업의 데이터 활용률은 2%~4%에 불과하다. 기업들이 데이터를 최대한 활용하지 않아 시간과 돈을 낭비했다. 따라서, 기계학습을 이용한 기업 데이터 분석은 기업의 비즈니스 의사결정, 경영관리 효율성 제고, 재무적 곤경 예측에 도움을 줄 수 있다.

2. 재무적 곤경

재무적 곤경(Financial Distress)은 재무적 위기(Financial Crisis)라고도 한다. 가장 심각한 재무적 곤경은 기업 파산(Bankruptcy)이다. 기업이 재무적 곤경으로 인해 파산을 초래하는 것은 사실상 위약 행위이므로, 재무적 곤경은 위약 위험(Default Risk) 이라고도 한다[5].

기업이 재무적 곤경에 빠지는 것은 점진적인 과정으로, 대다수 기업의 재무적 곤경은 모두 재무 상태가 정상에서 점진적으로 악화될 때까지이며, 결국에는 재무적 곤경에 빠지거나 파산을 초래한다.

따라서 기업의 재무적 곤경은 전조 적일 뿐만 아니

라 예측할 수 있다. 정확하게 기업의 재무적 곤경을 예측하는 것은 투자자와 채권자의 이익을 보호하고 경영자를 도와 재무위기를 방지하며 정부를 도와 상장기업의 품질과 증권시장의 위험을 감독하는 데 있어서 모두 중요한 현실적 의의가 있다.

상장기업 재무적 곤경 예측은 기업 재무제표 데이터에 따라 다른 데이터 특징을 뽑아 기업이 미래에 재무적 곤경에 빠질지 예측하는 것이다. 상장기업의 재무적 곤경 예측에 대한 연구는 주로 두 가지 방향으로 나뉘는데, 하나는 서로 다른 특징의 재무 데이터를 사용하는 것이고, 다른 하나는 예측모형을 구축하는 데 사용되는 다른 알고리즘 방법을 연구하는 것이다.

재무 데이터 특징을 선택하는 단계에서 연구자들은 상장기업의 재무적 곤경 예측모형에서 각종 예측변수를 고려하였다. Aitman(1968)은 회계변수를 사용하고 회계자료에 근거하여 재무비율을 구성하고 이를 기업의 채무불이행위험을 측정하는 지표로 구성되었다[6]. Shunway T(2001)는 파산예측모형에 시장기반 변수와 2개의 일반적인 회계변수를 포함하는 방법으로 모형을 개선하였다[7]. Servaas(2015)는 재무 데이터는 모든 실패 정보를 포함하지 않으며 시장 정보를 사용하여 예측 가능성을 높일 수 있다고 주장하였다[8]. 선택 알고리즘 단계에서 학자들은 이미 여러 가지 다른 알고리즘 모형을 사용했는데, 이 모형들은 주로 통계 모형과 기계학습 모형 두 종류로 나눌 수 있다. 본 연구는 주로 서로 다른 알고리즘에 기초하여 전통적인 통계 모형과 기계 학습 모형을 건립한다.

3. 전통적 통계모형

전통적인 통계 모형은 기업의 재무적 곤경을 예측하는데 최초로 사용된 방법이다. 그러나 전통적인 통계 방법은 데이터에 대해 엄격한 요구가 있어 전통적인 통계 예측모형의 응용을 제한하고 있다. 본 연구는 전통적인 통계 모형을 판별분석과 로지스틱 회귀로 중국 문화산업 기업의 재무적 곤경을 예측하였다.

3.1 판별분석 모형

판별분석은 선형함수값에 따라 판별하는 통계적 방법이다. 판별 분석의 기본 사상은 투영인데 적합한 투

영 방향을 선택하여 원래 고차원 공간에 있던 표본 데이터를 저차원공간에 투영하고 저차원 공간에서 다시 분류를 진행해야 한다. 투영을 통해 각각의 표본의 차이를 최소화하고, 투영된 표본의 간격은 가능한 한 크게 한다. 그런 다음 준칙에 근거하여 판별식을 구축하여 미지 유형의 표본을 판별하고 분류한다. 판별 분석을 응용하는 제한 조건이 적기 때문에 판별 분석의 응용 범위가 비교적 넓다[9]. William(1966)는 판별 분석 모형을 개척적으로 운용하여 재무비율로 건강한 회사와 파산한 회사를 분류하였으며 연구, 분석을 통해 현금 흐름양 부채비율은 회사의 파산여부를 가장 잘 예측할 수 있는 지표라는 결론을 내었다[10]. 徐永智(2013) 등은 재무적 곤경을 예측하는 판별 분석 모형을 중국 차스닥 상장기업에 처음으로 적용하였다. 그 결과 86%의 정확도를 보였다[11].

3.2 로지스틱 회귀(Logistic)

로지스틱 회귀(Logistic)는 두 개의 다른 범주를 나타내는 출력 변수의 결과만을 갖는 분류 모형이다. 샘플 카테고리가 플러스가 될 확률이 0.5 이상이면 샘플이 플러스로 보고, 그렇지 않으면 마이너스가 된다[12]. 姜國華(2014) 등은 로지스틱 회귀모형을 이용하여 ST(경고) 기업의 영향요인을 분석하였는데 많은 재무 지표 중에서 소수의 재무 지표만이 기업이 ST(경고)를 받을 가능성에 크게 영향을 미친다는 것을 발견하였다[13]. Ohlson(1980)은 논리 회귀 알고리즘을 디폴트 추정에 적용한 최초의 연구자 중 한 명으로, 로지스틱 알고리즘을 통해 105개 부도업체와 2,058개 건강회사를 모형으로 한 회사 재무파산 모형은 잠재적 차입자의 디폴트 확률을 정하였다[14]. 楊貴軍(2019) 등은 Benford 법칙을 로지스틱 회귀모형에 도입하였다. 효과적인 조기경보 변수를 증가시키고 재무적 곤경의 정확성을 제고하였다[15].

4. 기계학습 모형

전통적인 통계 방법을 사용하여 예측 모형을 구축하려면 복잡한 가설 조건과 엄격한 데이터 요구가 충족되어야 한다. 기계학습 알고리즘은 이처럼 엄격한 조건의 규제가 필요 없고, 새로운 데이터에 근거해 기존 모형

을 끊임없이 업데이트할 수 있다.

