

시계열 예측을 위한 LSTM 기반 딥러닝: 기업 신용평점 예측 사례

이현상* · 오세환**

〈목 차〉

I. 서 론	IV. 연구 결과
II. 문헌 연구	4.1 예측 모델 평가 및 선정
2.1 기업 신용등급 예측	4.2 5등급 분류 예측 및 신용 불량기업 예측
2.2 기업 부도 예측	4.3 예측 시점 확장
2.3 기존 연구와의 차별성	4.4 특성 그룹 평가
III. 연구 방법론	V. 결 론
3.1 데이터	참고문헌
3.2 분석	<Abstract>

I. 서 론

금융기관들과 투자 회사에게 다양한 투자자산의 신용 위험을 평가하는 것은 매우 중요한 일이다. 1997년 외환위기와 2007년의 서브프라임 모기지 사태(Subprime Mortgage Crisis)와 같은 금융위기가 발생하면서 기업부실예측 분야와 신용위험관리에 대한 중요성이 강조되고 있다(최진배, 2002). 최근 미중 무역 갈등, 홍콩 시위 사태 등 글로벌 정치경제 환경의 불확실성이 높아지고 있는 가운데 국내 경기침체 및 일본과의 무역마찰 등으로 글로벌 신용평가회

사인 무디스는 2020년 한국 기업들의 신용위험 수준이 높다고 평가했다(김진성, 2019). 갈수록 심화되고 있는 세계경제의 불확실성 확대와 더불어 국내 및 글로벌 기업들의 신용위험관리 중요성은 더욱 강조되고 있는 상황이다. 경영정보 분야의 연구에서도 이와 같은 이유로 머신러닝 기법을 활용한 기업 신용평가 시스템에 대한 연구들이 수행되고 있다(이영찬, 2005; 노태협 등, 2005, 홍태호와 신택수, 2007).

신용위험관리 분야 중 하나인 회사채 신용등급 예측모형에 대한 연구가 국내에서 활발하게 이루어지고 있는 가운데(김태정 등., 2003; 나

* 경북대학교 경영학부, coolwin20@knu.ac.kr(주저자)

** 경북대학교 경영학부, sehwan@knu.ac.kr(교신저자)

영과 진동민, 2003; 박형권 등, 2018), 최근에는 지능형 정보 시스템의 급부상으로 머신러닝(machine learning) 기법을 활용한 신용등급 예측모델 연구들이 등장하고 있다(안현철, 2014; 김성진과 안현철, 2016; 최병설과 김남규, 2019; 이륜경 등, 2019). 그러나 지금까지의 기업 신용위험 예측 모델 연구는 시계열적인 입력 특성을 활용하지 못하고 예측 시점이 한정적이라는 한계점이 있다(김성진과 안현철, 2016; Kim and Ahn, 2012; Chaudhuri and De, 2011; Guo et al., 2012; Barboza et al, 2017). 또한, 회사채 신용등급은 비교적 규모가 큰 기업들을 대상으로 부여되는 것이기 때문에 기업 선정에 있어서도 한정적이다.

이러한 이유로 신용위험에 대한 요인분석을 할 때 KIS 신용평점(기업 신용평점)을 활용하기도 한다(조현우 등, 2005; 전성일과 이기세, 2010). KIS 신용평점이란 한국신용평가에서 개발한 지표로써 중공업, 경공업, 건설업, 도소매업, 기타서비스업 등 5개 산업을 구분하여 부실 예측모형과 재무평점 모형을 결합하여 산출한다. 부실예측모형이란 추정 부도율, 역사적 부도율로 명명되기도 하며 무디스 등급 정의에 따라 포괄적으로 고려하여 계산한다. 재무평점 모형이란 기업 부도에 영향을 주는 재무변수들을 활용하여 안정성, 수익성, 현금흐름, 성장성, 활동성의 가중치를 고려하여 평점을 합산하여 등급화 한다(KISVALUE, 2019). KIS 신용평점은 1~10 등급으로 구분되는데 1등급에 가까울수록 기업의 신용 수준이 높다고 평가한다.

회사채 신용등급과 달리 기업 신용평점은 단순한 표기법을 사용하여 손쉽게 해석이 가능하고 적용되는 기업이 광범위하기 때문에 신용위

험관리 연구에 있어서 척도로써 유용하게 사용되고 있다(조현우 등, 2005; 전성일과 이기세, 2010). 그러나 국내에서 회사채 신용등급 예측 모델 연구는 활발하게 진행 중인 반면 신용평점 예측모델 연구는 현재까지 제한적인 상황이다. 또한, 기존 기업 신용위험 예측 모델들은 시계열적인 데이터를 활용하지 못했으며 이로 인해 예측 시점을 조절하지 못한다는 한계도 있는데(김성진과 안현철, 2016; Kim and Ahn, 2012; Chaudhuri and De, 2011; Guo et al., 2012; Barboza et al, 2017) 이에 따라 단기적 또는 장기적 관점에서 기업 신용위험 예측에 영향을 미치는 중요 지표들을 파악하기 어렵다는 문제점도 있다. 다년간 데이터를 활용하여 시계열적으로 예측할 수 있다면 기업의 신용위험을 단기적 및 장기적 관점에서 대비할 수 있으며 중요 지표의 분석을 통해 신용위험의 신호를 포착할 수 있다.

선행연구의 한계를 극복하기 위해 본 논문은 18년 간 코스피(KOSPI), 코스닥(KOSDAQ), 코넥스(KONEX)에 등록된 2,191개 기업 데이터에 머신러닝 기법을 적용하여 기업의 신용평점을 시점의 제한 없이 예측할 수 있는 모델 개발을 시도한다. 세부적으로는 먼저 다수 기업의 연도별 데이터를 재조정하여 하나의 시퀀스를 가지는 형태로 구현했다. 이를 슬라이딩 윈도우 기법을 활용하여 각 기업에 해당하는 3년까지의 기업 데이터를 통해 그 다음 해 기업 신용평점을 예측하는 학습 모델을 구현하고, 여러 가지 머신러닝 분류 예측과 시계열 예측 분석 기법을 적용하여 각각의 성능을 평가했다. 다음으로 재조정된 데이터로 슬라이딩 윈도우 기법 과정에서 목표 예측 값을 다음 연도의 신용평

점으로 할당하는 방식을 적용하여 예측 시점을 7년 후 까지 확장했다. 마지막으로 단기적 및 장기적 예측 시점에서의 중요 지표를 파악하기 위해 모델의 특성들을 그룹화하고 1년 후 예측 모델과 7년 후 예측 모델의 특성 그룹 평가를 진행했다.

본 연구는 KIS 신용평점을 활용하여 예측 시점을 조절할 수 있는 예측 모델을 구현하고 대상 기업을 확장함으로써 기존 연구의 시점적, 범위적 한계점을 극복했다. 시점적인 측면에서 투자자들과 연구자들은 장기적 및 단기적 관점에서 중요 지표를 파악할 수 있고 범위적인 측면에서 일반화된 기업 신용위험 예측 모델을 구현할 수 있다. 결론적으로 본 논문은 다수 개체에 대한 시계열 예측을 구현하는 연구 방안을 제시하고, 기존 연구에서 잘 사용되지 않았던 딥러닝 분석 방식을 적용하여 기존의 예측 모델보다 성능이 우수한 예측 모델을 구현했다는 점에서 공헌점이 있다.

II. 문헌 연구

2.1 회사채 신용등급 예측

기업 신용위험 예측에 대한 연구는 크게 회사채 신용등급 예측과 기업의 부도 예측연구가 있다(박형권, 2018). 이러한 연구들은 다양한 통계모형을 기반으로 진행되어 왔으며 최근에는 신용위험 예측모델에 머신러닝 분류 분석을 활용하는 다양한 연구들이 시도되고 있다. 머신러닝 분류(machine learning classification)란 지도학습(supervised learning) 기반으로 특성과

라벨(label) 값으로 구성된 분할 데이터를 학습하여 분류 모델을 만드는 것이다(Kotsiantis et al., 2007). 이 때 모델의 성능은 앞서 분할한 데이터 중 테스트 데이터를 활용하여 정확도, 재현율, 정밀도, f1 점수(f1 score) 등으로 검증한다. 딥러닝(deep learning) 분석은 머신러닝 기술에 포함되는 개념이다. 딥러닝이란 인공신경망(neural network) 기술을 기반으로 은닉층(hidden layer)을 2개 이상 사용하는 분석 기법으로 종류가 다양하고 그만큼 현재 다양한 연구분야에서 사용되고 있다(Glorot et al., 2011; Chen et al., 2014; Lecun et al., 2015).

회사채 신용등급 예측의 경우 1999년 유전자 알고리즘(genetic algorithm, 이하 GA) 기반 사례 기반 추론(case-based reasoning, 이하 CBR) 분석을 통해 예측모델을 개발한 연구가 등장했다(Shin and Han, 1999). Shin and Han.(1999)은 다양한 분석 기법을 예측모델에 적용한 결과 GA-CBR 모형이 75.5%의 정확도로 가장 우수한 성능을 보인다고 주장했는데 이 연구를 필두로 회사채 신용등급을 예측하기 위해서 다양한 머신러닝 분류 기법을 적용하는 연구들이 등장하기 시작했다.

