

# Forecasting Feature Selection based on Single Exponential Smoothing using Wrapper Method

Ani Dijah Rahajoe

Faculty of Engineering, University of Bhayangkara Surabaya  
Surabaya, Indonesia

**Abstract**—Feature selection is one way to simplify classification process. The purpose is only the selected features are used for classification process and without decreasing its performance when compared without feature selection. This research uses new feature matrix as the base for selection. This feature matrix contains forecasting result using Single Exponential Smoothing (FMF(SES)). The method uses wrapper method of GASVM and it is named FMF(SES)-GASVM. The result of this research is compared with other methods such as GA Bayes, Forward Bayes and Backward Bayes. The result shows that FMF(SES)-GASVM has maximum accuracy when compared of FMF(SES)-GA Bayes, FMF(SES)-Forward Bayes, FMF(SES)-Backward Bayes, however the number of selected features are more than if compared with FMF(SES)-GA Bayes and FMF(SES)-Forward Bayes.

**Keywords**—Single exponential smoothing; forecasting; feature selection; genetic algorithm

## I. INTRODUCTION

The number of variables of features would influence the time and performance in data mining process. One of the techniques to improve efficiency of mining data result especially in terms of process time was by selecting feature. In feature classification process, the selected features were those which had a better classification result or the same with the classification result without feature selection. The other benefit was faster process time because not all features were used. Feature selection had several methods namely filter, embedded and wrapper [1]. In filter method, the process of its feature selection was separated with its classification process. The qualified features were sought with the purpose to be able to influence the training result without looking at the mechanism of the training. Only the relevant features were used in the training. In wrapper method, learning machine was used as a black box to count feature-subset based on the level of its prediction [2],[3]. In embedded method, the evaluation process was to connect or link to a task (for example classification) from algorithm learning. In this embedded method, a sub-algorithm feature selection explicitly or implicitly was an integrated part in a more general algorithm learning [4].

In the field of forecasting, when there were more features were used, it certainly would influence its access time. Only feature with better accurate time would be selected. The research in this paper was different because the feature selection was conducted using machine learning process or wrapper method where the features were previously placed in a matrix of forecasting result. If the research [5] conducted

forecasting from the result of the selected features, in this research we used the forecasting result to select relevant features which resulted in a better accuracy than without feature selection. In this research, data daily demand forecasting orders from UCI machine learning repository was also used.

The use of Single Exponential Smoothing (SES) was especially to form its feature matrix. Feature selection method used wrapper method. The idea was how to transform forecasting data into a feature matrix which then was continued with feature selection process of wrapper method. Forecasting Algorithm of this research used Single Exponential Smoothing (SES) while its wrapper method used Genetic Algorithm – Support Vector Machine (GASVM). The use of wrapper method was because the accuracy was better compared with filter method, although the process time of filter method was faster compared with wrapper method [6]-[11]. The selection of GASVM Algorithm was because it was better than PSOSVM [12]. This research was also compared with other feature selection algorithm such as algorithm-Naive Bayes, forward-Naive Bayes approach, backward-Naive Bayes approach, and also without feature selection. The comparison without through feature selection meant that after through feature matrix formation, classification process was directly conducted.

The contributions of this research were:

- a) A proposal to form feature matrix through forecasting result was called Feature Matrix of Forecasting (FMF) using Single Exponential Smoothing (SES).
- b) Feature Selection using wrapper GASVM method was conducted after feature matrix formation called FMF(SES)-GASVM.
- c) The comparison of classification result with or without feature selection, and algorithm of other feature selections.

The rest of this paper was organized as follows: Section 2 presented the related work and Section 3 provided the materials and the proposed research method. Section 4 presented the experimental result and Section 5 concluded the paper.

## II. RELATED WORK

Several studies have used neural network to find the result of the selected features [5], [13]-[17]. However, the feature selection process is different, some use filter method, wrapper method and hybrid method.

Feature selection using filter method tends to process the original data first. The first feature with original data is conducted with clustering or classification using certain algorithm to select features [5], [13]. Then the selected features are classified to find their accurate result. For example, in the research [5] where the research used feature selection of filter method. The feature selections by using relief, PCA, discrete transformation of data (ReliefF, Information Gain, and Clustering K-means). The result of the selected features from different algorithm in the ranking has many similarities to the selected features, then the process using Elman Network (EN), Fuzzy System (Subtractive Clustering) (FSSC), Adaptive Neuro-Fuzzy Inference System. (ANFIS), FasArt (FART) is conducted. While the research with forecasting used multilayer perceptron (MLP) from selected features as its node [15]. Mutual information and K-nearest neighbors was also used as features selection process which would be predicted using neural network [15].