따라서 본 연구는 기업의 재무적 곤경을 예측하는 모형으로 기계학습 방법론을 검토한다. 총 3개의 기계학습 모형으로 재무적 곤경을 예측하고 성과를 비교한다. 기계학습 모형은 주어진 데이터 집합을 바탕으로 새로운 데이터가 어느 집단에 속할 것인지 판단하는 모형을 의미한다. 본 연구에서는 기계학습 모형으로 서포트 벡터 기계(SVM), 결정 트리(Decision Tree), 랜덤 포레스트(Random Forest) 모형을 선택하여 중국 문화산업 기업의 재무적 곤경을 예측하였다[16].

4.1 서포트 벡터 기계(SVM)

서포트 벡터 기계(SVM)는 패턴인식, 자료 분석을 위한 기계 학습의 분야이며, 주로 분류 또는 회귀 분석에 사용한다. 서포트 벡터 기계(SVM)는 두 개의 그룹 중 어느 하나에 속하는 데이터 집합에서, 주어진 데이터 집합의 특성을 분석하여, 이를 바탕으로 하여 새로이 입력되는 데이터가 어떤 분야에 속하는지를 판단할 수 있는 비확률적 이진 선형 분류 모형을 만든다. 만들어진 모형은 데이터가 매핑된 공간상에서 경계로 표시되며, 서포트 벡터 기계(SVM)는 가장 큰 폭의 경계를 찾아가는 알고리즘이다. 서포트 벡터 기계(SVM)는 선형 분류, 비선형 분류에서 복합적으로 사용할 수 있다[17]. 또한, 서포트 벡터 기계(SVM) 모형의 최적화는 일반적으로 핵함수의 전환으로 이루어진다. 따라서 핵함수 선택은 최종 서포트 벡터 기계(SVM) 모형의 좋고 나쁨을 결정한다. 상용하는 핵함수 모형은 Linear, Poly, Rbf, Sigmoid를 포함하고 있다. Mselmi, Lahiani and Hamza(2017)는 재무위기 발생 1년 전 기업의 재무 보고서 데이터를 이용하여 예측한 서포트 벡터 기계 모형의 정확도에 88.57%였다[18]. 張露(2021) 등은 중국 상장기업을 표본으로 서포트 벡터 기계(SVM) 모형과 판단분석 모형을 구축하였다. 결과에 따르면 서포트 벡터 기계(SVM)이 판단분석 모형보다 우수하다는 것을 보여준다[19].

4.2 결정 트리(Decision Tree)

결정 트리(Decision Tree)는 의사결정을 위한 규칙과 규칙에 따른 결과를 트리 구조로 구조화한 의사 결

정 지원 도구로 시작되었다. 결정 트리(Decision Tree)는 데이터 과학 분야에서 주로 활용되며, 의사 결정을 위해 목표에 가장 가까운 결과를 낼 수 있는 방안을 찾기 위해 사용된다. 결정 트리(Decision Tree)는 수치형 데이터와 분류형 데이터를 포함하는 모형을 구성한다[20]. 武彩霞(2020)은 결정 트리(Decision Tree)를 이용하여 26개 상장기업을 표본으로 하여 예측모형을 구축하고 예측모형에 대해 분석을 진행하였다. 연구결론은 결정 트리(Decision Tree)는 좋은 예측 효과가 있는 것으로 나타났다[21].

4.3 랜덤 포레스트(Random Forest)

랜덤 포레스트(Random Forest)는 다수의 결정 트리(Decision tree)를 학습하는 앙상블 종류이다. 앙상블은 여러 개의 서로 다른 예측모형을 생성한 후 모형들의 결과를 종합하여 하나의 최종 결과를 도출하는 방법을 의미한다. 이 방법론은 다수의 결정 트리(Decision trees)에 기반하여 정확도를 더욱 높인다. 이로부터 분류 또는 평균 예측치를 출력한다[22]. Ristolainen(2018)은 로지스틱과 랜덤 포레스트 모형을 비교한 결과 무작위 숲 예측 모형은 은행권의 재무적 곤경 예측을 위한 가장 좋은 도구 중 하나였다[23]. Tsai and Wu(2008)는 결정 트리 모형과 랜덤 포레스트 모형의 파산예측과 신용평점 기능을 비교한 결과 랜덤 포레스트 모형의 예측이 결정 트리 모형보다 나은 경우가 많았다[24]. 張培榮(2019)는 기계학습 모형으로 중국 제조업 상장기업을 대상으로 재무적 곤경 예측을 하고 있다. 랜덤 포레스트 모형이 로지스틱 모형보다 월등히 우수함을 발견하였다[25].

5. 모형의 평가

본 연구는 전통적인 통계 모형과 기계학습 모형의 재무적 곤경 예측 효과를 비교하기 위해 정확도, 리콜률, AUC 값과 ROC 곡선 등 4가지 평가지표를 도입하였다[26].

5.1 혼동행렬의 구성

[표 1]에서 기계학습 혼동행렬의 구성은 실제값과 예측값의 관계를 나타내는 행렬로서 행에는 실제값을 그

리고 열에는 예측값으로 구성하고, 이렇게 구성된 행렬의 각 값은 해당 경우를 만족하는 수를 나타낸다.

본 연구는 기계학습을 활용하여 기업의 재무 상태를 정상(=0)과 위험(=1)이라는 2가지로 분류한다. 표본의 실제값과 모형 예측값의 조합에 따라 4개의 카테고리가 형성되는데, 이 카테고리는 각각 TP(True Positive), FP(False Positive), TN(True Negative), FN (False Negative)이다. TP(True Positive)는 재무적 곤경의 예측값과 실제값이 모두 1(위험)인 경우의 개수를, TN(True Negative)은 재무적 곤경의 예측값과 실제값이 모두 0(정상)인 경우의 개수를 의미한다. FP(False Positive)는 예측값은 1이지만, 실제값은 0인 경우이며, FN(False Negative)은 예측값은 0이지만, 실제값은 1인 경우의 개수를 나타낸다. 이 4개 부류에 포함된 표본을 더하면 완전한 데이터 세트가 되며 분류 결과의 혼동행렬은 [표 1]에서 표시한 것과 같다[27].

