Journal of Physics: Conference Series **1456** (2020) 012026 doi:10.1088/1742-6596/1456/1/012026

Predicting consumer price index cities and districts in East Java with the gaussian-radial basis function kernel

M F Rohmah[1], I K G D Putra[2], R S Hartati[3] and L Ardiantoro[1]

[1] Informatics Engineering,Universitas Islam Majapahit, Indonesia
[2] Informatics Engineering, Universitas Udayana, Indonesia
[3] Electronics Engineering, Universitas Udayana

Email: miminfr@gmail.com

Abstract. An economic indicators regarding information on the prices of goods and services paid by consumers are known as the Consumer Price Index (CPI). Predicting CPI using the Gaussian-Radial Basis Function Kernel with Support Vector Regression (SVR) method because the algorithm produces functions with surging data trends following the data path formed, so the Prediction results are more accurate. As a reference in predicting the CPI used the Foodstuffs issued by Statistics Indonesia and as an input variable taken from the Foodstuffs commodity prices on the East Java Disperindag siskaperbapo website in five cities namely Kediri, Madiun, Malang, Probolinggo, Surabaya and in three districts namely Sumenep, Jember, Banyuwangi from the beginning of 2016 and the end of 2018. Input data is divided in the training and testing data. The results of the Kediri CPI prediction with Mean Square Error (MSE) is 0.8697 and Mean Absolute Percentage Error (MAPE) is 0.432, so that the city of Kediri have CPI prediction approaching reference. The average MSE and MAPE for the three districts and five cities is 3.3462 and 1.2336125.

1. Introduction

Prices on stable foodstuffs are an indicator of economic progress in a region both in cities and districts. Food prices can always be stabilized by knowing the value of the Consumer Price Index (CPI) in advance, by predicting the CPI using the Support Vector Regression (SVR) method with the Gaussian-Radial Basis Function Kernel. The index number used to calculate the average change in the price of goods and services that is always consumed by households and communities is called the CPI. The amount of CPI is affected by seven expenditure groups and several support groups. The seven expenditure groups consist of the Foodstuffs Group; Processed Food, Beverage, Cigarettes and Tobacco Groups; Housing, Water, Electricity, Gas and Fuel Group; Clothing Group; Health Group; Education, Recreation and Sports Groups; the last group is Transportation, Communication and Financial Services. The Foodstuffs Group consists of several supporting sub-groups, namely the Paddy-Grains, Tubers, Meat, Fresh Fish, Preserved Fish, Eggs and Milk, Vegetables, Beans, Fruits, Spices, Fat and Oil and other food ingredients taken from Siskaperbapo [1]. The calculation of CPI in Districts and Cities in East Java is carried out by Statistics Indonesia with conducting surveys in eight predetermined areas, namely five Cities (Surabaya, Madiun, Kediri, Probolinggo, Malang) and three Districts (Sumenep, Jember, Banyuwangi). Statistics Indonesia input data comes from the transaction price of goods at retail level, which is collected from several traditional and modern markets that are considered to be able to represent prices in the region concerned [2]. The CPI is displayed monthly on the website http://www.bps.go.id and can be accessed by the public. The formulation of the problem in this study to predict the Consumer Price Index of Cities and Districts in East Java with the Gaussian-Radial Basis Function Kernel. In predicting the value of the CPI, can take solving so that the stability of food commodity prices is maintained. Data input in this study uses the price data sources from the Foodstuffs group from the Department of Industry and Trade of East Java province, which is presented in www.siskaperbapo.com. The data is used as an input variable, with a data collection period of 3 years, which is the beginning of January 2016 - the end of December 2019, which consists of 34 attributes of foodstuffs, while the output variable is the City and Districs CPI in East Java taken at the same time. In predicting CPI researchers used the SVR method with the Gaussian Radial Basis Function Kernel. This method was chosen because SVR is able to find the f (x) function as a hyperplane, for all input data that has the greatest deviation ε from the actual target yi for all training data and can make errors as thin as possible. The hypothesis in this study is that the system is able to predict the CPI of the Foodstuffs group with a MAPE value below three percent.

