

기업 리뷰 정보를 활용한 주가 방향 예측 모델 비교 분석

임용택¹, 임희석^{2*}

¹고려대학교 빅데이터융합학과 석사과정, ²고려대학교 컴퓨터학과 교수

A Comparative Analysis of the Prediction Models for the Direction of Stock Price Using the Online Company Reviews

Yongtaek Lim¹, Heuseok Lim^{2*}

¹Master's Course, Department of Bigdata Convergence, Korea University

²Professor, Department of Computer Science and Engineering, Korea University

요약 텍스트 마이닝을 활용한 주가 방향 예측 연구에서는 대부분 뉴스, SNS 데이터를 사용하고 있다. 하지만 뉴스, SNS 데이터로부터 기업에 대한 솔직하고 생생한 정보는 얻기 어렵다는 약점이 존재한다. 본 논문에서는 실제 근무 경험 이 있는 내부 직원의 기업 리뷰를 반영하여, 종업원 만족도를 활용한 주가의 방향성을 예측하는 문제를 다룬다. 머신러닝 모델별 성능평가를 통해 예측 정확도를 비교, 분석한 결과 종업원의 기업 리뷰 데이터를 추가로 이용한 주가 방향 예측 모델은 그렇지 않은 모델 대비 뛰어난 분류 성과를 보였다. 본 연구는 금융 공학에 자연어처리기술을 활용한 융합 연구로서 주가 예측 분야에서 종업원 만족도를 활용한 기준에 없던 새로운 방법론을 추구하였다. 실무적으로 주가 방향 예측 분야에 유용한 정보를 제공할 것으로 기대된다.

주제어 : 주가 예측, 종업원 만족도, 기업 리뷰, 텍스트 마이닝, 기계 학습

Abstract Most of the stock price prediction research using text mining uses news and SNS data. However, there is a weakness that it is difficult to get honest and vivid information about companies from them. This paper deals with the problem of the prediction for the direction of stock price by doing text mining the online company reviews of internal staff indicating employee satisfaction. The comparative analysis of the prediction models for the direction of stock price showed the prediction model, which adds internal employee reviews, has better performance than those that did not. This paper presents the convergence study using natural language processing in financial engineering. In the field of stock price prediction, This paper pursued a new methodology that used employee satisfaction. In practice, it is expected to provide useful information in the field of forecasting stock price direction.

Key Words : Stock prediction, Employee satisfaction, Company reviews, Textmining, Machine learning

1. 서론

주가 움직임을 예측하는 문제는 오랜 기간 중요한 연구과제로 인식되고 있다. 금융공학 측면에서 데이터 마이닝(data mining) 기술은 금융투자산업에 활발히 적용되

고 있으며, 통계학적 예측 방법은 실무에서 널리 사용되고 있다[1-3]. 머신러닝을 활용한 재무 정보 분석 및 예측 방법은 주가 예측 분야의 발달에 큰 공헌을 하였다. 최근에는 수치형 데이터 분석 뿐만 아니라 뉴스, SNS에서 추출한 정성적 데이터를 가공해 주식투자 정보를 제

*Corresponding Author : Heuseok Lim(limhseok@korea.ac.kr)

공하는 연구 또한 활발히 진행되고 있다[4-6]. 기존 온라인상의 텍스트 데이터를 활용한 주가 예측 관련 연구들을 통해 텍스트 마이닝을 활용한 주가 방향 예측 방안이 유용함을 확인할 수 있다.

그러나 뉴스, SNS 데이터는 기업의 주가 가치 뿐만 아니라 전반적인 주제를 다룬다는 구조적인 한계로 인해 뉴스, SNS 데이터를 통해 실제적인 주가 관련 정보만을 마이닝 하는데 한계를 가지고 있다. 또한 뉴스에서 얻은 기업 정보는 실제 그 기업에서 근무한 경험이 있는 직원들이 작성한 정보가 아니기 때문에 기업에 대한 솔직하고 생생한 정보는 얻기 어렵다는 약점도 존재한다. 이 논문에서는 그러한 약점들을 극복하기 위하여 실제 근무 경험이 있는 내부 직원의 기업 리뷰를 반영하여 주가의 방향성을 예측하는 문제에 대해서 살펴보고자 한다.

