

**YOUTUBE USER SENTIMENT ANALYSIS ON METAVERSE-RELATED  
OPINION USING NEURAL NETWORK**

**THESIS**



by

**Muhammad Arkan Rayhan**

**18.61.0146**

**BACHELOR OF INFORMATICS  
FACULTY OF COMPUTER SCIENCE  
UNIVERSITAS AMIKOM YOGYAKARTA  
2023**

**YOUTUBE USER SENTIMENT ANALYSIS ON METAVERSE-RELATED  
OPINION USING NEURAL NETWORK**

**THESIS**

to fulfill requirements  
for a Bachelor's degree  
in the Informatics study program



by

**Muhammad Arkan Rayhan**

**18.61.0146**

**BACHELOR OF INFORMATICS  
FACULTY OF COMPUTER SCIENCE  
UNIVERSITAS AMIKOM YOGYAKARTA  
2023**

**APPROVAL**

**THESIS**

**YOUTUBE USER SENTIMENT ANALYSIS ON  
METAVERSE RELATED OPINION  
USING NEURAL NETWORK**

prepared and arranged by

**Muhammad Arkan Rayhan**

**18.61.0146**

has been approved by undergraduate thesis supervisor  
on 22 December 2022

Supervisor,



**Ferian Fauzi Abdalloh, M.Kom.**  
**NIK. 190302276**

VALIDATION

THESIS

YOUTUBE USER SENTIMENT ANALYSIS ON  
METAVERSE RELATED OPINION  
USING NEURAL NETWORK

prepared and arranged by

**Muhammad Arkan Rayhan**

**18.61.0146**

has been maintained by examiners

on 22 December 2022

**The Examiners**

**Examiner**

**Signature**

Ferian Fauzi Abdulloh, M.Kom.  
NIK. 190302276



Anggit Dwi Hartanto, M.Kom.  
NIK. 190302163



Donni Prabowo, M.Kom  
NIK. 190302253



This Thesis has been accepted as one of  
the requirements for obtaining a Bachelor of Computer Degree  
on 22 December 2022

**DEAN OF FACULTY OF COMPUTER SCIENCE**



Hanif Al Fatta, S.Kom., M.Kom.  
NIK. 190302096

## DECLARATION

I, the undersigned below, state that, this thesis is my own work (ORIGINAL) and the contents of this thesis have never been applied by any other person to receive an academic degree at a certain education institution, and as far as I know, there are no works or thoughts which have been written and/or published by anyone, except those in writing which are listed in this manuscript and which are mentioned in the reference list.

Anything that applies to the manuscripts and works that have been made is my own responsibility.

Yogyakarta, 22 December 2022



Muhammad Arkan Rayhan

NIM. 18.61.0146

## MOTTO

*"Ad astra!"*

-Publius Vergilius Maro-

*"The best climbers, know how to fall."*

-Pantheon-

*"You are suffering from self doubt, while others are intimidated by your full potential."*

-Unknown-

*"Artificial Intelligence, deep learning, machine learning—whatever you're doing if you don't understand it—learn it. Because otherwise, you're going to be a dinosaur within 3 years."*

-Mark Cuban-

## DEDICATION

All praise is due to Allah SWT, the Lord of the 'Alamin, the forgiver and the concealer of faults, the one who responds to duas. For His guidance, the author can complete the study entitled YouTube User Sentiment Analysis on Metaverse Related Opinion Using Neural Network. More appreciation and gratitude are also given to:

1. The author's parents, for their endless love, support, and encouragement throughout my pursuit of education.
2. The author's little brother, for all the company in helping around the family and as a reminder for me to be a great example that hopefully, you will always look up to.
3. The author's close friends, namely The Hunter Bros which consist of Agrifaldy, Arya, Eqhi, Ian, Oksa, and Via who have accompanied me through joy and sorrow and shared their words of advice with encouragement to finish this study.
4. The author's friends Rivin, Zan, Kun, Tiffu, Ayaya, and others in the Champion Club for being the absolute best in terms of helping me achieve better results throughout this study.

