한국지능정보시스템학회 [2022 춘계학술대회] 논문 투고

|  |  |  |
| --- | --- | --- |
| **논문 전문(Full Paper) 제출 희망 여부 (√표시)** | 예 [ ] | 아니오 [√] |
| **심사용 전문 제출 용도 (√표시) – 복수 선택 가능** | 우수논문 심사 [ ]  학술지 Fast Track 심사 [ ] | **해당사항 없음** |

제목

|  |  |  |  |
| --- | --- | --- | --- |
| 안승규 | (국민대학교 | 비즈니스IT전문대학원 석사과정, 저자 | tmdrb0415@kookmin.ac.kr) |
| 박종현 | (국민대학교 | 비즈니스IT전문대학원 석사과정, 저자 | ppjjhh1027@kookmin.ac.kr) |
| 송찬우 | (국민대학교 | 비즈니스IT전문대학원 석사과정, 저자 | cksdnthd2008@kookmin.ac.kr) |
| 안현철 | (국민대학교 | 비즈니스IT전문대학원 교수, 교신저자 | hcahn@kookmin.ac.kr) |

[저자 연락처]

◎　안승규 (010-8573-1296)

주소 : 서울특별시 성북구 정릉로77 국민대학교 국제관 305호

◎　박종현 (010-3610-8658)

주소 : 서울특별시 성북구 정릉로77 국민대학교 국제관 305호

◎　송찬우 (010-2303-9675)

주소 : 서울특별시 성북구 정릉로77 국민대학교 국제관 305호

◎　안현철 (010-3227-7310)

주소 : 서울특별시 성북구 정릉로77 국민대학교 경영관 313호

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scaler 종류에 따른 부도 예측 결과에 대한 영향성 분석   |  |  |  |  | | --- | --- | --- | --- | | 안승규  국민대학교  비즈니스IT전문대학원  tmdrb0415@kookmin.ac.kr | 박종현  국민대학교  비즈니스IT전문대학원  ppjjhh1027@kookmin.ac.kr | 송찬우  국민대학교  비즈니스IT전문대학원  cksdnthd2008@kookmin.ac.kr | 안현철  국민대학교  비즈니스IT전문대학원  hcahn@kookmin.ac.kr | |
|  |

Abstract – 머신러닝 알고리즘은 데이터의 스케일 조정에 매우 민감하며, 통상적으로 전처리 단계에서 각 특성 스케일을 조정해 데이터를 가공한다. 스케일러의 종류는 StandardScaler, MinMaxScaler, RobustScaler, Normalizer등이 있는데 부도예측에 있어 어떤 Scaler를 사용했을 때 가장 우수한 정확도를 보이는지는 알려진 바가 없다.

이에 본 연구는 부도 예측 데이터를 통해 4가지의 Scaler를 적용하여 어떠한 것이 가장 우수한 정확도를 보여주는지 알아보고자 한다.

Key Terms – 부도예측, StandardScaler, MinMaxScaler, RobustScaler, Normalizer

## **I. 서론**

‘Data Industry Promotion Strategy – I-KOREA 4.0 Data Field Plan, I-DATA+’(2018)이란 보고서에서는 4차 산업혁명을 견인하는 핵심 동인인 빅데이터를 통해 사회문제 해결 능력을 강화하는 것을 핵심 과제로 정했다. 또한 2018년 3월 4차산업혁명위원회의 첫 회의에서 문재인 전 대통령은 인공지능, 사물인터넷, 빅데이터를 위한 투자를 확대하여 혁신생태계를 조성할 것임을 밝히게 되면서 금융 분야에서의 빅데이터 분석에 관한 연구의 필요성이 높아지고 있다(차성재와 강정석, 2018).

그중 기업의 부도를 예측하는 것은 지속하여 연구되고 있고, 예측 정확도를 보다 높이고자 노력하고 있다.

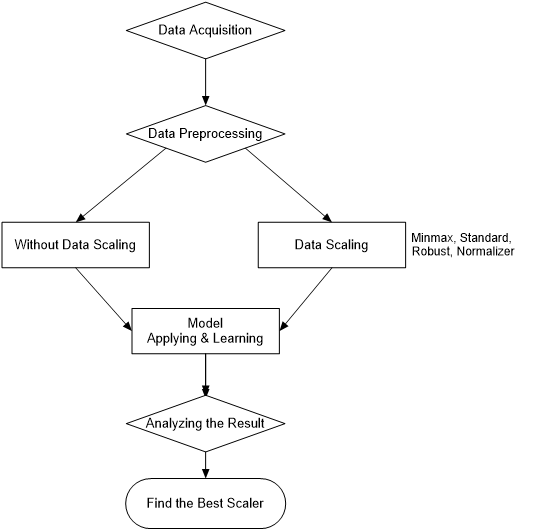
기업의 부도는 그 기업의 경영자, 노동자 측면에만 국한되는 것이 아니라 그 기업에 직∙ 간접적으로 연관관계를 가진 이해관계자(투자자, 금융기관, 거래 기업 등)에게도 연쇄적인 피해를 양산할 수 있다. 더 나아가 국민경제에도 심각한 타격을 미칠 수 있다(강치형과 신해수, 2015).

그러한 데이터 분석 기반 의사결정 지원 시스템을 개발하기 위해서는 데이터 전처리가 필요하다. 데이터 전처리 단계에는 data cleaning, pruning, feature selection, scaling이 존재하지만, 대부분의 부도예측 연구는 이에 집중하지 않고 다른 알고리즘을 적용하며 정확도를 높이고자 하였다(Ahsan, Md Manjurul, et al, 2021).

Scaling 기법 중에는 각 특성의 평균을 0, 분산을 1로 변경해주는 StandardScaler, 이와 유사하지만, 평균과 분산 대신 중간 값과 사분위 값을 사용하는 RobustScaler, 모든 특성이 정확하게 0과 1 사이에 위치하도록 변경하는 MinMaxScaler, 특성 벡터의 Euclidean Distance가 1이 되도록 데이터 포인트를 조정하는 Normalizer가 있다(Andreas C. Muller, Sarah Guido 2016).

본 연구에서는 기존 연구된 부도 예측 모델에 4가지 Scaling 기법을 적용하여 어떠한 영향을 미치는지 평가를 하는 것이 목적이다.

## **II. 연구 모델**



**<그림 1> 제안 모형**

본 연구에서 제안하는 모형은 <그림 1> 과 같다. 먼저 부도 예측 데이터는 2001~2007년 제조업 기업의 재무 정보를 활용한다.

부도 예측 데이터는 각 재무 비율이 모델에 미치는 영향이 다르기 때문에 데이터 전처리를 통해 편향을 줄여줄 것이다. 다음으로 상기 <그림1> 에 제시한 Scaler를 Model에 적용하여 결과를 비교해보고 최종적으로 부도 예측에 가장 적절한 Scaler를 도출하고자 한다.

## **III. 참고문헌**

차성재, 강정석. (2018). 딥러닝 시계열 알고리즘 적용한 기업부도예측모형 유용성 검증. 지능정보연구,24(4),1-32.

강치형, 신해수. (2015). 회원제 골프장기업의 부도예측 모형개발. 관광연구논총, 27(4), 241-269.

Ahsan, Md Manjurul, et al. (2021) "Effect of data scaling methods on machine learning algorithms and model performance." *Technologies* 9.3: 52.

Andreas C. Muller & Sarah Guido, “Introduction to Machine Learning with Python”