Forecast Analysis of Instant Noodle Demand using Support Vector Regression (SVR)

E Fradinata1*, Z M Kesuma2, S Rusdiana3, and N Zaman1
1Industrial Engineering Department, Faculty of Engineering, Syiah Kuala University, Banda Aceh, Indonesia.
2Statistics Department, Faculty of Mathematics and Natural Science, Syiah Kuala University, Banda Aceh, Indonesia.
3Mathematics Department, Faculty of Mathematics and Natural Science, Syiah Kuala University, Banda Aceh, Indonesia.

*Corresponding author: edinata69@gmail.com

Abstract. Support Vector Regression (SVR) is a part of Data Mining (DM) techniques where it can be used for forecasting the instant noodle. The cycle of the product demand is hard to predict. It will influence the resistant of the product quality where the product be expired easily and the other thing is the market demand. The objective of this research is approaching the predictive models with their performance measured with Mean Square Error (MSE) of SVR. The data was collected from the determinant of instant noodle demand dataset. The random normal generated data was explored to get the amount of specific data. Then, it used SVR to forecast the demand. The result of this study the MSE of standard is 1.612 and the SVR is 1.436, means it increases around 11% better the performance than the original dataset. Since, we conclude that the SVR method would be promising to be one of a forecast demand method.

1. Introduction
Nowadays, the demand of instant noodle might be hard to anticipate due to those self-destructive considerations, it was caused of the behavior of a customer and the groups, for example, the society and the government regulation. The technique of support vector machine (SVM) will turn into established to utilize to solve in this area. The precision of demanding from this practice will be promising to be a better approach.

The learning machine theory would minimize the structural risk where it is construct with the boundaries upper bound and lower bound before it is generalized error. The gap is prepared after the link to the predicting. SVM obtains to be an elective also a chance to be those capable strategies on unraveling those relapse issues eventually utilizing the loss function, risk minimization, and this called SVR. An additional point of interest of SVR might make discovered [1-2].

The SVM could be working on the risk minimization (SRM) concept; its idea will be on minimizing the error and reduce the issue for overfitting. This technic could utilize in both proposes. They are in classification and regression cases in many areas. A portion investigation on which utilization this technique connected to forecasting modeling. Some studies in which use this method applied for forecasting modeling. The concept of support vector regression is
finding the function that can be the best approximation the output of the vector \( y \) with a tolerance of error, to make it as flat as possible [3-5].

SVR may be a strategy on include the preparation focuses on the little subset of getting a yield result with the computer. The loss function of SVR is to pretend the optimization in a global minimum area. SVR can improve various interesting features and produce a better performance [6-8].

The contribution of this study might have invested the SVR method to forecast the points in the future for instant noodle industry. The indication of this matter, the SVR proves the small mean square error in prediction. This is the condition where the SVR to be one of the data mining method to predict the certain points in the future when it is used to be the predicting method.

2. Background

2.1. Support vector regression

SVR is a machine learning where it is called support vector machine (SVM) to use the useful computational load [9]. The dots results would constantly utilize to figure the variables in the unique space with portion work chosen by the SVM with gatherings made the mapping of the problem [10]. SVR will develop degeneration strategy that formed from SVM, the place might have been taken those hyperplanes closest to a measure of information the place it turns into a purpose. The little standard hyperplane needs to be gathered when the separation starting with information focuses on the hyperplane should keep minimizing [11]. It is demonstrated in figure 1.

![Figure 1. \( \varepsilon \)-sensitive loss function for linear of SVR [12].](image)

The optimization to an \( \varepsilon \)-insensitive loss function can be expressed in matrix notation as SVR finds the role of \( f(x) \) where has the biggest deviation \( \varepsilon \) from the original target \( y \), for all the training data with SVR, we got the something like a trench as in figure 1. Moreover, the formula is as follow:

\[
f(x) = w^T \varphi(x) + b
\]

(1)

Where \( \varphi(x) \) shows the dot in the feature space \( f \) from the map of \( x \) in the input space, coefficient, \( w \), and \( b \) is estimated the minimization of risk function, this is defined in the formula of regression risk as follow:

\[
\text{Min} = \frac{1}{2} \| w \|_2^2 + \frac{C}{\ell} \sum_i \max(0, 1 - \ell y_i f(x_i))
\]

(2)
Constraints to:

\[ Y_i - w \varphi(x_i) - b \leq \epsilon, i = 1, \ldots, \ell \]

\[ w \varphi(x_i) - y_i + b \leq \epsilon, i = 1, \ldots, \ell \]

The \( \epsilon \)-insensitive loss function is usually as a cost function, where the function as given by (3).

\[
L_{\epsilon}(y_i, f(x_i)) = \begin{cases} 
|y_i - f(x_i)| - \epsilon & \text{if } |y_i - f(x_i)| \geq \epsilon \\
0 & \text{otherwise}
\end{cases}
\] (3)