회사채 신용등급 예측에서 서포트 벡터 머신(support vector machine, 이하 SVM)과 인공신경망(artificial neural network, 이하 ANN) 분석의 성능을 비교한 연구에 의하면 SVM 분석 기법의 정확도가 대체적으로 인공신경망 분석보다 성능이 우수한 것으로 나타난다(Huang et al., 2004). 한편, 인공신경망 모형과 선형 판별 분석(linear discriminant analysis, 이하 LDA)을 비교 분석한 연구에서는 결과적으로 인공신경망 예측 모델의 성능이 종합적으로 우

수한 것으로 나타났다(Kumar and Bhattacharya, 2006). Kumar and Bhattacharya.(2006)는 인공 신경망 예측 모델이 복잡한 데이터 및 비선형 분석에 있어서 선형 판별 분석에 비교적 성능이 우수한 양상을 나타낸다고 주장했다. 기업 신용등급 예측을 위한 다중 클래스 분류에 관한 연구는 기존 연구의 계량 분석법의 한계를 지적하고 다양한 머신러닝 알고리즘을 멀티 클래스 분류로써 활용한다(Ye et al., 2008). Ye et al.(2008)은 기업 신용등급 다중 클래스 분류에서 PSVM(proximal SVM)이 예측 정확도 84%로 연산 시간을 줄이고 성능이 우수한 것으로 나타났다. 차원 축소 기반 통합 SVM 알고리즘을 통한 기업 신용평점 예측 연구는 입력 데이터의 차원을 줄이고 KGE(kernel graph embedding) 방식으로 SVM 분류의 성능을 향상시켰다(Huang, 2009). 결과적으로 다른 멀티 클래스 SVM이나 기존 분류 방식에 비해서 KGE를 활용한 일대일 SVM이 약 오차율 10% 내외로 가장 높은 성능을 나타냈다. SVM에서 OPP(ordinal pairwise partitioning) 전략으로 발전시킨 OMSVM (ordinal multi-class SVM) 기법을 사용하여 기존 분석기법과 비교 분석한 연구에 따르면 OMSVM 분석 방식이 다른 MDA, MLOGIT, CBR 및 ANN 방식보다 예측 정확도 67.98%로 우수한 성능을 보인다(Kim and Ahn, 2012).

다른 예측 모델 연구들과 마찬가지로 회사채 신용등급 예측 모델 연구에서도 2010년부터 하이브리드 방식이 적용된 머신러닝 기술을 활용하는 예측 모델 구현 연구들이 수행되기 시작했다. 퍼지 군집화(fuzzy clustering)와 SVM을 활용한 신용 예측 모형에 관한 연구는 다중 분

류를 위해 퍼지 군집화 알고리즘을 적용하고 SVM 분류 기법을 적용하는 방안을 제시했다(Guo et al., 2012). 퍼지 군집화 알고리즘을 활용한 데이터 전처리를 통해서 연산 과정을 줄이고 성능을 향상시키기 위해 경계 데이터 포인트(boundary data point)만을 트레이닝 샘플로 사용했다. 결과적으로 제안된 모델의 성능이 약 73%로 연산 시간을 줄이고 성능을 향상시키는 것을 확인했다. 기업 신용등급 2단계 예측에 관한 연구의 경우 다중 클래스 분류에서 2단계 분석 방식을 제안한다. 1단계에서는 특성 선정, 데이터 군집화 및 리샘플링을 통해 데이터를 전처리하고 2단계에서는 여러 가지 분류 기법 및 앙상블을 적용하여 예측 모델을 구현한다. 결과적으로 데이터 리샘플링(data resampling)을 적용한 Bagging-DT 방식이 예측 정확도 82.96%로 가장 우수하게 측정되었다(Wu et al., 2014). 최근 랜덤 포레스트(random forest, 이하 RF) 분석 기법을 예측모델에 적용한 연구에서는 랜덤 포레스트가 다른 머신러닝 분석기법에 비해 과적합에 자유롭고, 데이터의 잡음이나 이상치를 효과적으로 처리한다고 주장한다(김성진과 안현철, 2016). 이 연구에서는 1,295개의 국내 상장 기업을 대상으로 최고 75%에 해당하는 예측모델을 랜덤 포레스트 기법을 통해 개발했다.

2개 클래스 예측을 하는 기업 부도 예측과는 다르게 신용등급 예측은 다중 클래스 분류 예측이기 때문에 학습 모델의 비교적 높은 성능을 요구한다. 기업 신용등급의 종류가 다양하고 분포가 고르지 않기 때문에 일반적으로 5개 내외의 클래스로 등급을 통합한다(박형권 등., 2018). 2000년 전후로 머신러닝 분류 예측 기술

이 발전하면서 인공 신경망 분석, 서포트 벡터 머신과 같은 분석 기법으로 회사채 신용등급을 예측하는 연구들이 등장했다. 예측 모델의 성능을 높이기 위해 다양한 지표와 예측 모델의 하이브리드 방식이 적용되었다. 2010년 이후에는 다양한 서포트 벡터 머신 기술과 함께 랜덤 포레스트 기술의 도입으로 예측 모델의 성능이 향상되었다. 회사채 신용등급 연구들의 활용 지표와 분석 기법 및 구체적인 성능을 정리하면 <표 1>과 같다.

2.2 기업부도 예측

기업 부도 예측은 1968년에 재무비율을 활용

하는 판별 분석을 통해서 기업 부도를 예측한 연구 이후로 활발하게 수행되고 있다(Altman, 1968). Altman.(1968)은 통계적인 다중판별 분석을 사용하여 66개 기업의 초기 표본을 대상으로 5년 후까지의 기업 부도를 1년 후 정확도(accuracy) 기준 95%에서 5년 후 36%까지 해당하는 예측 모델을 만들었다. 이후 국내외에서 꾸준히 기업 부도 예측 모델 연구가 진행되었다. 실패한 기업들의 데이터를 활용하여 기업 부도를 예측한 연구는 기존 연구의 단순 기업 부도 여부 예측 모델의 한계점을 지적하고 재무 비율을 활용하여 이를 해결하려 시도했다(Flagg et al., 1991). 제안된 기법은 로지스틱 회귀 분석과 샘플 선정, 그리고 관찰적 접근법

<표 1> 회사채 신용등급 예측 연구

논문	활용 지표	분석 기법	모델 성능
Shin and Han. (1999)	규모 지표, 수익성 지표, 주가 지표	MDA, ANN, CBR, GA-CBR	GA-CBR(acc) 75.5%
Huang et al. (2004)	규모 지표, 수익성 지표, 안정성 지표, 주가 지표	LR, SVM, ANN	SVM, ANN(acc) 약 80%
Kumar and Bhattacharya. (2006)	규모 지표, 수익성 지표, 안정성 지표, 주가 지표	ANN, LDA	ANN(acc) 약 78%
Ye et al. (2008)	규모 지표, 수익성 지표 등	LR, Probit, BDT, SVM, PSVM	SVM, PSVM(acc) 약 84%(2notch)
Huang. (2009)	규모 지표, 수익성 지표 등	NN, LR, BN, SVM, MSVM, PCA, ICA, KGE	1vs1 KGE SVM(error rate) 약 10%
Guo et al. (2012)	규모 지표, 수익성 지표, 안정성 지표, 주가지표	ANN, SVM, SVDD, SVM with FCA	SVM with FCA(acc) 72.12%
Kim and Ahn. (2012)	규모 지표, 수익성 지표, 안정성 지표, 주가지표	MDA, MLOGIT, CBR, ANN, SVM, WW, CS, OMSVM	OMSVM(acc) 67.98%
Wu et al. (2014)	규모 지표, 수익성 지표, 안정성 지표, 주가 지표, 지배구조 지표, 시장 지표	DT, BN, ANN, SVM	Bagging DT(acc) 82.96%
Kim and Ahn. (2016)	규모 지표, 수익성 지표, 안정성 지표, 현금흐름 지표, 생산성 지표	LDA, ANN, MSVM, RF	RF(acc) 72.79%

을 활용했는데 결과적으로 기존 연구의 기업 부도 예측 모델에 비교적 우수한 예측 모델을 구현할 수 있었다.

2000년 전후로 인공지능망을 비롯한 머신러닝 기술이 연구에 적용되기 시작하면서 확률적 인공지능망 분류 기술을 활용한 연구가 등장하기 시작했다(Yang et al., 1999). Yang et al. (1999)의 연구에서 확률론적 신경망과 퍼지 판별 분석 모델의 성능이 각각 74%, 71%로 우수한 것으로 나타났으며 판별분석의 경우 하나의 회사에 해당하는 예측에 우수한 성능을 보였다. Duffie et al.(2007)은 확률적 공변량(stochastic covariates)을 활용한 다년간 기업 부도 예측 연구에서 1980년부터 2004년까지 총 39만개 이상의 월별 데이터를 활용하여 기업의 채무 불이행을 예측했다. 제안된 기법은 통해 기업 고유의 공변량과 거시경제 공변량을 통합하여 기업 부도 조건부 확률의 기간에 따른 최대우도 추정치를 제공했는데 결과적으로 기업 부도 확률은 기업의 변동성 조정 레버리지 척도(a volatility-adjusted measure of leverage), 주식 수익률, S&P 500 수익률 및 미국의 금리에 영향을 받는 것이 나타났다. 이 예측 모델은 다른 분석 기법에 비해 비교적 성능이 개선된 것이다.

