# Optimization of SV-kNNC using Silhouette Coefficient and LMKNN for Stock Price Prediction

1st Frans Mikael Sinaga
*Information Technology Department*
*STMIK Mikroskil*
Medan, Indonesia
frans.sinaga@students.mikroskil.ac.id

2nd Pahala Sirait
*Information Technology Department*
*STMIK Mikroskil*
Medan, Indonesia
pahala@mikroskil.ac.id

3rd Arwin Halim
*Information Technology Department*
*STMIK Mikroskil*
Medan, Indonesia
arwin@mikroskil.ac.id

2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) | 978-1-7281-8406-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ISRITI51436.2020.9315516

***Abstract*— A broker often finds it difficult to decide to buy or sell shares. Due to the large amount of stock data and errors in analyzing stock indicators, it can cause the broker to suffer losses. Predicting stock prices with high accuracy becomes onerous. Although unnecessary variables have been excluded but there are still outliers. It is still difficult to attain the most optimal class value at the data clustering stage. The objective of this research is to propose a model to refine the accuracy of stock prediction and to produce the best class at the data clustering stage. The proposed model is named SV-kNNC - Silhouette Coefficient - SOM (Self Organizing Map) - LMKNN (Local Mean-based K Nearest Neighbor). First, the best class value is obtained using the Silhouette Coefficient, then the data is weighted, finally the stock data is classified using the LMKNN to secure the prediction results. In the testing section, SV-kNNC - SOM + Silhouette Coefficient - LMKNN is compared against SV-kNNC - SOM - KNN (K Nearest Neighbor). Confusion Matrix is used during the test to get the predictive accuracy value. It can be seen from the results of the prediction accuracy of BBCA's shares is 86.99% with the best K is 3, ASII company is 78.30% with the best K is 3, TLKM company is 85.10% with the best K is 5, BBRI company is 84.73% with the best K is 6, PGAS company is 71.36 % with the best K is 5.**

***Keywords*— *SV-kNNC, SOM, Silhouette Coefficient, LMKNN, Support Vector, Technical Analysis, Nonlinear Time Series.***

## I. INTRODUCTION (*HEADING 1*)

The capital market is a place where companies sell stocks and bonds to raise fund for strengthening company's capital [1]. In every capital market transaction, a broker is faced with the choice of selling or buying shares. Although it can be predicted by technical analysis or fundamental analysis, errors often occur in the process of reading and correlating stock values [2].

Predicting stock prices with high accuracy becomes onerous. Although unnecessary variables have been excluded but there are still outliers [3]. Data clustering techniques are very widely used in data mining. The K-Means clustering technique unable to produce good quality clusters with high dimensional data. If it is observed from the standard deviation and structural errors, the SOM (Self Organizing Map) are smaller [4]. However, it is still difficult to know the optimal class value at the data cluster stage [5].

Previously, research on stock price predictions had also been carried out by using the SV-kNNC (Support Vector - k Nearest Neighbor Clustering) and SOM (Self Organizing Map) techniques to predict if stock prices going up or down in the presence of comparison at the clustering stage. The results of this research show that SV-kNNC and SOM are effective models for predicting the direction of stock price whether up or down and their performance is better than SV-kNNC and K-Means [6]. Research was also carried out using modified KNN, namely EEMD (Ensemble Empirical Mode Decomposition) - MKNN (Multidimensional k-Nearest Neighbor). The results obtained show that the EEMD-MKNN model has a higher accuracy of estimation results compared to the KNN, namely MAPE EEMD-MKNN 0.4997 and MAPE KNN 0.5955 [7].

To solve the problem of the SV-kNNC and SOM models, it is necessary to optimize the clustering stage to obtain the best class value when automatically clustering stock data [6]. Proposing the Silhouette Coefficient method, this technique uses the square of the distance between the sample points in each cluster and the center of mass of the cluster to provide a series of K values to improve SOM performance at the data cluster stage [8]. From the SV-kNNC and SOM models, it is also necessary to optimize the data classification stage to increase the accuracy of stock price predictions. To enhance the value of the accuracy of stock predictions, a classification method of LMKNN (Local Mean-Based K Nearest Neighbor) has been proposed. This method is called average-based closest K-neighbor which is a simple and effective nonparametric classification. [9].

