

머신러닝 기반의 기업가치 예측 모형: 온라인 기업리뷰를 활용하여[☆]

Machine Learning based Firm Value Prediction Model: using Online Firm Reviews

이 한 준¹ 신 동 원² 김 희 은^{3*}
Hanjun Lee Dongwon Shin Hee-Eun Kim

요 약

빅데이터 분석의 유용성이 주목을 받으면서 경영학 분야에서도 이를 활용하여 기업의 성과를 예측하고자 하는 다양한 연구들이 진행되고 있다. 이러한 선행연구들은 주로 뉴스 기사나 SNS 등 기업 외부의 자료에 의존하고 있다. 직원의 만족도나 기업에 대한 직원의 인식, 장단점 평가와 같은 기업 내부의 목소리는 기업가치에 대한 잠재적인 영향력에도 불구하고 상대적으로 확보가 어려워 관련 연구가 아직 충분치 못하다. 이에 본 연구에서는 국내 유가증권시장 상장 기업을 대상으로 임직원의 기업리뷰가 기업가치에 미치는 영향을 살펴보고, 이를 기반으로 기업가치를 예측하는 모형을 구축하고자 한다. 이를 위해 온라인 기업리뷰 사이트인 잡플래닛(Jobplanet)에 2014년부터 2019년까지 전·현직원이 남긴 97,216건의 기업리뷰를 수집하고 동 데이터에 근거하여 머신러닝 기반의 예측 모형을 제안하였다. 제안한 모형 중 LSTM 기반 모형의 정확도가 73.2%로 가장 높았고 MAE 또한 0.359로 가장 낮은 오차를 보였다. 본 연구는 국내에서 머신러닝을 활용한 기업가치 연구 분야에 유용한 사례가 될 것으로 기대한다.

☞ 주제어 : 기업가치, 머신러닝, 온라인 리뷰, 기업리뷰, 잡플래닛

ABSTRACT

As the usefulness of big data analysis has been drawing attention, many studies in the business research area begin to use big data to predict firm performance. Previous studies mainly rely on data outside of the firm through news articles and social media platforms. The voices within the firm in the form of employee satisfaction or evaluation of the strength and weakness of the firm can potentially affect firm value. However, there is insufficient evidence that online employee reviews are valid to predict firm value because the data is relatively difficult to obtain. To fill this gap, from 2014 to 2019, we employed 97,216 reviews collected by JobPlanet, an online firm review website in Korea, and developed a machine learning-based predictive model. Among the proposed models, the LSTM-based model showed the highest accuracy at 73.2%, and the MAE showed the lowest error at 0.359. We expect that this study can be a useful case in the field of firm value prediction on domestic companies.

☞ keyword : Firm value, Machine learning, Online review, Firm review, Jobplanet

1. 서 론

빅데이터를 수집하고 분석하는 기술이 발전함에 따라

경영학 분야에서도 이를 활용하여 기업가치를 예측하고자 하는 다양한 연구들이 진행되고 있다. 과거 기업의 가치를 평가하기 위해 전통적으로 가장 널리 사용되어 온 것은 기업의 회계 정보이다. 즉 재무제표로 공시되는 이익과 장부가에 근거하여 기업의 가치를 판단하는 것이다. 기존 연구에서는 기업의 회계적 성과가 기업가치에 대해 설명력을 가진다는 것을 보이고 이를 활용한 기업가치 평가 모형을 제시한 바 있다[1][2]. 그리고 후속 연구에서는 기업의 재무상태나 규모, 재무건전성 등 회계정보의 가치 관련성(value-relevant)에 영향을 주는 여러 추가적인 요인을 분석하여 기업가치 평가 모형을 정교화하는 노력을 기울여왔다.

빅데이터 분석 방법론이 다양한 분야에 폭넓게 적용되

¹ Department of Management Information Systems, Myongji University, Seoul, 03674, Korea.

² Industry Academic Cooperation Foundation, Myongji University, Seoul, 03674, Korea.

³ College of Business Administration, Myongji University, Seoul, 03674, Korea.

* Corresponding author (gracehekim@mju.ac.kr)

[Received 24 August 2021, Reviewed 8 September 2021, Accepted 23 September 2021]

☆ 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2019R1G1A1084863)

면서 기업가치 연구 분야에서도 새로운 시각에서 연구가 진행되고 있다. 트위터의 메시지에 담긴 감성을 분석하여 주가지수 등락에 대한 설명력을 확인한 연구[3]를 필두로, 뉴스나 SNS의 비정형 데이터들을 수집하여 기업성과와의 상관관계를 분석하거나 기업의 성과를 예측하는 등의 연구들이 활발히 진행되었다[4]. 이러한 연구들은 기업 외부의 데이터를 활용하여 기존 기업 회계정보 위주의 데이터에 근거한 접근 방법이 갖는 예측력의 한계를 극복하는 결과를 제시하였다는 점에서 의미가 있다[3][4].