In [14], features selection research was conducted for time series forecasting using combination of filter wrapper. An interactive neural filter was proposed for feature evaluation to automatically identify the frequency of the time series, embedded in wrappers for feature construction, feature transformation and architecture selection. Time series data was conducted to calculate penalized Euclidean distance and the result was applied in wrapper method neural network to identify its seasonality. The use of neural network would require longer access time so that the combination between feature selections of filter method could cover the lack of access time. This was because the features used were less so that it can reduce the access time. Selection of features with a neural network is included in the wrapper method [18]. The use of neural network could also be used in the research of [13]. The research uses Improved Principal Component Analysis (IPCA) and the selected features were conducted through forecasting process using Neural Network Ensemble (NNE) and it was called NNEIPCABag. The feature selection process was done first so that the method was filter method, while the forecasting process uses NNE, because if the method was wrapper, the feature selection process was included in its learning machine.

The hybrid method can be a combination of a filter-wrapper method or wrapper-filter method. In [19], The hybrid method used is the filter-wrapper method. ACF combination (AutoCorrelation Function) and LSSVM (Least Squares Support Vector Machines) are used to build the electricity load forecast. In [20], researchers proposed a hybrid model based on features of WOA for the prediction of  $PM_{2.5}$  concentration.

### III. MATERIALS AND METHODS OF RESEARCH PROPOSED

Feature selection of wrapper method means that the selected features use machine learning as black box to count the level of its prediction [2], [3]. Therefore, it requires a process related to its machine learning.

#### A. Feature Selection of Wrapper Method

In this research, its wrapper method uses GASVM algorithm. Generally, the researchers select features from first data and then the result is conducted with forecasting process.

This kind of research, its selected features are not based on the result of forecasting accuracy but the result of feature selection is based on variables which may be considered relevant. In this paper, the result of forecasting is used as the bases of feature selection

#### B. Dataset

This research used data daily demand forecasting orders from VCI Machine learning repository (<https://archive.ics.uci.edu/ml/datasets/DailyDemandforecastingOrders>). The dataset was collected during 60 days; this is a real database of a Brazilian logistics company. This dataset has 13 attributes / features and 60 data. The dataset has 12 predictive attributes and a target that is the total of orders for daily treatment. The variable is week of month, day of the week, non-urgent order, urgent order, order type A, order type B, order type C, fiscal sector, orders from the traffic controller sector, banking orders 1, banking orders 2, banking orders 3, and target (total orders). Variable week of the month and day of the week are the data of observing day, while other variables contain value and target orders. Variables of days and months are extracted into data per day. Variable means the same as feature.

#### C. Single Exponential Smoothing

Forecasting is an action to read a behavior of future time from observation of the past time. Forecasting the value of the future time nowadays is extensively required to prepare anything happens, like in forecasting population, energy, economy. In civil population, the forecasting is conducted to estimate the number of people for several years in the future. For energy, estimation is needed to be able to know how much energy reserve available. In economy, global or local economy trend is needed to predict market behavior is the future time. Various techniques of forecasting except classical statistic, many now use soft computing [21]-[23].

Exponential Smoothing is simple method in its operation, low cost, extensive in adaptability and good in performance [24]. The exponential Smoothing method is frequently used as prediction method. Exponential Smoothing is a simple approach for forecasting. Forecasting is built from an average weight exponentially from past observations [25].

Principally, there are 3 techniques of exponential smoothing [26]. Exponential Smoothing technique is simple exponential technique [27], Holt's exponential smoothing [27], and Holt-Winters method [28]. Simple exponential smoothing needs a little computation and is used at the time of data pattern neither have not a cyclic variation nor trend in the historical data [29], [26]. Holt's method is known as double exponential smoothing method, using time series contains a trend. Holt-Winters technique can capture trend and seasonality in historical data [26].

Single Exponential Smoothing Method gets smoothing curve using one past data and one prediction of past data. Let an observed time series be  $y_1, y_2, \dots, y_n$ . The equation of simple exponential smoothing is shown as follows:

$$\hat{y}_{i+1} = \alpha y_i + (1-\alpha) \hat{y}_i \quad (1)$$

Where  $\hat{y}_{i+1}$  represents the predicted value at the time that  $i+1$ .  $\hat{y}_i$  is forecast value of variable  $Y$  at the time of period  $i$  and  $y_i$  is actual value at the time of period  $i$ ,  $\alpha$  is smoothing constant [27]. When starting algorithm, it needs initial forecast namely an actual value and a smoothing constant [25]. Smoothing constant has value between 0 and 1. This is because  $\alpha = 1$ , so the actual value and smoothed version of series are identical, while if  $\alpha = 0$ , the series is smoothed flat. When there is a difference between forecast value during the period of  $i + 1$  and the period of  $i$ , so it is shown in the following (2).

