Journal of Convergence for Information Technology
Vol. 11. No. 1, pp. 28-33, 2021
e-ISSN 2586-4440
DOI : https://doi.org/10.22156/CS4SMB.2021.11.01.028

A study on stock price prediction through analysis of sales growth performance and macro-indicators using artificial intelligence

Sunghyuck Hong
Associate Professor, Division of ICT, Baekseok University

인공지능을 이용하여 매출성장성과 거시지표 분석을 통한 주가 예측 연구

홍성혁
백석대학교 ICT학부, 부교수

Abstract Since the stock price is a measure of the future value of the company, when analyzing the stock price, the company's growth potential, such as sales and profits, is considered and invested in stocks. In order to set the criteria for selecting stocks, institutional investors look at current industry trends and macroeconomic indicators, first select relevant fields that can grow, then select related companies, analyze them, set a target price, then buy, and sell when the target price is reached. Stock trading is carried out in the same way. However, general individual investors do not have any knowledge of investment, and invest in items recommended by experts or acquaintances without analysis of financial statements or growth potential of the company, which is lower in terms of return than institutional investors and foreign investors. Therefore, in this study, we propose a research method to select undervalued stocks by analyzing ROE, an indicator that considers the growth potential of a company, such as sales and profits, and predict the stock price flow of the selected stock through deep learning algorithms. This study is conducted to help with investment.

Key Words : Stock price analysis, Big data, Macro-indicative artificial intelligence, Prediction system, LSTM

요 약 주가는 그 기업의 미래 가치의 척도이기 때문에 주가를 분석할 때 기업의 성장성인 매출과 이익 등을 고려하여 주식을 투자한다. 기관투자자들은 종목 선정 기준을 잡기 위해서 현재 산업의 트렌드와 거시경제 지표를 보고 성장 가능한 관련 분야를 먼저 정하고 관련 기업을 선정한 후 기업에 대한 분석을 하고 목표가를 설정 후에 매수를 하고 목표가에 도달하면 매도하는 방식으로 주식 매매를 실시한다. 하지만, 일반 개인 투자자들은 경제에 대한 지식이 기관이나 외국인 투자자에 비교하여 부족하고, 기업에 대한 재무재표 분석이나 성장성에 대한 분석 없이 전문가나 지인의 추천종목을 따라 투자를 하여 기관투자자나 외국인 투자자들 보다 수익률 면에서 낮은 편이다. 따라서, 본 연구에서는 기업의 성장성인 매출과 이익 등을 고려한 지표인 ROE를 분석하여 저평가된 종목을 선택하고, 선택된 종목의 주가 흐름을 딥러닝 알고리즘을 통하여 예측하는 연구방법을 제안하여 투기가 아닌 건전한 투자에 도움이 되기 위해 본 연구를 진행한다.

주제어 : 주가분석, 빅데이터, 거시지표 인공지능, 예측 시스템, LSTM

*This research was supported by 2020 Baekseok University Research Fund.
*Corresponding Author : Sunghyuck Hong (shong@bu.ac.kr)
Received November 26, 2020
Revised December 15, 2020
Accepted January 20, 2021
Published January 28, 2021

1. Introduction

The PE ratio is a common method of measuring the relative value of a stock based on earnings per share. Companies that say that high stocks show consensus in the PE ratio market have strong prospects for future earnings growth. Because the PE ratio does not reflect future earnings growth, the PEG ratio is often used to determine whether the market valuation is supported by the expected future earnings growth rate. The PEG ratio was developed to compensate for the disadvantages of the PE ratio[1]. In particular, it was created to adjust the ratio to the relative future expected return growth rate of other companies. In other words, a small PEG Ratio (Price Earning to Growth Ratio) 〈1 means a company has a good net profit growth rate but a low stock price. The low PEG ratio means that this growth rate has not yet been reflected in the stock price.

Choosing a company with a high operating margin is important for stock selection. It is desirable to choose a stock price of a company with an operating margin of 20% or more. Sales profit is sales minus cost, and net profit is obtained by excluding sales activity expenses. Net income may not be viewed as operating income without distortion because it may not be possible to accurately determine the company's operating performance.

Select companies with a recent sales growth rate of 15% or more. The stock price reflects the future value of a company, but a good sales growth rate over the past five years can be a high probability of good future performance. It is necessary to select a company with a positive EPS(Earning Per Share) growth rate for next year and invest in stocks. Therefore, if we find a good PEG ration company by using AI analysis, stock buyers can have a profit to sell or buy a good stock on the trading market.