본 연구는 기존 연구와 다음과 같은 차별성을 가진다. 첫째로 기존 텍스트 마이닝을 활용한 주가 방향 예측연구가 뉴스, SNS의 텍스트 정보를 활용했다면 본 연구는 종업원 만족도라 할 수 있는 내부 직원의 기업 리뷰를 학습 데이터로 사용했다는 점이다. Lakin(2015)의 연구에 따르면 사람들은 기업에 대한 정보를 얻을 때 뉴스나 기업 가이드라인과 같은 전통적인 정보보다 실제 그 기업에서 근무한 경험이 있는 직원들이 웹서비스를 통해 작성한 온라인 기업 리뷰를 더 신뢰함을 밝혔다[12]. 한국에서 이러한 웹서비스를 제공하는 플랫폼으로 잡플래닛이라는 웹 서비스가 있으며, 잡플래닛에서는 내부 직원의 기업 리뷰, 면접 리뷰 등을 제공하고 있다.

잡플래닛에서 제공하는 온라인의 기업 리뷰의 장점은 정보 제공자의 익명성이 보장되기 때문에 직원들이 재직하고 있는 기업에 대해서 실제 경험했던 솔직한 체감사례와 정보가 제공될 가능성이 높다는 것이다. 위와 같이 제공된 기업 리뷰는 정성적인 직원의 만족도 평가 결과로 볼 수 있으며 그 결과는 기업의 생산성과 효율성을 측정할 수 있는 중요한 척도가 될 수 있다. 따라서 이러한 기업 리뷰의 특징은 종업원 만족도를 나타내는 데이터로서 충분히 가치 있음을 시사하며, 기업 리뷰를 이용하면 기업의 재무성과, 즉 주가 방향성을 예측할 수 있는 가능성을 보여준다. 이번 연구에서는 잡플래닛에서 제공하는 내부 직원의 74,291개 온라인 기업 리뷰 데이터를 웹크롤링을 통해 수집하고 본 연구에 활용함으로써 기존 주가 방향 예측 연구에서 시도하지 않았던 정성적인 종업원 만족도 데이터의 활용을 극대화 하였다.

둘째로 기존의 연구에서는 종업원 만족도가 기업의 주식가치에 직접적인 영향력이 있음을 보여줬다는 점에서

의의를 찾아볼 수 있으나, 실제 모형을 통해 주가 방향성 예측력을 비교, 분석하려는 시도가 없었다. 이번 연구에서는 비정형 데이터인 온라인 기업 리뷰를 텍스트 마이닝을 통해 스코어링 하여 머신러닝 기법을 통해 수치적 결과값을 도출하고 모델간 주가 방향 예측력을 비교, 분석 하였다. 기존의 뉴스 데이터를 활용한 주가 방향 예측 모형[4]의 경우 평균적으로 약 56%의 예측률을 보였으나, 본 연구에서 사용한 내부 직원의 기업 리뷰를 사용한 모델은 최대 78.8%의 예측 정확도를 보이며 기존 연구와 차별성을 가진다.

본 논문의 구성은 다음과 같다. 2장에서는 이 논문과 관련된 논문들에 대해 설명한 뒤 3장에서는 본 연구의 가설과 연구모형에 대하여 상세히 설명할 것이다. 4장에서는 모델별 성능을 비교, 분석하고 연구 결과를 해석한다. 마지막 5장에서 이 논문의 결과를 정리하고 연구가 가지는 의의와 향후 미래연구 방향에 대한 설명과 함께 이 논문을 마친다.