## ACKNOWLEDGEMENT

Praise and deep gratitude to Allah SWT for the profusion of grace, and guidance of Him given to the writer that made the undergraduate thesis could be complete. Greetings and sholawah might always be devoted to the prophet of Muhammad SAW. The undergraduate thesis entitled YouTube User Sentiment Analysis on Metaverse Related Opinion Using Neural Network, designed to fulfill the undergraduate program (S1) in the Study Program of Informatics, Faculty of Computer Science, Universitas AMIKOM Yogyakarta. The author would like to thank you profusely for all the help that has been given, either directly or indirectly, to complete this final undergraduate thesis during the preparation process. In particular, gratitude is due to:

1. Prof. Dr. M. Suyanto, M.M. As Rector of Universitas AMIKOM Yogyakarta.
2. Mr. Hanif Al Fatta, S.Kom., M.Kom. As Dean of Faculty of Computer Science Universitas AMIKOM Yogyakarta.
3. Mrs. Windha Mega PD, M.Kom. As Head of Study program S1 Informatika Universitas AMIKOM Yogyakarta.
4. Mr. Ferian Fauzi Abdulloh, S.Kom., M.Kom. As the author's thesis supervisor who provides guidance and encouragement in the preparation of this thesis.
5. The entire lecturer staff of the Study Program of Informatika Universitas AMIKOM Yogyakarta has guided me with the knowledge.
6. All the parties who could not have been listed one by one by the author, who contributed a lot of support to complete this study.



The authors concluded that this work has not been flawless, either in terms of content or presentation. Ideas and positive suggestions are expected at the conclusion of this thesis. Lastly, the writers hope that this research will offer a further contribution to science.