2. Method

2.1 Gaussian-radial base function kernel

Because of the nonlinearity, data mining / machine learning techniques have been developed. So the resulting algorithm is not limited to linear algorithms that are also non-linear. Therefore, if a prediction case shows nonlinearity, algorithms such as perceptron cannot overcome them. For example, consider Figure 1, this data is difficult to separate linearly. The kernel method (Scholkopf and Smola, 2002) [3] is one way to overcome it. With the kernel method, an x data in the input space is mapped to feature space F with a higher dimension through the ϕ map as follows ϕ: x → ϕ (x). Therefore, the x data in the input space becomes ϕ (x) in the feature space. Often the function ϕ (x) is not available or cannot be calculated, but the dot product of two vectors can be calculated both in the input space and in the feature space. In other words, while ϕ (x) may not be known, dot product (ϕ (x1), ϕ (x2) can still be calculated in the feature space. To be able to use the kernel method, the objective and constraint functions need to be expressed in the dot product form of vector data xi. As a consequence, the objective function that explains the problem in classification must be reformulated so that it becomes a dot product. In this feature space dot product (.) becomes (ϕ (x), ϕ (x) ′). A kernel function, k (x, x ′), can be used to replace the dot product (ϕ (x), ϕ (x) ′). Then in feature space, we can find a linear separating function that represents nonlinear functions in the input space. Figure 1, describes an example feature mapping from two-dimensional space to two-dimensional feature space. In input space, data cannot be separated linearly, but we can separate in feature space. Therefore, mapping data to feature space makes classification tasks easier [3].

Table 1. Type of Kernel function

No	Function	Formula
1	Linear	$K(x_i, x)= x_i^T x$
2	Polynomial	$K(x_i,x)= (x_i^T x + 1)^d$ d=1,2,………
3	Gaussian - Radial Basis Function (RBF)	$K(x_i,x)= \exp(-\gamma \|x - x_i\|^2)$ $K(x_i,x)=\exp(-\frac{1}{2\sigma^2}\|x - x_i^T\|^2)$
4	SPlines	$K(x_i, x) = \prod_{m=1}^{n} K_m (X_m, X_{im})$

In this study the kernel used is the Gaussian-Radial Basis Function type kernel, due to previous studies [4], with Gaussian-RBF MSE and MAPE values smaller and approaching the reference CPI.

2.2 Support vector regression

Support Vector Regression is an application of data mining. The SVR method of development of the Support Vector Machine (SVM) discovered by Vladimir N Vapnik in 1999. The SVM that describes Regression is known as Support Vector Regression (SVR). The difference between SVM and SVR lies in the output and application of the system [5], [6]. The SVR output is a continuous number, while the output of SVM is an integer. SVM is used to find the best hyperplane (separator function) in separating two classes, namely class +1 and class -1, by maximizing the margin or distance between two different objects. The advantage of SVR over linear regression is that the linear regression algorithm produces a linear function, while the SVR algorithm produces functions with wavy data trends following the data path formed, so that the prediction results are more accurate. In the regression function f (x) if the deviation limit (ε) is equal to 0, a perfect regression function is obtained, shown in figure 1, while the SVR linear model for the regression function matches the equation below [7], [8] :

$$f(\,[x) = w^T \varphi(x) + b \tag{1a}$$

φ(x) : Point inside feature space F, the result of mapping x inside the input space
w: Weight vector with one dimensions
b: Bias
x: Input vector
f (x): Regression function