표 1. 혼동행렬의 구성

실제값	예측값	
	1(위험)	0(정상)
1(위험)	TP(True Positive)	FN(False Negative)
0(정상)	FP(False Positive)	TN(True Negative)

5.2 정확도(Accuracy)

정확도(Accuracy)는 정확하게 분류된 표본이 전체 표본을 차지하는 비율을 차지한다. 분류 모형의 분류 효과를 가장 직관적으로 판단하는 평가 지표이다. 정확도(Accuracy)는 (1)과 같이 계산할 수 있다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

5.3 리콜률(Recall)

리콜률(Recall)은 포지티브(Positive)를 정확하게 분류한 경우의 확률이다. 리콜률(Recall)은 (2)와 같이 계산할 수 있다.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

5.4 ROC 곡선과 AUC 값

ROC 곡선은 x축이 TPR(True Positive Rate)이고 y축이 FPR(False Positive Rate)인 평면에 그려진다. TPR과 FPR은 각각 (3), (4)과 같이 계산할 수 있다. ROC 곡선은 0 또는 1의 값을 반환하는 것이 아니라 [0, 1] 구간 사이의 확률값을 계산한다. AUC 값은 ROC 곡선 아래의 면적을 나타낸다. AUC의 최대값은 1이며, 좋은 모형일수록 1에 가까운 값을 가진다. AUC 값은 (5)와 같이 계산할 수 있다[28].

$$TPR = \frac{TP}{TP + FN} \quad (3)$$

$$FPR = \frac{FP}{FP + TN} \quad (4)$$

$$AUC = \frac{1}{2} \sum_{i=1}^{n-1} (x_{i+1} - x_i) * (y_i - y_{i+1}) \quad (5)$$

III. 표본 선정 및 지표 구축

1. 표본 선택

본 연구는 중국 문화산업 기업의 주식이 ST로 표기되어 있는지를 기업의 재무 상태를 판단하는 기준으로 구성되었다. 즉 상장기업이 재무 상황으로 ST로 표기되어 있으면 재무적 곤경이 있는 것으로, ST로 표기되어 있지 않으면 정상인 것으로 판단하였다. 기업이 ST로 표기되는 당년은 T 년도이며, 해당 기업의 T-3 년의 기업 재무제표를 예측 데이터로 선택한다. 따라서 본 연구는 2018년~2020년 중국 문화산업 기업의 재무 상황에 따라 2015년~2017년의 재무제표를 예측 데이터로 하였다.

본 연구에서 분석에 사용된 표본을 추출하기 위한 구체적인 기준은 다음과 같다.

첫째, 2018년부터 2020년까지 중국 상하이(上海)와 선전(深圳) 증권거래소에 상장된 중국 문화산업 기업.

둘째, 중국 CSMAR(國泰安數據庫)의 데이터베이스에

서 재무제표자료를 추출할 수 있는 기업.

이러한 표본 추출 기준을 모두 충족한 최종 분석 표본은 128개 기업의 384개 표본이다.

2. 데이터베이스의 구축

본 연구는 2015년~2020년 재무적 곤경 때문에 증시 퇴출 경고(ST로 표시됨)를 받은 32개 중국 문화산업 상장기업과 재무적 곤경이 발생하지 않은 96개 중국 문화산업 상장기업을 연구 대상으로 표본을 선정하였다. 본 연구에서는 모형의 예측성능을 검증하기 위해 전체 데이터를 70%의 훈련(Train) 데이터와 30%의 테스트(Test) 데이터로 나누었다. 훈련(Train) 데이터는 주로 재무적 곤경 예측모형을 예측하는데 사용된다. 테스트(Test) 데이터는 주로 재무적 곤경 예측모형의 예측성능을 검증하는데 사용된다.

3. 변수의 확정

변수의 선택 원칙은 연구에서 사용하는 데이터를 결정하고, 데이터는 예측 모형을 구축하는 기초가 된다. 재무적 곤경 예측모형의 연구에서 연구 문제가 다름에 따라 선택하는 변수도 다르다. 따라서 본 연구는 蔣盛益(2010) 등의 재무적 곤경 예측 모형변수를 선택하는 방법을 참고하여 3개 방면에서 변수를 선택한다[29].

첫째, 선택한 변수는 상장기업의 재무 상황을 전면적으로 반영할 수 있어야 한다.

둘째, 선택한 변수의 데이터가 진실해야 한다.

셋째, 재무적 곤경 예측 모형 구축의 목적은 상장기업의 데이터에 근거하여 상장회사의 재무적 곤경 발생 확률을 예측하는데 있으므로 선택한 변수는 반드시 예측에 사용할 수 있어야 한다.

이상의 원칙에 따라 본 연구에서는 종속 변수 1개와 설명 변수 25개를 선택하여 기업 재무적 곤경 예측모형을 구축하였다.

3.1 종속 변수

본 연구의 종속 변수는 기업의 재무적 곤경이며, 이 지표로 중국 문화산업 기업의 재무적 곤경을 판단한다. 이 지표는 분류변수로서 0과 1이 있다. 지표의 값이 0이면 기업에 재무적 곤경이 존재하지 않음을 나타내고

지표의 값이 1이면 기업에 재무적 곤경이 있음을 의미한다.

3.2 설명 변수

설명 변수의 경우 크게 기업의 채무상환능력, 기업의 채산성, 기업의 경영능력, 기업의 발전 능력, 기업의 현금 회수 능력 그리고 기업의 지분구조로 총 6가지 분류로 나뉜다[30]. 세부적인 변수는 기존 연구에서 다루는 변수들을 참고하여 총 25 개의 변수로 구성된다.

채무 상환능력은 기업의 재무 상황을 가늠하는 중요한 지표로서 기업이 채권자의 합법적 이익을 보장할 수 있는가, 기업이 부채 수준을 합리적인 범위 내로 통제하여 부채경영의 위험을 피면할 수 있는가를 반영한다. 따라서 본 연구에서는 기업의 채무상환능력을 평가하는 지표에는 유동비율, 현금비율, 권익 승수와 소유권 비율을 포함하였다.

채산성은 기업이 경영을 통해 이익을 낼 수 있는 능력을 가늠하는 지표이다. 이윤을 얻는 것은 기업 존재의 근본적인 목적이며, 이윤을 얻는 능력은 기업의 발전에 매우 중요하다. 따라서 본 연구에서는 기업의 채무상환능력을 평가하는 지표에는 영업이익증가율, 영업 이익률, 영업순이익률, 원가비용이익률, 주주당수익과 주당순자산을 포함하였다.