Yogyakarta, 22 December 2022



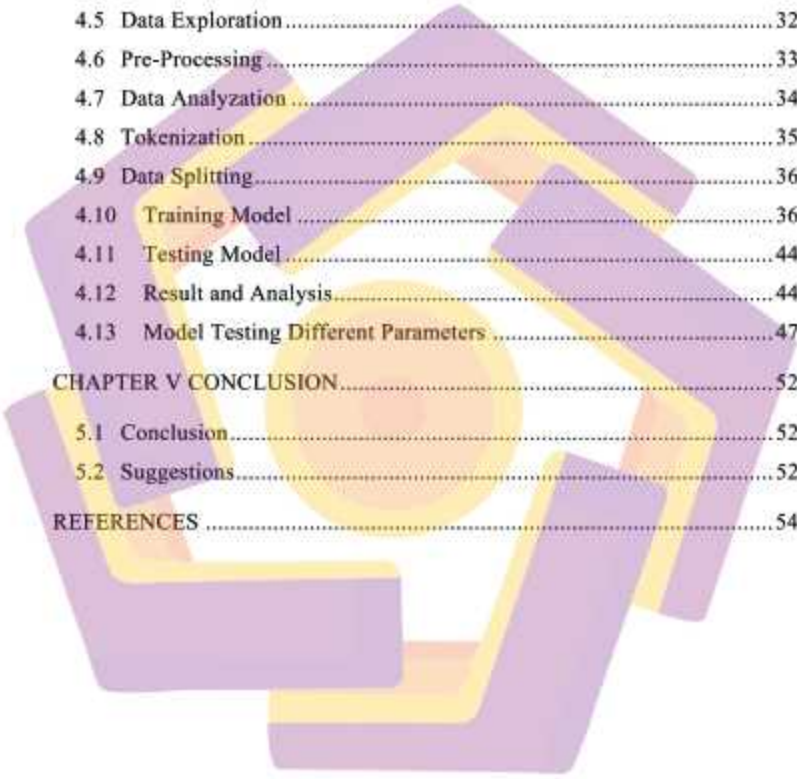
Muhammad Arkan Rayhan



## TABLE OF CONTENTS

|                                    |      |
|------------------------------------|------|
| COVER.....                         | II   |
| APPROVAL .....                     | III  |
| VALIDATION.....                    | IV   |
| DECLARATION .....                  | V    |
| MOTTO.....                         | VI   |
| DEDICATION.....                    | VII  |
| ACKNOWLEDGEMENT .....              | VIII |
| TABLE OF CONTENTS .....            | X    |
| LIST OF TABLES .....               | XIII |
| LIST OF FIGURES.....               | XIV  |
| GLOSSARY .....                     | XVI  |
| ABSTRACT.....                      | XVII |
| CHAPTER I INTRODUCTION.....        | 1    |
| 1.1 Background.....                | 1    |
| 1.2 Problems.....                  | 3    |
| 1.3 Problems Limitation .....      | 4    |
| 1.4 Research Purpose.....          | 4    |
| 1.5 Research Benefits .....        | 4    |
| 1.6 Research Methodology .....     | 5    |
| 1.6.1 Data Collection Method ..... | 5    |
| 1.6.2 Literature Review .....      | 5    |
| 1.6.3 Data Scraping .....          | 5    |
| 1.6.4 Analysis Method.....         | 5    |
| 1.6.5 System Design.....           | 5    |
| 1.6.6 Implementation.....          | 5    |

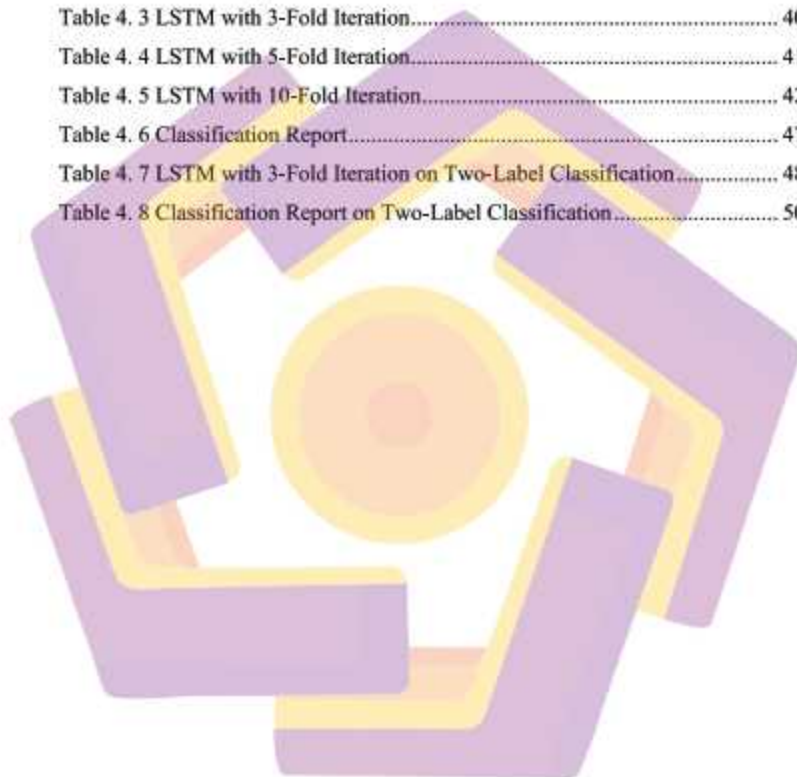
|   |           |
|---|-----------|
| 1.6.7 Testing.....                                    | 6         |
| 1.7 Systematics Writing.....                          | 6         |
| <b>CHAPTER II RELATED WORKS .....</b>                 | <b>8</b>  |
| 2.1 Literature Review .....                           | 8         |
| 2.2 Theoretical Basis .....                           | 13        |
| 2.2.1 Sentiment Analysis.....                         | 13        |
| 2.2.2 Machine Learning.....                           | 14        |
| 2.2.3 Deep Learning.....                              | 14        |
| 2.2.4 Recurrent Neural Networks.....                  | 14        |
| 2.2.5 Long Short-Term Memory .....                    | 16        |
| 2.2.6 Softmax .....                                   | 18        |
| 2.2.7 Keras .....                                     | 18        |
| 2.2.8 Adam Optimizer.....                             | 19        |
| 2.2.9 Python.....                                     | 19        |
| 2.2.10 TensorFlow .....                               | 19        |
| 2.2.11 Pandas.....                                    | 20        |
| 2.2.12 Scikit-learn.....                              | 20        |
| <b>CHAPTER III RESEARCH METHODOLOGY.....</b>          | <b>22</b> |
| 3.1 Research Stages .....                             | 22        |
| 3.2 Research Pipeline .....                           | 23        |
| 3.2.1 Data Scraping Method .....                      | 24        |
| 3.2.2 Data Labelling.....                             | 28        |
| 3.2.3 Pre-Processing.....                             | 28        |
| 3.2.4 Tokenization .....                              | 28        |
| 3.3 Data Splitting.....                               | 29        |
| 3.4 Data Training & Testing .....                     | 29        |
| 3.5 Result Analysis.....                              | 29        |
| <b>CHAPTER IV IMPLEMENTATION AND DISCUSSION .....</b> | <b>30</b> |
| 4.1 System Implementation.....                        | 30        |