기업에 대한 직원들의 만족도나 충성도, 동기 등 기업 내부의 정보들은 여러 선행연구에서 기업의 성과에 영향을 미치는 변수로 다뤄져 왔다[5][6]. 하지만 이러한 연구들은 주로 설문을 기반으로 진행되어 표본의 대표성을 확보하기 어렵고 소수의 선행 요인에만 초점을 두고 있어 한계점을 갖는다. 기업의 성과에 영향을 미치는 요인은 다양하고 복잡적이므로 이들 간의 관계를 폭넓게 살펴볼 필요가 있다.

최근 온라인 리뷰가 사회 전반에 걸쳐 보편화되면서 기업리뷰를 제공하는 플랫폼들이 등장하게 되었다. 이러한 플랫폼에서는 과거에 수집하기 어려웠던 기업 내부의 정보를 손쉽게 대량으로 확보하는 것이 가능하여 다양한 연구의 좋은 자료로 활용되고 있다. 이에 본 연구에서는 그동안 잘 다뤄지지 못했던 직원의 만족도, 기업에 대한 직원들의 인식과 전망, 기업에 대한 장단점 평가와 같은 기업 내부의 데이터를 활용하여 기업가치를 예측할 수 있는 예측 모형을 제안하고자 한다.

이를 위해 국내 온라인 기업리뷰 사이트로부터 직원들의 의견을 수집하고 이를 기업 특성 변수와 통합하여 데이터셋을 구성한다. 그리고 해당 데이터에 다양한 머신러닝 기법을 적용하여 기업가치 예측 모형을 구축하고 성능평가를 통해 최적의 모형을 제안한다. 또한 변수 중요도 평가를 통하여 기업가치에 영향력을 갖는 변수를 도출하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 연구들을 소개하고, 3장에서는 본 연구의 접근방법에 대하여 설명한다. 4장에서는 연구 결과를 제시하며, 5장에서 결론과 함께 시사점을 전달한다.

2. 선행 연구

2.1 온라인 기업리뷰 정보와 기업가치

빅데이터 처리 기술이 발전함에 따라 온라인에 존재

하는 비정형 데이터들을 활용하여 기업가치를 예측하고자 하는 시도들이 늘어나고 있다. [7]의 연구에서는 소셜 미디어에 나타난 투자자 의견이 주가 수익률을 예측할 수 있다는 점을 밝혔다. [8]의 연구에서는 소셜 미디어의 고객 평점(web blogs and consumer ratings)에 내재된 정보가 주가치의 선행지표로 활용될 수 있다고 하였으며, [9]의 연구에서는 온라인 고객 의견이 금융시장에서 정보로서의 가치를 지닌다는 것을 입증하였다. 이러한 선행연구들은 집단지성(wisdom of crowd)의 활용이 가치 관련성이 높은 정보를 제공할 수 있다는 점을 강조하고 있으나 주로 온라인상에 존재하는 기업 외부 자료에 의존하고 있다는 단점이 있다.

한편 직원만족도와 같은 회사 내부의 의견을 이용하여 기업 성과 및 가치를 예측하는 연구들도 진행되고 있다. 최근 온라인 리뷰가 사회 전반에 걸쳐 보편화되면서 글래스도어(Glassdoor)와 같이 직원 리뷰를 제공하는 플랫폼들도 생겨나고 있다. 이러한 플랫폼을 활용하면 과거에 수집하기 어려웠던 기업 내부의 정보를 손쉽게 대량으로 확보하는 것이 가능해져 연구자료로서의 가치를 지닐 수 있다. [10]의 연구에서는 10만 건이 넘는 글래스도어 리뷰 정보를 활용하여 가족기업을 중심으로 직원만족도와 기업가치 간에 양(+)의 관계가 있다고 밝혔다. 또한, [11]의 연구에서는 직원만족도 평점이 높아진 기업이 평점이 낮아진 기업 대비 주식 수익률이 우수하다는 점을 보였으며, [12]의 연구에서는 기업의 미래전망에 대한 직원들의 인식이 기업의 미래 수익성을 예측할 수 있다고 하였다.