표 2. 변수의 측정

변수 종류	변수 이름	변수의 측정
종속 변수	재무적 곤경	재무적 곤경은 1이다 재무적 곤경 없음은 0이다
	유동비율	유동자산/유동부채
채무 상환 능력	현금비율	현금/유동부채
	권익 승수	자산/소유자권익
	소유권 비율	부채/소유자권익
	영업이익증가율	(영업이익당기-영업이익 상기)/영업이익 상기
설명 변수	채산성	영업이익률 영업순이익률 원가비용이익률 주주당수익 주당순자산
	미수금회전율	영업수익/미수금
	재고 회전율	영업원가/재고
	경영 능력	유동자산 회전율 고정자산회전율
	총자산 회전율	영업수익/총자산
	유동자산 회전율	영업수익/유동자산
고정자산회전율	영업수익/정자산	

발전 능력	자본축적률	(소유자권익당기-소유자권익상기)/소유자권익합계 당기
	총자산증가율	(총자산당기-총자산상기)/총자산당기
	주당 수익증가율	(주당순이익당기-주당 순이익상기)/주당주식수익상기
현금 회수 능력	영업수입증가율	(영업수입당기-영업수입 상기)/영업수입상기
	영업수익현금점유율	현금/영업수익
	현금 회수율	현금/자산
지분 구조	현금 재투자 비율	현금 / (자산-유동부채)
	지분집중도	1대 주주 지분율
지분 구조	Z 지수	기업 1대 주주 지분 비율/기업 2대 주주 지분 비율
	S 지수	기업 2대 주주와 10대 주주 지분 비율의 합

경영 능력은 기업의 경영 관리 수준을 가늠하는 지표이다. 경영능력은 기업이 자원을 합리적으로 배분하고 운용하는 능력을 반영한다. 따라서 본 연구에서는 기업의 경영 능력을 평가하는 지표에는 미수금회전율, 재고 회전율, 유동자산 회전율, 고정자산회전율과 총자산 회전율을 포함하였다.

발전 능력은 기업의 미래 발전 잠재력을 가늠하는 지표이다. 기업의 발전 능력은 기업이 앞으로 몇 년 동안의 재무 상황을 반영한다. 기업의 성장력이 강할수록 재무적 곤경에 직면할 가능성은 적다. 따라서 본 연구에서는 기업의 발전 능력을 평가하는 지표에는 자본축적률, 총자산증가율, 주당 수익증가율과 영업이익증가율을 포함하였다.

현금 회수 능력은 기업의 현금이 충분하지를 가늠하는 지표이다. 기업의 현금 회수 능력이 높을수록 기업의 재무 위험이 적다는 의미이다. 따라서 본 연구에서는 기업의 현금 회수 능력을 평가하는 지표에는 영업수익 현금점유율, 현금 회수율과 현금 재투자 비율을 포함하였다.

지분구조는 기업의 안정 여부를 가늠하는 중요한 지표로서 기업의 지분구조가 안정할수록 기업의 재무위험이 적어진다. 따라서 본 연구에서는 기업의 지분구조를 평가하는 지표에는 지분집중도, Z 지수와 S 지수를 포함하였다.

[표 2]는 이 논문의 종속 변수, 설명 변수를 측정하는 방법이다.

3.3 변수의 기술통계

[표 3]은 표본 데이터를 평균, 표준편차, 최대치와 최

소치로 기술한 통계 분석이다.

표 3. 변수의 기술통계

변수	종류	평균	표준 편차	최 소치	최 대치
유동비율	0(정상)	1.656	0.983	0.318	4.7
	1(위험)	1.741	1.424	0.66	6.787
현금비율	0(정상)	0.584	0.65	0.042	3.737
	1(위험)	0.64	0.879	0.017	3.732
권익승수	0(정상)	1.386	3.729	-26.8	4.698
	1(위험)	2.129	1.607	1.077	7.304
소유권비율	0(정상)	0.386	3.729	-27.8	3.698
	1(위험)	1.129	1.607	0.077	6.304
자본축적률	0(정상)	0.515	1.324	-2.62	5.555
	1(위험)	0.85	2.341	-0.24	9.457
총자산증가율	0(정상)	0.446	0.841	-0.50	3.792
	1(위험)	0.771	2.234	-0.14	9.077
주당수익증가율	0(정상)	0.915	4.565	-17.1	16
	1(위험)	6.095	13.76	-1.5	51.38
영업이익증가율	0(정상)	0.855	3.571	-0.60	28.53
	1(위험)	0.519	0.845	-0.58	2.894
미수금회전율	0(정상)	9.049	15.38	0.927	90.37
	1(위험)	2.99	1.6	0.843	6.867
재고회전율	0(정상)	456.7	3465	0	27744
	1(위험)	5.584	7.075	0	22.07
유동자산회전율	0(정상)	1.217	1.092	0.218	7.474
	1(위험)	0.678	0.361	0.073	1.614
고정자산회전율	0(정상)	35.72	78.19	0	383.7
	1(위험)	155.4	556.1	0.982	2238.
총자산회전율	0(정상)	0.566	0.922	0.078	7.215
	1(위험)	0.275	0.166	0.048	0.692
지분집중도	0(정상)	25.65	11.89	6.351	58.36
	1(위험)	23.10	8.219	11.60	35.64
Z지수	0(정상)	4.525	5.615	1.004	31.03
	1(위험)	4.69	8.415	1.085	35.91
S지수	0(정상)	28.58	14.04	4.388	53.88
	1(위험)	27.85	11.50	4.03	42.64
영업수익현금점유율	0(정상)	1.013	0.214	0.519	2.076
	1(위험)	0.934	0.121	0.714	1.163
현금회수율	0(정상)	0.032	0.093	-0.29	0.214
	1(위험)	0.029	0.056	-0.12	0.126
현금재투자비율	0(정상)	0.091	0.923	-4.51	3.76
	1(위험)	0.242	0.629	-0.53	2.349
영업이익증가율	0(정상)	0.039	0.15	-0.92	0.324
	1(위험)	0.069	0.041	-0.03	0.146
영업이익률	0(정상)	0.108	0.365	-2.47	0.5
	1(위험)	0.163	0.521	-1.56	0.952

영업순이익률	0(정상)	0.094	0.382	-2.58	0.498
	1(위험)	0.288	0.24	-0.05	0.913
원가비용이익률	0(정상)	0.2	0.307	-1.27	1.072
	1(위험)	0.454	0.462	-0.07	1.711
주당수익	0(정상)	0.351	0.762	-4.55	1.868
	1(위험)	0.278	0.316	-0.11	1.324
주당순자산	0(정상)	5.24	4.24	-0.16	23.78
	1(위험)	3.042	2.363	0.545	10.19

주) 0(정상): 정상적 기업, 1(위험): 재무적 곤경

[표 3]에 따르면 재무적 곤경 기업과 정상적 기업은 변수의 값에서 차이가 있었다.

