Universiti Teknologi MARA

Forecasting of Kelantan's Air Pollution Index (API) $PM_{2.5}$ using Support Vector Machine (SVM)

Nur Fithrinnisaa Binti Zamani

Report submitted in fulfilment of the requirements
For
Bachelor of Science (Hons.) Management Mathematics
Faculty of Computer and Mathematical Sciences

July 2020

STUDENT'S DECLARATION

I certify that this report and the research to which it refers are the product of my own work and that any ideas or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

……

NUR FITHRINNISAA BINTI ZAMANI

2017175335

AUGUST 5, 2020

ABSTRACT

Forecasting the air pollution index has become a popular topic in recent years due to the impact of air pollution on environmental and human health. The Support Vector Machine (SVM) is not only appropriate for object classification, regression analysis and pattern recognition, it can also be used in time series forecasting. The work presented here aims to compare the accuracy of different types of the kernel function in Support Vector Machine (SVM) and to build a forecasting system for Kelantan's air pollution index for $PM_{2.5}$ using Support Vector Machine (SVM). The data used was provided by the Department of Environment (DOE) and was recorded from two Continuous Air Quality Monitoring Stations (CAQM) located at Tanah Merah and Kota Bharu. The results of the model were analyzed by using mean absolute error (MAE) and root mean squared error (RMSE). It is found that the proposed model using Radial Basis Function (RBF) kernel function with its parameters of cost and gamma equal to 100 can effectively and accurately forecast the air pollution index based on the model testing with 0.03868583 (MAE) and 0.06251793 (RMSE) for API in Kota Bharu and 0.03857308 (MAE) and 0.05895648 (RMSE) for API in Tanah Merah.

Keywords: Air pollution index, Support Vector Machine (SVM), time series forecasting, kernel function, $PM_{2.5}$

TABLE OF CONTENTS

CONTENTS		PAGE
SUPERVISOR'S APPROVAL		ii
DECLARATION		iii
ACKNOWLEDGEMENT		iv
ABSTRACT		v
TABLE OF CONTENTS		vi
LIST OF FIGURES		viii
LIST OF TABLES		ix
LIST OF ABBREVIATIONS		x
CHAPTER ONE: INTRODUCTION		
1.1	Background of the Study	1
1.2	Problem Statement	2
1.3	Objectives of the Study	4
1.4	Scope of the Study	4
1.5	Significance of the Study	4
CHAPTER TWO: LITERATURE REVIEW		
2.1	Application to Forecast the Air Pollution	5
2.2	Previous Studies on Support Vector Machine (SVM)	6
2.3	Previous Studies on Prediction of Air Pollution using SVM	8
2.4	Summary	9
CHAPTER THREE: RESEARCH METHODOLOGY		
3.1	Method of Data Collection	10
3.2	Support Vector Machine (SVM)	10

3.3 Data Analysis 13
3.3.1 Data Normalization 15
3.3.2 Model Development 16
3.3.3 Data Implementation 18

CHAPTER FOUR: RESULTS AND DISCUSSIONS

4.1 Data Normalization 19
4.2 Model Development 24
4.2.1 Kernel Function Selection 24
4.2.2 Parameter Selection 25
4.2.3 Fitting on Model Training 28
4.2.4 Validation on Model Testing 28
4.3 Data implementation 29

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion 33
5.2 Recommendations 34

REFERENCES 35

APPENDICES

APPENDIX A: Actual data of API in Kota Bharu, Kelantan (Daily data) 36
APPENDIX B: Actual data of API in Tanah Merah, Kelantan (Daily data) 39
APPENDIX C: Command of R Programming for API in Kota Bharu, Kelantan 42
APPENDIX D: Command of R Programming for API in Tanah Merah, Kelantan 46
APPENDIX E: Summary on performance of tuning SVM 50
APPENDIX F: Combined data (normalized data, training data, testing data and predicted data) for API in Kota Bharu, Kelantan 52
APPENDIX G: Combined data (normalized data, training data, testing data and predicted data) for API in Tanah Merah, Kelantan 61