## II. LITERATURE STUDY

SV-KNNC is an algorithm consisting of SVM, K-Means, and KNN. The clustering output in this research still needs improvement because the performance of K-Means is unable to produce efficient data clusters which can be seen from the poor accuracy value obtained from several tests. SV-KNNC uses the Support Vector Machine (SVM) to obtain support vector (SV) from training data located near the hyperplane to classify test data. Therefore, the SV obtained is an ideal representation of training data [6]. A re-examination of the SV contribution on the classification is done in the initial dimensions before using it to perform the test data. Weights are used to determine the contribution of each SV. The instance (SV) with a higher weight is more reliable and has a greater influence on the KNN [10]. In the SV-kNNC and SOM models, comparison is carried out at the data clustering stage. From the results of this research, SOM is better for data clustering. The weakness of this research is that it is still difficult to determine the best class at the data clustering stage. Additionally, the prediction results produced are not optimal [6]. Before Support Vector is clustered, the best class is determined using the Silhouette Coefficient. Then this best class is used for the SOM process in data clustering. After the data is weighted, LMKNN is used for clustering in order to increase the accuracy of stock predictions. The proposed model can obtain the best K parameter at the data clustering stage. Therefore, the stock prediction results can be better.

### *A. Support Vector Machine (SVM)*

Statistical learning theory has methods which can provide better results compared to other methods. SVM performance is good on datasets with high dimensions. Kernel using SVM has to map the original data to a fairly higher dimension. SVM works differently with ANN because SVM does not select all data during the training process. SVM only selects some data during the model formation that will be used in the learning classification. Compared to Nearest Neighbor, SVM does not store all training data for predictions. Only a small portion of training data is stored by SVM for the prediction. These facts are the good quality of SVM because only some training data is used in each training iteration. The method is named Support Vector Machine because the contributing data is Support Vector. SVM contains some kernels which are:

*1) Polynomial Degrees*

In order to solve classification problem, the polynomial kernel trick is used. The kernel trick is stated as following:

$$K\ (x_i\ x_j) = (x_i\ x_j + 1)^h \quad (1)$$

*2) Radial Basis Function*

Some datasets are not linearly separated. This causes classification problem which is usually solved by using the kernel of the radial base function trick. This kernel is chosen because it can provide accuracy of training and prediction. The kernel radial base function is stated as following:

$$K\ (xi,\ x_j) = e^{-\|xi-xi\|\ \ 2/2a^2} \quad (2)$$

Where: Xi and Xj = Pair of two training data.

*3) Sigmoid (Hyperbolic Tangent)*

The sigmoid kernel is a kernel trick. This kernel is declared in Support Vector Machine.

$$K\ (x_i,\ x_j) = \tan\ (k\ x_i.x_j-\ \delta\ ) \quad (3)$$

In SVM to produce the best vector, it can be affected by the kernel used when handling nonlinear data. Radial Basis Function Kernel is the only kernel trick which can give best results in grouping data on SVM. This kernel is best for data that cannot be separated linearly.

## *B. An Elbow Method Algorithm*

To generate a series of K values, you can use the square rule using the square of the distance between the sample points in each cluster and the center of mass of the cluster. From the Sum of Squared Errors (SSE) value will determine the K value that will be used [8]. This will determine the convergence of the cluster values obtained. A rapid drop is displayed when the cluster number is set closer to the actual cluster number.

Algorithm 1: *Silhouette Coefficient* [3]
**Input:** *iris* = *datasets.load_iris*(), *X* = *iris.data* [:, 2 :]
**Output:** *d*, *k*
1: *d* = [];
2: **for** *k* = 1, *k in rang* (1, 9) **do**
3 : $d = \sum_{i=1}^{x} \sum dist\ (x, C_1)^2$
4: return *d*, *k*;

Choose the centroid of K = 9 pieces haphazardly from the dataset, compute the SSE by calculating the distance of the smallest square of each data to the centroid, normalizing the SSE list obtained, the elbow point (See the most nosedive) is at what K.