기업 부도 예측모델에서는 SVM을 다른 분석기법과 결합하여 하이브리드 방식으로 사용하기도 했는데, DEA(Data Envelopment Analysis) 와 RST(rough set theory)-SVM을 하이브리드 방식으로 활용한 연구의 경우 RST-BPNN(back propagation neural network)과 RST-SVM을 비교했을 때 82% 대 85%로 후자의 성능이 더 낫다는 것을 입증했다(Yeh et

al., 2010). 기업 부도 예측을 위해 퍼지 알고리즘 기반 SVM 예측 모델 구현 연구는 퍼지 집합을 통해 데이터의 불확실성을 통제하고 SVM을 통해 기업 부도를 예측했다(Chaudhuri and De, 2011). 제안된 기법은 입력 데이터의 퍼지 집합 알고리즘을 통해 예측 정확도를 향상시켰다. FSVM(fuzzy SVM)과 다른 분석 기법을 비교한 결과 FSVM은 최적의 특성 집합과 매개변수를 발견하는데 효과적이다. 역전파 기반 신경망 분석과 비교하여 FSVM의 성능이 우수하게 나타났다.

최근에는 모델의 성능을 높이고자 유전자 알고리즘과 퍼지 군집화의 하이브리드 방식을 예측모델에 적용한 연구와 같이 다양한 머신러닝 분석 기법들을 활용하여 하이브리드 방식으로 적용하는 연구들이 수행 중이다(Chou et al., 2017). 이 연구에 따르면 BPNN과 유전자 알고리즘 및 퍼지 군집화의 하이브리드 방식의 예측 성능을 비교했을 때 후자의 방식이 우수한 성능을 보였다고 주장한다. 기업 부도 예측을 위한 머신 러닝 예측 모델 연구는 기업 부도 사건 발생 1년 전에 시스템 학습 모델을 시험하여 평가했다(Barboza et al., 2017). 사용된 예측 모델은 총 4가지로 SVM, 배깅(bagging), 부스팅(boosting), 랜덤 포레스트에 해당한다. 1985년부터 2013년까지 북미 기업 데이터를 활용했으며 Salomon Center와 Compustat의 데이터베이스를 통합해서 약 10,000개 이상의 데이터를 적용했다. 기존 예측 모델 지표에서 새롭게 영업이익, 전일 대비 수익률 변화, 장부상 가격 변동, 자산, 매출, 직원 수 등을 활용했다. 결과적으로 로지스틱 회귀 분석과 선형 판별 모형에 비해 랜덤포레스트 분석 기법의 정확도가 87%

으로 우수한 성능을 나타냈다.

기업 부도 예측은 2개 클래스 분류 문제이기 때문에 통계적 기법에 의한 연구들이 신용등급 예측 연구에 비교적 먼저 등장했다. 회사채 신용등급 예측과 같이 2000년 전후로 머신러닝 기술이 도입되면서 분석 대상 기업이 확장되고 모델의 성능이 증가하는 양상을 보였다. 최근에는 SVM 뿐만 아니라 랜덤 포레스트 기법이 우수한 성능을 보이는 것으로 나타났다.

2.3 기존 연구와의 차별성

기존 기업 신용위험 예측 연구에서는 위와 같이 전반적으로 시간적인 요인을 고려하기 보다는 새로운 변수의 도입이나 예측 모델 방법론의 제시를 통해 예측 모델의 성능을 높이는 방향성을 가지고 있다(김성진과 안현철, 2016; Kim and Ahn, 2012; Chaudhuri and De, 2011; Guo et al., 2012; Barboza et al, 2017). 이러한 연구들과 달리 본 연구는 KIS 신용평점의 다년간 데이터를 활용하여 시계열적으로 기업의 신용위험을 예측하는 모델을 구현하고 시점에 따른 기업의 중요 지표를 식별하고자 한다. KIS 신용평점은 KISVALUE에서 제공하는 복합 요인을 고려한 기업 신용의 척도로 대상 기업의 범위가 넓고 등급 해석에 용이하여 기업 신용에 대한 요인 분석으로써 활용되기도 한다(조현우 등, 2005; 전성일과 이기세, 2010). 조현우 등(2005)은 사외이사의 특성과 감사 품질이 KIS 신용평점에 어떻게 영향을 미치는지 연구했다. 전성일과 이기세(2010)는 보수주의 회계처리가 KIS 신용평점에 어떤 영향을 미치는지 연구했다.

본 논문은 기업 투자자들과 관련 연구자들에게 새로운 기업 신용위험 예측 측정 지표와 단기적 및 장기적 중요 지표 분석 방안을 제안한다. 여러 기업들을 대상으로 시계열 예측 모델을 구현함으로써 다수 개체에 대한 시계열 예측 분석이 가능한 연구 프레임워크를 제시한다. 본 연구는 추가적으로 기존 연구들에서 일반적으로 사용되는 SVM과 랜덤 포레스트 기법을 신경망 분석과 LSTM과 같은 딥러닝 기법과 같이 적용하여 비교 분석한다.

Ⅲ. 연구방법론

3.1 데이터

본 연구에서는 기업의 재무적, 비재무적 데이터를 활용하기 위해 KISVALUE에서 KOSPI, KOSDAQ, KONEX에 속하는 비금융업 2,362개 기업을 대상으로 최근 18년 간의 데이터를 수집했다(KISVALUE, 2019). KIS 신용평점의 경우 2000년도부터 평가되기 시작했기 때문에 2000년부터 2017년까지의 데이터를 수집했으며, Python을 사용하여 전처리를 통해 2,191개의 기업 데이터를 예측모델에 적용할 수 있도록 했다. 2017년 KIS 신용평점이 출력되지 않은 기업들은 제외했으며 나머지 결측치 값(missing value)은 0으로 처리했다. 특성 선정(feature selection)은 KISVALUE 내의 기업 데이터 범주에 해당하는 개요 정보, 현황 정보, 신용 정보, 투자 지표를 활용했다. 투자 지표는 KISVALUE의 Financial 항목들을 참고하여 선정했으며 특성 그룹 분류는 <표 1>의 선행연구

에서 활용된 지표들을 참고했다. 모든 수치 데이터는 예측모델에 편향없이 적용시키기 위해 0~1 값으로 변환했으며 명목 데이터의 경우 더

미(dummy) 형태로 치환하여 사용했다. 자세한 특성 목록은 <표 2>와 같다.

<표 2> 특성 선정

특성 목록	특성 그룹 구분	데이터 형태		
총자산	규모 지표 (Scale Index)	수치 데이터		
무형자산				
유형자산감가상각비				
무형자산상각비				
유형고정자산증가율				
유동자산증가율				
재고자산증가율				
종업원 수				
시장구분				
기업 형태				
업력	수익성 지표 (Profitability Index)	수치 데이터		
총자산순이익률(ROA)				
총자본회전율				
부채대매출액비율				
순영업자본대매출				
매출채권회전율				
판매비와 관리비				
법인세비용차감전계속사업이익				
영업활동으로인한 현금흐름				
수출비중			무역 지표 (Trade Index)	수치 데이터
외화자산				
외화부채				
관계회사상호거래비용	관계회사 지표 (Affiliated Company Index)	수치 데이터		
관계회사상호거래수익				
관계회사상호거래매출				
관계회사상호거래매입				
단기차입금	안정성 지표 (Stability Index)		수치 데이터	
사채				
장기차입금				
회사채유효등급	신용 지표 (Credit Index)			명목 데이터
이전 KIS 신용평점				
CP유효등급				
감사의견				
현황 평가	텍스트 감성 지표 (Text Sentiment Index)	텍스트 데이터		
전망 평가				

본 연구에 사용된 텍스트 데이터는 KISVALUE의 투자 지표에 있는 현황 및 전망에 해당하는 기업 평가 항목을 텍스트 감성 분석(text sentiment analysis)을 통해 점수를 매긴 것이다. 텍스트 감성 분석이란 텍스트의 정보를 추출하기 위해 긍정, 부정 신호를 분석하는 것을 의미한다(Aggarwal and Zhai, 2012). 감성 분석은 텍스트 마이닝(text mining) 기법으로 다양한 분야의 예측 모델에서 사용되고 있다(Liu et al., 2007). KIS 신용평가 관련 텍스트 데이터는 라벨 값을 각각 선정하는 것이 데이터 특성상 불가능하다. 따라서 이를 구현하기 위해서 Python의 Tensorflow, Keras 패키지를 활용하여 심층 신경망 분석을 통해 영화 리뷰 데이터를 학습하여 텍스트 데이터에 적용한 다음 감성 점수를 추출했다.

수치 데이터, 명목 데이터, 텍스트 데이터를 결합하고 시계열적인 기업 신용평점 예측을 위해 데이터의 형태를 재조정(reshape)했다. 하나의 기업에 해당하는 1년의 데이터를 데이터 프레임의 행으로 지정하여 모든 기업의 모든 연도를 대상으로 분석을 진행했다. 시계열적으로 데이터를 인식하기 위해 슬라이딩 윈도우 기법(sliding window technique)을 활용했는데 슬라이딩 윈도우 기법이란 데이터를 순차적으로 윈도우 사이즈에 맞춰서 이동시키는 형태로 만드는 것이다(Koç, 1995). 이를 통해 3개의 연도에 해당하는 입력 데이터에 다음 해의 기업 신용평점을 라벨 값으로 할당하는 형태의 데이터셋을 만들었다. 슬라이딩 윈도우 기법은 시계열 예측에 있어서 시간 복잡성 문제를 상당량 해결해주기 때문에 일반적으로 사용된다(Frank et

al., 2001). 재조정된 데이터 셋과 슬라이딩 윈도우 기법을 통해 생성된 데이터 셋에서 머신러닝 분류 모델을 학습하고 평가하기 위해 임의로 기업 데이터를 7:3으로 구분했다. 분리된 데이터 셋을 가지고 모델 선정에 해당하는 본격적인 분석을 진행했다.