2. Stock Prediction System

Many researchers tried to predict the movement of stock prices through big-data. Since 2008, about 10 million Twitter sentiments have been classified into six sentiments, such as Calm, Alert, Sure, Vital, Kind, and Happy using psychology tools. As a result of comparing it with the US Dow Jones Index using emotion, the Dow Jones fluctuation was predicted with a hit rate of 86.7% [1].

Historic data try to determine the correlation between big-data and the stock market[2]. As a result of analyzing about 340 million tweeters, it was empirically analyzed that the more mentions of a specific company, the higher the probability of a stock price increase. As a result of model simulation for 4 months, the companies mentioned more in big-data showed about 2% less drop compared to the Dow index. In addition, it was shown that information such as a company's new product launch has an effect on the company's stock trading volume[3].

An empirical study on the relationship between price fluctuations in the future stock market, where the number of searches on the Internet for a specific term in the near future is a search frequency for emotional words [4] and Wikipedia related terms are performed. It has been announced that the weekly stock trading volume of S&P 500 companies is related to the number of Internet searches through the company's Google.

Bessembinder[5] compared and analyzed the relationship between the stock trading volume of companies listed on NADAQ and the number of times the terms related to the company were searched on the Yahoo search engine, and showed that there is a correlation between the two [2]. In February 2013, 98 words frequently used from 2004 to 2011 were extracted using the Google search engine, and as a result of

analyzing the words, the search query (words or phrases entered in the search box) and the stock market movement were correlated. The more word searches such as stocks, portfolios, and economics were found, the more the stock market fell, and as the financial market search decreased, the stock market rose [1].

In 2014, Naver, the number one search in Korea, also performed big data analysis using a search engine. 98 words were translated into Korean, and verbs were replaced with nouns, and duplicated words were replaced with other words, and converted to fit the Korean vocabulary. Frequency analysis was performed based on 100 vocabularies selected as terms suitable for domestic stock market analysis, and all search volumes were saved in the form of a big data Excel file. For the stock market analysis, companies selected 40 KOSPI companies and calculated the yield by obtaining stocks at the highest and lowest times of search volume by word during a certain period and selling them on a specific day. Results In the case of search terms for the government-related industry among 100 search terms, both the highest and lowest points showed a result showing plus return. The search term that showed the highest return on the highest point was lifestyle, which showed a 140.2% return. Transactions, gardens, and housing all recorded a yield of 133% at the lowest point. It was found that consumption search terms such as money, oil, and pork belly were not good for the rate of return, and search terms for economic recovery, happiness, culture, and futures were generally related to the rate of return[1].

3. AI Based Prediction System

In the 2013 study of Minsoo Kim[1], 98 words of Pries et al were extracted, 84 words to be analyzed out of 98 words were selected, and the rate of return was analyzed based on trends. The investment strategy used the cumulative stock rate basis (buy-and-hold) of a strategy of always buying shares at the closing price on the first trading day of the week and selling at the closing price on the first trading day of the following week. Among the search terms, banks showed a high rate of return of 146.5%, and headlines, cancers, and sales achieved a cumulative rate of over 100%. On the other hand, the trend investment strategy using society showed cumulative returns of -134% and religion -100%[2]. Researchers found a correlation between the sentiment of individual and institutional investors and stock returns[6]. Kim[2] presented an intelligent investment decision- making model through a big data sentiment technique called opinion mining, and showed that the result of sentiment analysis of news contents and index stock price fluctuations had a significant relationship. Park [11] can calculate consumer-oriented information by analyzing unstructured data that contains moods and emotions that correspond to human emotions and psychological information through research, and can be widely used not only in companies but also in public institutions. In recent years, a study on scoring and seizure analysis according to polarity distributions such as positive and negative sentences or vocabulary based on text using text mining techniques to supplement the limitations of existing structured data has been conducted.

As a result, the prediction accuracy of each company was different, and the average prediction rate was 56%. Although the more news about the company, the more information there is, as the information that is not related to the business activities of the company is also collected, the accuracy declined. The stock price prediction according to industry classification

showed relatively high accuracy of stock price prediction for energy/chemicals, consumer goods, and consumer goods, while information technology and shipbuilding/transport had low predictability. In the process of collecting and classifying information, there were limitations in research on the connection of direct, indirect news, and meaning. It examined whether it contains information that can predict the future by identifying whether the 9 sensibilities that contain stock price information from big data are self-correlated[7]. The 9 emotions, which are the main variables, collected information from blogs, Twitter, Facebook, bulletin boards, and news to remove spam and noise data. Afterwards, through natural language processing and text mining, (ANGER, HATE, DISLIKE, FEAR, LOVE, SHAME, SADNESS, and HOPE against the stock market), JOY, and classified into 9 sensations to identify the relationship between the 9 sensations[8-10]. After language processing, it automatically extracts important words using text mining, extracts frequency and correlation by date, and then uses visualization technology to help users easily understand data[11-13].

