

Implementation of Fund Recommendation System Using Machine Learning

Chae-eun Park¹, Dong-seok Lee², Sung-hyun Nam³, Soon-kak Kwon^{1*}

Abstract

In this paper, we implement a system for a fund recommendation based on the investment propensity and for a future fund price prediction. The investment propensity is classified by scoring user responses to series of questions. The proposed system recommends the funds with a suitable risk rating to the investment propensity of the user. The future fund prices are predicted by Prophet model which is one of the machine learning methods for time series data prediction. Prophet model predicts future fund prices by learning the parameters related to trend changes. The prediction by Prophet model is simple and fast because the temporal dependency for predicting the time-series data can be removed. We implement web pages for the fund recommendation and for the future fund price prediction.

Key Words: Fund recommend system, Machine learning, Time series data prediction

I. INTRODUCTION

Due to the COVID-19 pandemic, economic activity is contracting, and accordingly, a considerable amount of capital is flowing into the fund market. Accordingly, interest in forecasting for fund investment is also increasing. Capital Asset Pricing Model (CAPM) [1], a representative investment prediction model, calculates the expected return of an investment company using the risk-free rate of return, expected market return, and beta. However, forecasting through CAPM requires the premise that all investors have risk aversion. In addition, CAPM has a disadvantage in that the amount of information required for investment prediction is large. Recently, several methods of making investment predictions through machine learning have been proposed. There is a method of recommending products using existing investment data by showing the ranking of the products that have made the most profit in the last three months, six months, a year, and three years, or recommending them through weekly sales or set amount increase rate. The subjective recommendation method of

each platform can also be determined by the recommendation fund selection committee using quantitative and qualitative evaluation. In the case of alpha funds, different from the selection criteria for recommended funds, they are recommended according to market conditions. However, in the case of such an existing method, there is a problem that it is difficult for general investors to use [2].

Most of the existing fund recommendation systems are formula-based. The formula is designed with the fund to be recommended and the degree of usefulness and user information as parameters. The fund is recommended through the optimization of the estimation result through the formula by a specific performance problem. The formula used in the formula-based system is often customized based on the experience of a recommender. Therefore, the fund recommendation through the formula-based system is highly likely to be biased towards the recommender.

In this paper, we propose a method for predicting fund prices through a machine learning method which predicts time series data. The fund price is predicted by

Manuscript received September 10, 2021; Revised September 30, 2021; Accepted September 30, 2021. (ID No. JMIS-21M-09-032)

Corresponding Author (): Soon-kak Kwon, Address: (47340) Eomgang-ro 176, Busanjin-gu, Busan, Korea, +82-51-890-1727, skkwon@deu.ac.kr

¹ Dept. of Computer Software Engineering, Dongeui University, Korea

² AI Grand ICT Research Center, Dongeui University, Korea

³ Dept. of Artificial Intelligence, Dongeui University, Korea

Prophet model, which is an algorithm for predicting time series data [3]. The daily fund prices for 380 funds are learned by Prophet. The future fund prices are predicted based on given previous prices. The prediction results are visualized on the web page as graphs.

II. RELATED WORKS

2.1. Financial Product Investment Forecast

Existing methods for predicting financial products such as stocks and funds include Bollinger band, Dow theory, and Kostolani egg [4]. Bollinger band is an index that combines the average trend and volatility of stocks and was invented by an investment expert named J. Bollinger in the early 1980s. Bollinger band provides a standard for evaluating whether the current stock price is overvalued or undervalued. Bollinger band generates the bands of upper, lower, and center lines according to the movement of stock prices. The stocks are sold or bought as the stock depending on the type of band the stock price approaches.

Dow theory was devised by C. Dow, who created Dow Jones Industrial Average in 1882. Dow theory analyzes the movement of stock prices and the trend between stocks. Dow theory assumes that stock price movements are trending rather than random. If the stock price continues to rise, predict the stock price under the assumption that the trend will continue. In the Dow Theory, trends are divided into accumulation, boom, overheat, dispersion, panic, and recession.