2. 이론적 배경

2.1 텍스트 마이닝을 이용한 주가 예측 연구

주가 예측에 관한 연구는 전통적으로 재무적 데이터를 활용한 통계 기반 예측 접근법이 대부분을 이루고 있지만[1-3], 최근 자연어처리 기법이 발달함에 따라 비정형 텍스트 데이터로부터 주식시장에 영향을 미치는 다양한 지표들을 분석할 수 있게 되었다.

텍스트 마이닝을 이용한 주가 예측 연구들은 주로 뉴스 데이터를 기반으로 이루어졌다.

뉴스 감성분석을 통한 기업별 주가 방향 예측 모형[4]에서는 온라인 뉴스에서 '명사'를 추출한 뒤 단어의 긍정, 부정 점수를 부여한 후 기업별 일별 뉴스에 대해 스코어링 하고 실제 해당일의 주가 등락을 예측하였다. 그러나 수집된 뉴스 데이터에 해당 기업을 단순 언급하는 뉴스이거나 주가 방향 예측에 필요한 경제적 의미가 없는 뉴스가 포함됨으로써 예측 정확도가 다소 낮을 수 있는 한계점이 있었다.

뉴스와 SNS 데이터를 이용한 주가예측 모형 비교 연구[5]에서는 텍스트 마이닝을 통해 형태소의 극성을 분류하고 감성분석과 머신러닝 기법을 사용한 주가 예측 모형을 제시하였고, 주가지수 방향성 예측을 위한 도메인 맞춤형 감성사전 구축방안[6]에서는 말뭉치를 기반으로 주식시장에 특화된 감성사전을 구축을 진행하였다. 그러

나 기존 연구들은 범용적 뉴스 데이터에만 의존하는 점, 그리고 기업의 주가와 관련없는 불필요한 어휘들이 분석 데이터에 포함되는 한계점이 있었다.

2.2 종업원 만족도와 주가 관계 연구

많은 기업들은 기업의 수익성과 재무성과를 달성하기 위하여 고객만족도에 많은 관심을 가지며[7], 종업원 만족도는 기업들이 고객만족도를 높이기 위해 집중하는 지표이다[8].

근무여건이 직·간접적으로 주식가치에 미치는 영향에 대한 연구[9]에서는 종업원의 근무여건으로 인한 만족도의 차이에 따라 주가 관련성에 차이가 있음을 검증하였다. Best, R. J.(2008)는 직원 만족도가 높을수록 기업의 생산성이 높아지는 것을 검증하였고[10], Edmans, A(2011)는 직원 만족도가 주주 수익률과 양의 상관관계가 있다고 보고하였다[11].

앞선 연구들은 종업원 만족도와 주가의 관련성을 밝혔는데 시사점을 가지나, 종업원 만족도를 분석 데이터로 활용하여 실제 모형을 통해 주가 방향 예측력을 비교, 분석하지 못한 한계점을 가진다.

3. 연구모형 및 가설

본 논문에서 제안하는 연구 프로세스는 크게 머신러닝을 학습시키는 단계와 테스트 데이터를 통해 주가 방향을 예측하는 단계로 나뉜다. Fig. 1은 주가 방향 예측 연구 프로세스를 나타낸다.

본 연구의 가설은 다음과 같다.

가설: 내부 직원의 기업 리뷰를 활용한 주가 예측 모델은 재무 정보만 이용한 모델보다 우수한 예측력을 가질 것이다.

가설검증을 위해 아래와 같이 연구를 설계하였다.

3.1 데이터 수집

2020.2.12 기준 유가증권 상장법인으로 등록된 기업을 대상으로 웹 크롤링(web crawling)을 진행하였다. 텍스트 데이터인 기업 리뷰는 잡플래닛[13]에서 472개 기업의 74,291개 연도별 기업 리뷰를 수집하였다. Table 1은 웹 크롤링을 통해 얻은 '기업 리뷰' 데이터의 구조를 보여주고 있다.