국내에서도 온라인 직원 리뷰 정보가 비재무정보의 하나로서 기업가치에 대해 더 높은 예측력을 가질 수 있다는 연구가 있다. [13]의 연구에서는 국내 온라인 기업리뷰 사이트인 잡플래닛(Jobplanet)의 데이터를 이용하여 직원만족도 평점이 기업가치(Tobin's Q)에 유의한 양(+)의 영향을 미치는 것을 확인하였다. 이러한 결과는 직원만족도와 기업가치 간의 내생성을 통제한 2SLS(two-stage least square) 분석에서도 강건하게 나타났다. 머신러닝을 활용한 [14]의 연구에서는 기업의 재무 변수에 잡플래닛 리뷰 텍스트의 장단점 텍스트 변수를 추가로 활용할 경우 기업의 재무 정보만을 학습한 모델 대비 개선된 추가 예측력을 보인다는 것을 확인하였다.

본 연구에서는 이상의 연구에서 개별적으로 다루었던 변수들을 포괄적으로 활용하여 기업가치 예측 모형을 제안하고자 한다. 기존 연구들이 주로 전통적인 통계 모형에 기반한 반면 본 연구에서는 머신러닝 방법론을 활용

하여 접근하고자 한다. 아울러 기업가치에 영향력을 갖는 변수들을 도출하여 제시하고자 한다.

2.2 머신러닝 기반의 기업가치 예측

전통적으로 주가 등 금융 데이터를 분석할 때는 가중 이동평균(Weighted Moving Average), 지수평활법(Exponential Smoothing), ARIMA(Autoregressive Integrated Moving Average), 칼만필터링(Kalman Filtering) 등 다양한 시계열 분석 모형이 활용되어 왔다[15]. 이러한 모형들은 통계적 추론 방법에 기반을 두고 변수 간 관계성을 발견하거나, 예측하고자 하는 변수의 자기 상관성을 검증하고 변수의 과거값을 활용하여 예측하게 된다.

머신러닝 방법론이 여러 학문 분야에서 우수한 예측력을 보임에 따라 재무 분야에서도 머신러닝 기법을 적용하여 기존 시계열 기반의 접근방법이 갖는 한계를 뛰어넘으려는 시도들이 있어왔다[16][17][18]. [16]의 연구에서는 K-NN(K-Nearest Neighbor) 알고리즘을 사용하여 삼성전자의 주가를 예측하는 모형을 제안하였다. 결과에 따르면 학습데이터의 양이 많을 때 모형의 예측력이 높아지는 것을 확인하였다. [17]의 연구에서는 인도 주가 지수를 예측하기 위하여 SVM(Support Vector Machine)과 K-NN을 융합한 하이브리드 모형을 제안하였으며 기존 모델 대비 개선된 성능을 보인 바 있다. [18]의 연구에서는 주가 예측을 주가의 방향성을 예측하는 분류 문제로 정의하여 예측 모형을 제안하였다. 그동안 주가 예측에 잘 사용되지 않던 랜덤포레스트(Random Forest)와 같은 트리 기반의 알고리즘들을 사용함으로써 비선형적인 주가 방향성 분류 문제에 대한 트리 기반 모형의 유용성을 보였다.

최근에는 비선형적이고 시간 의존적인 특성을 보이는 데이터에 대해 우수한 예측 성능을 보이는 딥러닝 기반의 알고리즘을 활용한 연구가 늘고 있다[19][20][21][22]. [19]의 연구에서는 10년치의 테헤란 주식시장의 주가 데이터를 활용하여 예측 모형을 개발하였다. 의사결정나무(Decision Tree), 배깅(bagging), 랜덤포레스트(Random Forest), 그래디언트 부스팅(Gradient Boosting Machine), XG부스트(eXtreme Gradient Boosting), LSTM(Long Short Term Memory)을 포함하여 총 9개 알고리즘을 사용하였는데 그중 LSTM 모형이 가장 정확도가 높았다. [20]의 연구에서는 특정 키워드의 인터넷 검색량을 투자자의 관심 척도로 가정하고 이를 기반으로 코스닥 상장 기업의 주가 변동성을 예측하였는데 LSTM과 벡터자기회귀(Vector

Autoregression) 모형을 활용하였다. [21]의 연구에서는 주가 예측에 전통적 시계열 기법인 ARIMA 모형과 LSTM 모형의 예측 성능을 비교하였는데 결과에 따르면 LSTM 모형이 우수한 것으로 나타났다. [22]의 연구에서는 LSTM 모형에 마르코프(Markov) 통계 모형을 접목하여 개선된 LSTM 모형을 제안하였다. 실험 결과 기존의 통계 모형 기반의 예측 정확도를 뛰어넘는 정확도를 보였다.

이상에서와 같이 재무분야에서 시계열 데이터 예측을 위한 다양한 머신러닝 연구가 진행되고 있다. 딥러닝을 활용한 최근의 연구에서는 LSTM이 많이 활용되고 있는데 타 모형 대비 LSTM 모형의 예측 성능이 우수한 이유로 해석된다. 본 연구에서는 머신러닝 및 딥러닝 알고리즘을 활용하여 기업리뷰를 활용한 기업가치 예측 모형을 제안하고자 한다.

