Universiti Teknologi MARA

Implementation of Predictive Analytics for Future Risk Prediction of Malaysian Against Covid-19 Infection

Muhammad Haziq Jakaria

Thesis submitted in fulfilment of the requirements for Bachelor of information Technology (Hons.) Business computing Faculty of computer and Mathematical sciences

January 2022

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. Firstly, my special thanks go to my supervisor Madam Cik Ku Haroswati Cik Ku Yahaya for always giving me encouragement and guidance to reach the objectives of my final year project. Not to forget, I would also like to extend my gratitude towards Madam Norulhidayah Isa for her guidance to complete my thesis. Apart from that, my gratitude also goes to the experts such as Madam Zanariah, and Madam Zeti Darleena Eri for their expert views on the deployment of my final year project. Besides that, my respect also goes to UiTM Kuala Terengganu for giving me the opportunity to do the assignment work and providing me with all the necessity which made me complete this final year project. Special appreciation also goes to my beloved parent for giving me an emotional support and encouragement to complete this thesis. Last but not least, I would also like to give my gratitude to my dearest friend for their support and guidance to complete this thesis.

ABSTRACT

On the 31st December of 2019, the first cluster of. Later, the unknown disease was identified as Covid-19 which was a highly infectious virus that can cause chronic respiratory disease. After that, this virus spread throughout the world and was later declared a pandemic by the WHO. Malaysia was no exception and was also affected. Therefore, the MoH come out with an initiative to create a prediction model using the SEIR model to predict cases. However, it does not emphasize the prediction based on events such as the MCO implementation. Due to this, the impact of MCO relaxation has caused a drastic increase of covid cases up to 40,000 cases a day. This big figure also comes with a severe consequence to the safety of the public. Therefore, this work presented a forecast that can help to identify a unique pattern during the MCO period based on data from 18 March until 9 June of 2020. This project also employs the CRISP-DM methodology until the outcomes can be made into a dashboard. Variables such as date, infection, recovery, and fatality numbers are crucial to achieve better accuracy of ARIMA forecasting. With this research, the EDA on covid trends can be plotted and the forecast is also heavily affected by the component of time series. The result also revealed that ARIMA (1, 2, 8), (1, 2, 3) and (0, 3, 3) showed appropriate results. In short, the ARIMA model is a good model to forecast time series-related data with a proper parameter adjustment.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	i
STUDENT DECLARATION	ii
ACKNOWLEDGEMENT	iii
ABSTARCT	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	viii
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xiii
CHAPTER ONE: INTRODUCTION	1
1.1 Background of Study	1
1.2 Current Process	4
1.3 Problem Statement	5
1.4 Project Objectives	7
1.5 Project Scope	7
1.6 Project Significance	8
1.7 Project Framework	9
1.8 Conclusion	10
CHAPTER TWO: LITERATURE REVIEW	11
CHAPTER THREE: METHODOLOGY	37
3.1 Introduction to adapted CRISP-DM	37
3.2 Data and Business Understanding	39

3.2.1 Collect Initial Data 39
3.2.2 Describe Data 42
3.2.3 Explore Data 44
3.2.4 Verify Data Quality 56
3.3 Data Preparation 57
3.3.1 Clean Data 57
3.3.2 Construct Data 57
3.3.3 Integrate Data 57
3.3.4 Format Data 59
3.3.5 Attribute Selection 60
3.4 Modeling 61
3.4.1 Select Modeling Technique 61
3.4.2 Generate Test Design 62
3.5 Build Model 68
3.5.1 Decision Tree 68
3.5.2 ARIMA model 69

CHAPTER FOUR: RESULT AND DISCUSSION 71

4.1 Introduction 71
4.2 Current Situation 71
4.3 DM Experiment 1 75
4.3.1 Quick ARIMA 75
4.3.2 Custom ARIMA 77
4.4 DM Experiment 2 81
4.4.1 MCO forecasting 81
4.5 Discussion 89
4.6 Deployment 90
4.6.1 Data Import 91
4.6.2 Data analysis expression (DAX) 91
4.6.3 Visualize the Data Analysis Expression 93