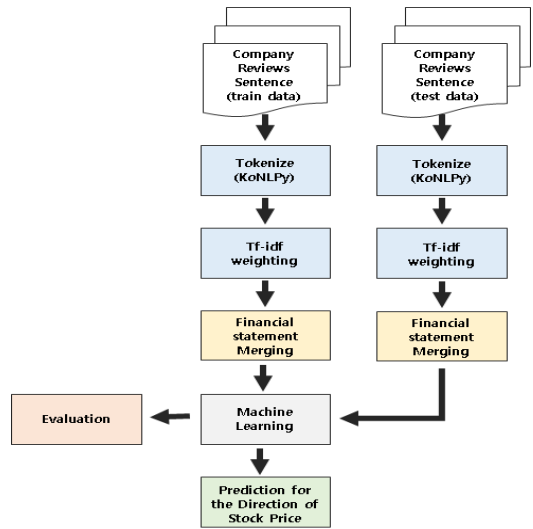


Fig. 1. Overall Research Process

Table 1. Examples of Company Review Web crawling "Samsung Electronics"

	corp	code	date	company review
2718	Samsung Elec.	30139	2020.02.16	"I think I have less money compared to the amount of work I do."
2719	Samsung Elec.	30139	2020.02.15	"Our company's welfare is excellent and well-organized."

수치형 데이터인 기업별 재무 정보는 네이버 금융[14]에서 기말기준 2016년부터 2019년까지 4개년도를 대상으로 776개 기업의 재무 정보를 수집하였다. 다양한 재무 정보 중 주식 가치를 나타내는 지표인 ROE(자기자본이익률), EPS(주당순이익), PER(주가수익비율), BPS(주당순자산가치), PBR(주가순자산비율) 등 5개 재무 정보를 사용하였다. 연도별 기업 주가는 2015년부터 2019년까지 798개 기업의 기말기준 주가를 수집하였다. 기업 주가는 주가 방향 예측의 타겟 레이블(target label)로 사용하기 위해 전년 대비 해당년도 주가 방향으로 전처리하여 사용하였다.

데이터셋은 재무 정보 특성 5개, 기업 리뷰 TF-IDF 키워드 500개로 총 505개 특성으로 구성하였고, 타겟 레이블은 전년 대비 해당년도 주가 방향(상승방향 +1, 하락방향 0)을 사용하였다.

3.2 텍스트 마이닝

머신러닝 모델을 학습시키기 위해 텍스트의 전처리 단계가 요구된다.

본 연구의 텍스트 마이닝은 KoNLPy[15]의 형태소 분석기 기반 tokenizer와 Scikit-learn의 TfidfVectorizer를 사용하여 웹 크롤링한 기업 리뷰를 TF-IDF 형식으로 변환하는 방법을 제안한다. 입력 데이터가 온라인 기업 리뷰라는 도메인 특성을 고려하여 인터넷 텍스트에 특화된 Twitter 형태소 분석기를 사용하였다. 74,291개 문장에서 2글자 이상으로 구성된 526,403개의 유니크한 단어를 추출하였으며 전체 문서에서 빈도순으로 최상위 500개의 단어를 TF-IDF 입력 데이터로 추출하였다. Table 2는 TF-IDF 입력 데이터의 예시를 보여주고 있다.

Table 2. Examples of TF-IDF “Samsung Electronics”

Year	Word1	Word2	...	Word499	Word500
2017	0.01648	0.00987	...	0.02197	0.06801
2018	0.03869	0.00543	...	0.06391	0.02105

3.3 머신러닝 모델

전처리를 마친 기업 리뷰 텍스트 데이터와 기업 재무 정보를 활용하여 다양한 머신러닝 기법을 통해 주가 방향 예측을 실행한다. 본 연구에 사용한 머신러닝 기법은 의사결정나무(DT), 랜덤포레스트(RF), 서포트 벡터 머신(SVM), 그래디언트 부스팅 머신(GBM), XGBoost, LightGBM 등 6개 머신러닝 모델을 사용하였다.