## *C. Selft Organizing Map (SOM)*

SOM is one of the methods found in Neural Networks that is used to visualize data by reducing data dimensions using self-regulating neural networks. The goal is that we understand high dimensional data that is mapped into the form of low dimensional data [12]. In the SOM algorithm, the weight vector of each cluster unit serves as an example of the input pattern associated with the cluster. The unit group weights that resemble the closest input vector pattern (usually the minimum Euclidean distance square) will be taken as the winner. Winning units and neighboring units will continue to update brand weights. Each output will react to a certain input pattern. Thus, the results of Kohonen SOM will show similarities in characteristics between members in the same cluster [11]. The process in SOM Grouping contains two main processes, namely initialization and training. Here's the process at SOM:

*1) Initialization*

The starting vector's weight going into SOM is initialized using a random-based technique. Additionally, the map's size is calculated using the input parameters' width and height related to the dimension grid.

*2) Best Matching Unit (BMU)*

Identifying BMU is done through the closest node distance to the environment with Euclidean distance. The Euclidean distance is written as follows:

$$d_{ij}\sqrt{\sum_{k=1}^{n}(X_{ki} - X_{kj})^2} \quad (4)$$

Where: $d_{ij}$ = distance between instant i and j

$X_{ki}$ = sample values j and j

*3) Update Weight*

The node's weight is updated in the BMU's radius enviroment. Therefore, the environment's size around contracts for each iteration using a Gaussian.

$$\sigma\ (t) = \ \sigma_0\ \exp(-\frac{t}{\pi}\ .t) = 1,2,3, \ldots. \quad (5)$$

Where:
$\sigma\ (t)$ = radius at time $t_0$
t = Current time step, t = 1, 2, 3
π = constant time.

The neuron Update Value is defined.

$$Wi' = W2 + Wi(Xi\text{-}Wi) \quad (6)$$

Where: Wi is weight, Xi = neuron

*4) Repetition*

The formed clusters' quality has been checked by using some well-known measures for internal criteria which are Silhouette index, Davies-Moldin index, and Dunn index. Internal criteria are processed because process validation relies on information taken from the data itself. So, the internal validation's results are calculated according to the cluster elements' proximity (cohesion) and the distance between cluster groups (separation). An ideal cluster is formed from the minimum distance of cohesion and the maximum distance between the experimental distance.

## *D. Local Mean Based K-Nearest Neighbor (LMKNN)*

The K-Nearest Neighbor (KNN) clustering is a very well-known and simple non-parametric technique for pattern clustering. However, the clustering is easily influenced by existing outliers especially in situation with small sample size. Local Mean Based K- Nearest Neighbor (LMKNN) is the enhancement of KNN [9]. The LMKNN has been proven to improve clustering performance and also reduce the effects of existing outliers, especially in small data size. The LMKNN process can be described as follows [13].

Step 1: Determination of K Value

Step 2: Calculate the distance between the test data and each of the training data using the Euclidean distance

$$D(x,y) = ||x-y||_2 = \sqrt{\sum_{j=1}^{N} |x-y|^2} \quad (7)$$

Step 3: Sort data distances from smallest to largest as many as K for each data class

Step 4: Calculate the local mean vector of each class

$$w_c = argmin_{wj}\left(x, m_{w_j}^k\right), j = 1,2,\ldots,M \quad (8)$$

Step 5: Determine the test data class by calculating the closest distance to the local mean vector of each data class

$$m_{w_j}^k = \frac{i}{k}\sum_{i=1}^{k} yi^N, j^N \quad (9)$$

### E. Confusion Matrix

Confusion matrix is also named as contingency table which matrix could be randomly large. The amount of correctly clustered instances is the sum of diagonals in the matrix. Confusion matrix is applied to compute precision and recall [14]. Precision is equal to the ratio of correctly identified items (true positives) and all returned (sum of true positives and false positives). The equation is written as follows [15]:

$$\text{Fmeasure} = 2x\frac{precision\ x\ recall}{precision+recall} \quad (10)$$

### F. Technical Indicator

#### 1) Price Rate of Change (PROC)

Price Rate of Change (PROC) indicates the difference between recent prices and past prices. PROC is written in percentage. PROC is similar to momentum indicator. The difference is that momentum indicator is written in ratio. PROC formula is written as follows:

$$\text{PROC} = \left(\frac{((Today's\ Close - Close\ x\ Periods\ ago}{Close\ x\ Periods\ ago}\right) \quad (11)$$