3.2 분석

본 연구는 최근 예측모델 분야의 분석 경향에 따라 전통적인 머신러닝 기법 및 최근의 딥러닝 분류 분석 기법을 적용했다. 예측모델을 개발하기 위한 학습 방법으로 선행연구에서 주로 쓰였던 머신러닝 분석 기법 3개에 딥러닝 분석 기법 6개를 추가하여 총 9개의 예측모델을 구현하였으며 이중 가장 높은 성능을 갖는 모델을 선정했다(Huang et al., 2004; 김성진과 안현철, 2016). 본 연구에 사용된 분석기법을 정리하면 <표 3>과 같다. 모든 분석은 Python의 Tensorflow, Keras, Scikit-Learn 패키지를 활용했고 모델 선정(model selection) 전에는 기본적인 튜닝 설정만을 사용했다.

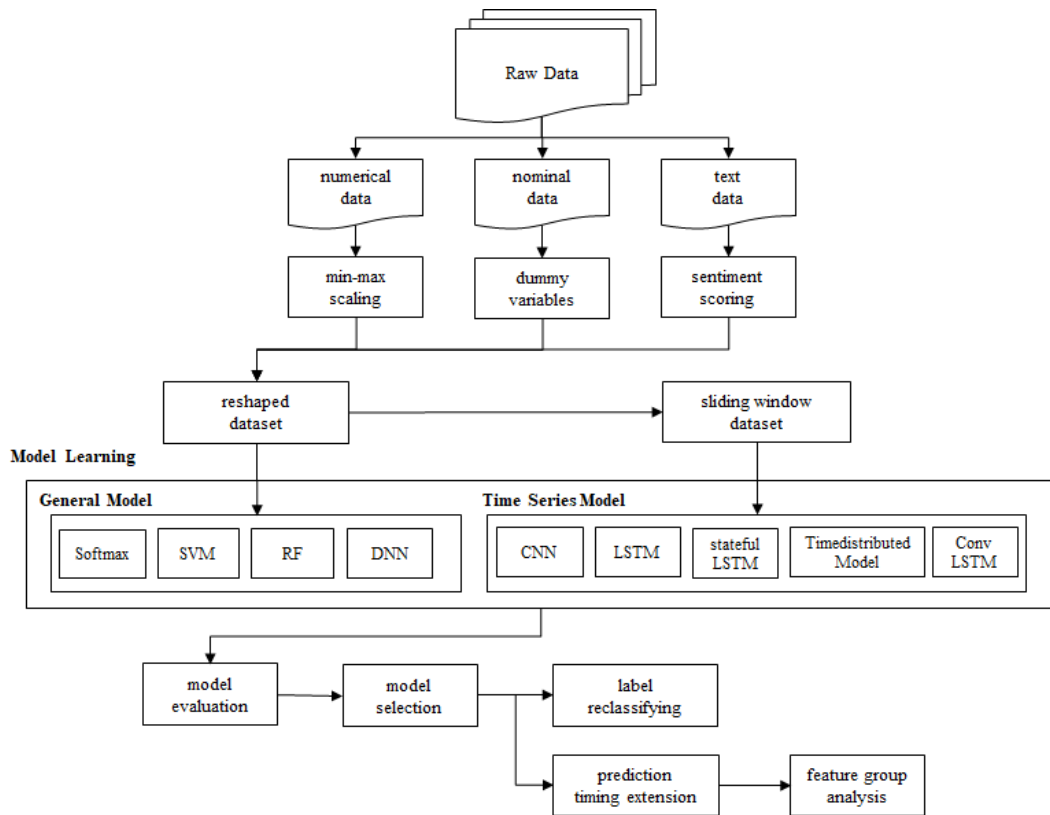
9개의 머신러닝 분류 모델의 성능을 평가하여 가장 예측 정확도가 우수한 예측 모델을 선정했다. 예측 모델 선정 후 예측 시점 조절을 위해서 시계열 분석을 통해 7년 후 까지 예측 시점을 확장했다. 7년 후 까지 예측 시점을 확장한 이유는 예측 시점을 확장하면서 5년 후 예측 모델부터는 시계열적 요인이 대부분 사라지는 것을 관찰했으며 7년 후까지 시점을 확장하여 성능 변화가 거의 존재하지 않는 것을 확인했기 때문이다.

<표 3> 분석 기법

모델 분류	분석 기법	개요
일반 예측 모델	Softmax	목표 라벨 값에 대해서 각 클래스에 대한 확률을 제공
	Support Vector Machine (SVM)	분류기점에서 서포트 벡터(support vector)와 가장 먼 거리를 가지는 초평면(hyperplane)을 찾는 기법
	Random Forest (RF)	다수의 의사결정 트리(decision tree)를 생성하여 분류 또는 평균 예측치를 출력하는 방식
	Deep Feed-forward Neural Network (DNN)	신경망 분석에서 은닉층을 2개 이상 사용하는 전방향(feed-forward) 네트워크 분석 기법
시계열 예측 모델	Convolutional Neural Network (CNN)	1차원적 입력 배열의 한계를 극복하기 위해 이미지의 공간정보를 유지한 채로 학습이 가능한 합성곱 계층(convolutional layer) 기반 신경망 학습
	Long Short Term Memory(LSTM)	장기적 연속성을 가지는 데이터에 대한 경사소실(vanishing gradient) 문제와 연산 지연문제를 곱셈게이트 유닛(multiplicative gate unit) 기술을 통해 다수해결
	stateful Long Short Term Memory (stateful LSTM)	각 배치의 학습이 진행될 때 은닉상태 벡터(hidden state vector)와 셀 상태 벡터(cell state vector)를 초기화 하는 기법
	Timedistributed Model	CNN 모델을 통해 입력데이터의 1차원 서브시퀀스(subsequence)를 해석하여 LSTM의 시퀀스로써 활용
	ConvLSTM	서브시퀀스를 2차원의 이미지로 간주하고 해석

예측 모델 연구에서는 성격별 특성의 영향을 파악하기 위해 특성 그룹이 모델 성능에 미치는 영향을 평가하기도 한다(Li et al., 2016). 이에 따라 1년 후 시점의 예측 모델과 7년 후 시점의 예측 모델을 적용하여 단기적 관점과 장기적 관점에서 특성 그룹이 어떻게 모델의 성능에 영향을 미치는 지 파악했다. 선행 연구의 신용 등급 예측 모델 연구들은 각 등급에 해당하는 표본 수의 부족으로 크게는 5등급 분류, 세부적으로는 20등급으로 분류 예측 모델을 구

현했다(박형권 등, 2018). 기업부도 예측 관련 연구는 기업의 부실 여부를 더미 변수로써 두 가지 클래스를 예측하는 연구 모형을 일반적으로 구현했다(Altman, 1968; Yang et al., 1999; 김영태와 김명환, 2001). 기존 연구들과의 비교를 위해 KIS 신용평점의 등급체계에 따라 다시 최상급, 상급, 중급, 하급, 불량으로, 그리고 불량인 기업들과 불량이 아닌 기업들을 분류하는 방식으로 예측 모델을 재구성했다. 전반적인 연구 분석 절차는 <그림 1>과 같다.



<그림 1> 분석 절차

IV. 연구 결과

4.1 예측 모델 평가 및 선정

분석결과는 <표 4>와 같다. 머신러닝 분석 기법 9개를 적용한 결과 10개의 신용 평점 예측에 대해서 stateful LSTM의 예측 정확도가 39.02%로 가장 높게 나타났다. 정밀도(precision), 재현율(recall), f1 점수(f1 score)도 각각 39%, 39%, 38%로 다른 분석 기법들에 비해 우수하게 측정되었다. 가장 우수한 성능이 측정된 stateful LSTM을 사용하여 추후 분석을 진행했다.

softmax 기법의 경우 정밀도는 높게 측정되었지만 f1 점수가 낮은 것을 고려했을 때 전반적으로 개체 분류에 문제가 있다고 볼 수 있다. SVM은 전반적으로 낮은 성능 지표를 나타냈고 RF의 경우 예측 정확도는 높지만 정밀도와 f1 점수가 크게 낮은 것으로 나타나 실제 분류한 양성 값에 심각한 오류가 있는 것으로 나타났다. DNN은 예측 정확도가 비교적 높게 측정되었지만 정밀도, 재현율, f1 점수 모두 낮은 것으로 봤을 때 개체 수가 많은 목표 라벨값으로 분류가 편중되었다.