Kostolany egg is a theory proposed by A. Kostolany, who worked as an investment analyst in the U.K., and divides the market into a boom and a recession depending on the interest rate situation. It is a theory that recommends taking a portfolio. When interest rates peak and interest rates fall, they invest in bonds, when interest rates approach a trough, invest in real estate, when interest rates rise after a trough, invest in stocks, and when interest rates rise near a peak, they invest in cash.

2.2. Time series prediction through machine learning

Time series data means data recorded sequentially according to the passage of time. The prediction problems through time series data are used in various fields ranging from forecasting major economic indicators to predicting the demand for a certain product.

Studies for predicting time series data by neural network have been conducted due to the success of recognition and prediction in various fields through the neural networks. However, the previous structure of neural network, DNN (Deep Neural Network) [5], does not suitable for predicting time series data because it is a structure that does not consider the order of data input

time. RNN (Recurrent Neural Network) [6] is introduced to predict the time series data. RNN has the characteristic of sending the result value obtained through the activation function at the node of the hidden layer to the output layer as well as the input of the next calculation of the node of the hidden layer. The RNN has the characteristic of outputting the result data processed at the node and inputting the corresponding output value to itself node along with the next input data. Therefore, the previous input data is also considered in calculating output data. However, the performance of RNN drops sharply for long time series data because of the gradient vanishing problem of RNN. It causes the problem of long-term dependencies that the data from the distant past rarely reflect.

LSTM [7] overcomes the shortcomings of RNN by introducing long-term and short-term data. The input and output of the short-term data is the same as RNN. The long-term data reflects current inputs through the corresponding weight for each step. In this way, important data from the distant past can also be reflected in the forecast.

III. FUND RECOMMENDATION BASED ON INVESTMENT PROPENSITY

The investment propensity of users is classified through the investment propensity diagnosis list [8] distributed by the Korea Council for Investor Education to diagnose the investment propensity. The investment propensity diagnosis list consists of questions about an age, investment experience, knowledge level of financial products, the amount of tolerable loss, an investment period, the ratio of investment funds to total assets, and the trend of increase or decrease in income as shown in Table 1. The answers of the user are scored and are classified into the following six types according to the score: an aggressive investment type, an active investment type, a risk neutral type, a stability-seeking type, and a stability type.

Table 1. Questions in propensity diagnosis list.

No.	Question
1	What is age range of the user?
2	What financial product has the user invested in?
3	What is the user's knowledge level of financial products?
4	What level of loss can user tolerate?
5	How long is the range of the investment duration of the user?
6	How is the ratio of investment funds to total assets?
7	Is user's income constant, and is it expected to remain constant in the future?

The funds are classified into six ratings based on the level of risk. The criteria of the fund risk classification refer to that of the established banks [9]. We recommend the fund with a risk rating that is appropriate for the investment propensity. The funds with low risk are recommended to the users with a safety-oriented investment propensity and the funds with high risk are to the user with offensive investment propensity. Table 2 shows the types of the recommended funds based on the user propensity.

Table 2. Matching between investment propensity and funds.

User propensity	Fund
Offensive-investment	Grade 1
	Grade 2
Active-investment	Grade 3
Risk-neutral	Grade 4
Stability-seeking	Grade 5
Stable	Grade 6

The investment propensity is identified through the questions in Table 1. The result of the investment propensity identification is represented to the users and is stored with user personal information into a database. The fund information is also stored in the database. The fund information is composed of a name, a risk level, and average, lowest, and maximum prices. Fig. 1 shows the user and fund information that are stored in the database. The user is matched based on the user and fund information in the database. First, the investment propensity is determined by scoring the responses of the user derived from the propensity diagnosis list. After that, the proposed system finds the suitable funds in the database based on the investment propensity. The system shows the found funds to the user. Fig. 2 shows the flow of the proposed fund recommendation method.

id	name	investment propensity
chaechae	Chaeun Park	Risk-neutral
dongsek	Dongseok Lee	Offensive-investment

(a)				
Fund No.	Average	Max	Min	Risk
1	929.753	1203.16	645.75	2
2	2061.494	2534.16	1502.1	1
3	1894.838	2053.05	1765.32	5