Where:

Today's Close = Current closing price
Close x Periods ago = The price of the past time is in accordance with the specified period

#### 2) Exponential Moving Average (EMA/XMA)

EMA/XMA is a variant of moving average (MA). EMA gives weight to the recent price relative to the starting price of the EMA calculation at a selected time span. Shorter time span is preferred in order to provide the newest average value of securities. [18]

$$\text{XMA} = K * (C-P) + P \quad (12)$$

$$K = \frac{2}{n+1} \quad (13)$$

Where:

*K = Smoothing constant*
*C = current close price*
*P = previous XMA*
N = number of periods XMA

#### 3) Relative Strength Index (RSI)

To obatain the ratio between the attractiveness of rising and falling prices, RSI can be used. RSI, ranging from 0-100, can check whether a price is overbought or oversold. When the RSI is worth very high (exceeding 70), it may indicate the market is overbought. Hence, the potential to fall is high. Consequently, it is a good moment to sell. On the contrary, when the RSI is worth very low (under 30), it may indicate the market is oversold. Thefore, it is a good moment to buy because there might be a potential to rise.

$$\text{RSI} = \left(\frac{X}{Y}\right) * 100 \quad (14)$$

Where:

X = current closing price
Y = The last Closing price according to the specified period

## III. Methodology

Only support vectors (SVs) is used in support vector machine to cluster unseen data. The SVs used for example are those closest to the hyperplane. Hence, the selected SVs represent training data which is enough to represent all training data. In the selection process, all of the training data are inserted into SVM. Only a set of SVs is produced as the result of SVM training. These SVs are then kept in memory. The process also eliminates repeating instances in each class from the feature space.

Re-checking the SVs' contribution to the clustering in the input space is needed before using them to cluster a query instance in order to avoid problem. In order to establish the contribution of each SV, a weight will be assigned. A higher weight instance is better so it will be more useful in LMKNN clustering.

$$w_i = \frac{n(class(x_i))}{Total} \quad (15)$$

where i = 1,…, m, for n (class (xi)) is the number of clusters that have the same class and Total is the number of samples in that cluster.

Before data grouping is carried out, the Silhouette Coefficient model will determine the best class, where the class values will be used for the data grouping process in the SOM. The SV that has been weighted will be classified using the LMKNN model to see the prediction results. Figure 1 (a) Figure before optimization that still uses SV-kNNC + SOM. Figure 1 (b) Figure after optimization, namely SV-kNNC + Silhoutte Coefficient + LMKNN.

SV-kNNC-Silhouette Coefficient-LMKNN method will be used to predict the first price, which is shown in Fig. 1.

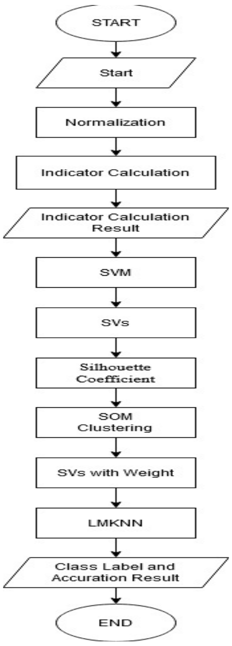


Fig. 1. SV-kNNC- Silhouette Coefficient -LMKNN Model

The steps for price prediction are:

Step 1: normalize stock data, namely (open, high, low, close)

Step 2: Calculate 10 (ten) technical analysis indicators used, namely Exponential Moving Average10. Exponential Moving Average20, Exponential Moving Average30, Stochastic Oscillator14, Stochastic Oscillator19, Stochastic Oscillator 30, Relative Strength Index14, Relative Strength Index19, Rate of change and momentum, ADO / CLV (Accumulation / Distribution).