CNN 기법은 트레이닝 과정에서 제대로 모델이 패턴을 학습하지 못하는 모습을 보였고

<표 4> 모델 평가 및 선정

학습 모델	accuracy	precision	recall	f1 score
softmax	0.2168	0.25	0.22	0.14
SVM	0.2104	0.08	0.21	0.11
RF	0.2281	0.09	0.23	0.12
DNN	0.2228	0.13	0.13	0.06
CNN	0.2036	0.21	0.14	0.08
LSTM	0.3271	0.31	0.31	0.3
stateful LSTM	0.3902	0.39	0.39	0.38
Timedistributed Model	0.2038	0.21	0.13	0.10
ConvLSTM	0.2021	0.22	0.15	0.11

예측 정확도, 정밀도, 재현율, f1 점수 모든 면에서 성능이 떨어지는 현상을 발견했다. CNN 모형이 트레이닝 데이터를 제대로 학습하지 못하는 이유는 이미지 분석을 목적으로 개발된 CNN 기법이 기업 신용평점 예측이라는 불완전하고 잡음이 많을 수 있는 데이터 분석에서 과적합 현상이 발생하는 것이라고 해석된다. RNN 기반 LSTM 기법은 예측 정확도, 정밀도, 재현율, f1 점수 모두 기타 분석들에 비교했을 때 상당히 우수한 성능을 나타냈다. 특히 stateful LSTM의 경우 예측 정확도가 약 39%에 육박하면서 같은 데이터 내에서 기존 연구에서 사용되는 다른 분석 기법에 비해 2배 가까운 성능이 측정되었다. 데이터 특성상 시계열적으로 입력된 다수의 기업 데이터에 대해서 하나의 시퀀스로 데이터를 해석하는 기본 LSTM 모델 보다 데이터의 배치 단위로 시퀀스를 기억하여 학습하는 stateful LSTM 기법의 성능이 우수하게 나타났다고 판단된다. 시간 분포 모형(time distributed model: CNN + RNN)과 ConvLSTM 기법의 경우 CNN이 입력 데이터의 서브시퀀스를 제대로 학습하지 못하는 이유로 예측 모델의 성능이 낮게 측정된 것으로 예

상된다.

기존 신용위험 예측 모델 연구들은 회사채 신용등급 예측의 경우 5개에서 20개, 기업 부도 예측에서는 2개의 클래스를 분류한다(김성진과 안현철, 2016; 박형권 등., 2018; Kim and Ahn, 2012; Chaudhuri and De, 2011; Guo et al., 2012; Barboza et al, 2017). 분류 문제에서는 클래스의 개수에 따라 예측 모델의 성능이 크게 차이가 날 수 있다(Kotsiantis et al., 2007). 따라서 기존 연구들과 직접적으로 비교하기 위해 예측 모델의 라벨을 재분류했다.

4.2 5등급 분류 예측 및 신용 불량 기업 예측

기존 연구에서 표본 데이터의 부족으로 적게는 5등급, 세부적으로는 20등급 까지 분류 예측을 하는 경우가 있다. 이와 비교하기 위해 <표 4>에서 가장 높은 예측 성능을 보이고 있는 stateful LSTM을 활용하여 라벨 재분류(reclassifying) 후의 예측 모델 평가를 진행했다(박형권 등, 2018). 모델 평가에서의 예측 모델이나 라벨 재분류를 통한 모델의 성능을 평가

했을 때, 데이터나 예측 기법의 튜닝(tuning)에 따라 예측 모형의 성능이 상당량 향상될 가능성이 존재한다. <표 5>의 결과는 본 연구의 기본 예측 모델에서 Scikit-Learn의 Classification Report 패키지를 활용하여 라벨을 재분류한 것을 적용하여 소수점 셋째 자리에서 반올림하여 나타낸 것이다. 예측 모형의 재분류 기준은 KIS 신용평점의 등급체계에 해당하는 최상급, 상급, 중급, 하급, 불량 5개 목표 라벨에 대해서 분석했다. 그 결과 라벨을 재분류한 모형의 예측 정확도는 68%로 측정되었다. 정밀도와 재현율, f1 점수는 각각 66%, 68%, 67%로 개체 분류에 있어서 성능이 양호하게 나타났다(김성진과 안현철, 2016; Kim and Ahn, 2012; Guo et al., 2012).

기존 회사채 신용등급 연구의 경우 목표 예측 값이 다르고 시계열 예측 분석이 아니기 때문에 본 연구의 결과와 다소 차이가 있을 수 있다. Kim and Ahn.(2012)은 4개의 클래스 분류에 대해서 OMSVM 분석 기법으로 67.98%의 예측 정확도가 나타났다. 본 연구의 5개 클래스 분류와 비교했을 때 예측 모형의 성능이 안정적인 것을 알 수 있다. Wu et al.(2014)의 연구에서는 9개의 신용위험 수준 클래스 예측 모델에서 3가지 데이터 처리 기법을 통한 성능을 비교했는데, 특성 선정이 사용된 모형에서는 예측 정확도 50.74%, k-means 모형을 사용하여 군집화를 적용했을 때 61%, 데이터 리샘플링을 적용했을 때 82.96%의 성능을 나타냈다. 하지만 이 연구에서 사용된 데이터 리샘플링의 경우는 실제 기업들의 분포 비율과 다르게 균일한 분포로 리샘플링을 수행했기 때문에 본 연구와 비교하기에 적합하지 않다. Wu et al.(2014)의

연구는 본 연구와 비교했을 때 데이터 전처리가 예측 모델에서 성능 향상을 나타낼 수 있다는 점을 시사한다. Kim and Ahn.(2016)의 경우 회사채 신용등급 예측에서 4개의 클래스 예측을 했을 때 랜덤 포레스트의 예측 정확도가 72.29%로 나타났다. 최근 랜덤 포레스트 분석 기법이 다양한 비정형 데이터에 대한 예측 모델 연구에서 우수한 성능을 나타내고 있다(Chen et al., 2015; 김성진과 안현철, 2016). 하지만 본 연구에서는 랜덤 포레스트 분석 기법이 시계열적인 입력 데이터를 활용하지 못하여 예측 성능이 낮은 것으로 나타났다.

<표 5> 5등급 분류 모델 평가

accuracy	precision	recall	f1 score
0.68	0.66	0.68	0.67

기업 부도 예측과 관련된 연구에서는 기업의 부실 여부를 두 가지 클래스로 예측한다(김영태와 김명환, 2001; Altman, 1968; Yang et al., 1999). KIS 신용평점의 등급체계에서 불량에 속하는 기업들은 등급체계의 ‘차입금 상환 능력 매우 낮음, 부실화 가능성 높음, 주의요구, 부실징후, 잠재적 부실화 요인에 따라 상환능력 지속적 저하 예상.’에 해당하는 8등급 이하라는 것을 의미한다. 기업의 부실화 징후를 예측하기 위해 KIS 신용평점의 8등급 이하와 7등급 이상을 분류하는 2개 클래스 분류 예측 모델을 구현했다. 결과적으로 <표 6>과 같이 기존 기업 부도 예측 모델 연구에 비해 비교적으로 우수한 성능을 나타내고 있다(Yand et al., 1999; Yeh et al., 2010; Chou et al., 2017). 예측 정확도가 약 90%, f1 점수가 약 89%로 안정적인 예측 모형의 성능이 측정되었다. 결론적으로 분류 클래스

스의 개수에 따라 각 예측 모델간의 성능 격차가 크게 발생하는 것으로 나타났다.

기존 기업 부도 예측 연구의 경우 2가지 클래스 분류로써 Chaudhuri and De.(2011)은 하이브리드 방식으로 퍼지 군집화 알고리즘 기반 SVM을 적용했을 때 예측 정확도가 83.37%였다. Chou et al.(2017)은 유전자 알고리즘과 퍼지 군집화를 하이브리드로 적용하여 결과적으로 예측 정확도가 약 92% 정도로 측정된다. Barboza et al.(2017)은 다양한 머신러닝 분석 기법을 적용한 결과 랜덤 포레스트가 기업 부도를 약 87%로 예측했다. 기업 부도 예측 연구들이 데이터 분포적인 측면에서 편차가 크기 때문에 단순 예측 정확도로 비교하기는 어렵지만 결과적으로 본 연구의 예측 모형이 기존 연구들과 비교적 양호한 예측 정확도를 가진다는 것을 알 수 있다.

<표 6> 신용불량 기업 예측 모델 평가

accuracy	precision	recall	f1 score
0.9	0.89	0.9	0.89

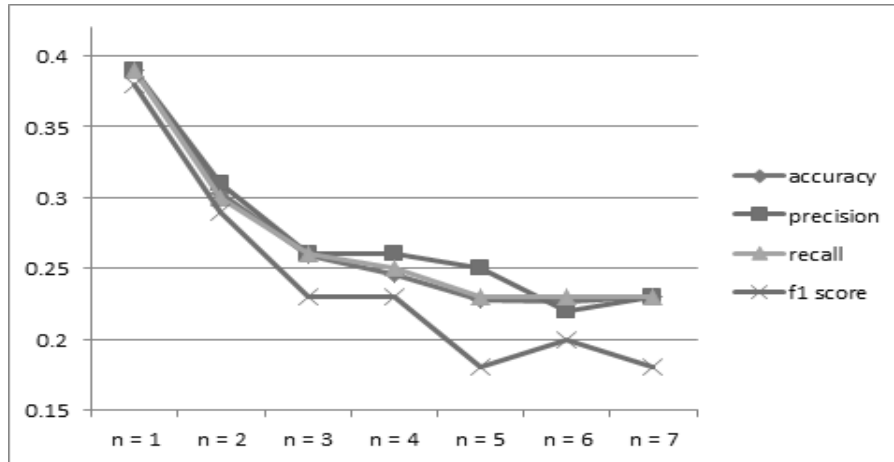
4.3 예측 시점 확장

예측 시점을 7년까지 확장한 결과 <표 7>과

같은 결과가 도출되었다. 3년 후 예측 시점부터는 성능 지표들의 하락 폭이 상당히 줄어들고 5년 후 예측 시점부터는 거의 예측 성능의 변화가 없는 것이 측정된다. 이는 본 연구의 예측 모델이 5년 후 부터는 입력 데이터의 시계열적인 요인이 대부분 사라진다는 것을 의미한다. 따라서 예측 시점이 5년 이상일 경우 각 예측 모델들은 비슷한 속성을 가질 것으로 판단된다. 다른 지표에 비해 f1 점수의 하락폭이 큰 것으로 보아 예측 시점이 미래일수록 특정 개체에 대한 분류가 다양하게 이루어지지 못하는 것을 알 수 있다. 정밀도와 재현율은 대부분 똑같은 추이를 보이는데, 이는 예측 모델 전반적으로 개체 분류에 대해 안정적인 형태를 가지고 있다고 볼 수 있다. 예측 시점이 확장된 학습 모델을 단기적 및 장기적 시점에서 특성 그룹화 평가를 하기 위해 1년 후 시점의 예측 모델과 7년 후 시점의 예측 모델의 특성 그룹별 예측 분석을 진행했다. 7년 후 예측 시점을 장기적 시점의 예측 모델의 기준으로 선정한 근거는 다음과 같다. 학습 모델의 예측 시점이 5년 후부터 입력 데이터의 시계열적 요인이 대부분 사라지기 시작하여 시점별 성능의 변화가 관찰되지 않았기 때문에 5년 후 시점의 예측 모델 이후 마지막 시점의 모형을 활용한 것이다.

