International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

LQ45 Stock Index Prediction using k-Nearest Neighbors Regression

Julius Tanuwijaya, Seng Hansun

Abstract*: The capital market is an organized financial system consisting of commercial banks, intermediary institutions in the financial sector and all outstanding securities. One of the benefits of the capital market is creating opportunities for the community to participate in economic activities, especially in investing. In daily stock trading activities, stock prices tend to have fluctuated. Therefore, stock price prediction is needed to help investors make decisions when they want to buy or sell their shares. One asset for investment is shares. One of the stock price indices that attracts many investors is the LQ45 stock index on the Indonesian stock exchange. One of the algorithms that can be used to predict is the k-Nearest Neighbors (kNN) algorithm. In the previous study, kNN had a higher accuracy than the moving average method of 14.7%. This study uses kNN regression method because it predicts numerical data. The results of the research in making the LQ45 stock index prediction application have been successfully built. The highest accuracy achieved reaches 91.81% by WSKT share.

***Index Terms*: k-Neighbors Regressor, LQ45, Prediction, Stock, Time Series.**

I. INTRODUCTION

Shares are one form of assets in investment. Investments in the form of shares are chosen by investors because stocks can provide attractive benefits [1]. Basically, stocks provide two benefits to investors who buy or own shares. First is dividends and second is capital gains. Stocks also have risks, including capital loss and risk of liquidation [2].

LQ45 Stock Index is the one most investors glance at. Stocks that are incorporated into the LQ45 stock index are stocks that have good financial status. The property sector is chosen as a focus in this study because Property industry has limited land supply, while demands continue to grow. So, property has a good prospect due to such conditions [3, 4].

Stock prices are fluctuating because there is always a condition of stock prices going up or down. Therefore, an application is needed that can predict stock prices properly. With the prediction application, investors can be helped to make the right decision [5].

k-Nearest Neighbors (kNN) is an algorithm that can be used to make a prediction. In the research conducted by Vainionpaa and Davidsson in 2014, produced that kNN has better accuracy than Moving Average method by 14.7%. [6, 7] Since in this study, we are trying to predict future stock prices, we use KNN regression method, which has a better accuracy level [8]. The next section will further explain the kNN regression method.

Manuscript published on 30 September 2019
* Correspondence Author
Julius Tanuwijaya*, Informatics Department, Universitas Multimedia Nusantara, Tangerang, Indonesia.
Seng Hansun, Informatics Department, Universitas Multimedia Nusantara, Tangerang, Indonesia.

II. K-NEAREST NEIGHBORS REGRESSION

The focus of regression problems is to predict the output based on variables given by a set of independent variables. k-Nearest Neighbors Regression (kNR) is done by simply assigning the property value for the object to be the average of its *k* nearest neighbors. It can be useful to weight the contributions of the neighbors so that the nearer neighbors contribute more to the average than the more distant ones [9, 10]. kNR makes predictions based on the outcome of the *k* neighbors closest to that point. Therefore, to make predictions with kNR, a metric to measure the distance between the query point and cases from the example samples is needed [9]. One of the most popular choices to measure this distance is the Euclidean distance. Euclidean distance is suitable for calculating numerical predictions. Eq. (1) is the formula for Euclidean distance [9, 11].

$$D(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \quad (1)$$

where D is distance between p and q, n is amount of data, and i is the index ith. After selecting the value of k, we can make predictions based on the kNR. For regression, the predictions are the average of the k nearest neighbors outcome. Eq. (2) is the formula to make the predictions [9].

$$y = \frac{1}{k}\sum_{i=1}^{k} y_i \quad (2)$$

where y is the predicted value, k is the number of neighbors, y_i is the stock price value at index i.

III. RESEARCH METHODOLOGY

The research methodology used in this study, as follows.

1) Data Gathering
2) Data Preprocessing
3) Data Splitting
4) Model Building
5) Model Prediction
6) Evaluation

The first step in conducting this research is the data gathering process. Stock data is obtained through the Alphavantage Application Programming Interface (API). Data taken through the Alphavantage API is still raw data, so it is necessary to do data cleaning or data preprocessing process.

After that, we do the data filtering process, where data that contain no trading activities will be removed. The data that we filtered was stock data from 2016 to 2018.

After the data has been cleaned, we do data splitting process. Data separated between data training and data test. Then, we made the data modeling and model prediction. Finally, we evaluated to check whether the results are optimal. If not, back to data modeling process.