$$\hat{y}_{i+1} - \hat{y}_i = \alpha(y_i - \hat{y}_i) \quad (2)$$

When residual  $e_i = y_i - \hat{y}_i$  is forecast error during  $i$ , so it is shown on the equation as follows (3):

$$\hat{y}_{i+1} = \hat{y}_i + \alpha e_i \quad (3)$$

Therefore, exponential smoothing is old forecast plus an adjustment for error that occurred in the last forecast [30][31][25]. Then with the past forecasting value, so it is shown in the following:

$$\begin{aligned} \hat{y}_{i+1} &= \alpha y_i + (1-\alpha)[\alpha y_{i-1} + (1-\alpha)\hat{y}_{i-1}] = \\ \hat{y}_{i+1} &= \alpha y_i + \alpha(1-\alpha)y_{i-1} + (1-\alpha)^2 \hat{y}_{i-1}, \\ \hat{y}_{i+1} &= \alpha y_i + \alpha(1-\alpha)y_{i-1} + \alpha(1-\alpha)^2 y_{i-2} + (1-\alpha)^3 \hat{y}_{i-2}, \\ \hat{y}_{i+1} &= \alpha y_i + \alpha(1-\alpha)y_{i-1} + \alpha(1-\alpha)^2 y_{i-2} \\ &+ \alpha(1-\alpha)^3 y_{i-3} + (1-\alpha)^4 \hat{y}_{i-3} \dots \end{aligned}$$

So that forecasting equation in general is as follows (4):

$$\hat{y}_{i+1} = \alpha \sum_{k=0}^{i-1} (1-\alpha)(1-\alpha)^k y_{i-k} \quad (4)$$

Where  $\hat{y}_{i+1}$  is the forecast value of Variable  $y$  during the period of  $i+1$  so that  $\hat{y}_{i+1}$  is the weighted moving average of all past observations [25].

#### D. Forecast Error Management

In forecasting result, it is certainly not always the same with the reality or in the real condition, however, from this forecasting result supporting factors can be prepared to anticipate the risky things. The difference between forecasting result and the result of the real condition is called forecasting error. The comparison of mass difference between forecasting and real result can be used as a reference to determine the needs in the future. Small mass difference shows that the forecasting result comes near the real condition. One of the methods is by using Mean Squared Error (MSE). MSE is a computation of the total forecasting data difference with the real data. MSE is shown as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 \quad (5)$$

Where  $f_i$  is value of forecasting result. Variable  $y_i$  is real value and  $n$  is number of data. MSE is generally used as criteria to select its smoothing constant [30].

#### E. Feature Matrix of Forecasting

Feature selection is based on feature columns which are ready to be used for prediction. In time series research involving many dimensions such as multivariate time series (MTS), data dimension reduction is needed in order to be fit in

the feature columns which then continued by feature selection process. The transformation process into this feature columns is called vectorization [32]. Various ways have been used in order to be fit into that vectorization such as common principle component analysis (CPCA), using matrix correlation [33]. This is because the use of the word "matrix vectorization" refers to long calculation of vector in the research, while the research of [32] mentioned about feature matrix because it is a matrix which contains feature columns. In this research uses the term feature matrix because those matrix contains the selected features.

Feature matrix contains forecasting result of each feature using single exponential smoothing which then the combination of features with high accuracy are selected using wrapper GASVM method. Therefore, preprocessing from dataset is conducted to form its feature matrix.

#### F. Feature Selection of Wrapper GASVM Method

GASVM is a feature selection method which has two steps conducted repeatedly until a certain limit. The first step is feature selection step at random (using genetics algorithm) and the second step is classification process with the calculation of accuracy level from the combination of selected features (using Support Vector Machine). At the time the limit of feature selection process has been fulfilled, so the final result is the selected features which have maximum accuracy value when compared without feature selection.

Genetics Algorithm (GA) used for the application of feature selection has different way with algorithm used for optimization [34]. In the application of feature selection, GA is aimed to select features which will be used for classification process using SVM. Genetics Algorithm has several processes namely encoding of chromosomes, population initialization, fitness function, parents selection, crossover, mutation, population replacement and criteria limit. Population in one generation is the collection of some chromosomes containing the selected and unselected features. The selected features are represented by 1 and 0 if not selected. In this research, fitness function in genetics algorithm for feature selection uses the accuracy value of SVM classification result like the following (6).

$$\text{fitness}(x) = \text{accuracy}(x) \quad (6)$$

Accuracy value is from classification calculation using SVM. When  $a$  is the total of data which has label according to the real data and  $n$  is the whole total of testing data, so the accuracy equation is as follows (7):

$$\text{accuracy}(x) = \left(\frac{a}{n}\right) * 100\% \quad (7)$$

One of classification techniques is by using Support Vector Machines. Its classification usually involves collection of data with several features or attributes and there is a class label [35]. This SVM finds and makes use of margin or optimal separating hyper plane which separates two classes and maximize the closest distance from point to point from different class [35], [36]. A hyper plane which separates two classes with the following (8) [36][37][32]:

$$g(x) = w'x + w_0 \quad (8)$$

Where  $w$  is vector hyper plane  $g(x)$  and  $w_0$  or  $\|w\|$  is the distance from the point of origin to its hyperplane. The details is in [36], [37].