<표 7> 예측 시점 확장

예측 연도 n	accuracy	precision	recall	f1 score
n = 1	0.3902	0.39	0.39	0.38
n = 2	0.3028	0.31	0.3	0.29
n = 3	0.2598	0.26	0.26	0.23
n = 4	0.2453	0.26	0.25	0.23
n = 5	0.2274	0.25	0.23	0.18
n = 6	0.2263	0.22	0.23	0.2
n = 7	0.2306	0.23	0.23	0.18



<그림 2> 예측 시점 확장 성능 변화 추이

4.4 특성 그룹 평가

기업 신용위험 예측 모델의 시점별 분석을 위해 1년 후와 7년 후 신용평점을 예측하는 모델의 특성 그룹 평가를 <표 8>, <표 9>와 같이 나타냈다. 신용평점(Credit Rating)은 예측 목표 값으로써 단순히 신용평점 등급만 가지고 시계열 예측을 했을 때 정확도가 36%로 나타났다. 전반적으로 특성 그룹이 추가된 예측 모델과 단일 변량 예측 모델의 성능이 큰 차이가 나지 않는다.

1년 후 시점의 예측모델에서 규모 지표(Scale Index)의 경우 특성 그룹들 중 예측 정확도 측면에서는 가장 큰 영향을 미치는 것으로 나타났다. 이 특성 그룹 평가 모델의 예측 정확도는 39.21%로 특성을 전부 적용한 모델보다도 높은 수치다. 시계열 분석의 특성상 다양한 지표들을 전부 활용하는 것보다 주요 지표들을 활용하는 것이 성능을 개선할 수 있는 것으로 해석된다. 하지만 규모 지표에서 f1 점수가 낮게 측정되는 것은 특정 목표 라벨에서는 분류가

잘 이루어지지 않는다는 것을 의미한다. 수익성 지표(Profitability Index)는 예측 모델에 대해서 준수한 영향을 미치는 것으로 나타났고, 정밀도, 재현율, f1 점수 모두 안정적인 수치가 측정되었다. 무역 지표(Trade Index)와 관계회사 지표(Affiliated Company Index), 안정성 지표(Stability Index)는 예측 모델의 성능에 큰 영향을 주지 못하는 것으로 나타났으며 특히 f1 점수가 낮게 측정되었는데, 이는 다중 클래스 분류에서 특정 클래스의 목표 값을 제대로 측정하지 못하는 것으로 추론된다. 신용 지표(Credit Index)의 경우 다른 지표들에 비해 가장 적은 영향력을 나타냈다. KISVALUE에서 제공하는 신용 지표 정보는 2002년 기준으로 공시를 하기 때문에 다른 지표들에 비해 큰 영향을 미치지 못한 것으로 예상된다. 텍스트 감성 지표(Text Sentiment Index)의 경우 예측 정확도가 39.12%로 상당히 높게 측정되면서 예측 모델에서 KISVALUE에서 제공하는 텍스트 데이터가 신용평점 예측 모델에 큰 영향을 주는 것을

확인했다. 특히 재현율의 측면에서 다른 성능 지표들 보다 높게 나타났는데, 이것은 데이터의 양성 값 중에서 예측 목표 값을 정확하게 예측 하도록 한다는 것이다.

7년 후 시점의 예측 모델에서도 <표 8>처럼 전반적인 예측 모델의 성능이 단일 변량 예측 모델에 비해서 크게 차이가 나지 않았다. 규모 지표의 경우 정밀도, 재현율, f1 점수 부분에서는 안정적인 형태를 보였지만 단기적 시점과는 다르게 전체적인 성능 향상이 크게 이루어지지 않았다. 수익성 지표는 예측 정확도가 23.91%로 가장 높은 성능 향상을 보였지만 정밀도가 낮게 측정되었다. 이는 수익성 지표를 활용한 예측 모델이 양성으로 분류한 값들이 높은 오차를 가지고 있다는 것을 의미한다. 단기적

시점에서의 수익성 지표는 개체 분류에 대해 안정적인 성능이 나타나는 반면에 장기적 시점에서는 크게 불안정한 모습이 나타났다. 무역 지표는 단기적 시점에서 낮은 성능 향상을 나타냈지만 장기적 시점에서는 안정적인 예측 성능이 측정되었다. 이는 무역 지표가 기업 신용 위험 예측에 단기적인 영향은 미미하지만 장기적으로 안정적인 개체 분류에 영향을 미칠 수 있다는 점을 시사한다. 관계회사 지표와 안정성 지표는 예측 정확도에는 양호한 성능 향상을 나타내지만 f1 점수적인 측면에서 불안정한 수치가 측정되었다. 신용 지표의 경우 단기적 시점에서는 영향이 다소 적었지만 장기적 시점에서 가장 안정적이고 큰 영향을 나타냈다. 이를 통해 KISVALUE에서 제공하는 기업 신용정보

<표 8> 단기 시점 특성 그룹 평가

Feature Group	accuracy	precision	recall	f1 score
Credit Rating(target label)	0.3605	0.30	0.35	0.32
Scale Index	0.3921	0.38	0.39	0.37
Profitability Index	0.3904	0.38	0.39	0.38
Trade Index	0.3844	0.37	0.38	0.36
Affiliated Company Index	0.3885	0.38	0.39	0.35
Stability Index	0.3850	0.36	0.39	0.36
Credit Index	0.3749	0.37	0.37	0.35
Text Sentiment Index	0.3912	0.37	0.39	0.36

<표 9> 장기 시점 특성 그룹 평가

Feature Group	accuracy	precision	recall	f1 score
Credit Rating(target label)	0.2110	0.13	0.21	0.13
Scale Index	0.2336	0.21	0.23	0.21
Profitability Index	0.2391	0.18	0.24	0.19
Trade Index	0.2331	0.21	0.23	0.20
Affiliated Company Index	0.2346	0.18	0.23	0.16
Stability Index	0.2347	0.19	0.23	0.19
Credit Index	0.2365	0.23	0.24	0.21
Text Sentiment Index	0.2327	0.18	0.23	0.19

가 신용위험 단기 예측모델에서는 큰 역할을 하지 못하지만 장기적 시점에서는 성능 향상과 함께 가장 안정적인 개체 분류에 큰 영향을 미치는 것을 알 수 있다. 반대로 텍스트 감성 지표는 단기적 시점에서는 비교적 큰 영향을 나타냈지만 장기적 시점에서는 큰 영향을 주지 못하며 특정 개체 분류에 대해서도 불안정한 형태를 보였다.

V. 결 론

본 논문은 KIS 신용평점의 시계열적 예측을 위해 9개의 머신러닝 기법을 사용한 결과 RNN 기반 상태저장 LSTM 기법이 기존 연구에 사용되었던 기법이나 다른 딥러닝 기법에 비해 2배 가까이 높은 예측 성능을 보이는 것을 관찰했다. 연구결과와 강건성을 높이기 위해 라벨 재분류를 통한 예측 모델을 다시 구현하여 분석한 결과, 라벨의 클래스(class)가 줄어들면서 예측 성능은 향상되었지만 예측 정확도 68%로 기존 연구의 신용 등급 예측에 비해서 큰 차이가 나지 않는 것을 확인했다(김성진과 안현철, 2016; Kim and Ahn, 2012; Guo et al., 2012). 예측 시점 확장의 경우 시계열 예측 모델을 통해 예측 시점을 7년 후 까지 확장했는데, 이는 기존 연구의 예측 시점 조절에 대한 문제점을 해결했다고 볼 수 있다(김성진과 안현철, 2016; Kim and Ahn, 2012; Chaudhuri and De, 2011; Guo et al., 2012; Barboza et al, 2017).