영업이익증가율, 미수금회전율, 재고 회전율, 유동자산 회전율, 총자산회전율 등 변수의 값에 따르면 재무적 곤경 기업의 변수의 값이 정상적 기업보다 작다. 이는 정상적 기업이 재무적 곤경 기업보다 재무 실적과 경영능력이 더 뛰어나다는 것을 보여준다. 지분집중도, S지수 등 변수의 값에 따르면 재무적 곤경 기업의 변수의 값이 정상적 기업보다 작으며, 이는 정상적 기업이 재무적 곤경 기업보다 지분 구조가 더 안정적이며 기업 발전에 더 유리함을 나타냈다. 영업수익현금점유율, 현금회수율, 현금재투자비율 등 변수의 값에 따르면 정상적 기업이 재무적 곤경 기업보다 자금회수 능력이 더 우수한 것으로 나타났다. 기업의 성장성에서는 주당수익, 주당순자산 등 변수의 값에 따르면 정상적 기업이 재무적 곤경 기업보다 더 성장할 수 있다. 이 밖에 다른 변수들 사이에서도 재무적 곤경 기업과 정상적 기업 간에 차이가 있었다.

3.4 다중 공선형 검사

변수 사이에 다중 공선형이 존재하여 이후의 예측모형에 오차를 가져오는 것을 방지하기 위하여 먼저 변수에 대해 공선형 검증을 해야 한다. 변수의 VIF 검사 결과는 [표 4]와 같다.

표 4. VIF의 검사 결과

변수	VIF	1/VIF
영업순이익률	7.78	0.128
원가비용이익률	6.22	0.161
총자산 회전율	6.15	0.163
유동 자산 회전율	5.90	0.170
유동비율	4.54	0.220
현금비율	4.22	0.237

영업이익률	4.14	0.242
주당수익	2.61	0.383
Z 지수	2.35	0.426
미수금회전율	2.28	0.468
지분집중도	2.26	0.443
주당순자산	2.10	0.477
S 지수	2.09	0.478
소유권 비율	1.86	0.539
현금 회수율	1.76	0.568
총자산증가율	1.47	0.680
자본축적률	1.42	0.704
권익 승수	1.42	0.706
영업수익현금점유율	1.14	0.875
현금 채무자 비율	1.13	0.889
고정자산회전율	1.09	0.916
재고 회전율	1.07	0.935
영업수입증가율	1.06	0.945
주당 수익증가율	1.05	0.949
영업이익증가율	1.03	0.973
Mean	VIF	2.790

본 연구에서는 선택된 변수에 대해 분산 확대 인자 방법을 사용하여 공통성을 검사하였는데, 검사 결과 VIF 값으로 판단하여 $0 < VIF < 10$ 이면 공선형이 존재하지 않는다[30]. [표 4]에 따르면 모든 변수의 VIF 값이 10보다 작기 때문에 변수 간 다중 공선형이 존재하지 않는다. 모형 예측이 가능하다.

3.5 데이터의 정규화

데이터를 사용하여 모형을 구축하기 전에 데이터 간의 차이를 해소하기 위해 데이터를 정규화 처리해야 한다[31]. 공식은 다음과 같다.

$$x^* = \frac{x - \mu}{\sigma}$$

이 중 μ 은 표본의 평균이고 σ 는 표본의 표준편차이다. 정규화 후 데이터의 평균은 0이고 표준편차는 1이다.

IV. 전통적 통계모형과 기계학습 모형의 예측 결과와 분석

1. 전통적 통계모형의 예측 결과와 분석

1.1 판별분석 모형의 예측 결과와 분석

본 연구는 25개의 종속변수를 예측변수로, 1개의 분류변수를 설명변수로 하여 결정모형을 구축하였다. [표

5]는 판정분석 모형의 예측 결과를 보여준다.

표 5. 판정분석 모형의 예측 결과

모형	정확도 Accuracy	리콜률 Recall	AUC
판정	0.89	0.89	0.47

[표 5]의 판정분석 모형의 예측 결과에 따르면 판정 분석 모형의 정확도(Accuracy)는 89%이며 리콜율(Recall)은 89%이다.

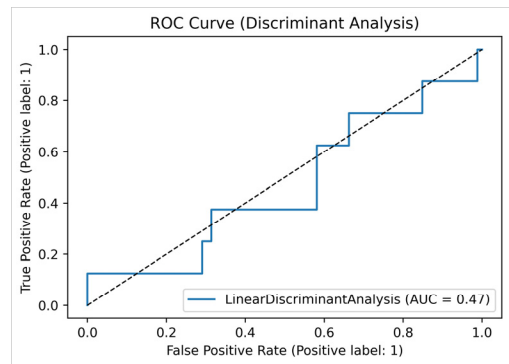


그림 1. 판정분석 예측 모형의 ROC 곡선

[그림 1]에 따르면 판정분석 모형의 AUC 값은 0.47이다.

1.2 로지스틱 회귀(Logistic) 모형의 예측 결과와 분석

로지스틱 회귀(Logistic) 모형은 기업이 재무적 곤경(ST로 표시됨)에 빠질 확률을 계산해 역량과 비교해서 판단한다. 또한 로지스틱 회귀(Logistic)은 격차 검색과 K-절차 검증을 통해 알고리즘을 최적화한다. 비교에 따라 가장 좋은 알고리즘을 선택한다. [표 6]은 로지스틱 회귀(Logistic) 모형의 예측 결과를 보여준다.

표 6. 로지스틱 회귀(Logistic) 모형의 예측 결과

모형	정확도 Accuracy	리콜률 Recall	AUC
결정트리	0.88	0.87	0.56

[표 6]의 로지스틱 회귀(Logistic) 모형의 예측 결과에 따르면 로지스틱 회귀(Logistic) 모형의 정확도

(Accuracy)는 88%이며 리콜율(Recall)은 87%이다.

[그림 2]에 따르면 로지스틱 회귀(Logistic) 모형의 AUC 값은 0.56이다.

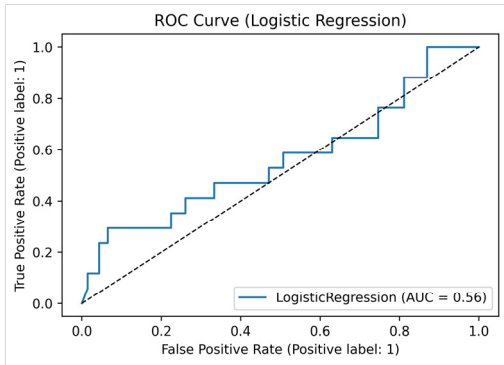


그림 2. 로지스틱 회귀(Logistic) 예측 모형의 ROC 곡선

2. 기계학습 모형의 예측 결과와 분석

2.1 서포트 벡터 기계(SVM) 모형의 예측 결과와 분석

본 연구는 서포트 벡터 기계(SVM)의 네 가지 핵함수에 대해 각각 분류기를 구축하고 비교 분석을 통해 최고의 서포트 벡터 기계(SVM) 모형을 선정한다. [표 7]은 네 가지 분류기 모형의 학습 결과를 보여준다[33].