Feature Selection of GASVM Wrapper method starts from the separated data from testing and training data. On the first data, there is initialization data which contains the selected features. Then this feature build classifier model uses SVM. The calculation of its accuracy value uses training data. Fitness evaluation in genetic algorithm process is based on maximum accuracy value (then the selected feature is used as parents), next using crossover process, mutation, population replacement to get a new generation and so on. The process will stop when the criteria restrictions have been fulfilled.

G. Feature Selection of GASVM Wrapper Method from Forecasting Feature Matrix

Before feature selection is conducted, forecasting feature matrix is needed. In every feature, the result of forecasting for some time in the future ( $\hat{y}_{i+1}$ ) is calculated. In one dataset there must be one feature representing the class/label to determine the relevant feature when using GASVM. This is because algorithm support vector machine is a classification which evaluates the selected features based on the highest accuracy value. The steps of this research are:

Step 1. Preparing class/label feature: GASVM is a feature selection method based on the result of SVM classification. Therefore, features containing class/label data are needed.

Step 2. Forecasting of every feature for  $\hat{y}_{i+1}$ : In every feature, forecasting process using single exponential smoothing is conducted. Therefore, there is forecasting feature matrix with one column as a label column. In this research, normalization data are used.

Step 3. Forming forecasting feature matrix: Feature matrix of forecasting is a matrix containing features which will be selected, in which the forecasting data are the result of forecasting.

Step 4. Selection of features with high accuracy level: The selection of features uses GASVM wrapper method. Feature matrix of forecasting contains the result of forecasting using SES. During the process of feature selection with GASVM, it means that the data used is the result of SES forecasting. These selected features will show the accuracy value from the highest classification. Therefore, to know the performance of the selected features, the comparison with the result of classification is required if without the selection of features. The description of this research process is shown in Fig. 1.

IV. RESULT AND DISCUSSION

Single Exponential Smoothing selects smoothing constant with the smallest MSE. In the dataset, there are 11 features with one feature containing label for classification with label value of 0 and 1 (see sub-chapter III.B). Therefore, 10 features which are joined in the selection of features use GASVM. Prior to feature selection process, the dataset needs to be prepared.

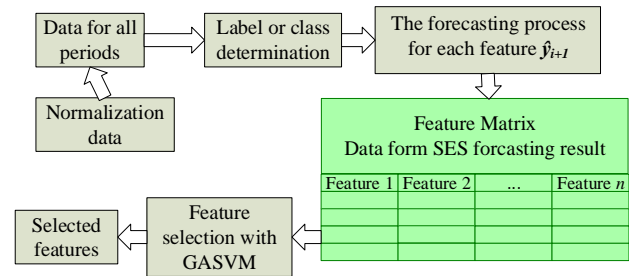


Fig. 1. The Proposed Work Block Diagram.

A. Preparation of Data

The dataset consists of the data of ordering for 5 weeks and 5 days (from Monday to Friday). The dataset is collected within 60 days. This is a real database of a Brazilian Logistic Company. Therefore, all data are used and prepared. This preparation of data includes data preprocessing and data normalization. The preparation of data is necessary because it is possible that the data without being processed first won't be complete and noisy, like there is field of redundancy, zero value or null, there is Outliner or inconsistent value with the valid rule [38]. In this research, the process of data preprocessing when features have no complete data, the empty data will be filled with average data from minimum value with maximum value of the same column data, while the normalization of the data uses MinMax data normalization.

B. Preparation of Class or Label Features

In this research, there are 11 features. One feature is a target or total orders. Feature or attribute of total orders is used as class/label. The selection of attribute or feature of total orders as label is because the attribute is the result of total orders at that time. Therefore, the total orders determine the amount of order which will be executed. The method used in the formation of class/label is by looking at values above or below average. The value of total orders below average is 0 and above average is 1. So there are 10 features left which will be processed for feature selection.

C. Forecasting with Single Exponential Smoothing

All data on features are calculated for their forecasting for some time in future  $\hat{y}_{i+1}$  (see sub-chapter III.G). Then smoothing constant is selected with the smallest MSE value. Table I shows the result of all features calculated their forecasting value using SES. The result can be seen from Fig. 2 to Fig. 6. Fig. 2 shows training data and the forecasting result using SES with smoothing constant having been selected for features of Non Urgent Order and Urgent Order. Fig. 3 shows training data and forecasting result using SES for feature data of Order Type A, Order Type B and Order Type C. Fig. 4 shows training data and forecasting result using SES for feature data of Fiscal Sector Orders and Orders from the traffic. Fig. 5 shows training data and forecasting result using SES for feature data of Banking Orders 1 and Banking Orders 2. Fig. 6 shows training data and forecasting result using SES for feature data of Banking Orders 3 and Target (Total Orders).