The coefficients w and b function to minimize the risk function, which is described as follows: $R = \min \frac{1}{2} ||w||^2 + C\,\frac{1}{l}\left(\left(\sum_{i=1}^{l} L\varepsilon(\,y_i\,, f(x_i)\right)\right)$ (1b)

with limitations:

$$y_i - w\varphi(x_i) - b \le \varepsilon$$
$$w\varphi(x_i) - y_i + b \le \varepsilon \qquad i = 1, 2, 3, \ldots$$

where :

$$L\varepsilon\,(\,y_i\,, f(x_i) \;= \begin{cases} 0\,, & \text{for } \;|y_i - f(x_i)| < \varepsilon \\ |y_i - f(x_i)| - \varepsilon\,, & \text{for } \;|y_i - f(x_i)| \ge \varepsilon \end{cases}$$

Lε: Loss function is type ε-insentive loss function
R: Risk Function
|| w || : Normalization w
ε: Epsilon is the deviation / degree of error tolerance
C: (deviation> error limit)

Equation (2) assumes all points are Lε (yi, f (xi)) the penalty function for the ε-incentive loss function, subject to a penalty if error | yi, f (xi) | > = ε.

Problem Optimization for feasible conditions as follows [7]:

Problem Optimasi_Feasible = $\mathrm{Min}\frac{1}{2}||w||^2$ (2)

With the following constraints:

$$y_i - w^T\varphi(x_i) - b \le \varepsilon \;;\quad w^T\,\varphi(x_i) - y_i + b \le \varepsilon\;;\quad i = 1, 2, \ldots, l$$

A decent limiting margin: f (x) ± ε.

SVM also looks for the best separator function on two infinite objects by increasing the distance between two unequal objects, while the SVR application determines functions that have deviations greatest ε of the actual target yi. The SVR method can find the f (x) function as a dividing line in the form of a regression function, for all data inputs that have the greatest deviation ε from the actual target y for all training data and make the smallest error according to Scholkopt and Smola, 2002 [3]. The goal of the SVR method is to map the input vector to a higher dimension and the error is ignored if error <epsilon. Epsilon (ε) is the margin of tolerance for errors. The SVR method has advantages over linear

regression, namely the linear regression algorithm, the output is a linear function / straight line, while the SVR output trend data follows the data path formed, for example, wavy, so that the predicted data output is more accurate. In previous studies, many CPI forecasting was done using the SVR method including Ye Wang et.all with the theme A new application of the support vector regression in the construction of financial conditions index to CPI prediction and Zhang at.all with the theme Inflation Forecasting Using Support Vector [10], [11].

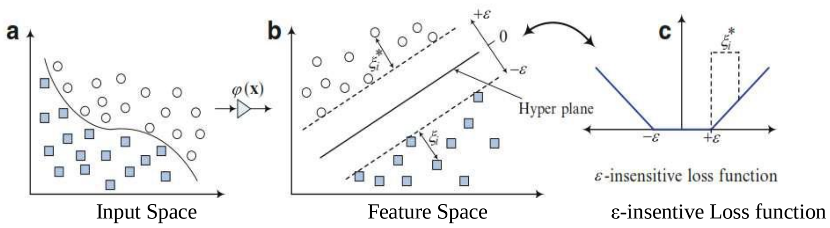

Figure 1. Support Vector Regression (SVR)

2.3 Consumer Price Index (CPI)

Information on the price of goods / services paid by users by comparing current and previous prices is known as the Consumer Price Index. How to determine CPI by knowing changes in the prices of goods and services in consumers (purchasing costs) that are commonly consumed. Raw data is input data that supports CPI from Foodstuffs from www.siskaperbapo.com managed by Disperindag East Java. The data consists of consumer / producer price data and data on availability that is updated every day. In this research, consumer price data are used.

Consumer Price Index can be calculated by the formula:

$$\mathrm{IHK} = \frac{\mathrm{Pn}}{\mathrm{Po}} \text{ x } 100\% \qquad (3)$$

Pn = current price

Po = base year price

In this study, the Consumer Price Index (CPI) of Foodstuffs as a reference is found in Statistics Indonesia (https://jatim.bps.go.id), which is in the CPI e-book in 82 cities in Indonesia for 2016, 2017, year 2018 [2].