표 7. 핵 함수 모형의 예측 결과

핵 함수 모형	정확도 Accuracy	리콜률 Recall	AUC
Linear	0.92	0.92	0.51
Poly	0.93	0.93	0.70
Rbf	0.90	0.90	0.58
Sigmoid	0.92	0.92	0.39

[표 7]의 핵 함수 모형을 예측한 결과에 따르면 핵 함수인 Poly의 정확도(Accuracy)가 가장 높고 그 다음이 핵 함수인 Linear와 Sigmoid, 마지막이 핵 함수인 Rbf이다. 리콜률(Recall)에 있어서 핵 함수인 Poly가 가장 좋으며 그 다음이 핵 함수인 Linear와 Sigmoid, 마지막은 핵 함수인 Rbf이다.

AUC 값은 모형의 좋고 나쁨을 가늠하는 중요한 지표로 모형의 AUC 값이 1에 가까울수록 모형의 표현이 가장 좋다고 평가한다. [표 7] 핵 함수 모형의 AUC 값

에 따르면 핵 함수인 Poly가 가장 좋으며, 그 다음이 핵 함수인 Rbf, 그 다음이 핵 함수인 Linear, 마지막은 핵 함수인 Sigmoid이다.

[표 7]에 따르면 핵 함수인 Rbf 서포트 벡터 기계(SVM) 모형은 기계학습 효과가 가장 좋으며 핵 함수인 Linear, 다시 핵 함수인 Sigmoid, 마지막 핵 함수인 Rbf가 뒤를 이었다. 따라서 서포트 벡터 기계(SVM) 모형의 핵 함수를 Poly로 선택한다.

2.2 결정 트리(Decision Tree) 모형의 학습 결과와 분석

본 연구는 나무 결정 트리를 통해 모형의 파라미터에 대한 K 절점기 검증과 격자 검색을 통해 최적화된 모형 평가 효과를 얻었다. [표 8]은 결정 트리(Decision Tree)의 학습 결과를 보여준다.

표 8. 결정 트리(Decision Tree) 모형의 예측 결과

모형	정확도 Accuracy	리콜률 Recall	AUC
결정 트리	0.92	0.91	0.67

[표 8]의 결정 트리(Decision Tree) 모형의 학습 결과에 따르면 결정 트리(Decision Tree) 모형의 정확도(Accuracy)는 92%이며 리콜률(Recall)은 91%이다.

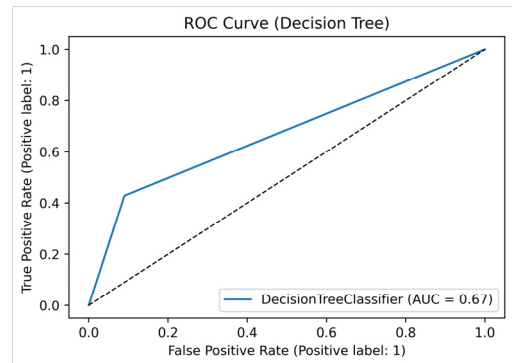


그림 3. 결정 트리(Decision Tree) 예측 모형의 ROC 곡선

[그림 3]에 따르면 결정 트리(Decision Tree) 모형의 AUC 값은 0.67이다.

2.3. 랜덤 포레스트(Random Forest) 모형의 예측 결과와 분석

본 연구는 불균형 데이터 집합의 개선된 임의의 삼림에 대해 예측 모형을 수립하는 것을 선택한다. [표 9]는 랜덤 포레스트(Random Forest) 모형의 예측 결과를 보여준다.

표 9. 랜덤 포레스트(Random Forest) 모형의 예측 결과

모형	정확도 Accuracy	리콜률 Recall	AUC
랜덤 포레스트	0.95	0.94	0.83

[표 9]의 랜덤 포레스트(Random Forest) 모형의 예측 결과에 따르면 랜덤 포레스트(Random Forest) 모형의 정확도(Accuracy)는 95%이며 리콜율(Recall)은 94%이다.

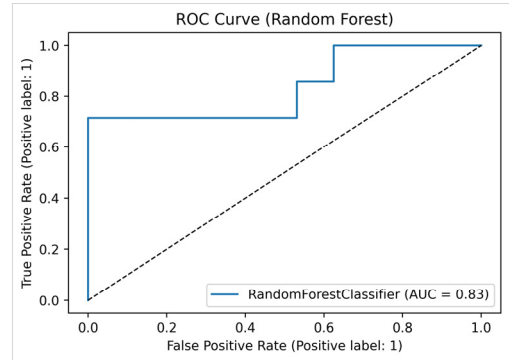


그림 4. 랜덤 포레스트(Random Forest) 예측 모형의 ROC 곡선

[그림 4]에 따르면 랜덤 포레스트(Random Forest) 모형의 AUC 값은 0.83이다.

3. 각 모형의 예측 결과 비교

본 연구에서는 두 개의 전통적인 통계모형과 세 개의 기계학습 모형을 이용하여 중국 문화산업 기업의 재무적 곤경을 예측하였다. 5개의 모형의 예측 결과는 [그림 5]과 같다.