TABLE I. THE RESULT OF SMOOTHING CONSTANT AND MSE OF EACH FEATURE

	Features					
	Non Urgent Order	Urgent Order	Order type A	Order Type B	Order Type C	Fiscal Sector Order
Smoo-thing constant (Alpha)	0.1	0.1	0.1	0.1	0.1	0.1
MSE	0.02772	0.0357	0.0279	0.038	0.0313	0.0159
	Features					
	Orders from the traffic	B. Orders 1	B. Orders 2	B. Orders 3	Total Orders	
Smoo-thing constant (Alpha)	0.1	0.5	0.1	0.6	0.1	
MSE	0.0363	0.0276	0.0515	0.0200	0.0318	

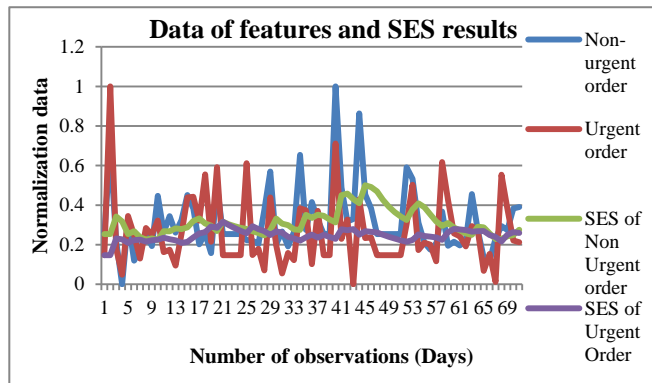


Fig. 2. Training Data and the Forecasting Result using SES with Smoothing Constant having been Selected for Features of Non urgent Order and urgent Order.

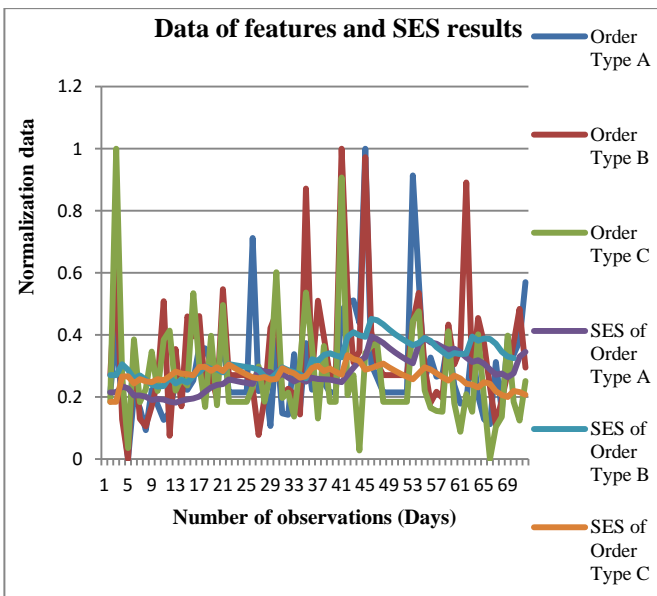


Fig. 3. Shows Training Data and Forecasting Result using SES for Feature Data of Order Type A , Order Type B and Order Type C.

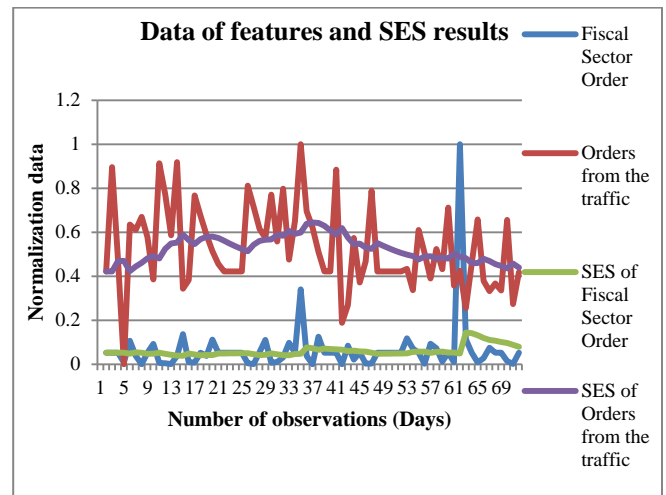


Fig. 4. Training Data and Forecasting Result using SES for Feature Data of Fiscal Sector Orders and Orders from the Traffic.