The factor of \( \|w\|^2 \) is named regularization. Minimize the \( \|w\|^2 \) will make the function flat so that it can control the function of capacity. The second factor in the objective function is an error calculation measured with the \( \epsilon \)-insensitive loss function, [13]. In the process, it should be minimized \( w \) value. It has the aim of to find the better generalization for the regression function \( f \). So it must be solved the optimization problem at (4), or another way to solve with a Matlab program is as follow:

\[
\min \frac{1}{2} x^T H x + c^T x
\] (4)

Where:

\[
H = \begin{bmatrix} XX^T & -XX^T \\ -XX^T & XX^T \end{bmatrix}, c = \begin{bmatrix} \epsilon + Y \\ \epsilon - Y \end{bmatrix}, x = \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix}
\]

Subject to:

\[
x, (1, \ldots, 1, -1, \ldots, -1) = 0, \alpha i, \alpha^* i \geq 0, i = 1, \ldots, \ell
\]

Where:

\[
X = \begin{bmatrix} x_1 \\ \vdots \\ x_{\ell} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_{\ell} \end{bmatrix}
\]

The \( \epsilon \)-insensitive can be described as a tube in the generation of precision on the individual situation of data preparation. SVR regression was used to estimate the capacity may be claimed where the information is located on the axis \((x_i, y_i)\) where \( x_i, y_i, \) and \( n \) need the support to input, output, then the others can focus on data, separately. Separate works under the direct jobs are also a linear and nonlinear function [14].

The specific essential ideas from SVR are giving the solution from the small subset of training points to contribute the huge computational advantages. The process uses the epsilon intensive loss function to maintain the global minimum and the same time produces the generalization bound for optimization. The better settings of parameters are \( C, \epsilon \), and the kernel parameters. The parameter will be determined by the degree of the larger deviation of \( \varphi \) and \( C \) parameter. Some very common of kernel functions are shown in table 1.
Table 1. Kernel functions.

| Kernel Function | Function |
|-----------------|----------|
| Linear          | x*y      |
| Polynomial      | [(x*x)+1]^d |
| Radial Basis Function (RBF) | Exp(-y | x-x_i |^2) |

2.2. Mean Square Error (MSE)
Mean square error (MSE) is the estimator of measure the average of squares of the error. It is a risk function where to measure the quality of an estimator and the value closer to zero are better. The MSE is a measurement of the estimator quality with incorporates both mean and square of error with the formula is:

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - f_i)^2 \]  

Where:
\( Y_i = \) the actual value.
\( F_i = \) the forecast value a period
\( n = \) the number of periods.

3. Methodology
The method of this study was used quantitative methods to process and analysis of data. The data were collected from the secondary data then created them with random normal generating. The instant noodle variables are price and demand were normalized. The method of this study was used quantitative methods to process and analysis of data. This study runs with the Steve Gunn's algorithm where it is used the parameter of kernel function is linear, the upper bond C (not separable case) is one, the loss function is e-insensitive [15]. SVR worked with 100 data, and it was separated into two-part data. The 90% was used for training and the rest 10% was used for testing, figure 2.
The SVR would present the solution by means of small subset of training points where it would contribute the massive computational advantages. The global minimum would be produce between the area points to determine the optimal condition of reliable prediction. The algorithm explanation of SVR is as followed:

i. Define the parameters that for sales (x) and demand (y), kernel, insensitive, upper bond, Number of support vector, Lagrange multiplier, and the bias.

ii. Check the correct number of arguments.

iii. Tolerance for Support Vector Detection.

iv. Construct the Kernel matrix.

v. Set up the parameters for the optimization problem.

vi. Set the bounds: alphas ≥ 0 and ≤ C.

vii. Configure the number of equality constraints (1 or 0).

viii. Set the constraint Ax = b.

ix. Add a small amount of zero order regularization to avoid problems when Hessian is badly. Rank is always less than or equal to n.

x. Solve the optimization problem.

xi. Compute the number of Support Vectors.

xii. Find bias from an average of support vectors with interpolation error e (0 < alpha < C).

xiii. Calculate the SVR for number support, beta and bias and svr output as follow:

xiv. $\text{[nsv beta bias]} = \text{svr (x,y,Metker, C,lossf, insen)}$.

xv. $\text{tstData = svroutput (x, y, Metker, beta, bias)}$. 

---

**Figure 2.** Block diagram of SVR algorithm.
xvi. SVR output
xvii. SD and MSE
xviii. Forecast the demand for the pattern from the characteristic of data
xix. Plot the comparative pattern

4. Results and Discussion

From this study, we prepared the data to check the cleaning data. In this research used two variables. They are price and demand variable of the Instant noodle industry, (Dy). The parameters used in this study with kernel function were linear, loss-function was insensitive, the number of insensitive was 1, and upper-bound (non-separable case) was 1.