본 논문의 공헌점은 다음과 같다. 첫 번째, 기업 신용 예측의 대상 기업 범위를 확장하면서 비교적 이해하기 쉬운 지표인 KIS 신용평점을

기업 신용위험의 척도로써 제시한다. 회사채 신용등급의 경우 대상 기업의 범위도 한정적이고 연도별 데이터 또한 확보하기 쉽지 않은 상황이다. 본 연구는 정보시스템연구 분야의 신용평가 연구에도 새로운 기업 신용위험의 척도를 제시함과 동시에 다수 개체에 대한 시계열 예측 모델에 대한 연구 방안을 제시하고 있다. 기업 신용평점을 대상으로 신용위험의 요인 분석을 수행한 연구들이 있지만 신용평점을 목표 라벨 값으로 예측 모델을 구현하는 연구는 제한적인 상황이다(조현우 등, 2005; 전성일과 이기세, 2010; 김성진과 안현철, 2016; Kim and Ahn, 2012; Guo et al., 2012). 본 논문은 KIS 신용평점을 분석함으로써 KOSPI, KOSDAQ, KONEX에 속하는 2,191개의 기업으로 확장하고 기업 신용평점을 목표 라벨 값으로 예측함으로써 다양한 기업의 데이터를 통해 예측 모델을 구현했다.

두 번째, 본 논문은 슬라이딩 윈도우 기법을 활용한 시계열 분석을 통해 시간적 단계에 따라 신용평점을 예측하는 모델을 구축했다. 기존 연구의 경우 다음 연도의 회사채 신용등급이나 기업 부도를 예측하는 연구들이 대부분이다. 이는 예측 시점을 확장하지 못하고 시계열적 시퀀스 정보를 활용하지 못한다는 한계가 있다(김성진과 안현철, 2016; Kim and Ahn, 2012; Chaudhuri and De, 2011; Guo et al., 2012; Barboza et al, 2017). 본 논문은 국내 2,191개 기업의 18년 간 데이터를 활용하여 7년 후의 신용평점을 예측할 수 있는 시계열 분석 모델을 실증적으로 구현했다는 것에 의의가 있으며 단기적 및 장기적 시점에서 특성 그룹 평가를 수행함으로써 채무 회사나 채권자들에게 중요 지

표들을 식별할 수 있게 하고 있다. 본 연구의 예측 모델을 통해 기업 신용 위험 예측에 있어서 단기적 관점에서의 예측뿐만 아니라 장기적 관점에서의 예측 또한 가능해진 것이다. 기업 신용평점 시계열 예측 모델은 회사채 신용등급 예측과 연계하여 회사채를 발행하지 않는 기업의 신용평점을 예측하여 채무 기업에 대한 시장이자율을 조정하는데 활용하거나, 채권자나 투자자가 대상 회사에 대한 신용 위험을 단기적 및 장기적 관점에서 예측하여 대응할 수 있는 점 등으로 활용이 가능한 측면이 있다.

세 번째, 본 연구는 기존 국내 신용등급 예측 모형 연구에 비해 최신 기술인 딥러닝 분석 방식을 적용했으며 기존 연구에서 주로 사용되던 SVM이나 랜덤 포레스트 기법에 비해 우수한 시계열 예측 모델을 구현했다(김성진과 안현철, 2016; Glorot et al., 2011; Kim and Ahn, 2012; Guo et al., 2012; Chen et al., 2014; Lecun et al., 2015). 가장 예측 정확도가 높았던 시계열 예측 분석 기법은 stateful LSTM 기법으로 예측 정확도가 39%에 달하고, 라벨 재분류 시에는 68%의 예측 정확도를 나타냈다. 기존 신용등급 예측에 비해서 성능이 크게 차이나진 않지만 2,191개의 기업을 대상으로 18년에 해당하는 목표 라벨, 즉 약 3만 9천개에 대한 예측 모델을 구현했다는 점, 그리고 기존 연구의 머신러닝 분석 기법보다 2배 가까이 예측 정확도가 높았다는 점에서 시사점이 있다고 볼 수 있다(김성진과 안현철, 2016; Kim and Ahn, 2012; Chaudhuri and De, 2011; Guo et al., 2012; Barboza et al., 2017). 기업 불량 예측의 경우 예측 정확도가 90%에 달하여 다른 연구들과 비교적 높은 수준의 예측 성능이 나타났다

(Yand et al., 1999; Yeh et al., 2010; Chou et al., 2017). f1 점수의 경우도 89%로 안정적으로 측정되었으며 이는 학술적 시사점과 동시에 실무적 시사점을 가진다고 볼 수 있다. 부실 기업 예측은 신용위험 관리 연구 분야의 한 영역으로서 해당되는 중요한 연구 분야이다. 본 연구는 성능적 측면에서 우수하면서 안정적인 시계열 예측 모델을 제안함으로써 관련 연구자들에게 참고점이 될 수 있으며 기업의 투자자나 채권자들에게 기업 부실 징후를 예측할 수 있는 방안을 제시한다.

결과적으로 본 논문은 기업 신용평점이라는 불완전한 데이터에 대해 다수의 개체들을 하나의 연속형 데이터로써 활용하여 시계열 예측 분석을 하는 연구 방안을 제시한다. 다수의 개체에 대한 시계열 예측 모형은 부동산, 주식, 암호화폐 등 다양한 분야에서 시계열 데이터를 활용한 예측 모델 구현에 활용될 수 있을 것으로 기대된다. 반면에 본 연구는 다음과 같은 한계점을 가지고 있다. 텍스트 분석의 경우 기업에 대한 평가를 영화 리뷰 데이터 기반 감성 분석을 활용했기 때문에 감성 점수를 할당하는 부분에서 적합하지 못한 측면이 있다. 입력 특성 데이터의 추가 및 다양한 샘플링 기법을 통해 더 나은 예측 모델을 만들 수 있는 가능성이 존재한다. 이에 따라 다수의 시계열 예측 모델에 대해서 다양한 방식의 접근 방법이 필요하다. 신용평점 예측은 국내 기업을 대상으로 KISVALUE 측에서 평가한 것으로 국제적인 관점으로 봤을 때 일반화하기 어렵다는 한계도 있다. 추후 연구에서는 글로벌 기업들을 대상으로 기업 신용위험을 일반화하여 시계열적으로 예측할 수 있는 연구 모형을 구현할 수 있다면

더욱 유용하게 사용될 수 있을 것이다.

참고문헌

- 김성진, 안현철, “기업신용등급 예측을 위한 랜덤 포레스트의 응용,” 산업혁신연구, 32권, 1호, 2016, pp. 187-211.
- 김영태, 김명환, “인공신경망을 활용한 부실기업 예측모형에 관한 연구,” 회계연구, 6권, 1호, 2001, pp. 275-294.
- 김진성, “무디스 ‘한국, 무더기 신용강등’ 경고,” 한국경제, <https://www.hankyung.com/finance/article/2019111986551>, 2019
- 김태정, 문남희, 문성주, “회사채 신용등급 예측에 관한 연구,” 회계정보연구, 21권, 2003, pp. 25-59.
- 나영, 진동민, “IMF 이후 신용등급예측에 있어서 재무정보의 유용성,” 회계정보연구, 21권, 2003, pp. 211-235.
- 노태협, 유명환, 한인구, “러프집합이론과 사례 기반추론을 결합한 기업신용평가 모형,” 한국정보시스템학회, 정보시스템연구, 14권, 1호, 2005, pp. 41-65.
- 박형권, 강준영, 허성욱, 유동현, “국내 회사채 신용 등급 예측 모형의 비교 연구,” 응용통계연구, 31권, 3호, 2018, pp. 367-382.
- 이륜경, 정남호, 홍태호, “딥러닝을 이용한 온라인 리뷰 기반 다속성별 추천 모형 개발,” 28권, 1호, 2019, pp. 97-114.
- 이성호, “M&A 대상기업의 재무 상태에 관한 실증적 연구,” 경영경제연구, 1997.
- 이영찬, “DEA와 Worst Practice DEA를 이용한 정보통신기업의 신용위험평가,” 한국정보시스템학회, 2005년 한국정보시스템학회 추계 학술발표논문집, 2005, pp. 334-346.
- 전성일, 이기세, “보수주의 회계처리가 KIS 신용평점에 미치는 영향,” 국제회계연구, 30권, 2010, pp. 245-263.
- 조현우, 박연희, 송혁준, “사외이사의 특성과 감사품질이 KIS 신용평점에 미치는 영향” 한국회계학회 학술연구발표회 논문집, 2005, pp. 489-515.
- 최병설, 김남규, “감정 딥러닝 필터를 활용한 토픽 모델링 방법론,” 28권, 4호, 2019, pp. 271-291.
- 홍태호, 신태수, “부도확률맵과 AHP를 이용한 기업 신용등급 산출모형의 개발,” 한국정보시스템학회, 정보시스템연구, 16권, 3호, pp. 1-20.
- Aggarwal, C. C., and Zhai, C., “Mining text data,” *Springer Science & Business Media*, 2012.
- Altman, E. I., “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy,” *The journal of finance*, Vol. 23, No.4, 1968, pp. 589-609.
- Barboza, F., Kimura, H., and Altman, E., “Machine learning models and bankruptcy prediction,” *Expert Systems with Applications*, Vol. 83, 2017, pp. 405-417.
- Breiman, L., “Random forests,” *Machine*