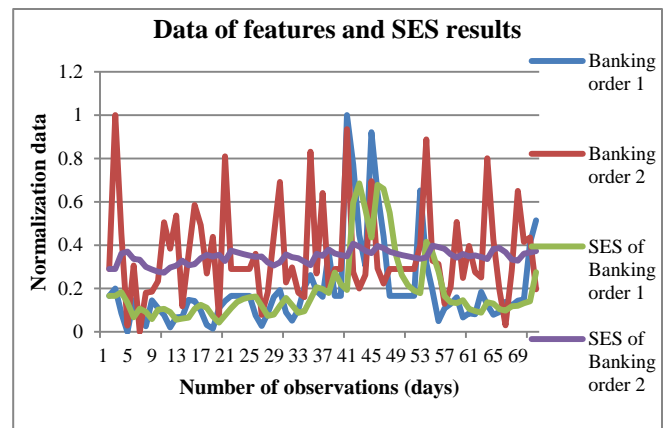


Fig. 5. Training Data and Forecasting Result using SES for Feature Data of Banking Orders 1 and Banking Orders 2.

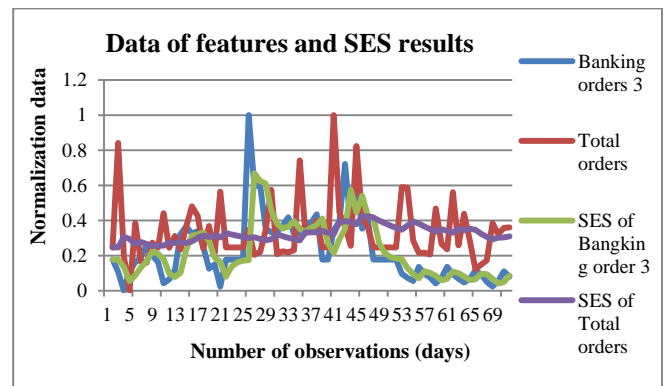


Fig. 6. Training Data and Forecasting Result using SES for Feature Data of Banking Orders 3 and Target (Total Orders).

#### D. Feature Matrix of Forecasting

Feature Matrix of forecasting here is a matrix containing the result of forecasting except for the result of feature forecasting used for label or class. Thus there are 10 features where lines/rows are observation days and the columns are features and one feature of total order is as class or label. The formation of class or label has been stated in sub-chapter IV.B.

### E. Selection of GASVM Feature

Several Parameters of genetic algorithm used in this research are as follows:

- 1) Population uses binary number where 0 is an unselected feature and 1 is the selected one.
- 2) Elitism is conducted in 2 ways, if the number of chromosomes is in odd population, so one duplicate of the best chromosome is kept, but if it is in even number two duplicates of the best one are kept.
- 3) The selection method used is roulette wheel weighting with rank weighting technique.
- 4) Crossover uses one-point crossover with crossover probability equal to 0.8.
- 5) Mutation probability used is 0.05. Mutation is done to all genes in chromosomes for random number GA grown smaller than the mutation probability.
- 6) New population for the next generation is the population as the result of GA process.
- 7) Maximum total of generation is 100; the size of population is 50.
- 8) The condition of terminal is used based on the maximum amount of population or generation evaluated and the maximum level of accuracy (100%).
- 9) The calculation of accuracy level with SVM is based on classification with two classes.

Special for Naïve Bayes on GA Bayes is helped by WEKA application version 3.6.11, whereas FSBLF and GASVM use Matlab version R211b. Parameter for Naïve Bayes uses default Weka, namely, fold 5, seed 1, threshold 0.01, start set-, and for genetic algorithm in Weka application uses the same parameter with the one in Matlab.

This reasearch uses algorithm and comparative method to know the result of several methods of feature selection. Feature selection of wrapper method used in this research is GASVM, GABayes, ForwardBayes and BackwardBayes. GASVM has been previously explained. GABayes is feature selection of wrapper method where features at random are selected using genetic algorithm and their Fitness is based on forward approach. This forward approach is an approach started from first feature in which its classification result is counted using naïve Bayes, then the combination of first feature and second, next is first feature, second and third. After that the features selected are only those with the highest accuracy value. On the contrary with ForwardBayes, on BackwardBayes the features counted, their accuracy starts from combination of all features, then decreasing one by one.