2.4 Evaluation methods and performance measures for regression models

In the prediction/forecast method, trying to obtain a forecast value that is close to the actual value in the same period or the difference between the predicted value and the actual value is as small as possible. The difference between the forecast value and the actual value is called error.

$$\mathrm{e_t} = \mathrm{X_t} - \mathrm{F_t} \qquad (4)$$

et = forecast error at time t

Xt = reference value at time t

Ft = predicted value at time t

Evaluation methods used to measure prediction accuracy include Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). The accuracy of forecasting is better if the MSE, RMSE and MAPE values are getting smaller [10], [11], [12].

2.4.1 Mean Square Error (MSE)

MSE is a forecast evaluation method where each error is squared then summed and divided by the number of observations as follows:

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{X_i} - \mathrm{F_i})^2 \qquad (5)$$

Xi = data of the actual value of observation period i;

Fi = period prediction data value i

n = Total amount of data in the data set testing, period i

2.4.2 Mean Absolute Percentage Error (MAPE)
The percentage technique on average absolute error is calculated by finding the absolute error of each period / time, and dividing by the observation value in that period, also averaging this absolute percentage.

$$\text{MAPE} = \frac{\sum_{i=1}^{n} \frac{|X_i - F_i|}{X_i}}{n} \text{x}100\% \quad (6)$$

Xi = period true value data i

Fi = data forecasts of period i, n = forecast time period

Sourced from research (Chang, Wang et.all, 2007) [12], produces a standard Prediction Criteria for Mean Absolute Percentage Error as detailed as follows, for values less than 10 percent have excellent prediction criteria, between 10 percent to 20 percent good prediction criteria, then 20 percent to 50 percent criteria Predictions are sufficient, lastly 50 percent greater have bad prediction criteria.

2.5 Research step
The stages in the research are as follows:
Identifying problems to predict CPI foodstuffs in five cities (Kediri, Madiun, Malang, Probolinggo, Surabaya) and three districts (Sumenep, Jember, Banyuwangi).

- Study of prediction concepts with RBF kernel and SVR methods, CPI theory, data utilization from the website of the Central Statistics Agency office in the city of Surabaya.
- Input variables from the siskaperbapo web and CPI data target variables, the results of the test are used to analyze the CPI forecast for the following year.
- Taking data online from the siskaperbapo website from January 1, 2016 to December 31, 2018, the data taken is daily updated data every day.
- Collection of CPI data for monthly foodstuffs for the period of January 2016 to December 2018, which is taken monthly from the website www.bps.go.id for the five Cities and Three Districs
- Preprocessing data, there are a total of 35 attributes consisting of 34 attributes as input variables, namely the price of basic commodities, with the names X1, X2, X3,... X34 and one attribute as a target (CPI).
- Analysis of the SVR method with MATLAB software, by dividing data into training data and testing data; SVR analysis method with the Kernel RBF function to get the value of MSE, MAPE, iteration value of the Gaussian RBF function by predicting the CPI value.

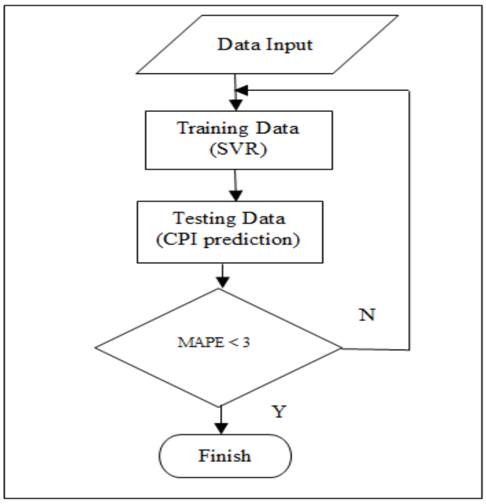

Figure 2. Design System

3. Results and discussion

The training data used in the research as input variable is data on consumer commodity prices for basic commodities in five cities (Kediri, Madiun, Malang, Probolinggo, Surabaya) and three districts (Sumenep, Jember, Banyuwangi), which have monthly CPI that can be viewed on the website (https://jatim.bps.go.id) as variables target and secondary data / input variables can be seen on the website (www.siskaperbapo.com). The website presents Pricing data for Consumers, Producers and Availability. The data on consumer prices for these staple commodities is related to previous data, current data and future data. The following table is an input variable taken on the Disperidag website (www.siskaperbapo.com), with Attributes (X1 - X34), like the following table:

Table 2. Antribute Variabel Input

Label	Atribut-Variabel Input	Label	Atribut-Variabel Input
X1	Bengawan Rice	X18	Refined iodized salt
X2	Mentik Rice	X19	Rice Blue Triangle Flour
X3	IR64 Rice	X20	Rice Imported Soybeans
X4	Domestic Sugar	X21	Indomie Instant Noodle Curry Chicken
X5	Bottled Edible Oil / Pack 2 l	X22	Regular Chillies
X6	Bulk Chili Oil	X23	Bulk Chili Oil
X7	Pure Beef	X24	Shallots
X8	Broiler Chicken	X25	Garlic
X9	Kampung Chicken	X26	Chicken Meat Anchovies
X10	Broiler Eggs	X27	Green Beans
X11	Kampung eggs	X28	Chicken Eggs Peanut
X12	Sweetened Condensed Milk Brand Flag	X29	Cassava Tree
X13	Sweetened Condensed Milk Brand Indomilk	X30	Vegetable Cabbage
X14	Flag Brand Milk Powder (Instant)	X31	Vegetable Potatoes
X15	Indomilk Milk Powder Brand (Instant)	X32	Vegetable Tomato
X16	Dried Pipilan Corn	X33	Vegetable Carrots
X17	Type iodized salt Brick	X34	Vegetable Beans

Input variables used are 34 (X = X1, X2, X3 to X34). Taken from the siskaperbapo website in three districts, data collection was taken for three years from 2016 to 2018. Target Variable Y = CPI is the value of the Consumer Price Index of the Foodstuffs Group.

3.1 Testing result

First by testing the training data for 2016-2017 it is set to 2018-2019. Where for the input data in 2019 use input data in 2018.
Then the result is as follows:

3.1.1 For five cities including Surabaya, Madiun, Kediri, Probolinggo, Malang:

- Surabaya:

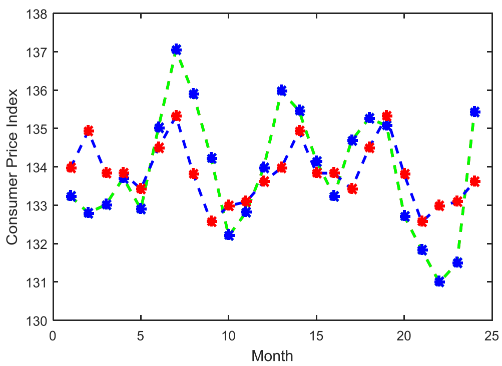

Figure 3. Graphic for Surabaya CPI Predicton Vs Month

From the graph, it can be seen that in the 4th, 11th, 15th month, the CPI prediction result approach the CPI referenc, as bellow:
4th Month: Reference Data = 133.71, Prediction Data = 133.8349
5th Month: Data Reference = 132.82, Data Prediction = 133.0876
15th Month: Reference Data = 134.15, Prediction Data = 133.8349
With the RBF Kernel simulation results can be determined the value of
MSE = 1.4919, MAPE = 0.1716, C = 50