Fig. 1. Example of storing information to DB: (a) user information and (b) fund information

IV. FUND PRICE PREDICTION BY MACHINE LEARNING

For the funds, the K-Means algorithm can be applied to classify funds according to the relationship between risk and

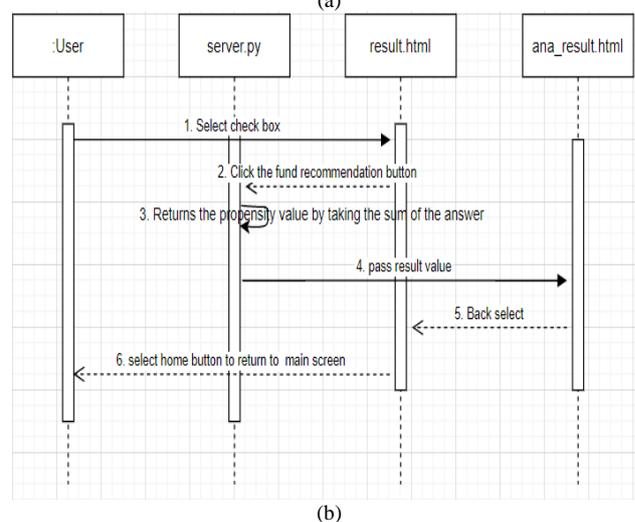
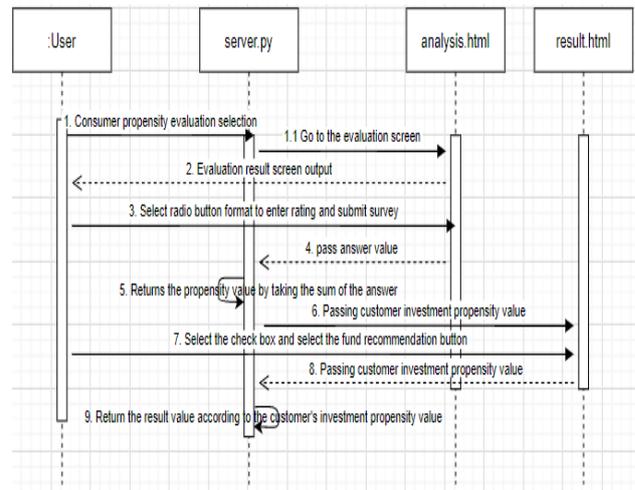


Fig. 2. Sequential diagram of the proposed fund recommendation method: (a) overall flow and (b) fund matching process.

return. In K-Means, the elbow method is applied to determine K which is the number of the fund clusters. As the number of the clusters increases, the average distance between centroids, which is the center point of each classified cluster, decreases. The decrease in the average distance is significantly reduced when the number of clusters is greater than or equal to a specific number. The specific number is determined as K . Fig. 3 shows that the results are divided into four groups, meaning low risk + low return, low risk + high return, high risk + low return, and high risk + high return. Although the K-Means algorithm can cluster fund propensity, it is difficult to clearly determine the variables that affect the fund price.

The Prophet model is applied to more accurately predict the future fund price through the relationship between the fund price and the fund's variables. Prophet model is a time series prediction library developed by Facebook and can be used in the form of a Python package. A prediction by Prophet model is made through the following equation.

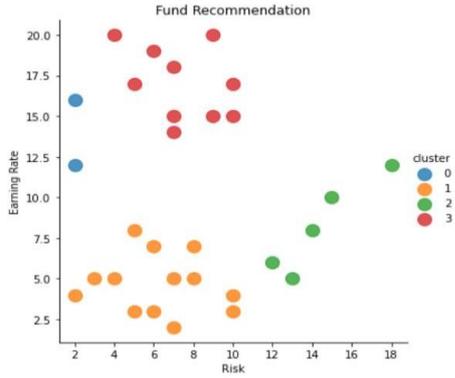


Fig 3. Clustering of relationship between risk and return by K-Means algorithm.