3.4 실험

이 논문에서는 Fig. 2 연구모형과 같이 재무 정보만을 학습시켰을 경우와 재무 정보 외 기업 리뷰 텍스트 정보를 추가로 학습 시켰을 경우의 예측 정확도를 비교 분석하기 위해 3가지 실험이 진행된다.

먼저 단일 모델로 하이퍼패러미터 튜닝을 진행한 후 모델별 예측 성능을 비교 분석한다. 하이퍼패러미터 튜닝 방법은 Coarse & Finer Search 방법을 사용하였다.

Coarse Search에서 좋지 않은 하이퍼패러미터를 필터링(filtering)하고 일정기준 이상 성능이 나온 하이퍼패러미터 범위를 가지고 Finer Search를 진행하였다. 1차 성능 개선을 마친 뒤에는 특성 선택(Feature Selection)을 통해 성능 개선 후 모델별 예측력을 비교, 분석 한다. 특성 선택을 통해 머신러닝 모델 성능에 더 많은 영향을

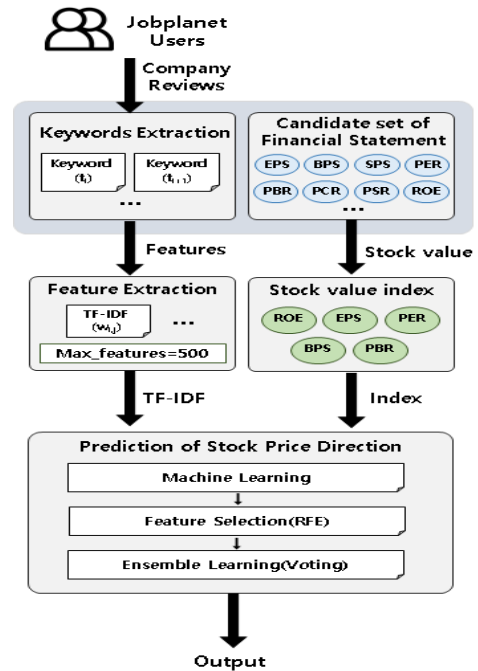


Fig. 2. Prediction Research Model

주는 특성을 우선적으로 선택하여 노이즈를 제거하고 모델 성능을 향상시킬 수 있다. 본 연구에서는 트리형 모델에서 제공하는 특성 중요도의 단점을 보완하기 위해 순열 특성 중요도(Permutation Importance)를 사용하였다. Table 3은 랜덤포레스트로부터 얻은 순열 특성 중요도의 일부를 보여준다.

Table 3. Permutation Importance of RandomForest

	Weight	Feature
1	0.0579±0.0289	pbr
2	0.0173±0.0082	eps
3	0.0117±0.0089	per
4	0.0111±0.0070	word159
5	0.0095±0.0045	word14
6	0.0095±0.0067	word167