Step 3: Apply SVM to obtain the best vector for each stock data. The obtained Support will determine each stock data's class.

Step 4: Before data grouping is done, the Silhouette Coefficient process will be done first to find out stock data's best class. This class will be used automatically for the data grouping process, as shown in Fig. 2.

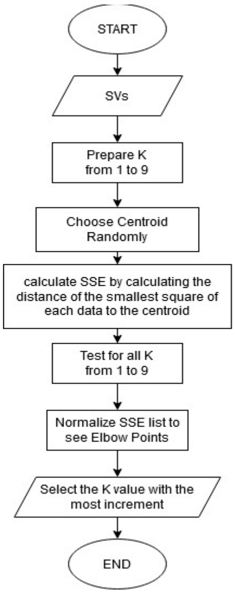


Fig. 2. Flowchart to Determine K Value using the Silhouette Coefficient

Fig. 2 shows a flowchart to get the K Value by using the Silhouette Coefficient. The input data for the Silhouette Coefficient process is the Support Vector of each stock data obtained from the previous preprocessing results. Then, a random K = 9 centroid is chosen from the dataset, SSE is computed by calculating each data's smallest square to the centroid. The SSE (Sum Squared Errors) list that has been obtained will be normalized to find out which elbow point (See the most nosedive) is located at which K.

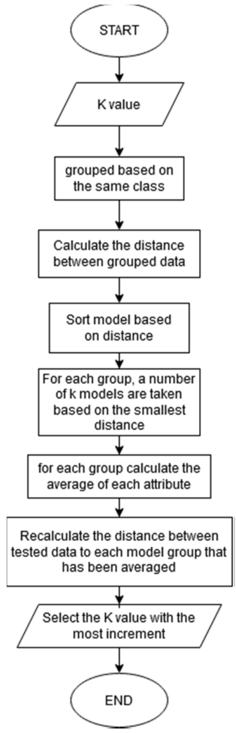


Fig. 3. LMKNN Flowchart

Fig.3 is the flowchart for data classification process using LMKNN. Support Vector is equipped with weight value for each data. Each data has a class too. The distance between these data will be computed. The mean is computed, and the class value is calculated in LMKNN process to enhance the stock price prediction's value.

Adaptive linear combiner formula is used in the normalization process. The daily stock exchange price taken from the dataset serves as the process' input.

$$y = \frac{2 * x - (\max + \min)}{(\max + \min)} \quad (16)$$

After normalization, daily stock price's technical indicators will be calculated. Ten indicators are used in this research which are: Exponential Moving Average10, Exponential Moving Average20, Exponential Moving Average30, Stochastic Oscillator14, Stochastic Oscillator19, Stochastic Oscillator30, Relative Strength Index14), Relative Strength Index19, Rate of change and momentum and ADO/CLV (Accumulation / Distribution). Data training is the combination of raw dataset and technical indicator. Then, SV's vector from the SVM model will be calculated. After the best vector is obtained, the best cluster parameters are determined in the data clustering process using the Silhouette Coefficient. After the best cluster is obtained, the data is clustered using SOM. After the stock data is clustered, the weight of each stock data will be set. Stock data that has been weighted will be classified using the LMKNN model to increase stock price prediction's accuracy.

## IV. Experiment and Result

Stock information from 2rd January 2015 until 17th June 2020 (totally 1376 days) of 5 blue-chip stocks in Indonesia Stock Exchange is used in this experiment. The stocks used are Telkom Indonesia, Astra Internasional, Bank Central Indonesia, Perusahaan gas Negara and Bank Rakyat Indonesia. They are chosen due to the stable income and large assets.

The dataset consists of the daily opening price, closing price, highest price and lowest price. The data is taken from Yahoo Finance website. BBCA stock rates' sample data is shown in Table I.

TABLE I. THE SAMPLE DATA OF BBCA STOCK RATES

| Date | Open | High | Low | Close |
|---|---|---|---|---|
| 02/01/2015 | 13275 | 13275 | 13150 | 13225 |
| 05/01/2015 | 13150 | 13200 | 13125 | 13200 |
| 06/01/2015 | 13000 | 13200 | 13000 | 13100 |
| 07/01/2015 | 13050 | 13200 | 13050 | 13125 |
| 08/01/2015 | 13125 | 13150 | 12975 | 12975 |

The model proposed to solve this problem is a combination of the SV-kNNC + Silhouette Coefficient + LMKNN model to predict the rise or fall of stock prices.