$$y(t) = g(t) + s(t) + h(t) + e_i, \quad (1)$$

where $y(t)$, $g(t)$, $s(t)$, and $h(t)$, and e_i are a predicted result at time t , a growth function, a seasonality function, a holidays function, and the error assumed to be normally distributed, respectively. The growth function represents a change trend, that is a continuous change of measurement target. Parameters in the function indicate whether the value is continuously increasing or decreasing. The parameters may change when checkpoint is reached because the trend of change can change at a specific point. It is probabilistically determined whether a particular point is the checkpoint. The seasonality function means periodic changes within the overall change trend. The periodic changes are represented by Fourier series function. The holidays function means the effect of an event occurring at a specific point. For example, sales of products increase significantly with each holiday. Prophet model treats the prediction as a curve fitting problem so it can avoid temporal dependence. Therefore, the prediction equation of Prophet model consists only of the sum of several linear or nonlinear functions. This characteristic of Prophet model makes the prediction learning easy and fast. The parameters of component functions in Equation (1) are determined by training through the past data. Prophet model predicts not only the future values but also the prediction value ranges.

The Prophet model requires the data information and numerical values as input. If the price of a commodity over a period of time is known, the future price of the fund can be predicted. The change points are automatically detected to predict growth, and patterns that appear periodically as user behavior patterns can be applied. Cross-validation is used to measure prediction accuracy. Cross-validation is a variety of similar model validation techniques for evaluating how statistical analysis results generalize to independent data sets. The cross-validation procedure is performed automatically for historical cutoff ranges using the cross-validation function. Specify the prediction period and then optionally the size of the initial training period and the

interval between the cutoff date. The output of cross-validation is a data frame with actual and predicted values for each simulated predicted date and each cutoff date.

In the proposed system, the prediction model is trained from the fund prices over the past 4 months and predicts the future prices. The predicted results are presented as graphs on the web page as shown in Fig. 4. In Fig. 4, a black line and a gray line are the measured past prices and the predicted prices, respectively, and a gray plane is the range of prediction error.

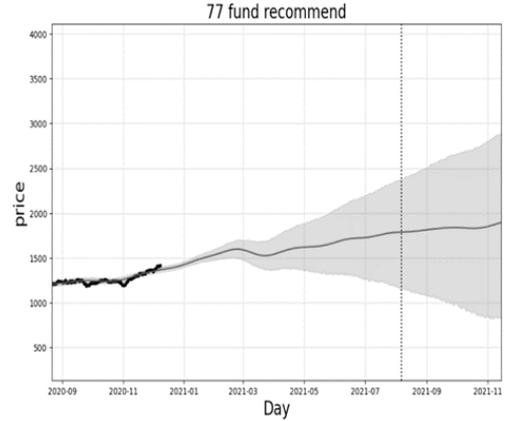


Fig 4. Graph of fund price prediction by the proposed method.

We measure the prediction accuracy of the proposed method through 380 fund data. In the fund data, the fund price for 981 days is stored. The prediction model is learned by the first 400 days and the fund prices after 100 days are predicted. Fig. 5 shows the results of the prediction results for various funds. In Fig. 5, black dash lines, black solid lines, and gray dash lines are the input fund prices which are used to learn the prediction model, predicted fund prices, and actual fund prices, respectively. The results show that the proposed method is insensitive to the actual price change, but accurately predicts an up- or down-trend in price.

The proposed method is compared with the prediction method with LSTM. LSTM can be also used in predicting the future fund price [10-11]. Fig. 6 and table 3 show the comparison between the proposed method and LSTM. The error of the fund prediction is calculated through the RMSE calculated as follow.

$$\text{RMSE} = \sqrt{\frac{1}{(981-294)} \sum_{i=294}^{981} (y_i - \hat{y}_i)^2}, \quad (2)$$

where y_i and \hat{y}_i mean the actual fund price and the predicted fund price on the i -th day, respectively. LSTM also predicts the future fund price well. However, LSTM requires a high-performance processor and a GPU, and requires a lot of data and time to learn. The proposed method is faster than the prediction method by LSTM.

