

# FORECASTING OF TIME SERIES DATA USING SUPPORT VECTOR REGRESSION BASED ON KERNEL TYPES

**OKTARINA SAFAR NIDA**



**DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS AND NATURAL SCIENCES  
BOGOR AGRICULTURAL UNIVERSITY  
BOGOR  
2016**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



**@Hak cipta milik IPB University**

**IPB University**



**IPB University**  
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## THE UNDERGRADUATE THESIS STATEMENT AND SOURCES OF INFORMATION AND COPYRIGHT DEVOLUTION

I hereby declare that the undergraduate thesis entitled Forecasting of Time Series Data Using Support Vector Regression based on Kernel Types is my work under the guidance of the supervisory committee and it has not been submitted in any form to any college. Resources derived or quoted from works published and unpublished from other writers are mentioned in the text and listed in the References at the end of this undergraduate thesis.

I hereby bestow the copyright of my papers to the Bogor Agricultural University.  
Bogor, June 2016

*Oktarina Safar Nida*  
NIM G14120052

@Hak cipta milik IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
  2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## ABSTRACT

OKTARINA SAFAR NIDA. Forecasting of Time Series Data Using Support Vector Regression based on Kernel Types. Supervised by ASEP SAEFUDDIN and HARI WIJAYANTO.

Support vector regression (SVR) is a nonlinear method that uses the principle of risk minimization. SVR model uses the kernel function with regularization constant ( $C$  and  $\epsilon$ ). The kernel function which is commonly used in the SVR are linear, polynomial and radial basis function (RBF). Determination of kernel with its constants is required in SVR model. Grid search algorithm determines the best regularization constant combination based on the smallest error rate. SVR is able to replace the case of linear regression and forecasting. The accuracy of forecasting is very important. The type of time series data is divided into several models, i.e. constant, trend, polynomial, and seasonal model. The study simulated various model of time series data using SVR. Model that had been tested was evaluated by Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The simulation found that SVR model which used RBF kernel on the four types of time series data performed the best forecasting accuracy. In the application data, the RBF kernel with  $C$  of 10 and  $\epsilon$  of 0.04 was implemented to build the SVR model. In the training data, the best SVR model produced MAE of 7.15436, RMSE of 10.03607 and MAPE of 0.86904%, while in the testing data the model produced MAE of 6.67387, RMSE of 9.41013 and MAPE of 0.865755%. Even if the testing data produced smaller error compared to training data, the difference of errors in the training and testing data was not statistically significant.

Keywords: kernel, stocks, SVR, time series.

## ABSTRAK

OKTARINA SAFAR NIDA. Peramalan Data Deret Waktu Menggunakan *Support Vector Regression* berdasarkan Berbagai Tipe Kernel. Dibimbing oleh ASEP SAEFUDDIN dan HARI WIJAYANTO.

Support vector regression (SVR) adalah metode nonlinear yang menggunakan prinsip minimisasi resiko. Model SVR menggunakan fungsi kernel dengan regularisasi konstanta ( $C$  dan  $\epsilon$ ). Fungsi kernel yang umum digunakan pada SVR adalah linear, polinomial dan *radial basis function* (RBF). Penentuan kernel dengan regularisasi konstantanya yang optimal dibutuhkan dalam model SVR. Algoritma *grid search* menentukan kombinasi konstanta terbaik berdasarkan error terkecil. SVR mampu menggantikan regresi linear dan peramalan. Akurasi dalam peramalan sangat penting. Data deret waktu terbagi menjadi beberapa model yaitu konstan, tren, polinomial, dan musiman. Penelitian ini menyimulasikan berbagai model data deret waktu menggunakan SVR. Model yang dibentuk dievaluasi menggunakan *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Simulasi pada penelitian ini

Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.  
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

menemukan bahwa model SVR yang menggunakan kernel RBF pada keempat data deret waktu memberikan akurasi terbaik. Pada data aplikasi, kernel RBF dengan  $C$  sebesar 10 dan  $\epsilon$  sebesar 0.04 diimplementasikan untuk membangun model SVR. Pada data percobaan, model SVR terbaik menghasilkan MAE sebesar 7.15436, RMSE sebesar 10.03607 dan MAPE sebesar 0.86904%, sementara pada data pengujian dihasilkan MAE, RMSE, dan MAPE sebesar 6.67387, 9.41013 dan 0.865755%. Walaupun data pengujian menghasilkan error yang lebih kecil dibanding data percobaan, perbedaan error tersebut tidak signifikan secara statistik.