IV. RESULTS

This research conducted using Python programming language and the Flask framework. The stock data used is stock index of LQ45 property sector. They are ADHI, BKSL, BSDE, LPKR, PTPP, WIKA, and WSKT. For more information, you can access from Github with this link https://github.com/juliusHin/KNR_Stock_Prediction.

A. Application Results

Figure 1 is the wireframe of main menu. User can choose options from the combo box to select stock code then click submit to see the prediction results.

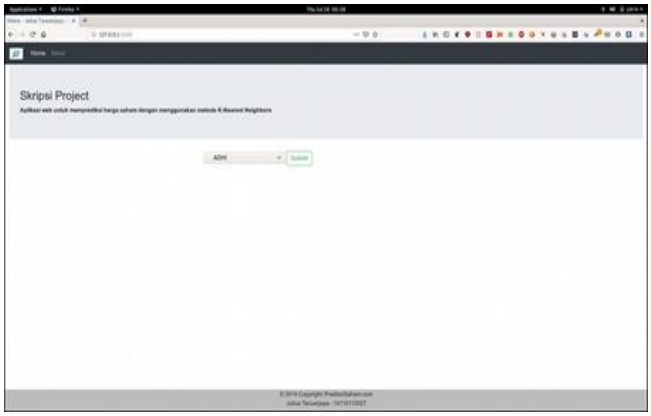

Figure 1. Main Menu

Figure 2. WSKT Prediction Results

Figure 2 shows an example of prediction results for WSKT stock. The result is represented by a plotting chart. The cursor can be directed by the user to show the price on that day directly. Other than that, below the chart there is a table containing the day, predicted price, and actual price. The table figure is shown in Figure 3.

Figure 3. WSKT Prediction Result with Table

B. Evaluation Results

This research using some variant values of *k* and using Root Mean Square Error (RMSE) and R^2 (R Squared Error) to measure the accuracy level. Table 1 to Table 7 show the results of RMSE values for different *k* for each stock index.

Table 1. RMSE values for ADHI

k	RMSE
2	246.19998887857312
3	250.59398132123366
4	255.78142650482124
5	260.70157938160537
6	264.51866684646143
7	269.44020105929616
8	273.522027751292

9	277.40328592031597
10	280.4968328061579

Table 2. RMSE values for BKSL

k	RMSE
2	23.989622425686154
3	24.289833917629647
4	24.551644224511403
5	24.719955308944034
6	24.825457719244934
7	24.89786392680934
8	24.949292573153123
9	24.989267733602166
10	25.070127040788407

Table 3. RMSE values for BSDE

k	RMSE
2	279.70222261064237
3	283.4838535756405
4	287.0208546118955
5	289.14033776115525
6	291.64958850227265
7	293.44628879713434
8	296.0290081615568
9	298.0419270837975
10	299.9926536798939

Table 4. RMSE values for LPKR

k	RMSE
2	130.37875989495387
3	130.39074196747606
4	130.373608440427
5	130.67739448187442
6	131.15607246298336
7	131.503858575381
8	131.95219191957727
9	132.4824762222419
10	133.06970634232172

Table 5. RMSE values for PTPP

k	RMSE
2	448.8889502789523
3	459.811888824176
4	467.0269204080546
5	474.0364032608212
6	479.80975219018023
7	484.89296560262926
8	489.57392864425645
9	494.73356945665836
10	499.56491549315547

Table 6. RMSE values for WIKA

k	RMSE
2	163.774489235186
3	165.86315275868589
4	168.93348687720584
5	170.81430725237257
6	172.5112742147065
7	173.80024246984976
8	174.6547953046146
9	176.47718204338858
10	177.399201413451

Table 7. RMSE values for WSKT

k	RMSE
2	131.0662924472689
3	132.23843604736848
4	133.25067192123393
5	134.14908195912557
6	134.83392113069306
7	135.39219895571105
8	135.83194679580095
9	136.2125607240411
10	136.30404490964366

Beside RMSE, we used R Square to have a more interpretable accuracy result. The highest result came from WSKT with 91.81%. As for the other stock values are 63.57% for BKSL, 68.6% for WIKA, 58.25% for ADHI, 38.92% for PTPP, 4.5% for BSDE, and -109.25% for LPKR.

The accuracy levels generated by the application were not optimal. Because the data were not scaled, and to find the best *k* value we only used a range from 2 to 10. These parameters are not able to make an elbow shape.