The past fund price learning can be adjusted by changing p , which is a parameter of $g(t)$ in Equation (1). The update

cycle of $g(t)$ is faster as the parameter p is larger, so the prediction adapts more quickly to the latest changing trends. However, too large p causes the overfitting problem that the prediction is too dependent on the past fund prices, so the prediction accuracy is rather drop. Therefore, finding an optimal p is important for the prediction accuracy. Table 4 shows the prediction accuracy according to p . The prediction is most accurate when p is 0.3.

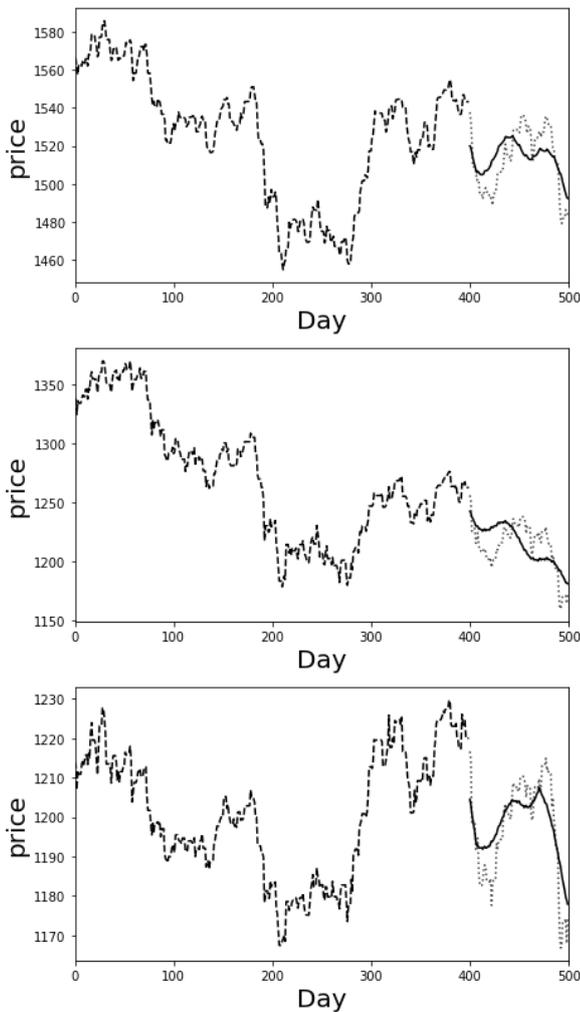


Fig 5. Fund price prediction by proposed method.

Table 3. Comparison between proposed method and LSTM.

Prediction	RMSE	Prediction time (s)
Proposed method	45.181	1.4
LSTM	44.918	2.1

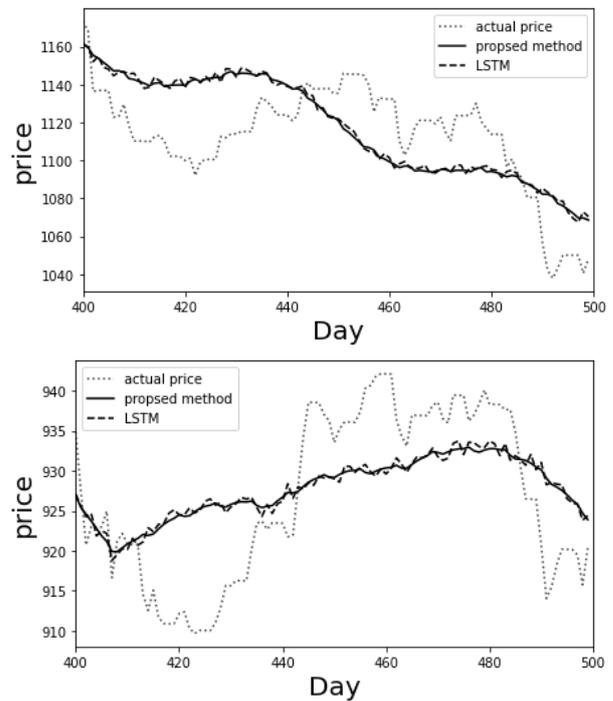


Fig 6. Graph of fund price prediction by proposed method and LSTM

Table 4. Prediction accuracy according to p .