Kata kunci: deret waktu, kernel, saham, SVR.

@Hak cipta milik IPB University

IPB University



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
  2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



### *@Hak cipta milik IPB University*

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

# FORECASTING OF TIME SERIES DATA USING SUPPORT VECTOR REGRESSION BASED ON KERNEL TYPES

**OKTARINA SAFAR NIDA**

An undergraduate thesis  
in partial fulfilment of the requirements for the degree of  
Bachelor of Statistics  
in  
Department of Statistics

**DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS AND NATURAL SCIENCES  
BOGOR AGRICULTURAL UNIVERSITY  
BOGOR  
2016**

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



*@Hak cipta milik IPB University*

**IPB University**



**IPB University**  
— Bogor Indonesia —


Hak Cipta Dilindungi Undang-undang


1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Title : Forecasting of Time Series Data Using Support Vector Regression based on Kernel Types  
Name : Oktarina Safar Nida  
NIM : G14120052

Approved by

  
Prof Dr Ir Asep Saefuddin, MSc  
Supervisor I

  
Dr Ir Hari Wijayanto, MSi  
Supervisor II

Acknowledged by

  
Dr Anang Kurnia, MSi  
Head of Department

Graduation Date: 22 JUN 2018

Hak Cipta Dilindungi Undang-undang  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.  
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## PREFACE

All praise and grateful to Allah SWT who give me the strength in completing my undergraduate thesis. This study theme is about Support Vector Regression with the title is Forecasting of Time Series Data Using Support Vector Regression based on Kernel Types. It is conducted from January 2016 until Juni 2016.

Special appreciation goes to my supervisors, Prof Dr Ir Asep Saefudin, MSc and Dr Ir Hari Wijayanto MSi, for their supervision and support. I also express my gratitude to my dearest parents and my beloved little sister for their prayers and support. Thanks to all my friends for their kindness and moral support during my study. I hope this study will give benefit.

Bogor, June 2016

*Oktarina Safar Nida*

Hak Cipta Dilindungi Undang-undang  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.  
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

## CONTENTS

LIST OF TABLE	viii
LIST OF FIGURE	viii
INTRODUCTION	1
Background	1
Objectives	2
LITERATURE REVIEW	2
Time Series Regression Model	2
Support Vector Regression	2
The LQ45 Index	4
DATA AND METHODS	5
Data	5
Methodology	5
RESULT AND DISCUSSION	8
Simulation Study	8
The Application	11
CONCLUSION AND RECOMMENDATION	14
Conclusion	14
Recommendation	15
REFERENCES	15
BIOGRAPHY	17



**LIST OF TABLE**

1	The values of $C$ , $\epsilon$ , MAE, RMSE, and MAPE of constant data	8
2	The values of $C$ , $\epsilon$ , MAE, RMSE, and MAPE of trend data	9
3	The values of $C$ , $\epsilon$ , MAE, RMSE, and MAPE of polynomial data	9
4	The values of $C$ , $\epsilon$ , MAE, RMSE, and MAPE of seasonal data	10
5	The values of mean of MAE, RMSE, and MAPE	11
6	The t-test for MAE in training and testing data	14
7	The t-test for MAPE in training and testing data	14

**LIST OF FIGURE**

1	Flow chart of simulation method	7
2	Line plot of actual and prediction of constant model	8
3	Line plot of actual and prediction of trend model	9
4	Line plot of actual and prediction of polynomial model	10
5	Line plot of actual and prediction of seasonal model	10
6	Graph of performance combination of epsilon and cost at the first grid search algorithm	12
7	Graph of performance combination of epsilon and cost at the second grid search algorithm	12
8	Line plot of actual and prediction with the best SVR model in the training data	13
9	Line plot of actual and prediction in testing data	13

Hak Cipta Dilindungi Undang-undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.