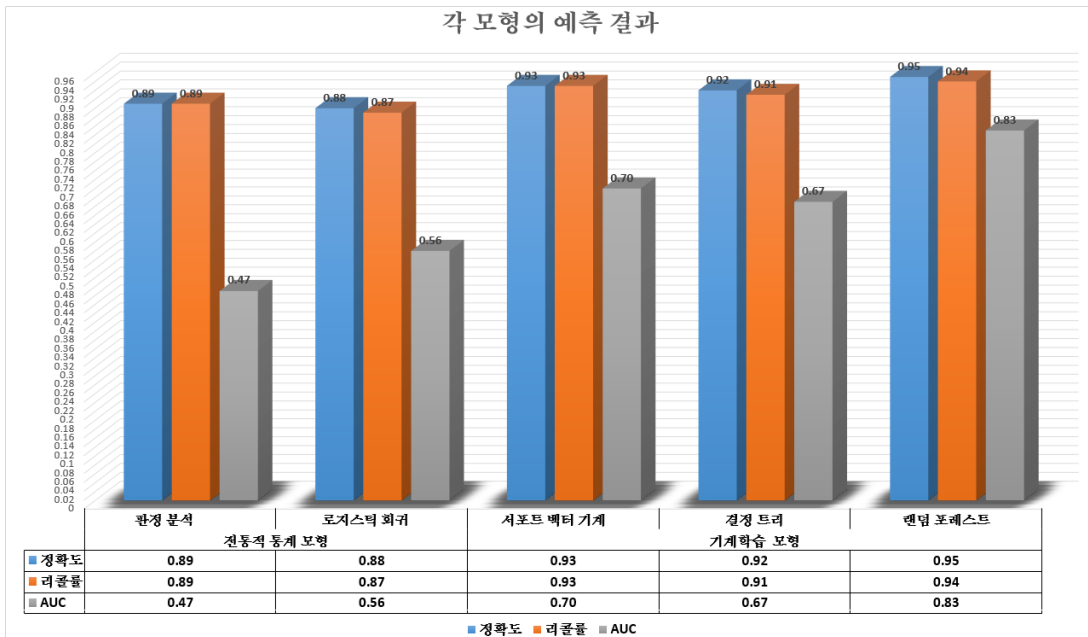


그림 5. 각 모형의 예측결과

3.1. 두 개의 전통적인 통계모형의 예측 결과 비교

[그림 5]에 따르면 전통적인 통계모형에서 판별분석 모형의 정확도(Accuracy)가 가장 높고 그 다음이 로지스틱 회귀(Logistic) 모형이다. 리콜률(Recall)에 있어서 판별분석 모형은 가장 높고 그 다음이 로지스틱 회귀(Logistic) 모형이다. AUC 값에 있어서 로지스틱 회귀(Logistic) 모형이 가장 좋으며, 그 다음이 판별분석 모형이다.

3.2. 세 개의 기계학습 모형의 예측 결과 비교

[그림 5]에 따르면 기계학습 모형에서 랜덤 포레스트(Random Forest)의 정확도(Accuracy)가 가장 높고 그 다음이 핵함수인 서포트 벡터 기계(SVM), 마지막은 결정 트리(Decision Tree)이다. 리콜률(Recall)에 있어서 랜덤 포레스트(Random Forest)는 가장 높고 그 다음이 서포트 벡터 기계(SVM), 그마지막은 결정 트리(Decision Tree)이다. AUC 값에 있어서 랜덤 포레스트(Random Forest)가 가장 좋으며, 그 다음이 서포트 벡터 기계(SVM), 마지막은 결정 트리(Decision Tree)이다.

3.3. 전통적인 통계모형과 기계학습 모형의 예측 결과 비교

[그림 5]에 따르면 정확도(Accuracy), 리콜률(Recall)과 AUC 값 등 예측모형의 평가 효과를 가늠하는 지표에서 3개 기계학습 모형의 예측 결과가 모두 2개 전통적인 통계모형보다 우월하다. 이는 기계학습 모형의 예측 효과가 전반적으로 전통적인 통계모형보다 우월하다는 것을 설명한다. 기계학습 모형은 중국 문화산업 기업의 재무적 곤경을 예측함에 있어서 전통적인 통계모형보다 더 좋은 예측 효과를 얻을 수 있다.

V. 결론

본 연구의 목적은 전통적인 통계모형과 기계학습(Machine Learning) 모형을 통해 중국 문화산업 기업의 재무적 곤경을 정확하게 예측하는 분석 모형을 탐색하는 데 있다. 예측모형을 구축하기 위하여 중국 128개

문화산업 상장 기업의 데이터를 수집하였다. 25개 설명변수로 이뤄진 데이터베이스를 토대로 판별분석과 로지스틱 회귀(Logistic) 등 전통적인 통계 방법과 서포트 벡터 기계(SVM), 결정 트리(Decision Tree)와 랜덤 포레스트(Random Forest) 등 기계학습을 이용한 예측모형을 구축하고 각 모형의 성능 평가를 위해 Python 소프트웨어를 사용한다. 분석 결과, 예측 성능이 가장 좋은 모형은 랜덤 포레스트(Random Forest) 모형으로 95%의 정확도를 보였다. 그 다음은 서포트 벡터 기계(SVM) 모형으로 93%의 정확도를 보였다. 그 다음은 결정 트리(Decision Tree) 모형으로 92%의 정확도를 보였다. 그 다음은 판정분석 모형으로 89%의 정확도를 보였다. 예측 효과가 가장 낮은 모형은 로지스틱 회귀(Logistic) 모형으로 88%의 정확도를 보였다. 이는 중국 문화산업 기업의 재무적 곤경을 예측함에 있어서 기계학습 모형이 전통적인 통계 모형보다 더 좋은 예측 효과를 얻을 수 있음을 설명한다.

본 연구는 각 측면에서 기여한 것은 다음과 같다. 첫째, 본 연구에서는 중국 문화산업 기업의 재무적 곤경을 예측하는 문제에 대하여 사용하는 전통적인 통계 예측모형과 기계학습 예측모형을 비교하고 검증하였다. 기계학습 모형은 예측에 있어서 전통적인 통계 모형보다 우수한 성능을 보유하고 있음에도 불구하고 그동안 기업의 재무적 곤경 예측모형 연구에서는 적극적인 도입을 하지 않았다. 둘째, 이 연구는 우수한 예측모형을 제시함으로써 학자들 이 기업의 재무적 곤경을 예측하는 연구를 수행하는 데 있어서 기존 방법론이 지닌 한계를 극복하며 보다 정확한 예측 분석 모형을 사용할 수 있도록 한다. 이러한 결과는 기업의 재무적 곤경 예측연구의 방법론 측면에서 연구 시각을 확대하는 데 기여한다. 이 연구에서 제시한 혁신성에 대한 예측모형은 기업의 재무적 곤경 예측모형을 평가하거나 기업의 가치를 분석하는 투자자에게도 유용한 방법론으로 활용될 수 있을 것으로 기대한다.

본 연구는 여러 한계점을 존재한다. 첫째, 기업의 재무적 곤경에 대한 측정에 25개의 설명변수를 사용하였다. 이는 객관적 평가 방법으로 측정하는 연구 수단이지만 주관적 측정 방법이 요구될 수 있다. 예를 들어 해당 기업의 재무적 곤경에 대한 평가에 대한 기업 경영

자나 투자자의 평가연구 등 주관적 측정을 통해 추가적 분석이 진행하면 더 정확한 시사점을 제시할 수 있다. 둘째, 본 연구는 2015년~2020년 중국 문화산업 상장 기업을 대상으로 모형 학습 및 평가를 진행하였다. 이는 본 연구에서 중국 상장된 128개 문화산업 기업만 포함하기 때문에 데 큰 범위의 데이터를 수집하지 못했다. 다른 문화산업 기업의 데이터를 함께 분석할 수 있으면 모형 결과의 일반화가 더 수월해질 수 있을 것이다. 향후 연구에서 주관적인 설명 변수와 더 많은 기업의 데이터를 추가하면 더 구체적인 시사점을 제시할 수 있을 것이다. 본 연구에서 기계학습 접근으로 시도한 새로운 방법이 기업의 재무적 곤경 예측모형이라는 중요한 주제의 지평을 넓히는 데 기여하기를 바라며, 본 연구가 기업의 재무적 곤경 예측모형 분야에서의 데이터 과학적 예측연구에 대한 관심을 확대하는 역할을 할 수 있기를 희망한다.