- learning, Vol. 45, No. 1, 2001, pp. 5-32.
- Bridle, J. S., "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition," *Neurocomputing Springer*, 1990, pp. 227-236.
- Chaudhuri, A., and De, K., "Fuzzy support vector machine for bankruptcy prediction," *Applied Soft Computing*, Vol. 11, No. 2, 2011, pp. 2472-2486.
- Chen, S.-Y., Chen, C.-N., Chen, Y.-R., Yang, C.-W., Lin, W.-C., and Wei, C.-P., "Will Your Project Get the Green Light? Predicting the Success of Crowdfunding Campaigns," *Paper presented at the PACIS*, 2015.
- Chen, X., Wei, L., and Xu, J., "House Price Prediction Using LSTM," *arXiv preprint arXiv:1709.08432*, 2017.
- Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y., "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected topics in applied earth observations and remote sensing*, Vol. 7, No. 6, 2014, pp. 2094-2107.
- Chou, C.-H., Hsieh, S.-C., and Qiu, C.-J., "Hybrid genetic algorithm and fuzzy clustering for bankruptcy prediction," *Applied Soft Computing*, Vol. 56, 2017, pp. 298-316.
- Claesen, M., and De Moor, B., "Hyperparameter search in machine learning," *arXiv preprint arXiv:1502.02127*, 2015.
- Corbett, P., and Boyle, J., "Improving the learning of chemical-protein interactions from literature using transfer learning and specialized word embeddings," *Database*, 2018.
- Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., and Darrell, T., "Long-term recurrent convolutional networks for visual recognition and description," *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.
- Duffie, D., Saita, L., and Wang, K., "Multi-period corporate default prediction with stochastic covariates," *Journal of Financial Economics*, Vol. 83, No. 3, 2007, pp. 635-665.
- Flagg, J. C., Giroux, G. A., and Wiggins Jr, C. E., "Predicting corporate bankruptcy using failing firms," *Review of financial Economics*, Vol. 1, No. 1, 1991, pp. 67-75.
- Frank, R. J., Davey, N., and Hunt, S. P., "Time series prediction and neural networks," *Journal of intelligent and robotic systems*, Vol. 31, No. 1-3, 2001, pp. 91-103.
- Glorot, X., Bordes, A., and Bengio, Y., "Domain adaptation for large-scale sentiment classification: A deep

- learning approach,” *Paper presented at the Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011.
- Guo, X., Zhu, Z., and Shi, J., “A corporate credit rating model using support vector domain combined with fuzzy clustering algorithm,” *Mathematical Problems in Engineering*, 2012.
- Hajek, P., and Michalak, K., “Feature selection in corporate credit rating prediction,” *Knowledge-Based Systems*, Vol. 51, 2013, pp. 72-84.
- Hall, M. A., and Smith, L. A., “Feature subset selection: a correlation based filter approach,” 1997.
- Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., and Scholkopf, B., “Support vector machines,” *IEEE Intelligent Systems and their applications*, Vol. 13, No. 4, 1998, pp. 18-28.
- Ho, S.-L., Xie, M., and Goh, T. N., “A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction,” *Computers & Industrial Engineering*, Vol. 42, No. 2-4, 2002, pp. 371-375.
- Hochreiter, S., and Schmidhuber, J., “Long short-term memory,” *Neural computation*, Vol. 9, No. 8, 1997, pp. 1735-1780.
- Huang, S.-C., “Integrating nonlinear graph based dimensionality reduction schemes with SVMs for credit rating forecasting,” *Expert Systems with Applications*, Vol. 36, No. 4, 2009, pp. 7515-7518.
- Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., and Wu, S., “Credit rating analysis with support vector machines and neural networks: a market comparative study,” *Decision Support Systems*, Vol. 37, No. 4, 2004, pp. 543-558.
- Kim, K.-j., and Ahn, H., “A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach,” *Computers & Operations Research*, Vol. 39, No. 8, 2012, pp. 1800-1811.
- KISVALUE, “KISVALUE Homepage,” <https://www.kisvalue.com/web/index.jsp>, 2019.
- KISVALUE, “KISVALUE MANUAL,” 2019.
- Koç, C. K., “Analysis of sliding window techniques for exponentiation,” *Computers & Mathematics with Applications*, Vol. 30, No. 10, 1995, pp. 17-24.
- Kotsiantis, S. B., Zaharakis, I., and Pintelas, P., “Supervised machine learning: A review of classification techniques,” *Emerging artificial intelligence applications in computer engineering*, Vol. 160, 2007, pp. 3-24.
- Kumar, K., and Bhattacharya, S., “Artificial neural network vs linear discriminant analysis in credit ratings forecast: A comparative study of prediction

- performances,” *Review of Accounting and Finance*, Vol. 5, No. 3, 2006, pp. 216-227.
- LeCun, Y., Bengio, Y., and Hinton, G., “Deep learning,” *Nature*, Vol. 521 No. 7553, 2015, pp. 436.
- LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., and Jackel, L. D., “Handwritten digit recognition with a back-propagation network,” *Paper presented at the Advances in neural information processing systems*, 1990.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., “Gradient-based learning applied to document recognition,” *Proceedings of the Ieee*, Vol. 86, No. 11, 1998, pp. 2278-2324.
- Lee, Y.-C., “Application of support vector machines to corporate credit rating prediction,” *Expert Systems with Applications*, Vol. 33, No. 1, 2007, pp. 67-74.
- Li, Y., Rakesh, V., and Reddy, C. K., “Project success prediction in crowdfunding environments” *Paper presented at the Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, 2016.
- Liu, Y., Huang, X., An, A., and Yu, X., “ARSA: a sentiment-aware model for predicting sales performance using blogs” *Paper presented at the Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007.
- Liu, Y., Zheng, H., Feng, X., and Chen, Z., “Short-term traffic flow prediction with Conv-LSTM,” *Paper presented at the 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2017.
- Mohammadi, K., Shamshirband, S., Tong, C. W., Arif, M., Petković, D., and Ch, S., “A new hybrid support vector machine - wavelet transform approach for estimation of horizontal global solar radiation,” *Energy Conversion and Management*, Vol. 92, 2015, pp. 162-171.
- Pineda, F. J., “Generalization of back-propagation to recurrent neural networks,” *Physical Review Letters*, Vol. 59, No. 19, 1987, pp. 2229.
- Price, D., Knerr, S., Personnaz, L., and Dreyfus, G., “Pairwise neural network classifiers with probabilistic outputs,” *Paper presented at the Advances in neural information processing systems*, 1995.
- Shin, K.-s., and Han, I., “Case-based reasoning supported by genetic algorithms for corporate bond rating,” *Expert Systems with Applications*, Vol. 16, No. 2, 1999, pp. 85-95.
- Sunasra, Mohammed., “Performance Metrics for

Classification problems in Machine Learning,” THALUS AI in medium (online publishing platform), <https://medium.com/thalus-ai/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b>, 2015.

Wu, H.-C., et al., “Two-stage credit rating prediction using machine learning techniques,” *Kybernetes*, Vol. 43, No. 7, 2014, pp. 1098-1113.

Yang, Z., Platt, M. B., and Platt, H. D., “Probabilistic neural networks in bankruptcy prediction,” *Journal of Business Research*, Vol. 44, No. 2, 1999, pp. 67-74.

Ye, Y., Liu, S., and Li, J., “A multiclass machine learning approach to credit rating prediction,” *Paper presented at the 2008 International Symposiums on Information Processing*, 2008.

Yeh, C.-C., Chi, D.-J., and Hsu, M.-F., “A hybrid approach of DEA, rough set and support vector machines for business failure prediction,” *Expert Systems with Applications*, Vol. 37, No. 2, 2010, pp. 1535-1541.

이 현 상 (Lee, Hyun-Sang)



경북대학교에서 경영학 학사와 석사 학위를 취득했으며 현재 경북대학교 경영학부 박사과정에 재학 중이다. 주요 관심 분야는 머신러닝, 예측 모델, 텍스트 마이닝 등이다.

오 세 환 (Oh, Sehwan)



현재 경북대학교 경영학부에서 조교수로 재직 중이다. 서울대학교 경제학부(학사)를 졸업했으며 카네기멜론대에서 e-비즈니스 석사, 서울대학교에서 경영학(경영정보) 박사 학위를 받았다. *International Journal of Mobile Communications*, *Internet Research*, *Journal of Electronic Commerce Research* 등에 논문을 게재했으며 주요 연구 관심분야는 공유경제, 전자상거래, 온라인 구전 등이다.

<Abstract>

LSTM-based Deep Learning for Time Series Forecasting: The Case of Corporate Credit Score Prediction

Lee, Hyun-Sang · Oh, Sehwan

Purpose

Various machine learning techniques are used to implement for predicting corporate credit. However, previous research doesn't utilize time series input features and has a limited prediction timing. Furthermore, in the case of corporate bond credit rating forecast, corporate sample is limited because only large companies are selected for corporate bond credit rating. To address limitations of prior research, this study attempts to implement a predictive model with more sample companies, which can adjust the forecasting point at the present time by using the credit score information and corporate information in time series.

Design/methodology/approach

To implement this forecasting model, this study uses the sample of 2,191 companies with KIS credit scores for 18 years from 2000 to 2017. For improving the performance of the predictive model, various financial and non-financial features are applied as input variables in a time series through a sliding window technique. In addition, this research also tests various machine learning techniques that were traditionally used to increase the validity of analysis results, and the deep learning technique that is being actively researched of late.

Findings

RNN-based stateful LSTM model shows good performance in credit rating prediction. By extending the forecasting time point, we find how the performance of the predictive model changes over time and evaluate the feature groups in the short and long terms. In comparison with other studies, the results of 5 classification prediction through label reclassification show good performance relatively. In addition, about 90% accuracy is found in the bad credit forecasts.

Keyword: KIS credit score, machine learning, deep learning, time series forecasting, sliding window technique

* 이 논문은 2020년 2월 20일 접수, 2020년 2월 29일 1차 심사, 2020년 3월 16일 게재 확정되었습니다.