순열 특성 중요도를 이용하여 순열 특성 중요도가 높은 상위 90% 특성을 대상으로 rfe(recurrent feature elimination)로 머신러닝 모델을 반복, 학습시키고 최적의 특성 선택을 통해 예측 정확도를 측정하였다. 랜덤포레스트 알고리즘에 rfe를 정의한 코드는 다음과 같다.

```
def rf_rfe(x_data,X_train,y_train,X_test,y_test,method,ratio=0.9,min_f
eats=40):
    feats = x_data.columns.tolist()
    archive=pd.DataFrame(columns=['model', 'n_feats', 'feats', 'rm
se', 'accuracy'])
    while True:
        model=RandomForestClassifier()
        model.fit(X_train[feats], y_train)
        pred = model.predict(X_test[feats])
        rmse = rmse(pred, y_test)
        accuracy = accuracy_score(y_test,pred)
        n_feats = len(feats)
        print(n_feats, rmse, accuracy)
        archive=archive.append({'model':model,'n_feats':n_feats,'f
eats':feats,'rmse':rmse,'accuracy':accuracy},ignore_index=
True)
        if method == 'perm':
            perm=PermutationImportance(model,random_state=0).f
it(X_test[feats], y_test)
            feat_imp=pd.Series(perm.feature_importances_,index=f
eats).sort_values(ascending=False)
            next_n_feats = int(n_feats * ratio)
            if next_n_feats < min_feats:
                break
            else:
                feats=feat_imp.iloc[:next_n_feats].index.tolist()

        archive_rf=pd.DataFrame(archive)
        archive_rf=archive_rf.sort_values(by="rmse")
        filename=f"randomforest_{method}.csv"
        archive_rf.to_csv(filename,encoding='utf-8-sig')

    return archive
```

마지막으로 모델들의 결합을 통한 앙상블 러닝(ensemble learning)을 통해 모델 성능을 개선하고 예측력을 평가한다. 본 연구에서는 앙상블 학습의 방법 중 투표 기반 분류기(Voting Classifier)를 사용하였다. 투표 기반 분류기는 2개 이상의 머신러닝 모델을 학습시킨 후, 각 모델의 예측을 모아서 다수결 투표를 통해 가장 많이 선택된 클래스를 예측(hard voting)하는 결과를 도출한다.

4. 연구결과

TF-IDF 방식으로 벡터화한 기업 리뷰 텍스트 데이터와 기업별 재무 정보를 이용하여 주가 방향을 예측, 분류하였다. 모델 성능 평가 척도는 정확도(accuracy)와 rmse(root mean squared error)를 사용하였다.

Table 4는 Coarse & Finer Search 방식의 하이퍼패러미터 튜닝을 통해 1차 성능 개선을 진행한 후 단일

모델에 대해서 재무 정보만을 학습시켰을 경우(Financial statements)와 재무 정보 외 기업 리뷰 텍스트 정보를 추가로 학습 시켰을 경우(+Company Reviews)의 예측 정확도를 비교 분석한 표이다.

Table 4. Predicting by Machine Learning Model

	Financial statements		+Company Reviews	
	accuracy	rmse	accuracy	rmse
DT	0.5933	0.6377	0.6128	0.6222
RF	0.6490	0.5924	0.6992	0.5485
SVM	0.6184	0.6178	0.6240	0.6132
GBM	0.6574	0.5853	0.6630	0.5806
XG Boost	0.6323	0.6064	0.6797	0.5660
Light GBM	0.6240	0.6132	0.6741	0.5709

Table 4의 결과를 보면 기업 리뷰와 재무 정보를 학습한 모델의 예측 정확도가 재무 정보만을 학습한 모델의 예측정확도 보다 평균 +2.97%(rmse -0.0252) 높은 것을 확인할 수 있다. 특히 RF 알고리즘은 가지치기에 의해 선정된 기업 리뷰 TF-IDF 키워드 변수를 다른 모델보다 잘 일반화하여 과적합 없이 잘 분류하였으며 예측 정확도가 69.92%(rmse 0.5485)로 가장 높았다.

Table 5는 단일 모델에 대해서 특성 선택을 통해 재무 정보 외 기업 리뷰를 추가로 학습한 모델(+Company Reviews)의 예측 정확도를 비교 분석한 표이다.

Table 5. Prediction using Feature Selection

	accuracy	rmse
RF	0.7827	0.4661
SVM	0.6462	0.5948
GBM	0.7688	0.4808
XGBoost	0.7744	0.4750
LightGBM	0.7326	0.5171

Table 5의 결과를 보면 특성 선택을 통해 학습시킨 모델은 입력 데이터의 노이즈 정보가 제거되었기 때문에 1차 성능 개선 때 보다 모델 성능이 큰폭으로 개선되었다. 단일 모델로 RF 알고리즘이 예측 정확도가 78.27%(rmse 0.4661)로 가장 높았다.