- Madiun

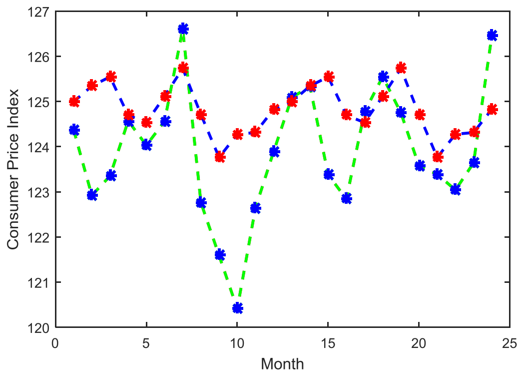

Figure 4. Graphic for Madiun CPI Predicton Vs Month

From the graph, it can be seen that in the 5th, 14th , 17th month, the CPI prediction result approach the CPI reference, as bellow:
5th Month: Reference Data = 124.03, Prediction Data = 124.5337
14th Month: Reference Data = 125.33, Prediction Data = 125.3419
17th Month: Reference Data = 124.77, Prediction Data = 124.5337
With the RBF Kernel simulation results can be determined the value of
MSE = 2.2893, MAPE = 2.6418, C = 50

- Kediri

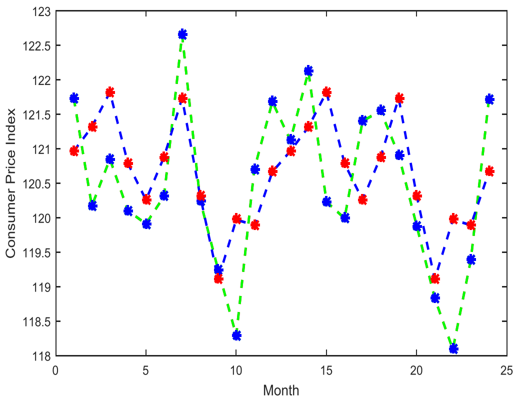

Figure 5. Graphic for Kediri CPI Predicton Vs Month

From the graph, it can be seen that in the 8th, 9th , 13th month, the CPI prediction result approach the CPI reference, as bellow:
8th Month: Reference Data = 120.25, Prediction Data = 120.2655
9th Month: Reference Data = 119.24, Prediction Data = 119.0842
Month 13th: Reference Data = 121.12, Prediction Data = 120.9459
With the RBF Kernel simulation results can be determined the value of
MSE =0.8697, MAPE = 0.432, and C = 50

- Probolinggo

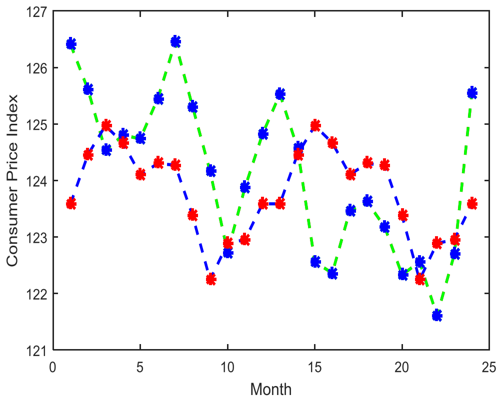

Figure 6. Graphic for Probolinggo CPI Predicton Vs Month

From the graph, it can be seen that in the 4th, 10th, 14th month, the CPI prediction result approach the CPI reference, as bellow:
4th Month: Reference Data = 124.8, Prediction Data = 124.6667
10th Month: Reference Data = 122.73, Prediction Data = 122.8881
Month 14th: Reference Data = 124.58, Prediction Data = 124.4587
With the RBF Kernel simulation results can be determined the value of
MSE = 2.0628, MAPE = 0.9667 C = 50