참 고 문 헌

[1] 中國國家統計局, <http://www.stats.gov.cn/>
 [2] 百度百科, <http://www.baidu.com/>
 [3] G. Wiederhold and J. Mccarthy, "Arthur Samuel: Pioneer in Machine Learning," *Ibm Journal of Research & Development*, Vol.36, No.34, pp.329-331, 2010.
 [4] E. Chen, "Financial Ratio Selection for Business Crisis Prediction," *Expert Systems With Applications*, Vol.38, No.12, pp.1094-1102, 2011.
 [5] 趙麗瓊, 張慶芳, "國內外財務困境研究綜述," *特區經濟*, Vol.09, pp.265-267, 2007.
 [6] Altman, "Financial Ratios Discriminant Analysis And The Prediction of Corporate Bankruptcy," *Journal of Finance*, Vol.23, No.4, pp.589-609, 1968.
 [7] T. Shumway, "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," *Journal of Business*, Vol.74, 2001.
 [8] C. Servaas Houben, "Variavle Selection and Corporate Bankruptcy Forecasts," *Cfa*

Digest, Vol.45, No.8, pp.63-73, 2015.
 [9] 王慶, 姚康, "機器學習方法在中小企業信用評估中的應用研究," *特區經濟*, Vol.1, pp.145-147, 2018.
 [10] William H. Beaver, "Alternative Accounting Measures as Predictors of Failure," *The Accounting Review*, Vol.50, 1968.
 [11] 徐永智, 秦培培, "基於Logistic回歸的中國創業 板上市公司信用風險度量研究," *經濟論壇*, Vol.2, pp.75-79, 2013.
 [12] 譚春波, 劉誌強, 徐偉, "現代企業利用機器學習 技術的方法探索," *計算機產品與流通*, Vol.12, pp.121-125, 2017.
 [13] 姜國華, 官小燕, "非經常性損益對ST公司的淨利潤影響研究," *商*, Vol.15, pp.37-38, 2014.
 [14] J. A. Ohlson, "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accouting Research*, Vol.18, No.1, pp.109-131, 1980.
 [15] 楊貴軍, 周亞夢, 孫玲莉, "基於Logistic模型的企業財務風險預警方法," *數量經濟技術經濟研究*, Vol.10, No.36, pp.149-165, 2019.
 [16] 김형준, 류두진, 조훈, "기업부도예측과 기계 학습," *금융工學研究*, 제18권, 제3호, pp.131-152, 2019.
 [17] 洪欣琪, 阮素梅, "基於SVM和Logistic算濃對比 的中小企業財務困境預測研究," *齊齊哈爾大學學報*, Vol.37, No.3, pp.80-88, 2021.
 [18] N. Mselmi, A. Lahiani, and T. Hamza, "Financial Distress Prediction: The Case of French Small and Medium-Sized Firms," *International Review of Financial Analysis*, Vol.50, pp.67-80, 2017.
 [19] 張露, 劉家鵬, 江敏祺, "集成機器學習模型在不平衡樣本財務預警中的應用," *電子技術應用*, Vol.47, No.8, pp.34-38, 2021.
 [20] 陳煒宏, "基於決策樹模型的構建方法及預測能力的分析 --以釣魚網站為例," *通訊世界*, Vol.25, No.12, pp.233-235, 2018.
 [21] 武彩霞, "機器學習在通信企業用戶流失預警中的應用," *中國新通信*, Vol.22, No.19, pp.9-10, 2020.
 [22] 李林傑, "機器學習在財務欺詐風險預警問題中的應用研究 --基於製造業上市公司財務資料的實證分析," *今日財富*, Vol.19, pp.190-192, 2021.
 [23] K. Ristolainen, "Predicting Banking Crises

with Artificial Neural Networks: The Role of Nonlinearity and Heterogeneity,” The Scandinavian Journal of Economics, Vol.120, No.1, pp.31-62, 2018.

- [24] C. Tsai and J. Wu, “Using neural network ensembles for bankruptcy prediction and credit scoring,” Expert Systems With Applications, Vol.34, No.4, pp.2639-2649, 2008.
- [25] 張培榮, “基于XGBoost模型的企業財務危機預警研究,” 財會通訊, Vol.35, pp.109-112, 2019.
- [26] 喬源, 陳夢帆, “基於多種機器學習算濃的員工離職預測模型對比及解釋研究,” 商訊, Vol.27, pp.189-191, 2021.
- [27] 李艷紅, “基於機器學習的企業產品評論數據的情感分析研究,” 微型電腦應用, Vol.35, No.11, pp.35-38, 2019.
- [28] 楊國慶, 袁椿昊, 陳康耀, 王豔芳, “基於資料分析與機器學習的銀行貸款策略研究,” 高師理科學刊, Vol.41, No.8, pp.27-31, 2021.
- [29] 蔣盛益, 汪珊, 蔡余冲, “基於機器學習的上市公司財務預警模型構建,” 統計與決策, No.11, pp.166-167, 2010.
- [30] 전해정, “EWS 기법을 사용한 중국주택시장위기에 대한 영향요인 분석: 정책영향을 위주로,” 중국지역연구, 제11권, 제2호, pp.11-109, 2014.
- [31] 김진영, “기업가치향성이 중소기업의 기업가적 성과에 미치는 영향,” 벤처창업연구, 제14권, 제2호, pp.83-93, 2019.

왕 쿤(Kun Wang)

정회원



- 2015년 6월 : 중국화북수리수전대학교 경영학과(경영학석사)
- 2020년 9월 ~ 현재 : 세종대학교 경제학과(박사과정)

<관심분야> : 산업경제, 응용미시

란 희(Luan Xi)

정회원



- 2019년 9월 : 세종대학교 경제학과(경제학석사)
- 2021년 9월 ~ 현재 : 세종대학교 경제학과(박사과정)

<관심분야> : 관광산업, 응용미시

배 기 형(Ki-Hyung Bae)

정회원



- 1984년 2월 : 세종대학교 경제학과(경제학석사)
- 1992년 2월 : 세종대학교 경제학과(경제학박사)
- 1998년 3월 ~ 현재 : 세종대학교 경제학과 교수

<관심분야> : 문화산업, 응용미시

저 자 소 개

원 도(Tao Yuan)

정회원



- 2020년 9월 ~ 현재 : 세종대학교 경제학과(박사과정)

<관심분야> : 기계학습, 실증연구