마지막으로 모델들의 결합을 통한 앙상블 러닝으로 모델 성능을 개선하였다. 앙상블 러닝을 위해 재무 정보 외 기업 리뷰를 추가로 학습한 모델의 평균 예측정확도

Top4 모델(RF, GBM, XGBoost, LightGBM)을 사용하여 투표 기반 분류기를 학습시켰다. Table 6은 투표 기반 분류기를 학습시켰을 경우(Voting Classifier)와 투표 기반 분류기와 특성 선택을 동시에 진행했을 경우(Voting + Feature Selection)의 예측 정확도이다.

Table 6. Ensemble Learning

	accuracy	rmse
Voting Classifier	0.7019	0.5459
Voting + Feature Selection	0.7883	0.4601

Table 6의 결과를 보면 투표 기반 분류기의 예측 정확도는 단일 학습 모델 이상의 성능을 보이는 것으로 나타났으며, 투표 기반 분류기와 특성 선택을 동시에 진행했을 경우 본 연구에서 분석한 예측정확도 중 가장 높은 예측정확도 78.83%(rmse 0.4601)의 결과가 나온 것을 확인할 수 있다. 이 결과는 앙상블을 구성하는 분류기 중 하나의 분류기가 분류해내지 못한 것을 다른 분류기들이 분류함으로써 예측 능력이 향상되어 나온 것으로 판단된다. 따라서 앙상블 러닝을 통해 안정적으로 모델 성능을 개선할 수 있는 가능성을 보인다.

본 논문의 실험결과는 주가 방향성 예측을 위해 시계열 모형을 이용한 연구(I. C. Park, O. J. Kwon, T. Y. Kim. 2009)와 재무 정보를 이용한 주가 예측 연구(J. Y. Heo & J. Y. Yang. 2015), 뉴스와 감성분석을 이용한 연구(J. S. Jeong, D. S. Kim & J. G. Kim. 2015; H. N. Moon & J. W. Kim. 2014)의 결과와 비교했을 때 상당히 진전된 결과로 볼 수 있다.

5. 결론

본 연구에서 머신러닝을 통하여 종업원 만족도를 나타내는 온라인 기업 리뷰 데이터를 기반으로 주가 방향을 예측하였다. 실험 결과 내부 직원의 기업 리뷰를 활용한 모델의 주가 예측력은 기업의 재무 정보만을 학습한 모델과 뉴스와 SNS 정보를 사용한 선행연구 모델의 예측에 비하여 뛰어난 예측력을 보여주었다.

본 연구의 시사점은 다음과 같다. 첫째로 기존 주가 방향 예측 연구에서 시도하지 않았던 종업원의 기업 리뷰 데이터를 활용함으로써 주가 방향 예측 모델의 예측력이 향상될 수 있음을 보였다. 둘째로 실제 모형을 통해 종업

원 만족도를 나타내는 기업 리뷰 데이터가 기업의 수익성 및 재무성과를 반영한다는 점을 밝혔으며, 잡플래닛의 기업 리뷰가 주가 방향 예측 연구의 데이터로서의 안정성 및 일반화가 가능함을 확인하였다. 실무적으로 이번 연구가 텍스트 마이닝을 활용한 주가 방향 예측 분야에 유용한 정보를 제공할 것으로 기대된다.

단, 본 연구의 한계점은 기말 기준의 주가의 상승, 하락을 분류하는 예측 결과를 도출하였기 때문에 시계열분석 관점에서 연속적인 날짜의 데이터는 사용하지 않은 점과 KOSPI 상장기업에 대해서만 실험이 이루어졌다는 점이다. 또한 이번 연구에서는 빈도 기반의 TF-IDF를 통해 기업 리뷰를 구성하는 키워드 중 연도별로 이슈화되고 많이 언급되는 키워드의 집합을 추출하여 사용하였으나, word2vec을 통한 단어 간 유사도까지 연구 범위에 포함하여 고려하지 못한 한계가 존재한다.