➢ Malang

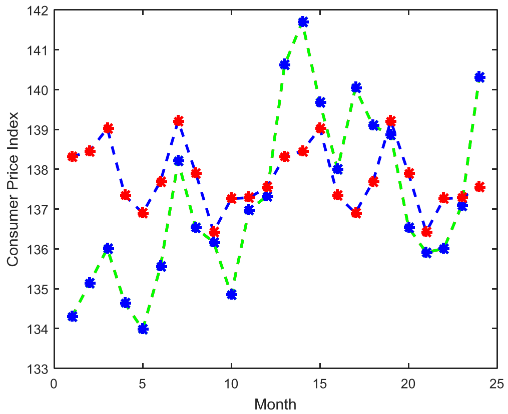

Figure 7. Graphic for Malang CPI Predicton Vs Month

From the graph, it can be seen that in the 9th, 12th , 19th , 23th month, the CPI prediction result approach the CPI reference, as bellow:
19th Month: Reference Data = 136.17, Prediction Data = 136.4278
12th Month: Reference Data = 137.32, Prediction Data = 137.5424
Month 19th: Reference Data = 138.87, Prediction Data = 139.2137
23th Month: Reference Data = 137.09, Prediction Data = 137.2868
With the RBF Kernel simulation results can be determined the value of MSE = 4.4027, MAPE = 1.5727, C = 50

3.1.2 *For three Districs: Sumenep, Jember, Banyuwangi*

➢ Sumenep

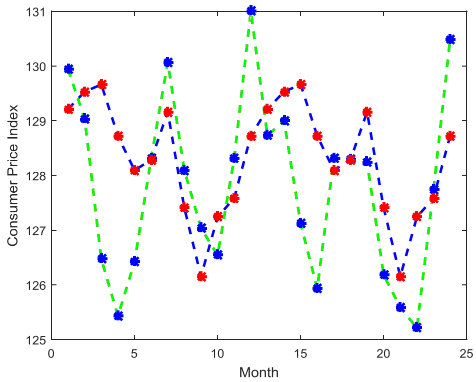

Figure 8. Graphic for Sumenep CPI Predicton Vs Month

From the graph, it can be seen that in the 17th month, 23rd month, the CPI prediction result approach the CPI reference, as bellow:
17th Month: Reference Data = 128.31, Prediction Data = 128.1916
23rd Month: Reference Data = 137.09, Prediction Data = 137.2868
With the RBF Kernel simulation results can be determined predictions Sumenep Distrcs with value of MSE = 2.5182, MAPE = 1.5508, C = 50

➢ Jember

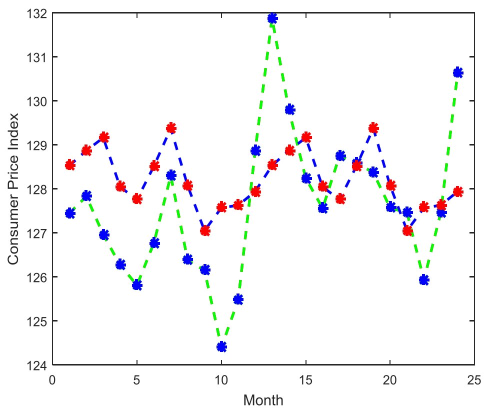

Figure 9. Graphic for Jember CPI Predicton Vs Month

From the graph, it can be seen that at 18th month, 23rd month , the CPI prediction result approach the CPI reference, as bellow:
Month 18th: Reference Data = 128.59, Prediction Data = 128.5154
23rd Month: Reference Data = 127.46, Prediction Data = 127.6268
With the RBF Kernel simulation results can be determined prediction of CPI for Jember Districs with the value of MSE = 2.6219, MAPE = 1.6823, C = 50

➢ Banyuwangi

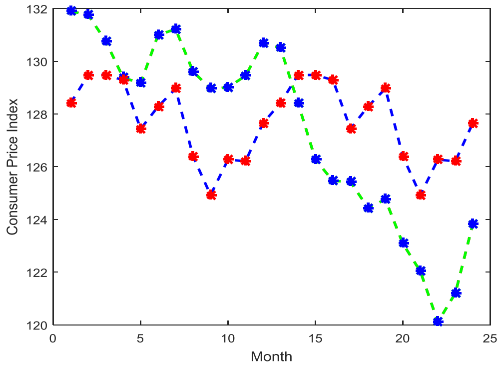

Figure 10. Graphic for Banyuwangi CPI Predicton Vs Month

From the graph, it can be seen that in the 4th month, the CPI prediction result approach the CPI reference, as bellow:
4th Month: Reference Data = 129.38, Prediction Data = 129.2894
With the RBF Kernel simulation results can be determined prediction of CPI for Banyuwangi Districs with the value of MSE = 10.5131, MAPE = 0.851 and C = 50