p	RMSE
0.1	48.140
0.3	45.181
0.5	46.413
0.8	47.692
1.0	47.824

V. FUND RECOMMENDATION AND PREDICTION SYSTEM

We implement a web page for the fund recommendation and the future fund price prediction. First, the user input his personal information. The user can understand his investment propensity by answering the questions in Table 1. The user can be recommended to the suitable funds based on his investment propensity. The user can see the price prediction of the fund he wants. Fig. 7 shows the implementation of web pages for the fund prediction through the proposed method.

VI. CONCLUSION

In this paper, fund information was collected through crawling, followed by analysis, and funds were visually classified through the K-Means algorithm by recruitment method, investment target, and management strategy. Based on the results, a price prediction program using Prophet was built after that. If you use this technology, you can check the

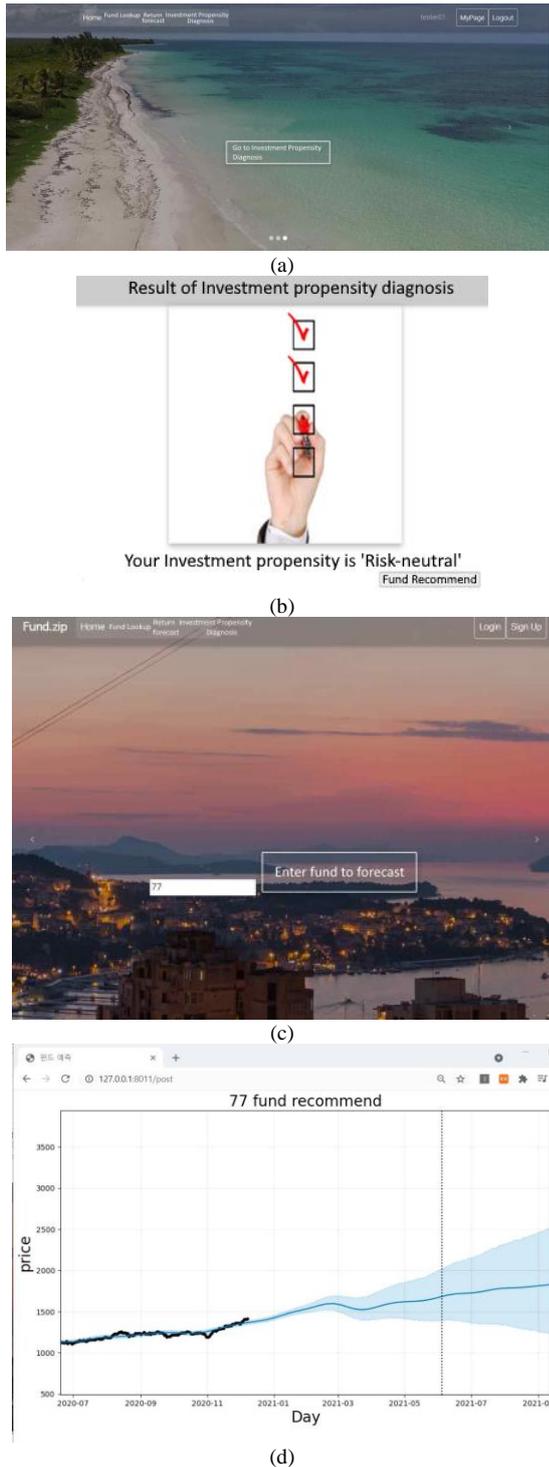


Fig 7. Implemented web page for proposed system. (a) main page, (b) investment propensity result page, (c) fund input page, and (d) prediction result page

future return of the entered fund. It provides economic feasibility that saves time and money by allowing you to understand your investment propensity without spending extra time and recommends a suitable fund. In addition, in the existing fund-related website, only fund information inquiry was possible, but this paper has originality by storing the inspection product. The website was built in consideration of UI/UX so that users can intuitively view and use it with high satisfaction. In the case of price

prediction, since the learning result is not the same every time, a highly reliable recommendation is made through multiple learning and shows different results for each person. This will help hyper-personalization in the era of digital transformation [12].

