

**LAKE BERA AND LAKE CHINI WATER QUALITY MONITORING
USING SUPPORT VECTOR MACHINE**

SITI FATIHAH ASY SYURA BT MAT JUBIT

**INSTITUTE OF GRADUATE STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2012

**LAKE BERA AND LAKE CHINI WATER QUALITY MONITORING
USING SUPPORT VECTOR MACHINE**

SITI FATIHAH ASY SYURA BT MAT JUBIT

**THESIS SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF
BIOINFORMATICS**

**INSTITUTE OF GRADUATE STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2012

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Siti Fatimah Asy syura bt Mat Jubit

(I.C/Passport No:) 870915355416

Registration/Matric No: SGJ 100008

Name of Degree: MASTER OF BIOINFORMATICS

TITLE(“this Work”): Lake Bera and Lake Chini Water Quality Monitoring Using Support Vector Machine (SVM)

Field of Study:

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any Copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature

Date:

Subscribed and solemnly declared before,

Witness’s Signature

Date:

Name:

Designation:

ABSTRACT

Water quality monitoring is very important to control the quality of water. Lake Bera and Lake Chini which are known as a very important wetland are used to apply SVM method to predict its water quality. The output used to predict the classification of high medium and low is the dissolved oxygen according to the standard provided by the Interim National Water Quality Standard of Malaysia and Department of Environment. The training and test data is divided to 80% for training data and 20% for testing data. The SVM is implemented using R software package kernlab which used ksvm as its implementation to do prediction. Kernel Anova was used to create the model. The result shows that the predicted accuracy is about 74%.

ABSTRAK

Kawalan kualiti air memainkan peranan yang sangat penting kerana ia membolehkan kualiti air dapat dikawal daripada dicemari. Tasik Bera dan Tasik Chini yang dikenali sebagai kawasan tanah lembap yang besar peranannya digunakan di dalam projek ini untuk dibuat ramalan kualiti air dengan mengaplikasikan kaedah SVM iaitu salah satu kaedah automatik. Kualiti air ditentukan dengan menggunakan oksigen terlarut sebagai kelas penentu untuk tasik ini berdasarkan sama ada oksigen terlarut berada di kelas yang tinggi, sederhana atau rendah berdasarkan Standard Kualiti Air Kebangsaan di Malaysia dan Jabatan Alam Sekitar. Data yang digunakan dibahagikan kepada data latihan dan data untuk percubaan di mana 80% digunakan sebagai data latihan dan 20% untuk percubaan. Kaedah SVM diaplikasikan dengan menggunakan perisian R di mana pakej kernlab digunakan dan ia mengaplikasikan ksvm yang dapat membuat ramalan untuk kaedah SVM. Berdasarkan kaedah kernel yang dapat digunakan kernel Anova digunakan dalam kaedah ini. Ketepatan ramalan yang di peroleh adalah sebanyak 74%.

ACKNOWLEDGEMENT

I would like to thank my supervisor and my co supervisor Dr. Sorayya Bibi Malek binti Malek Abd Rashid and Dr. Sharifah Mumtazah binti Syed Ahmad Abdul Rahman whom have guided me in completing this project especially on using the R to do SVM. I also would like to thank my friends who have helped me understanding the SVM. Lastly I would like to express my thanks to my family who has kept supporting me throughout this project.

Table of Contents

List of Figures.....	viii
List of Tables.....	ix
List of Abbreviations.....	x
1. INTRODUCTION	
1.1. Lake Bera and Lake Chini.....	1
1.2. Water Quality Monitoring.....	3
1.3. Support Vector Machine (SVM).....	5
1.4. Scope.....	5
1.5. Objectives.....	6
2. LITERATURE REVIEW.....	7
3. METHODOLOGY	
3.1. Support Vector Classification.....	12
3.2. Kernel-based Machine Learning Lab(Kernlab).....	16
3.3. Data.....	20
3.4. Data Prediction Using SVM.....	23
4. RESULTS.....	25
5. DISCUSSIONS.....	33
6. CONCLUSION.....	35

INSTRUMENTATION.....	36
APPENDIX	
Appendix A.....	43
Appendix B.....	59
Appendix C.....	61
BIBLIOGRAPHY.....	65

List of Figures

Figure 3.1 Lake Bera map.....	20
Figure 3.2 Lake Chini map.....	21
Figure 4.1 DO Levels at Lake Bera and Lake Chini.....	25
Figure 4.2 Levels of DO by year.....	26
Figure 4.3 Predictions and actual value of testing data.....	29
Figure 4.4 Sensitivity, Specificity and Accuracy Graph.....	31

List of Tables

Table1.1 Interim National River Water Quality Standards.....	4
Table1.2 Class and parameters standard.....	4
Table 3.1 Monitored station of the lakes.....	22
Table 4.1 Parameter ranking.....	26
Table 4.2 Accuracy for kernels.....	28
Table 4.3 Prediction result of training data.....	30
Table 4.4 Calculated sensitivity, specificity and accuracy.....	31

List of Abbreviation

ANN	Artificial Neural Network
ASMA	AlamSekitar Malaysia Sdn. Bhd
BOD	Biochemical oxygen demand
C	Constant
COD	Chemical oxygen demand
COND	Electrical conductivity
CPU	Colony Forming Units
CSV	Comma delimited
DNA	Deoxyribonucleic acid
DO	Dissolved Oxygen
DOE	Department of Environment
DS	Dissolved solid
ERM	Empirical Risk Minimization
INWQS	Interim National Water Quality Standard, Malaysia
Kernlab	Kernel-based Machine Learning Lab
mg/l	Milligram per liter
MI	Miligram
NH ₃ N	Ammonia nitrogen
NO ₃	Nitrate nitrogen
NTU	Nephelometric turbidity units
PO ₄	Phosphate
Ppt	Parts per thousand
QP	Quadratic Problem
SLT	Statistical learning theory
SMO	Sequential Minimization Optimization
SS	Suspended solid
SVM	Support Vector Machine