FMF(SES)-GASVM has the most maximum accuracy among the other four methods (without selection of feature, GABayes, ForwardBayes, Backward Bayes), even though the total of the selected features is more than GABayes and ForwardBayes. Table II shows accuracy value of 100 % for FMF(SES)-GASVM with five total number of selected features namely features 3,4,5,6 and 8. If without selection of feature, so there are 10 features with accuracy value of 97.2%. FMF(SES)-GABayes and FMF(SES)-ForwardBayes have less number of selected features that are 3 features, but the accuracy

value is lower for 97%. The less number of features, the faster the classification process, however, the accuracy is more emphasized. The lowest accuracy value is obtained from FMF(SES)-BackwardBayes with accuracy value 96%. Therefore, only features 3,4,5,6 and 8 are used to predict the future. This FMF-GASVM Method can be used in all forecasting dataset and has feature as class or label.

TABLE II. THE RESULT OF FEATURE SELECTION AND ITS ACCURACY PERCENTAGE

Feature selection method	Selected features	Number of selected features	Accuracy (%)
FMF(SES)-GASVM	3,4,5,6,8	5	100
Without feature selection (SVM classification)	-	-	97.2
FMF(SES)-GaBayes	1,4,5	3	97
FMF(SES)-ForwardBayes	1,4,5	3	97
FMF(SES)-BackwardBayes	1,3,4,5,9	5	96.9

### V. CONCLUSION AND SUGGESTION

Features Matrix of Forecasting (FMF) is one of many ways to form features matrix of forecasting result as the first basis of the process of feature selection. So the purpose is to produce selected features with better accuracy value if without selection of feature. The other benefit is faster process time because not all features are used. This research used Single Exponential Smoothing (SES) to determine forecasting value in the future and is used the value on features matrix. The selection of its features used method of Wrapper Genetic Algorithm Support Vector Machine (GASVM). In this research, FMF(SES)-GASVM is compared with several other methods, namely, FMF(SES)-GABayes, FMF(SES)-FowardBayes, FMF(SES)-BackwardBayes and without selection of features. FMF(SES)-GASVM has maximum accuracy (100%) compared with the other four, however the number of selected features is more than FMF(SES)-GABayes and FMF(SES)-ForwardBayes. The result shows that only with features 3.4.5.6 and 8 can be used as forecasting features for future value with 100 % accuracy compared to the other four algorithms.

In this research, forecasting algorithm can be used besides Single Exponential Smoothing such as Double Exponential Smoothing. But it is still forming its matrix forecasting as the basis of feature selection. In addition, it can also use other methods of feature selection like filter or embedded method.

### ACKNOWLEDGMENT

Thank you to the University of Bhayangkara Surabaya for supporting this research.

### REFERENCES

- [1] H. Chouaib, O. R. Terrades, S. Tabbone, F. Cloppet, N. Vincent, and B. P. Nancy, "Feature selection combining genetic algorithm and Adaboost classifiers," in 1th International Conference on Pattern Recognition, 2008, vol. 5, no. Ea 2517, pp. 1-4.
- [2] I. S. Oh, J. S. Lee, and B. R. Moon, "Hybrid Genetic Algorithms for Feature Selection," IEEE TRansaction on Pattern Analysis And Machine Intelligence, vol. 26, no. 11, pp. 1424-1437, 2004.