Acknowledgement

This research was supported by the BB21+ Project in 2021, by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2021-2020-0-01791) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation), and by Dong-eui University Grant (202101950001).

REFERENCES

- [1] A. F. Perold, "The capital asset pricing model," *Journal of Economic Perspectives*, vol. 18, no. 3, pp. 3-24, 2004.
- [2] E. K. Seo, S. G. Jun, and M. Y. Han. "Fund recommendations and fund performance," *Korean Journal of Finance Association*, vol. 32, no. 1, pp. 93-123, 2019.
- [3] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37-45, 2018.
- [4] S. M. You and D. H. Kim, "Combined index for mean trend and volatility of stock price and buy-sell strategy," *Journal of The Korean Data Analysis Society*, vol. 6, no. 5, pp. 1281-1290, 2004.
- [5] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2642-3649, 2012.
- [6] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for improved unconstrained handwriting recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 855-868, 2009.
- [7] H. Sak, H. A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," In *Proceedings of the Interspeech*, pp. 338-342, 2014.
- [8] Korea Council for Investor Education, "Investment propensity diagnosis list," https://www.kcie.or.kr/guide/24/31/web_view?series_idx=&content_idx=760(Access on September 09, 2021)
- [9] Shinhan Bank, "Risk classification table for financial investment product," https://www.shinhaninvest.com/siw/customer-center/protection/protect2_tab2/contents.do (Access on September 29, 2021)
- [10] I. T. Joo and S. H. Choi, "Stock Prediction Model based on Bidirectional LSTM Recurrent Neural

Network,” *Journal of Korea Institute of Information, Electronics, and Communication Technology*, vol. 11, no. 2, pp. 204-208, 2018.

- [11] J. S. Han, H. S. Choi, S. W. Kim, and S. H. Oh, “A study on financial time series data volatility prediction method using AI’s LSTM method,” *Journal of Knowledge Information Technology and Systems*, vol. 14, no. 6, pp. 665-673, 2019.
- [12] J. W. Kwon and T. S. Song, “Platform-based digital transformation promotion trend for manufacturing innovation,” *Journal of Electronic Engineering*, vol. 46, no. 12, pp. 34-46, 2019.

Department of Computer Software Engineering. From 2003 to 2004, he was a visiting professor of the Department of Electrical Engineering in the University of Texas at Arlington. From 2010 to 2011, he was an international visiting research associate in the School of Engineering and Advanced Technology in Massey University. His research interests are in the areas of image processing, video processing, and video transmission.

Authors



Chae-Eun Park (koo77223@gmail.com) received the bachelor’s degree in computer software engineering from Dong-Eui University, in 2021. As an undergraduate research student in the Artificial Intelligence Lab, I conducted machine learning-related research using Python. Her research interests are in the areas of

recommendation system, using data, and digital transformation of finance.



Dong-Seok Lee (ulsan333@gmail.com) received the B.S., M.S., and Ph.D. degrees in Computer Software Engineering from Dong-eui University in 2015, 2017, and 2021, respectively, and is currently a research professor in AI Grand ICT Research Center at Dong-eui University. His research interest is in the areas of image

processing and video processing.



Sung-hyun Nam (hkwmnsh@gmail.com) is received the B.S. degrees in Computer Software Engineering from Dong-eui University in 2021 and is currently a Ph.D. course in the Department of Artificial Intelligence at Dong-eui University.



Soon-kak Kwon (sskwon@deu.ac.kr) received the B.S. degree in Electronic Engineering from Kyungpook National University, in 1990, the M.S. and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), in 1992 and 1998, respectively. From 1998 to 2000, he was a

team manager at Technology Appraisal Center of Korea Technology Guarantee Fund. Since 2001, he has been a faculty member of Dongeui University, where he is now a professor in the