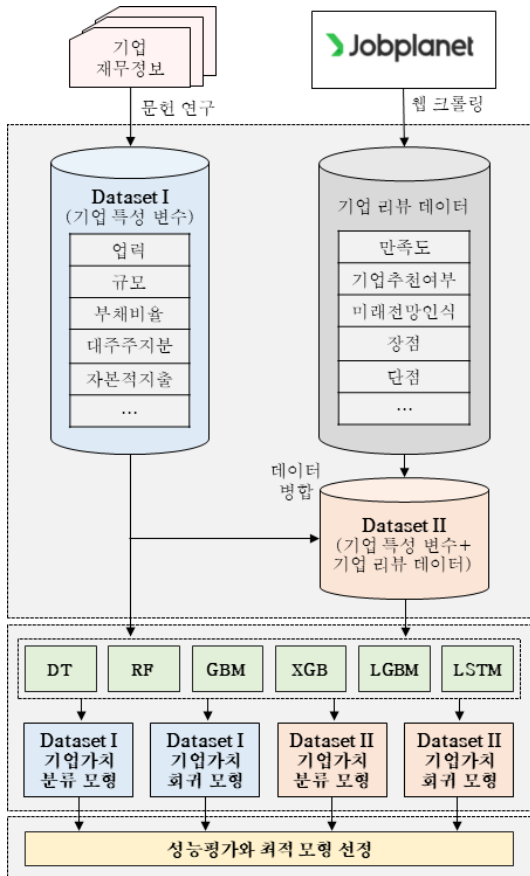
3. 연구 방법

3.1 연구 절차

본 연구의 개략적인 진행 절차는 다음과 같다. 우선 기업 재무정보를 수집하여 첫 번째 데이터셋을 구성한다. 그리고 국내 온라인 기업리뷰 사이트인 잡플래닛(<http://jobplanet.com>)에서 기업리뷰를 크롤링하고 이를 전처리한 뒤 기업 특성 변수와 통합하여 두 번째 데이터셋을 구성한다. 두 데이터셋에 다양한 머신러닝 알고리즘을 적용하여 기업가치 예측을 위한 분류 및 회귀 모형을 구축한다. 그리고 구축된 머신러닝 모형들의 성능을 평가하고 최적의 모형을 선정하는 순으로 진행한다.

본 연구에서 예측하려는 기업가치는 재무 분야 연구에서 가장 많이 사용되고 있는 Tobin's Q 값을 대용변수로 사용한다. Tobin's Q는 기업 자산의 장부가치 대 시장가치의 비율로서 값이 클수록 시장에서 해당 기업의 가치를 높게 평가하고 있다는 것을 의미한다. Tobin's Q는 연간 데이터를 기초로 하고 있어 기업 재무변수와 자료 갱신 주기를 일치시킬 수 있으며, 주가 수익률 변수보다 변동성이 작아 안정적인 예측이 가능하다.

그림 1은 본 연구의 실행 과정을 도식화하여 나타내고 있다. 이어지는 절에서 데이터 구성과 세부 연구 절차에 대해 살펴보도록 한다.



(그림 1) 연구흐름도

(Figure 1) Research Procedure

3.2 데이터 수집과 데이터셋 구성

본 연구에서 학습에 활용한 데이터는 두 종류이다. 첫 번째 데이터는 기업 재무정보 데이터로서 2015년부터 2019년까지 국내 상장 기업에 대하여 Dataguide DB로부터 수집하였으며 기업 규모, 업력, 부채비율, 대주주지분, 연구개발, 자본적 지출, 매출액 성장률 등 기업가치에 대한 영향력을 갖는 것으로 알려진 변수들로 구성하였다 [13]. 각 변수를 살펴보면, 규모는 기말 자산총액의 자연 로그 값, 업력은 상장일 이후 2019년 말까지 일 수의 자연 로그 값, 부채비율은 부채총액을 자산총액으로 나눈 값이다. 대주주지분은 최대 주주 1인 및 특수관계인의 지분율, 연구개발은 당기 R&D 지출여부, 자본적 지출은 당기 유형자산의 증분과 감가상각비의 합을 자산의 장부 가치로 나눈 값이다. 매출액 성장률은 전기 대비 당기의 매

출액 증가율이다.