V. CONCLUSION

Based on the experimental results, we can conclude that the LQ45 stock index prediction application using kNR had successfully been built. The accuracy which generated by the application was not too good. Some of the reasons are because the data were not scaled and there is no method to choose the best *k* value. For further research, the data can be scaled, and the *k* value can be optimized. A comparison study also can be done by using other prediction methods, such as WEMA [12], B-WEMA [13], and H-WEMA [14].

REFERENCES

1. Atikah, N. H., Djunaidy, A., and Mahananto, F. (2017). Pembuatan Aplikasi Prediksi Harga Saham Berbasis Web Menggunakan Metode Holt's: Studi Kasus di PT Bank Central Asia Tbk. *Jurnal Teknik ITS*, 6(2), 333-337.
2. Sia, E. R. and S. Hansun. (2014). Rancang Bangun Aplikasi Peramalan Nilai Saham Menggunakan Algoritma Kalman Filter. *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, 3(2), 74-79.
3. Afrianto, R. B., Tjandrasa, H., and Arieshanti, I. (2013). Prediksi Pergerakan Harga Saham Menggunakan Metode Back Propagation Neural Network. *SimanteC*, 3(3), 132–141.
4. Nurhidayat, R. (2009). Pengaruh Variabel Ekonomi Makro Terhadap Return Saham Properti pada Bursa Efek Indonesia. *Kajian Ekonomi dan Keuangan*, 13(2), 1-14.
5. Hansun, S. (2016). A Comparative Study on WEMA and H-WEMA Forecasting Methods in Time Series Analysis (Case Study: JKSE Composite Index Data). *Proc. of The 6th International Annual Engineering Seminar (InAES)*, Yogyakarta, Indonesia.
6. Andrada-Félix, J., Fernadez-Rodriguez, F., Garcia-Artiles, M. D., and Sosvilla-Rivero, S. (2003). An Empirical Evaluation of Non-linear Trading Rules. *Studies in Nonlinear Dynamics and Econometrics*, 7(3).
7. Vainionpaa, I. and Davidsson, S. (2014). Stock Market Prediction Using the K Nearest Neighbours Algorithm and a Comparison with the Moving Average Formula. *Thesis*, KTH.

8. Ban, T., Zhang, R., Pang, S., Sarrafzadeh, A., and Inoue, D. (2013). Referential kNN Regression for Financial Time Series Forecasting. In: Lee M., Hirose A., Hou ZG., Kil R.M. (eds) Neural Information Processing. ICONIP 2013. *Lecture Notes in Computer Science*, vol 8226. Springer, Berlin, Heidelberg.
9. Imandoust, S. B. and Bolandraftar, M. (2013). Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background. *International Journal of Engineering Research and Applications*, 3(5), 605-610.
10. Alkhatib, K., Najadat, H., Hmeidi, I., and Shatnawi, M. K. A. (2013). Stock Price Prediction Using K -Nearest Neighbor (k-NN) Algorithm. *International Journal of Business, Humanities and Technology*, 3(3), 32–44.
11. Hutami, R. and Astuti, E. Z. (2016). Implementasi Metode K-Nearest Neighbor untuk Prediksi Penjualan Furniture pada CV. Octo Agung Jepara. *Thesis*, Universitas Dian Nuswantoro, Semarang.
12. Hansun, S., Kristanda, M.B., and Winarno. (2019). Big 5 ASEAN Capital Markets Forecasting using WEMA Method. *TELKOMNIKA*, 17(1), 314-319.
13. Hansun, S. (2017). WEMA versus B-WEMA Methods in Forex Forecasting. *Proc. of 9th International Conference on Machine Learning and Computing (ICMLC)*, Singapore.
14. Hansun, S. (2019). Hybrid Double Exponential Smoothing Forecasting Method on Forex Data Transaction. *CIENCIA e TECNICA Vitivinicola*, 34(8), 18-25.

AUTHORS PROFILE

Julius Tanuwijaya had just recently graduated from Universitas Multimedia Nusantara in 2019 and received his Bachelor degree in Computer Science. He has actively participated in many events during his study at UMN and finished it without any problem.

Seng Hansun had finished his Bachelor and Master degree from Universitas Gadjah Mada, majoring Mathematics and Computer Science program. Since 2011, he has been a lecturer and researcher at Universitas Multimedia Nusantara and published more than 75 papers both nationally and internationally. His research interests mainly in time series analysis and machine learning domain where he has successfully granted some research grants from the government and UMN institution.