During the SOM step, class / cluster parameter values are generated automatically using Silhouette Coefficient model. Only iteration parameter values are needed in order to understand the obtained results' effect. To know which parameters is best for each stock value, scenarios will be implemented to each dataset.

In this research, predicted and actual values are compared using F-Measure. It is common to obtain high score of F-measure. The F-measure's comparison of LMKNN and KNN is shown in Table II. The F-measure's comparison compares the prediction accuracy results. Then, the best class score is shown for each company after the usage of Silhouette Coefficient model. The Silhouette Coefficient model generates the cluster values automatically. The iteration values is the multiplication of five.

TABLE II. THE F-MEASURE'S COMPARISON OF LMKNN DAN KNN

| Iteration | LMKNN | | | | |
|---|---|---|---|---|---|
| | TLKM | ASII | BBCA | PGAS | BBRI |
| 5 | 86.23 | 79.41 | 87.99 | 72.52 | 85.96 |
| 10 | 86.23 | 79.41 | 87.99 | 72.52 | 85.96 |
| 15 | 86.23 | 79.41 | 87.99 | 72.52 | 85.96 |
| 20 | 86.23 | 79.41 | 87.99 | 72.52 | 85.96 |
| 25 | 86.23 | 79.41 | 87.99 | 72.52 | 85.96 |
| 30 | 86.23 | 79.41 | 87.99 | 72.52 | 85.96 |
| 35 | 86.23 | 79.41 | 87.99 | 72.52 | 85.96 |
| 40 | 86.23 | 79.41 | 87.99 | 72.52 | 85.96 |
| 45 | 86.23 | 79.41 | 87.99 | 72.52 | 85.96 |

| Iteration | KNN | | | | |
|---|---|---|---|---|---|
| | TLKM | ASII | BBCA | PGAS | BBRI |
| 5 | 81.47 | 75.69 | 73.82 | 55.07 | 82.47 |
| 10 | 82.48 | 75.55 | 83.02 | 55.77 | 81.25 |
| 15 | 80.83 | 75.66 | 74.08 | 55.40 | 82.45 |
| 20 | 82.89 | 65.68 | 82.53 | 71.90 | 82.13 |
| 25 | 82.59 | 65.95 | 72.78 | 72.18 | 82.85 |
| 30 | 82.52 | 75.73 | 73.96 | 71.78 | 82.46 |
| 35 | 82.29 | 75.66 | 72.87 | 55.65 | 82.12 |
| 40 | 82.65 | 75.93 | 82.36 | 55.07 | 82.00 |
| 45 | 81.08 | 75.21 | 73.14 | 55.07 | 85.23 |

LMKNN model shows convergent result. The accuracy of LMKNN is better in each company used with different parameter iteration. The K value used in KNN and LMKNN model is set equals to two. Cluster parameter produced by Silhouette Coefficient from each company data are: BCA company best cluster = 3, ASII company best cluster = 3, TLKM company best cluster = 5, BBRI company best cluster = 6, and PGAS company best cluster = 5.

## V. Conclusion

In this research, the proposed model of SV-kNNC + Silhouette Coefficient + LMKNN can overcome the weakness of SV-KNNC + SOM. By adding Silhouette Coefficient in the data clustering step, SOM (Self Organizing Map) performance can be maximized to get the best cluster. By implementing LMKNN in the data clustering step, the stock prediction's accuracy obtained can be improved.

## References


[1] Riyanto, E. (2017) 'Peramalan Harga Saham Menggunakan Jaringan Syaraf Tiruan Secara Supervised Learning Dengan Algoritma Backpropagation', Jurnal Informatika Upgris, 3(2), pp. 137–142. doi: 10.26877/jiu.v3i2.1899.