두 번째 데이터는 잡플래닛으로부터 수집한 기업리뷰 데이터이다. 잡플래닛은 2014년 4월 개설된 온라인 기업리뷰 사이트로서 국내 기업에 대한 전·현직원의 리뷰를 수집·제공한다. 본 연구에서는 파이썬 기반의 웹크롤러를 구현하여 2014년 4월부터 2019년 12월까지 한국 유가증권시장에 상장된 440개 기업에 대한 리뷰 총 97,216건을 수집하여 활용하였다. 그림 2는 잡플래닛에 게시된 리뷰 예시를 보여준다. 그중 붉은 점선 안의 내용을 활용하여 변수를 추출하였다. 추출된 변수는 만족도 점수와 기업추천여부, 기업의 미래전망에 대한 인식, 장단점 비율이다. 그중 만족도의 경우 기업에 대한 전체만족도 및 5개 세부항목(승진 기회 및 가능성, 복지 및 급여, 업무와 삶의 균형, 사내 문화, 경영진)에 대한 만족도로 나뉘며 각 5점 만점의 점수로 표현된다. 기업추천여부는 해당 기업을 다른 사람에게 추천할 것인지 여부를 더미 변수로 활용하였다. 미래전망인식은 해당기업의 1년 뒤 성장 전망을 어떻게 보는지에 대한 내용으로서 성장, 비슷, 하락의 세 가지 값 중 하나를 갖는다. 장단점 비율은 장점 글자 수를 단점 글자 수로 나눈 값으로 계산하였다. 이상의 변수들을 기업 재무정보 데이터와 병합하여 두 번째 데이터셋으로 구성하였다.



(그림 2) 잡플래닛에 게시된 기업리뷰 예시

(Figure 2) Example of Firm Review posted on the JobPlanet

3.3 머신러닝 예측 모형 구축

본 연구에서 구축하고자 하는 기업가치 예측 모형은 회귀모형과 분류모형의 두 종류이다.

먼저 회귀모형에서는 과거 3년치의 데이터를 활용하여 기업가치값을 예측하는 모형을 만들었다. 이를 위해 3.2에서 준비한 두 가지 데이터셋 각각에 대하여 의사결정나무(DT), 랜덤포레스트(RF), 그라디언트 부스팅(GBM), XG부스트(XGB), Light GBM(LGBM), LSTM 등 총 여섯 가지 머신러닝 알고리즘을 적용하였다. 모형 구축에는 scikit-learn, lightgbm, xgboost 및 keras 라이브러리를 활용하였다. LSTM의 경우에는 데이터 전처리로서 모형의 안정성 확보와 성능 제고를 위하여 MinMaxScaler로 스케일링을 하였다. 학습시 데이터를 7:3의 비율로 훈련 데이터와 테스트 데이터를 무작위 분할하여 학습하였는데 이를 총 5회 반복하였다.

다음으로 분류모형에서는 과거 3년치의 데이터를 활용하는 것은 회귀모형과 같되 기업가치의 증감은 과거 3년치 평균 기업가치 수치 대비 기업가치의 증감여부를 예측하는 모형을 만들었다. 기업가치가 증가한 기업 수와 감소한 기업 비율이 각각 69.3%, 30.7%로 클래스 불균형 문제가 있어 SMOTE(Synthetic Minority Oversampling Technique) 기법을 적용하여 이를 해소하였다. 학습시 훈련 데이터와 테스트 데이터의 분할 비율과 반복 학습은 회귀모형 학습시와 동일하게 진행하였으며 머신러닝 알고리즘도 회귀모형과 동일한 여섯 가지의 알고리즘을 적용하여 모형을 구축하였다.

4. 연구 결과

4.1 성능 평가

회귀모형과 분류모형 모두 각각 두 데이터셋을 활용하여 데이터셋마다 여섯가지 알고리즘을 사용하여 모형을 구축하였다. 결과적으로 회귀모형과 분류모형 각각 열두 개의 모형이 만들어졌으며 이들에 대한 성능 평가를 각각 진행하였다.

먼저 회귀모형에 대한 평가는 MAE(Mean Absolute Error)와 RMSE(Root Mean Square Error)를 기준으로 평가하였다. 표 1에서 보는 바와 같이 기업 특성 변수만으로 구성된 첫 번째 데이터셋을 이용한 모형 대비 기업리뷰 데이터가 포함된 두 번째 데이터셋을 이용한 예측 모형이 더 우수한 성능을 나타냈다. 모형 간 성능을 비교

해보면 Dataset II를 이용한 LSTM 기반 모형의 경우 MAE가 0.3587, RMSE가 0.4654로 다른 알고리즘 대비 가장 성능이 우수한 것을 확인할 수 있다. MAE 값을 기준으로 결과를 해석해보면 본 연구에서 제안한 모형을 통해 평균 0.3587의 오차를 갖는 기업가치 예측치를 얻을 수 있다는 의미가 된다.