As a result of examining the correlation of each emotion in detail through the VAR analysis, it was found that the autocorrelation analysis resulted in that the emotions were every 7 days, and the past values of each emotion contained useful information for predicting the future. In addition, it could be confirmed that the nine sensations can be largely bound by positive and negative. In the VAR estimation, as in the results of the autocorrelation analysis, the auto regressiveness of the emotions was found [13].

Based on the above model, the simulation was conducted from January 2011 to October 2014, and the result exceeded about 20% of the KOSPI return. Also, from January 2015 to August 2015, actual investment results exceeded 12% of the KOSPI return. Through the research, it was possible to confirm the possibility that robot trading using big data analysis based on artificial intelligence could show a higher rate of return compared to actual market indicators.

4. Profitable Company

Based on criteria in Section 1, Table 1 shows the companies which meet the criteria such as PEG ration 〈1, profit 〉 20%, good ROA, and ROI.

Table 1. Companies which meet the criteria

Company	ROA	ROE	ROI	Profit M	Price
RenaissanceRe Holdings Ltd.	2.00%	9.90%	14.70%	12.10%	178.22
Exelixis, Inc.	14.70%	16.40%	17.30%	28.20%	22.29
Credit Acceptance Corporation	4.70%	15.50%	12.40%	21.80%	339.64
TCF Financial Corporation	0.40%	3.60%	19.30%	10.50%	29.62
Turkcell Iletisim Hizmetleri A.S.	7.10%	17.80%	14.30%	12.20%	5.13
SLM Corporation	1.70%	23.60%	1.70%	21.40%	9.56
Pinnacle Financial Partners, Inc.	1.00%	6.80%	12.60%	28.10%	45.19
Triton International Limited	3.00%	14.20%	7.60%	20.90%	38.24
Axos Financial, Inc.	1.50%	15.70%	25.40%	29.40%	27.81
Noah Holdings Limited	9.00%	11.20%	9.90%	26.80%	27.66
Grupo Financiero Galicia S.A.	4.50%	29.50%	0.70%	26.20%	6.65
Banco BBVA Argentina S.A.	4.80%	29.70%	47.30%	23.70%	2.69
Hope Bancorp, Inc.	**0.90%**	**6.80%**	**17.30%**	**21.30%**	8.65
Ebix, Inc.	5.80%	16.80%	12.00%	16.70%	20.02
The Bancorp, Inc.	1.00%	11.00%	21.50%	28.10%	9.57
PBF Logistics LP	13.20%	118.80%	17.60%	36.90%	**8.45**
KNOT Offshore Partners LP	2.60%	7.60%	7.90%	16.20%	13.9

Ebix, Inc. provides on-demand software and e-commerce services and solutions to insurance, financial, healthcare, and e-learning industries in the United States and internationally. Once, a company is selected by the criteria, it is time to predict the future stock price by using LSTM which is neural network. Figure 1 shows Python code for analysis Ebix stock price based on Yahoo Financial data from Internet.

```
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense, LSTM, Dropout
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
data = pd.read_csv("./EBIX.csv")
data
```

	Date	Open	High	Low	Close	Adj Close	Volume
0	2015-11-25	37.130001	38.290001	37.130001	37.790001	36.595394	344500
1	2015-11-27	38.000000	38.259998	37.639999	37.810001	36.614750	84800
2	2015-11-30	37.820000	38.000000	37.259998	37.509998	36.324238	465500
3	2015-12-01	37.549999	37.900002	36.049999	37.020000	35.849731	842400
4	2015-12-02	37.049999	37.549999	36.560001	36.720001	35.559216	229000
...	...	...	...	...	...	...	...
1254	2020-11-18	31.450001	34.330002	31.270000	32.590000	32.590000	547000
1255	2020-11-19	32.360001	33.279999	30.400000	30.540001	30.540001	469400
1256	2020-11-20	30.580000	31.709999	30.420000	31.520000	31.520000	216100
1257	2020-11-23	32.299999	34.189999	31.809999	33.939999	33.939999	337200
1258	2020-11-24	35.000000	37.970001	34.759998	35.939999	35.939999	587400

Fig. 1. EBIX Historical Stock Price from yahoo finance with Python Code

4.1 Analysis Methods

Keras, Tensorflow, and Jupyter notebookare used as a testing software and library for prediction of stock price. TensorFlow is a library which gives various functions to ease implement deep learning programs made by Google Inc. TensorFlow itself is fundamentally executed in Mircrosoft C++ and provides many different languages for examples, Go, Java, and Python, but it is convenient to develop in Python as it supports Python first and most convenient functions are implemented only with the Python library.