Analysis of MSE, MAPE, for the five cities and three districts if tabulated as in tables 4a and 4b below:

Table 3a. MSE, MAPE values with Kernel RBF for Iteration C = 50 predictions for 5 Cities

City	MSE	MAPE
Surabaya	1.4919	0.1716
Madiun	2.2893	2.6418
Kediri	0.8697	0.432
Probolinggo	2.0628	0.9667
Malang	4.4027	1.5727

Table 3b. MSE, MAPE values with Kernel RBF for Iteration C = 50 predictions for 3 districts.

Districs	MSE	MAPE
Sumenep	2.5182	1.5508
Jembers	2.6219	1.6823
Banyuwangi	10.5131	0.851

4. Conclusion
This study proposes a new forecasting approach that is built on a forecasting model based on the Consumer Price Index from Statistics Indonesia East Java Province using the RBF kernel with the SVR method consisting of 34 variable price staples in five Cities and three Districts and the output is the forecast of CPI values for Foodstuffs in three The district. The MSE value satisfies because it is less than 3%.

The result of the Kediri CPI prediction with Mean Square Error (MSE) of 0.8697 and the Mean Absolute Percentage Error (MAPE) of 0.432, so that the Kediri city predicted its CPI to be close to the reference CPI. The average MSE and MAPE for the three districts and five cities is 3.3462 and 1.2336125.

5. References
[1] http://siskaperbapo.com/harga/tabel Sistim *Informasi Ketersediaan dan Perkembangan Harga Bahan Pokok di Jawa Timur*
[2] https://sirusa.bps.go.id/sirusa/index.php/indicator/ index-Indeks Harga Konsumen (IHK)
[3] Mimin F Rohmah, et al 2019*Meramal Indeks Harga Konsumen Kabupaten di Jawa Timur dengan Metode Support Vector Regression Data Mining*, Prosiding Snati, UII, *ISSN: 1907 – 5022* P C30 – C36
[4] Scholkopf B., and Smola A 2002 *Learning With Kernel* MIT Press
[5] Cristianini N., Taylor J.S 2000 *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge Press University
[6] Bin Gui, et al 2014 *Financial Time Series Forecastin Using Support Vector Machine* Tenth International Conference on Computational Intelligence and Security

[7] Hossein Mahjub et.all 2016 *Performance Evaluation of Support Vector Regression Models for Survival Analysis: A Simulation Study* International Journal of Advanced Computer Science and Ap*plications, Vol. 7, No. 6*
[8] Hsu, C.W.et al 2004 *A Practical Guide to Support Vektor Classification, Department of Computer Science and Information Engineering* National Taiwan University
[9] Wei, W.W.S 2016 *Time series Analysis Univariate and Multivariate Methods* New York: Pearson education, Inc
[10] Ye Wang, Bo Wang, Xinyang Zhang 2012 *A new application of the support vector regressionon the construction offinancial conditions index to CPI prediction* , International Conference on Computational Science, ICCS, Procedia Computer Science 9 p 1263 – 1272
[11] Zhang, Jinchang LI 2012 *Inflation Forecasting Using Support Vector Regresion* ,fourth International Symposium on Information Science and Engineering, 978- 7695-4951-4/13/2013 IEEE, DOI 10.119/ISISE.2012.37
[12] Chang, P.-C. at al 2007 *The Weighted Evolving Fuzzy Neural Network for PCB Sales Forecasting* Elsevier, 32 (Expert Systems with Applications) , p 86-96