(표 1) 회귀모형 성능 비교

(Table 1) Performance Comparison among Regression Predictive Models

모형	Dataset I		Dataset II	
	MAE	RMSE	MAE	RMSE
DT	1.7591	1.3262	1.0375	1.0134
RF	0.4471	0.9878	0.3945	0.7102
GBM	0.5008	0.8802	0.4490	0.7739
XGB	0.4141	0.8256	0.4054	0.6702
LGBM	0.3993	0.7624	0.3760	0.6603
LSTM	0.3882	0.6545	0.3587	0.4654

분류모형에서는 정확도 및 정밀도, 재현율, F1점수를 이용하여 모형의 성능을 비교하였다. 표 2에서 보는 바와 같이 분류모형에서도 기업리뷰 데이터를 함께 활용한 모형(Dataset II)에서 상대적으로 우수한 정확도를 보였다. 그리고 제안한 모형 중 LSTM 기반 모형의 성능이 정확도와 F1점수를 근거로 가장 좋은 결과를 나타냈다. 이 모형을 활용하면 73.2%의 정확도로 기업가치의 증감 여부를 예측할 수 있는 것으로 해석할 수 있다.

(표 2) 분류모형 예측 정확도 비교

(Table 2) Performance Comparison among Classification Predictive Models

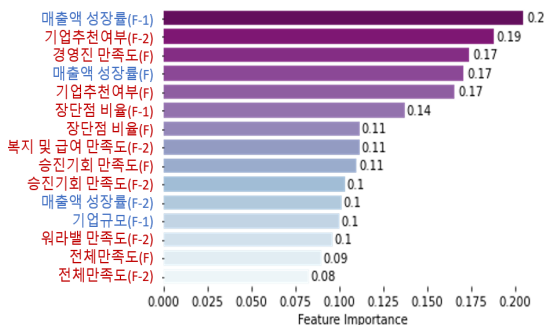
모형	Dataset I				Dataset II			
	정확도	정밀도	재현율	F1점수	정확도	정밀도	재현율	F1점수
DT	60.0	62.2	49.4	55.0	61.4	64.1	51.8	57.3
RF	58.0	59.0	52.1	55.4	57.5	59.8	45.5	51.7
GBM	63.5	66.0	55.4	60.2	70.7	74.8	62.5	68.0
XGB	66.1	69.9	56.7	62.6	68.5	72.7	59.2	65.3
LGBM	58.4	61.8	44.1	51.5	61.3	68.4	42.3	52.3
LSTM	55.1	70.7	61.0	65.5	73.2	75.8	85.1	80.2

4.2 변수중요도 추출

구축한 예측 모형에서 입력 변수들의 설명력이 얼마나 되는지 확인하기 위하여 변수중요도를 추출하였다. 이를 위해 본 연구에서는 순열(permutation) 방법을 활용하였다. 순열 방법에서는 각 변수의 값을 무작위로 하나

씩 치환하면서 순열을 사용하고 해당 순열이 모형의 정확도에 얼마나 영향을 미치는지 측정한다. 즉, 학습된 모형이 특정 변수 없이 기본 모형에 대비하여 얼마나 성능 차이가 발생하는지 그 차이의 절대값의 평균으로 계산하는 방법이다[23].

본 연구에서 구축한 모형 중 가장 우수한 성능을 보였던 LSTM 분류 모형에 대한 변수중요도를 계산하였다. 결과는 그림 3에서 보는 바와 같다. 참고로 과거 3년치의 데이터를 사용하였기 때문에 동일 변수에 대해서도 F년도, F-1년도, F-2년도의 세 변수값이 존재한다. 변수중요도 값 기준 상위 15개의 변수를 추출하였는데 푸른색 변수는 기업 특성 변수에 속하는 변수이고, 붉은색 변수는 기업리뷰 데이터에서 추출된 변수이다. 기업 특성 변수 중에는 매출액 성장률이 절대적으로 중요도가 높은 변수로 식별된 반면, 기업리뷰 변수 중에는 기업추천여부, 만족도 변수들과 장단점 비율 등 기업리뷰에 속한 변수들 전반적으로 중요도가 높은 것을 확인할 수 있었다.



(그림 3) LSTM 분류모형의 변수중요도

(Figure 3) Feature Importance of LSTM Classification Predictive Model

5. 시사점 및 결론

본 연구에서는 국내 상장 기업의 기업가치를 예측하기 위한 머신러닝 기반의 예측 모형을 제안하였다. 이를 위해 10만여 건에 달하는 기업리뷰를 수집한 뒤 머신러닝 알고리즘을 적용하고 성능평가를 통하여 최적의 예측 모형을 제시하였다. 연구 결과, LSTM 기반의 모형이 기업가치에 대해 가장 우수한 예측력을 가지고 있다는 것을 확인하였다. 또한 본 연구 결과를 통해 기업 추천여부, 만족도, 장단점 평가 등 기업 내부 정보가 기업가치 예측에 대해 유의미한 영향력을 갖고 있음을 확인할 수 있었다.

