PROJECT REPORT

VISUALIZATION AND PREDICTION OF SARS-COV-1 SPREADING BASED OF MOBILIZED URBANIZATION PATTERN USING NAÏVE BAYES THEOREM

MARSHEYLA OLIVIA DEBORA BR. S
16.K1.0051

Faculty of Computer Science
Soegijapranata Catholic University
2022

HALAMAN PENGESAHAN

Judul Tugas Akhir:	:	Visualization and Prediction of SARS-CoV-1 Spreading Based of Mobilized Urbanization Pattern using NaÃ ve Bayes Theorem
Diajukan oleh	:	Marsheyla Olivia Debora Br. Si
NIM	:	16.K1.0051
Tanggal disetujui	:	24 Mei 2022
Telah setujui oleh		
Pembimbing	:	Yonathan Purbo Santosa S.Kom., M.Sc
Penguji 1	:	Yonathan Purbo Santosa S.Kom., M.Sc
Penguji 2	:	R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D
Penguji 3	:	Y.b. Dwi Setianto S.T., M.Cs.
Penguji 4	:	Yulianto Tejo Putranto S.T., M.T.
Penguji 5	:	Rosita Herawati S.T., M.I.T.
Penguji 6	:	Hironimus Leong S.Kom., M.Kom.
Ketua Program Studi	:	Rosita Herawati S.T., M.I.T.
Dekan	:	Dr. Bernardinus Harnadi S.T., M.T.

Halaman ini merupakan halaman yang sah dan dapat diverifikasi melalui alamat di bawah ini.

sintak.unika.ac.id/skripsi/verifikasi/?id=16.K1.0051

STATEMENT OF ORIGINALITY

I, the undersigned:

Name : Marsheyla Olivia Debora Br. S

ID : 16.K1.0051

Certify that this project was made by myself and not copy or plagiarize from other people, except that in writing expressed to the other article. If it is proven that this project was plagiarizes or copy the other, I am ready to accept a sanction.

Semarang, May, 24, 2022

MARSHEYLA OLIVIA DEBORA BR. S

16.K1.0051

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : Marsheyla Olivia Debora Br. Silitonga

ID : 16.K1.0051

declare that this work, titled "Visualization and Prediction of SARS-CoV-1 Spreading based of Mobilized Urbanization Pattern using Naïve Bayes Theorem", and the work presented in it, is my own. I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at Soegijapranata Catholic University
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given.
5. Except for such quotations, this work is entirely my own work.
6. I have acknowledged all main sources of help.
7. Where the work is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Semarang, May, 24, 2022

MARSHEYLA OLIVIA DEBORA BR. S

16.K1.0051

ACKNOWLEDGMENT

I have received a myriad of support, advice, and assistance throughout this document writing. I would like to thank my supervisor who have guided me every step along the way and greatly encouraged and helped me to finish this writing. Without the help of my supervisor along with the staff of my faculty, I wouldn't have been able to finish this thesis.

I would also like to thank every single individual who have given me immense support and accompanied me, whose lessons have helped me immensely in finishing my studies at Soegijapranata Catholic University.

You gave me great escape to rest my mind from my thesis and advances further in my life.

Semarang, May, 24, 2022

MARSHEYLA OLIVIA DEBORA BR. S

16.K1.0051

ABSTRACT

Epidemiology is a study and analysis of distribution, pattern and determinant of health based on a disease form in a population. Epidemiology offers a strong data to calculate how risk factor and intervention can affect population's health in a crisis condition. One of the key matric in epidemiological is the base reproduction of virus transmigration, which means one sick individual can infect multiple other individuals. Time series data in epidemiology is a critical aspect to be used for analyzation and visualization to predict on-coming waves of epidemic waves. Data collected in a time-series format are crucial keys to prevent further spreading of the virus as it is dependent on the time which the event took place. Forecasting these data helps to detect future epidemics by understanding the spread of disease related by factors such as environments.

Keyword: Epidemiology Prediction, Time Series Data, Naïve Bayes, Data Mining

TABLE OF CONTENTS

COVER ..... **i**

HALAMAN PENGESAHAN ..... **ii**

STATEMENT OF ORIGINALITY ..... **iii**

DECLARATION OF AUTHORSHIP ..... **iv**

ACKNOWLEDGMENT ..... **v**

ABSTRACT ..... **vi**

TABLE OF CONTENTS ..... **vii**

LIST OF FIGURE ..... **ix**

LIST OF TABLE ..... **xi**

CHAPTER 1 INTRODUCTION **1**

1.1. Background 1

1.2. Problem Formulation 2

1.3. Scope 2

1.4. Objective 2

CHAPTER 2 LITERATURE STUDY **3**

CHAPTER 3 RESEARCH METHODOLOGY **8**

CHAPTER 4 ANALYSIS AND DESIGN **9**

4.1. Naïve Bayes and Epidemiology Analysis and Relation 9

4.2. Virus Spreading Illustration 10

4.3. Time Series Data of SARS 2003 11

4.4. Time Series Design 12

4.5. Naïve Bayes Function 15

CHAPTER 5 IMPLEMENTATION AND TESTING **16**

5.1. Implementation of formula and Designs 16

5.2. Method Testing and Data Results 19
5.3. RMSE, MAPE and Cross Validation 22
5.4. Implementation of Map Formula and Designs 25
CHAPTER 6 CONCLUSION 31
REFERENCES 32

LIST OF FIGURE

Figure 4-1 Example of Virus Spreading 10

Figure 4-2 Time Series Graph on Total Cases 12

Figure 4-3 Time Series Graph on Total Deaths 13

Figure 5-1 Import Libraries 16

Figure 5-2 Import data 16

Figure 5-3 Results of Imported Data 17

Figure 5-4 Plot Data 17

Figure 5-5 Separating the Data 17

Figure 5-6 Training the data 18

Figure 5-7 Plotting the Results 18

Figure 5-8 Result of t-1 19

Figure 5-9 RMSE and MAPE of t-1 19

Figure 5-10 Result of t-2 19

Figure 5-11 Result of t-3 20

Figure 5-12 Result of train_len= 20, t-1 20

Figure 5-13 RMSE and MAPE of t-2 20

Figure 5-14 RMSE and MAPE of t-3 21

Figure 5-15 Rolling Basis Cross Validation 23

Figure 5-16 Importing Libraries 25

Figure 5-17 Pre-Processing Data 25

Figure 5-18 Output of Data Pre-processing 25

Figure 5-19 Data Adjustment 25

Figure 5-20 Output of Data Adjustment 26

Figure 5-21 Assigning Colors 26

Figure 5-22 Creating dictionary for TimeSliderChoropleth 26

Figure 5-23 Map Configuration 27

Figure 5-24 Finished Map 28

Figure 5-25 Comparison Between Dates and Data 28

Figure 5-26 Comparison on another Area 29

Figure 5-27 Forecasted Map 29

LIST OF TABLE

Table 4-1 Time Frame Data 11

Table 5-1 Rolling Basis Validation 24