- [3] R. Kohavi and H. John, "Artificial Intelligence Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 97, pp. 273–324, 2011.
- [4] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Morgan Kaufmann, 2005.
- [5] R. G. Pajares, "Feature Selection for Time Series Forecasting: A Case Study," in *International Conference on Hybrid Intelligent Systems*, 2008, pp. 555–560.
- [6] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, pp. 273–324, 1997.
- [7] C. L. Huang and W. C. J., "A GA-based feature selection and parameters optimization for support vector machines," *Expert Systems with Applications*, vol. 31, pp. 231–240, 2006.
- [8] K. Z. Mao, "Feature subset selection for support vector machines through discriminative function pruning analysis," *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics*, vol. 34, no. 1, pp. 60–67, 2004.
- [9] E. Yu and S. Cho, "Ensemble based on GA wrapper feature selection," *Computers and Industrial Engineering*, vol. 51, no. 1, pp. 111–116, 2006.
- [10] J. Bi, K. P. Bennett, M. a Embrechts, and C. M. Breneman, "Dimensionality Reduction via Sparse Support Vector Machines," *Journal of Machine Learning Research*, vol. 1, pp. 1–16, 2002.
- [11] R. Sikora and S. Piramuthu, "Framework for efficient feature selection in genetic algorithm based data mining," *European Journal of Operational Research*, vol. 180, no. 2, pp. 723–737, 2007.
- [12] E. Alba, L. Jourdan, and E. Talbi, "Gene Selection in Cancer Classification using PSO / SVM and GA / SVM Hybrid Algorithms," *IEEE Congress on Evolutionary Computation (CEC)*, vol. 07, pp. 284–290, 2007.
- [13] L. Jian, "Neural Network Ensemble Based on Feature Selection," in *IEEE International Conference on Control and Automation*, 2007, vol. 00, pp. 1844–1847.
- [14] S. F. Crone and N. K. Å, "Neurocomputing Feature selection for time series prediction – A combined filter and wrapper approach for neural networks," *Neurocomputing Elsevier*, vol. 73, no. 10–12, pp. 1923–1936, 2010.
- [15] S. F. Crone and S. Häger, "Feature Selection of Autoregressive Neural Network Inputs for Trend Time Series Forecasting," in *International Joint Conference on Neural Networks*, 2016, pp. 1515–1522.
- [16] L. I. U. Tianhong, W. E. I. Haikun, Z. Kanjian, and G. U. O. Weili, "Mutual information based feature selection for multivariate time series forecasting," in *Proceedings of the 35th Chinese Control Conference*, 2016, pp. 7110–7114.
- [17] M. Maca and L. Lhotsk, "Wrapper Feature Selection Significantly Improves Nonlinear Prediction of Electricity Spot Prices," in *IEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 1171–1174.
- [18] M. Rana, I. Koprinska, and A. Khosravi, *Artificial Neural Networks*, Springer S., vol. 4. Switzerland: Springer International Publishing, 2015, pp. 445–446.
- [19] A. Yang, W. Li, and X. Yang, "Knowledge-Based Systems Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines," *Knowledge-Based Systems*, vol. 163, pp. 159–173, 2019.
- [20] P. M. Concentration, F. Zhao, and W. Li, "A Combined Model Based on Feature Selection and WOA for PM2.5 Concentration Forecasting," *Atmosphere, MDPI*, no. 2, pp. 1–20, 2019.
- [21] L. A. M, M. Beni, D. Nieto, C. Di, and F. A. Sa, "Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models," *Elsevier*, vol. 32, pp. 1218–1225, 2007.
- [22] D. Sonika, S. D. S, and K. Daljeet, "Long-Term-Load-Forecasting-Using-Soft-Computing-Techniques," *International Journal of Scientific & Engineering Research*, vol. 6, no. 6, pp. 450–457, 2015.
- [23] S. Chen, S. Member, and J. Hwang, "Temperature Prediction Using Fuzzy Time Series," in *IEEE Transactions on system, Man, and Cybernetics*, 2000, vol. 30, no. 2, pp. 263–275.
- [24] A. M. Systems, "Variable smoothing parameter of the double exponential smoothing forecasting model and its application Chen Feiyan, Li Qian, Liu Junhong and Zhang Jiuyong," in *International Cobference on Advance Mechatronic Systems*, 2012, pp. 386–388.
- [25] E. Ostertagova and O. Ostertag, "The simple Exponential Smoothing Model," in *The 4th International conference Modelling of Mechanical and Mechatronic Systems*, 2011, pp. 380–384.
- [26] L. Abderrezak, "Very Short-Term Electricity Demand Forecasting using Adaptive Exponential Smoothing Methods," in *15th International Conference on Science and Techniques of Automatic Control and Computer Engineering*, 2014, pp. 553–557.
- [27] R. G. Brown, *Statistical forecasting for inventory control*. New York: McGraw-Hill, 1959.
- [28] P. R. Winters, "Forecasting Sales by Exponentially Weighted Moving Average," *Management Science*, vol. 6, no. 3, pp. 231–362, 1960.
- [29] M. Mordjaoui, B. Boudjema, and M. Bouabaz, "Electric load consumption prediction using statistical approaches," *International Conference on Applied Analysis and Algebra*, pp. 20–24, 2012.
- [30] A. D. Azcel, *Complete Business Statistics*. Irwin, 1989.
- [31] E. Ostertagova, *Applied Statistic (in Slovak)*. Elfa, Kosice, 2011.
- [32] K. Yang, H. Yoon, and C. Shahabi, "A Supervised Feature Subset Selection Technique for Multivariate Time Series," in *International Workshop on Feature Selection for Data Mining*, 2005.
- [33] H. Yoon, K. Yang, and C. Shahabi, "Feature Subset Selection and Feature Ranking for Multivariate Time Series," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 9, pp. 1186–1198, 2005.
- [34] A. D. Rahajoe, E. Winarko, and S. Guritno, "A Hybrid Method for Multivariate Time Series Feature Selection," *International Journal of Computer Science and Network Security*, vol. 17, no. 3, pp. 103–111, 2017.
- [35] V. Vapnik, "Statistical Learning Theory," *Adaptive and learning Systems for Signal Processing, Communications and Control*, pp. 1–740, 1998.
- [36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: data mining, inference and prediction*. New York: Springer-Verlag, 2001.
- [37] R. . Duda, P. . Hart, and D. . Stork, *Pattern Classification*, Second. Wiley Interscience, 2001.
- [38] D. T. Larose, *An Introduction to Data Mining*. Canada: A John Wiley & Sons, Ltd, 2005.