본 연구는 실무적 활용이 가능한 기업가치 예측 모형을 제안하였다는 점에서 의미를 갖는다. 또한 직원들의 기업추천 여부와 같이 선행연구에서 다루지 않았던 새로운 변수들을 식별하여 제시한 것도 본 연구의 기여점으로 판단된다. 본 연구에서 제시한 결과는 국내 기업들을 대상으로 하는 기업가치 연구 분야에 유용한 사례가 될 것으로 기대되며 본 연구에서 제시한 방법론을 활용하면 기업의 재무 정보와 기업리뷰를 기업가치 예측을 위한 유용한 데이터로 활용할 수 있을 것이다.

본 연구에서는 제한된 기간의 데이터를 활용하였지만, 추후 데이터가 더 오랜 기간 축적되어 보다 긴 기간의 데이터를 활용할 수 있게 된다면 데이터의 크기에 따른 모형의 예측력을 검증하고 기업가치 예측을 위한 최적의 데이터 크기를 검증할 수 있을 것이다. 이는 거시적인 데이터의 필요성에 대한 논의로 이어질 수 있을 것이다. 또한 향후 다양한 텍스트 분석 기법을 적용하여 기업가치에 영향력을 갖는 변수들을 추가적으로 추출한다면 보다 높은 정확도를 갖는 예측 모형을 개발할 수 있을 것이다.

참고문헌(Reference)

- [1] J. A. Ohlson, "Earning, Book Values and Dividend in Equity Valuation: An Empirical Perspective," Contemporary Accounting Research, Vol. 8, No. 1 pp. 107-120, 2001.
<http://dx.doi.org/10.1506/7tpj-rxqn-tqc7-ffae>
- [2] S. Penman, "Combining Earnings and Book Value in Equity Valuation," Contemporary Accounting Research, Vol. 15, No. 3, pp. 291-324, 1998.
<http://dx.doi.org/10.1111/j.1911-3846.1998.tb00562.x>
- [3] J. Bollen, H. Mao and X. Zeng, "Twitter Mood Predicts the Stock Market," Journal of Computational Science, Vol. 2, No. 1, pp. 1-8, 2011.
<http://dx.doi.org/10.1016/j.jocs.2010.12.007>
- [4] A. K. Nassirtoussi, S. Aghabozorgi, T. Y. Wah and D. C. L. Ngo, "Text Mining for Market Prediction: A Systematic Review," Expert Systems with Applications, Vol. 41, No. 16, pp. 7653-7670, 2014.
<https://doi.org/10.1016/j.eswa.2014.06.009>
- [5] A. Edmans, "Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices," Journal of Financial Economics, Vol. 101 No. 3 pp.