The SVR value worked under its loss function was $\varepsilon$-insensitive, $C$ was 1 and kernel function was linear. Those parameters affected by the amount of SV to build the regression function. The generated dataset from the characteristic of dataset after testing step to generalized the dataset in the area of upper bond and lower bond, it should be produced the forecast dataset in the certain point for predicting. The data produced the statistical characteristic of dataset where it can be compared the performance through the MSE for both running. Since, the resulted of MSE show that the actual data of standard was 1.612 and the predicting method with SVR was 1.436. Furthermore, the SVR decreased the MSE from the standard value to predict the certain point in the future, it is shown in table 2 as follow:

| Time (Moths) | Actual Data (Box/period) | Predicted (Box/Period) |
|--------------|--------------------------|------------------------|
| 91           | 343,141,145              | 342,089,212            |
| 92           | 339,503,098              | 358,153,780            |
| 93           | 362,888,381              | 358,153,780            |
| 94           | 339,800,508              | 348,113,425            |
| 95           | 326,827,872              | 361,768,308            |
| 96           | 343,468,963              | 347,711,811            |
| 97           | 359,389,685              | 350,121,496            |
| 98           | 353,704,355              | 366,989,293            |
| 99           | 347,794,147              | 345,302,126            |
| 100          | 366,989,293              | 326,827,872            |

| Standard Deviation | 0.0554 | 0.0271 |
|--------------------|-------|-------|
| Mean               | 0.3722| 0.3619|
| MSE                | 1.612 | 1.436 |

SVR was used Steve Gunn’s algorithm code and then it was modified to calculate the output of SVR function in the program code [15]. The 90% of data was used for training and the rest of 10% as testing. The testing resulted from SVR-output and then calculates the statistical output to get the characteristic of data. It was drawn in Figure 3.
Figure 3 shows the result of SVR output for predicting point compare between the actual dataset to the forecast with the SVR.

5. Conclusions
This paper proposes the methods of SVR as a promising technique of forecasting. It is controlled by the combine generalisation control and other technique to obtain the curse dimensionality. This condition produces the result in the global quadratic optimisation with using the kernel function to view the model. The optimization points drop into the border of e-insensitive between positive and negative. After it is generalized compute the support vector and calculate the SVR for number support, beta, bias and SVR output, then this study obtain the value to compare between the standard and the SVR method. The result of the MSE is 1.436. from 1.612. It means that the performance of SVR method is better 11% than the standard. The accuracy of SVR would be able to be an alternative for the factory of instant noodle to predict the condition of customer demand. This method also can be used to predict the length of time (expired date) for instant noodle is consumed by the customer.

References
[1] Tay F E and Cao L 2001 Application of support vector machines in financial time series forecasting Omega 29 309-17
[2] Drucker H, Burges C J, Kaufman L, Smola A J and Vapnik V 1997 Support vector regression machines. In: Advances in neural information processing systems, pp 155-61
[3] Yu P-S, Chen S-T and Chang I-F 2006 Support vector regression for real-time flood stage forecasting Journal of Hydrology 328 704-16
[4] Santosa B 2007 Data Mining (Teori dan Aplikasi) Graha Ilmu, Yogyakarta
[5] Jiang H, Zhang Y, Muljadi E, Zhang J J and Gao D W 2018 A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization IEEE Transactions on Smart Grid 9 3341-50
[6] Drucker H, Wu D and Vapnik V N 1999 Support vector machines for spam categorization IEEE Transactions on Neural Networks 10 1048-54
[7] Basak D, Pal S and Patranabis D C 2007 Support vector regression Neural Information Processing-Letters and Reviews 11 203-24
[8] Santamaria-Bonfil G, Reyes-Ballesteros A and Gershenson C 2016 Wind speed forecasting for wind farms: A method based on support vector regression Renewable Energy 85 790-809
[9] Burges C J 1998 A tutorial on support vector machines for pattern recognition Data mining and knowledge discovery 2 121-67
[10] Ghaedi M, Reza Rahimi M, Ghaedi A, Tyagi I, Agarwal S and Gupta V K 2016 Application of least squares support vector regression and linear multiple regression for modeling removal of methyl orange onto tin oxide nanoparticles loaded on activated carbon and activated carbon prepared from Pistacia atlantica wood *Journal of colloid and interface science* **461** 425-34

[11] Hearst M A, Dumais S T, Osuna E, Platt J and Schölkopf B 1998 Support vector machines *IEEE Intelligent Systems and their applications* **13** 18-28

[12] Santosa B 2010 Tutorial Support Vector Machine *Teknik Industri, ITS*.[Online].

[13] Cortes C and Vapnik V 1995 Support-vector networks *Machine learning* **20** 273-97

[14] Smola A J and Schölkopf B 2004 A tutorial on support vector regression *Statistics and computing* **14** 199-222

[15] Gunn S R 1998 Support vector machines for classification and regression *ISIS technical report* **14** 5-16