UNIVERSITI TEKNOLOGI MARA CAWANGAN PULAU PINANG

DEVELOPMENT OF REGRESSION MODELS FOR PREDICTING WATER QUALITY INDEX BASED ON DISSOLVED OXYGEN FOR RIVER POLLUTION ASSESSMENT

MUHAMAD IRFAN BIN AHMAD SUHKRI

BACHELOR OF ENGINEERING (HONS) ELECTRICAL AND ELECTRONIC ENGINEERING

February 2025

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student	:	Muhamad Irfan Bin Ahmad Suhkri	
Student I.D. No.	:	202278	
Programme :		B. End. (Hons.) Electrical and Electronic Engineering	
		- CEEE200	
Faculty	:	Faculty of Electrical Engineering	
Thesis	:	Development of Regression Models for Predicting Water Quality Index Based on Dissolved Oxygen for	
		River Pollution Assessment	
Signature of Student	:		
Date	:	23 February 2025	

ABSTRACT

Water is a critical resource in Malaysia, but river water quality has been significantly impacted by industrialization, urbanization, and rapid development. Traditional water quality monitoring methods are time-consuming and often result in delayed detection of pollution. This study addresses these challenges by developing regression models to predict the Water Quality Index (WQI) based on Dissolved Oxygen (DO) measurements, enabling timely and efficient river pollution assessment. Data were collected from Kaggle, consisting of 219 clean and polluted river water samples from spanning June to November 2023. After data categorization, only 44 of clean and polluted DO data were used respectively. The data were categorized into clean and polluted water conditions and validated through statistical analyses, including normality tests and error bar plots. Multiple regression techniques, such as Linear Regression and Robust Linear Regression, were implemented using MATLAB and Python. Among the models tested, MATLAB's Linear Regression achieved the highest R² value of 0.95397 and the lowest RMSE of 7.2728, demonstrating superior performance. These findings highlight the potential of regression modelling as a reliable and proactive approach to water quality monitoring, supporting environmental authorities and stakeholders in safeguarding freshwater ecosystems and improving resource management.

ACKNOWLEDGEMENT

Gracious and the Most Merciful, for His abundant blessings that have guided and sustained me throughout my academic journey and in the completion of this proposal. May Allah's blessings be upon His final Prophet Muhammad SAW, his family, and his companions.

I extend my sincere thanks to my supervisor, Dr. Mohd Suhaimi Sulaiman and my esteemed panels, Dr. Muhammad Khusairi Osman and Dr. Rozan Boudville for their invaluable guidance, unwavering support, and constructive feedback that have significantly contributed to the development of this proposal.

Special appreciation is extended to my colleagues and friends, for their continuous assistance and collaborative efforts in overcoming challenges encountered throughout this final year project. Their support has been instrumental in the successful completion of this endeavour.

Lastly, heartfelt gratitude is extended to my parents and family members, who have been a constant source of spiritual and financial support. Their encouragement during challenging times and unwavering motivation played a crucial role in ensuring that I persevered and successfully completed this proposal.

TABLE OF CONTENTS

AUT	THOR'S DECLARATION	i
ABS	STRACT	ii
ACF	KNOWLEDGEMENT	iii
TAE	BLE OF CONTENTS	iv
LIST	Г OF TABLES	vi
LIST	Г OF FIGURES	vii
LIST	Г OF APPENDICES	viii
LIST	Г OF SYMBOLS	ix
LIST	T OF ABBREVIATIONS	x
CHA	APTER 1 INTRODUCTION	1
1.1	Research Background	1
1.2	Problem Statement	2
1.3	Objectives	3
1.4	Scope of Work	4
1.5	Significance of Study	5
СНА	APTER 2 LITERATURE REVIEW	6
2.1	Introduction	6
2.2	DO PARAMeter	7
2.3	Statistical Analysis	8
	2.3.1 Test for Normality	8
	2.3.2 Error Bar Plot Analysis	9
2.4	Learning Algorithm	10
CHA	APTER 3 RESEARCH METHODOLOGY	16
3.1	Introduction	16
3.2	Data collection and categorization	18
3.3	Statistical Analysis	20
	iv	