향후 이러한 연구 한계점을 보완한 연구를 통해 기업 리뷰를 활용한 주가 예측 모델의 성능 개선이 가능할 것이다.

REFERENCES

- [1] J. Y. Park, J. P. Ryu & H. J. Shin. (2016). Predicting KOSPI Stock Index using Machine Learning Algorithms with Technical Indicators. *Journal of Information Technology and Architecture*, 13(2), 331-340.
- [2] D. W. Hah, Y. M. Kim & J. J. Ahn. (2019). A study on KOSPI 200 direction forecasting using XGBoost model. *Journal of the Korean Data And Information Science Society*, 30(3), 655-669
- [3] J. Y. Heo & J. Y. Yang. (2015). SVM based Stock Price Forecasting Using Financial Statements. *KIISE Transactions on Computing Practices*, 21(3), 167-172.
- [4] J. S. Jeong, D. S. Kim & J. G. Kim. (2015). A Study on the Prediction of Individual Stock Prices Using Online News Sensitivity Analysis. *Korea Intelligent Information Systems Society*, 45-58.
- [5] D. Y. Kim, J. W. Park & J. H. Choi. (2014). A Comparative Study between Stock Price Prediction Models Using Sentiment Analysis and Machine Learning Based on SNS and News Articles. *Journal of Information Technology Services*, 13(3), 221-233.
- [6] J. B. Kim & H. J. Kim. (2017). A domain-specific sentiment lexicon construction method for stock index directionality. *Journal of Digital Contents Society* 18(3), 2017.6, 585-592
- [7] Y. J. Yi & C. L. Lee. (2006). The Effects of Customer

- Satisfaction on Firm's Profitability and Value. *Korean Journal of Marketing*, 21, 85-113.
- [8] Zeithaml, V. A., Berry, L. L. & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of marketing*, 60(2), 31-46.
- [9] O. R. Oh. (2017). The Effects of Working Conditions on Stock Value. *Journal of Taxation and Accounting*, 18(6), 107-128.
- [10] Best, R. J. (2008). Employee satisfaction, firm value and firm productivity. Retrieved August, 23, 2013.
- [11] Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial economics*, 101(3), 621-640.
- [12] D. W. Kim, J. Y. Kang & J. I. Lim. (2016). Comparative Analysis of Job Satisfaction Factors, Using LDA Topic Modeling by Industries. *Journal of the Korean IT Service*. 15, 157-171.
- [13] *Jobplanet*. <https://www.jobplanet.com>
- [14] *NAVER finance*. <https://finance.naver.com/>
- [15] E. J. Park & S. J. Cho. (2014). KoNLPy: Korean natural language processing in Python. *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology*, 1-4.

임 용 택(Yongtaek Lim)

[정회원]



- 2009년 2월 : 상명대학교 수학교육과 (이학사)
- 2018년 9월 ~ 현재 : 고려대학교 빅데이터융합학과 석사과정
- 관심분야 : 데이터마ining, 딥러닝, 자연어처리, AI
- E-Mail : clarklim@korea.ac.kr

임 희 석(Heuseok Lim)

[종신회원]



- 1992년 : 고려대학교 컴퓨터학과(이학 학사)
- 1994년 : 고려대학교 컴퓨터학과(이학 석사)
- 1997년 : 고려대학교 컴퓨터학과(이학 박사)
- 2008년 2월 ~ 현재 : 고려대학교 컴퓨터학과 교수
- 관심분야 : 자연어처리, 인공지능, 딥러닝
- E-Mail : limhseok@korea.ac.kr