- 621-640, 2011.
<https://doi.org/10.1016/j.jfineco.2011.03.021>
- [6] O. Guillon and C. Cezanne, "Employee Loyalty and Organizational Performance: A Critical Survey," *Journal of Organizational Change Management*, Vol. 27, No. 5, pp. 839-850, 2014.
<https://doi.org/10.1108/jocm-02-2014-0025>
- [7] H. Chen, P. De, Y. J. Hu, and B. H. Hwang, "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media," *The Review of Financial Studies*, Vol. 27, No. 5, pp. 1367-1403, 2014.
<https://doi.org/10.1093/rfs/hhu001>
- [8] X. Luo, J. Zhang, and W. Duan, "Social Media and Firm Equity Value," *Information Systems Research*, Vol. 24, No. 1, pp. 146-163, 2013.
<https://doi.org/10.1287/isre.1120.0462>
- [9] J. Huang, "The Customer Knows Best: The Investment Value of Consumer Opinions," *Journal of Financial Economics*, Vol. 128, No. 1, pp. 164-182, 2018
<https://doi.org/10.1016/j.jfineco.2018.02.001>
- [10] M. Huang, P. Li, F. Meschke, and J. P. Guthrie, "Family Firms, Employee Satisfaction, and Corporate Performance," *Journal of Corporate Finance*, Vol. 34, pp. 108-127, 2015.
<https://doi.org/10.1016/j.jcorpfin.2015.08.002>
- [11] T. C. Green, R. Huang, Q. and Wen, D. Zhou, "Crowdsourced Employer Reviews and Stock Returns," *Journal of Financial Economics*, Vol. 13, No. 1, pp. 236-251, 2019.
<https://doi.org/10.1016/j.jfineco.2019.03.012>
- [12] K. Huang, M. Li, and S. Markov, "What Do Employees Know? Evidence from a Social Media Platform," *The Accounting Review*, Vol. 95, No. 2, pp. 199-226, 2020.
<https://doi.org/10.2308/accr-52519>
- [13] H. E. Kim and H. Lee, "Online Employee Reviews and Firm Value," *Management & Economic Research Institute*, Vol. 43, No. 2, pp. 27-52, 2021.
<http://doi.org/10.22828/meri.2021.43.2.002>
- [14] Y. Lim and H. Lim, "A Comparative Analysis of the Prediction Models for the Direction of Stock Price Using the Online Company Reviews," *Journal of the Korea Convergence Society*, Vol. 11, No. 8, pp. 165-171, 2020.
<https://doi.org/10.15207/JKCS.2020.11.8.165>
- [15] G. E. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, "Time Series Analysis: Forecasting and Control," John Wiley & Sons, 2015.
<https://doi.org/10.1002/9781118619193>
- [16] S. H. Chun, "The Effect of Data Size on the k-NN Predictability: Application to Samsung Electronics Stock Market Prediction," *Journal of Intelligence and Information Systems* Vol. 25, No. 3, pp. 239-251, 2019.
<http://dx.doi.org/10.2307/2284112>
- [17] R. K. Nayak, D. Mishra, and A. K. Rath, "A Naïve SVM-KNN based Stock Market Trend Reversal Analysis for Indian Benchmark Indices," *Applied Soft Computing*, Vol. 35, pp. 670-680, 2015.
<https://doi.org/10.1016/j.asoc.2015.06.040>
- [18] S. Basak, S. Kar, S. Saha, L. Khaidem and S. R. Dey, "Predicting the Direction of Stock Market Prices using Tree-based Classifiers," *The North American Journal of Economics and Finance*, Vol. 47, pp. 552-567, 2019.
<https://doi.org/10.1016/j.najef.2018.06.013>
- [19] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi and E. Salwana, "Deep Learning for Stock Market Prediction," *Entropy*, Vol. 22, No. 8, 840, 2020.
<https://doi.org/10.3390/e22080840>
- [20] Y. Cho, K. Sohn and O. Kwon, "Comparison of Models for Stock Price Prediction Based on Keyword Search Volume According to the Social Acceptance of Artificial Intelligence," *Journal of Intelligence and Information Systems*, Vol. 27, No. 1, pp. 103-128, 2021.
<https://doi.org/10.13088/jiis.2021.27.1.103>
- [21] S. Siami-Namini, N. Tavakoli and A. S. Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series", 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1394-1401, 2018.
<https://doi.org/10.1109/ICMLA.2018.00227>
- [22] Rundo, F., "Deep LSTM with Reinforcement Learning Layer for Financial Trend Prediction in FX High Frequency Trading Systems," *Applied Sciences*, Vol. 9, No. 20, 4460, 2019.
- [23] C. Strobl, A. L. Boulesteix, T. Kneib, T. Augustin and A. Zeileis, "Conditional Variable Importance for

Random Forests,” BMC bioinformatics, Vol. 9, No.

1, pp. 1-11, 2008.

<https://doi.org/10.1186/1471-2105-9-307>

● 저 자 소 개 ●



이 한 준(Hanjun Lee)

2001년 서울대학교 컴퓨터공학과(공학사)

2004년 서울대학교 대학원 컴퓨터공학과(공학석사)

2016년 고려대학교 경영학과(경영학 박사)

2007년~2018년 한국국방연구원 국방획득연구센터 선임연구원

2018년~2020년 한남대학교 글로벌IT경영학과 조교수

2020년~현재 명지대학교 경영정보학과 조교수

관심분야 : 텍스트 마이닝, 빅데이터, 소셜미디어, 정보화 정책, etc.

E-mail : hjlee1609@gmail.com



신 동 원(Dongwon Shin)

2021년 명지대학교 경영정보학과(경영학사)

2021년~현재 명지대학교 산학협력단 연구원

관심분야 : 데이터 분석, 텍스트 마이닝, 자연어 처리, 빅데이터

E-mail : petter1286@gmail.com



김 희 은(Hee-Eun Kim)

2007년 이화여자대학교 중어중문학(학사)

2008년 성균관대학교 중국대학원 중국경제관리학(석사)

2019년 Peking University 광화관리학원 재무학(박사)

2009년~2012년 미래에셋대우 리서치센터 애널리스트

2012년~2015년 한국금융연구원 연구원

2020년~현재 명지대학교 경영학과 조교수

관심분야 : 빅데이터, ESG 투자, 중국 자본시장

E-mail : gracehekim@mju.ac.kr