[2] Roy, G. W. (2016) 'Analisis Teknikal Saham Menggunakan Indikator Bollinger Bands Dan Relative Strength Index Untuk Pengambilan Keputusan Investasi', Jurnal Manajemen, 6(1), pp. 63–68. doi: 10.26460/jm.v6i1.202..

[3] Jeon, S. et al. (2016) 'Stock price prediction based on stock big data and pattern graph analysis', IoTBD 2016 - Proceedings of the International Conference on Internet of Things and Big Data, (IoTBD), pp. 223–231. doi: 10.5220/0005876102230231.

[4] Nair, B. B. et al. (2017) 'Clustering stock price time series data to generate stock trading recommendations: An empirical study', Expert Systems with Applications. Elsevier Ltd, 70, pp. 20–36. doi: 10.1016/j.eswa.2016.11.002.

[5] Bansal, A., Sharma, M. and Goel, S. (2017) 'Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining', International Journal of Computer Applications, 157(6), pp. 35–40. doi: 10.5120/ijca2017912719.

[6] Sinaga, F. M. et al. (2019) 'Stock Trend Prediction using SV-kNNC and SOM', Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019. doi: 10.1109/ICIC47613.2019.8985731.

[7] Zhang, N., Lin, A. and Shang, P. (2017) 'Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting', Physica A: Statistical Mechanics and its Applications. Elsevier B.V., 477, pp. 161–173. doi: 10.1016/j.physa.2017.02.072.

[8] Yuan, C. and Yang, H. (2019) 'Research on K-Value Selection Method of K-Means Clustering Algorithm', J, 2(2), pp. 226–235. doi: 10.3390/j2020016.

[9] Mehta, S. et al. (2018) 'A new nearest centroid neighbor classifier based on k local means using harmonic mean distance', Information (Switzerland), 9(9). doi: 10.3390/info9090234.

[10] Srisawat, A., Phienthrakul, T. and Kijsirikul, B. (2006) 'SV-kNNC: An algorithm for improving the efficiency of k-nearest neighbor', Lecture Notes in Computer Science

[11] A. Hafiludien and D. Istiawan, "Penerapan Algoritma Self Organizing Maps Untuk Pemetaan Penyandang Kesejahteraan Sosial (PMKS) di Provinsi Jawa Tengah Tahun 2016," 7th Univ. Res. Colloqium, pp. 84–92, 2018.

[12] J. Boedecker, "Self-Organizing Map ( SOM ) Topological Maps in the Brain," 2015.

[13] Syaliman, K. U., Nababan, E. B. and Sitompul, O. S. (2018) 'Improving the accuracy of k-nearest neighbor using local mean based and distance weight', Journal of Physics: Conference Series, 978(1). doi: 10.1088/1742-6596/978/1/012047.Local Mean Based K- Nearest Neighbor (LMKNN)

[14] S. Visa, B. Ramsay, A. Ralescu, and E. Van Der Knaap, "Confusion matrix-based feature selection," CEUR Workshop Proc., vol. 710, pp. 120–127, 2011

[15] M. Czakon, D. Heymes, and A. Mitov, "High-Precision Differential Predictions for Top-Quark Pairs at the LHC," Phys. Rev. Lett., vol. 116, no. 8, pp. 1–5, 2016

[16] P. Willett and S. J. Clarke, "Estimating the recall performance of Web search engines," Aslib Proc., vol. 49, no. 7, pp. 184–189, 1997.

[17] Y. Shynkevich, T. M. McGinnity, S. A. Coleman, A. Belatreche, and Y. Li, "Forecasting price movements using technical indicators: Investigating the impact of varying input window length," Neurocomputing, vol. 264, pp. 71–88, 2017.

[18] V. A. Sobreiro et al., "The profitability of moving average trading rules in BRICS and emerging stock markets," North Am. J. Econ. Financ., vol. 38, pp. 86–101, 2016.

[19] Henrique, B. M., Sobreiro, V. A. and Kimura, H. (2018) 'Stock price prediction using support vector regression on daily and up to the minute prices', The Journal of Finance and Data Science. Elsevier Ltd, 4(3), pp. 183–201. doi: 10.1016/j.jfds.2018.04.003