Window size is 20 which means weekly predictoin, and 70% of data set is training set. The rest of 30% data set is testing set. After these setting, the result shows in Figure 2. The red line is real price which is historical data, and the blue line is prediction graph.

Fig. 2. Prediction Result for EBIX Stock

The based of the result, deep learning algorithm can predict the price very well.

5. Conclusion

In this research, The stock price prediction is implemented using a LSTM circulatory neural network and measured the performance of the proposed method through experiments. As a result, this research is not only using a deep learning algorithm, but also, the analysis of sales growth performance and macro-indicators such as PEG ratio and etc. Therefore, the result is much more accurate the prediction price than any other prediction based on the price itself. Also, it is a risk free compare to other AI based prediction. In the future, the possibility of developing a learning model with improved performance through various optimization studies as bidirectional LSTM circulatory neural networks is expected to increase.

REFERENCES

[1] P. D. Easton. (2004). PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The accounting review, 79(1),* 73-95.

[2] M. Kim & P. Koo. (2013). A Study on Big Data Based Investment Strategy Using Internet Search Trends. *Journal of the Korean Operations Research and Management Science Society, 38(4),* 53-63.
DOI : 10.7737/jkorms.2013.38.4.053

[3] Y. S. Kim, N. G. Kim & S. R. Jeong. (2012). News and stock prices: Intelligent investment decision-making model through sentiment analysis of big data. *Intelligence Information Research, 18(2),* 143-156.
DOI : 10.13088/JIIS.2012.18.2.143

[4] T. N. Sahu, K. Bandopadhyay & D. Mondal. (2014). *An empirical study on the dynamic relationship between oil prices and Indian stock market.* Managerial Finance.

[5] H. Bessembinder & H. M. Kaufman. (1997). A comparison of trade execution costs for NYSE and NASDAQ-listed stocks. *Journal of Financial and Quantitative Analysis,* 287-310.

[6] J. Kwak & J. Hong. (2018). Analysis of the YOLO phenomenon using big data: focusing on travel consumption. *Tourism Research Journal, 32(2),* 21-34.

[7] J. S. Jeong, D. S. Kim & J. W. Kim. (2015). A study on individual stock price prediction using online news sentiment analysis. *Journal of the Korean Intelligent Information Systems Society,* 45-58.

[8] J. C. Jeon. (1995). Analysis of development and use cases of expert systems for stock investment. *Journal of the Korean Intelligent Information Systems Society,* 163-175.

[9] S. H. Jang & J. Y. Yun. (2018). *A.I Design Strategy for Investment Judgment Support Aid.* 62-65

[10] J. H. Ki. (2016). A Study on Urban Statistical Expression Using Big Data Analysis-Focusing on the expression of local community health indicators in Korea using Chernov's face. *Space and Society, 55,* 336-358.

[11] H. C. Choi. (2017). Big data environment and humanities platform. *Culture and Convergence, 39(5),* 177-202.

[12] S. H. Lee, H. G. Kang, S.-H. Kim & C. M. Lee. (2013). Sentiment analysis in big data. *Financial Engineering Research, 12(2),* 79-96.

[13] S. H. Song, S. H. Hwang, Y. H. Lee, H. K. Lee, K. S. Han & J. B. Kim. (2016). Social big data analysis model for trading. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, 6(3),* 91-100.

[14] J. M. Perkel. (2018). Why Jupyter is data scientists' computational notebook of choice. *Nature, 563(7732),* 145-147.

[15] M. Abadi ea al. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265-283).

[16] S. Hong. (2020). Research on Stock price prediction system based on BLSTM. *Journal of the Korea Convergence Society, 11(10),* 19-24.

홍 성 혁(Sunghyuck Hong) **[종신회원]**

- 2007년 8월 : Texas Tech Univ., Computer Science (공학박사)
- 2012년 3월 ~ 현재 : 백석대학교 ICT학부 부교수
- 관심분야 : 딥러닝, 블록체인, 사물인터넷 보안, 경량보안프로토콜
- E-Mail : shong@bu.ac.kr