|       |  |    |
|-------|--|----|
| 4.2   | Research Hardware and Software .....     | 30 |
| 4.2.1 | Hardware Specification .....             | 30 |
| 4.2.2 | Software Specification .....             | 30 |
| 4.3   | Importing Dependencies .....             | 31 |
| 4.4   | Load Datasets .....                      | 31 |
| 4.5   | Data Exploration .....                   | 32 |
| 4.6   | Pre-Processing .....                     | 33 |
| 4.7   | Data Analysis .....                      | 34 |
| 4.8   | Tokenization .....                       | 35 |
| 4.9   | Data Splitting .....                     | 36 |
| 4.10  | Training Model .....                     | 36 |
| 4.11  | Testing Model .....                      | 44 |
| 4.12  | Result and Analysis .....                | 44 |
| 4.13  | Model Testing Different Parameters ..... | 47 |
|       | CHAPTER V CONCLUSION .....               | 52 |
| 5.1   | Conclusion .....                         | 52 |
| 5.2   | Suggestions .....                        | 52 |
|       | REFERENCES .....                         | 54 |

## LIST OF TABLES

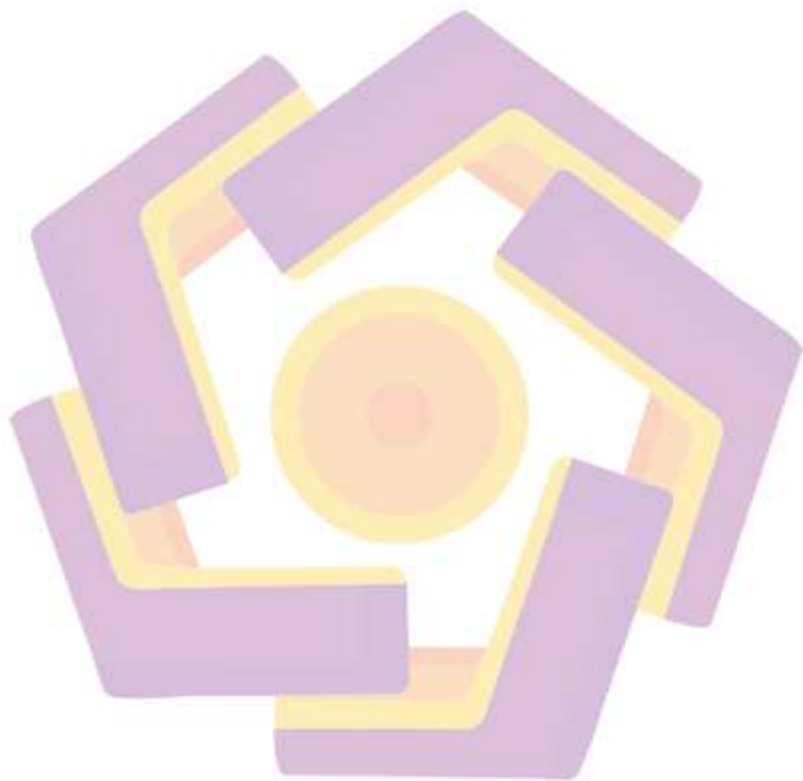
|  |    |
|--|----|
| Table 2. 1 Comparison table between previous studies .....             | 11 |
| Table 4. 1 Hardware Specification.....                                 | 30 |
| Table 4. 2 Software Specification .....                                | 30 |
| Table 4. 3 LSTM with 3-Fold Iteration.....                             | 40 |
| Table 4. 4 LSTM with 5-Fold Iteration.....                             | 41 |
| Table 4. 5 LSTM with 10-Fold Iteration.....                            | 42 |
| Table 4. 6 Classification Report.....                                  | 47 |
| Table 4. 7 LSTM with 3-Fold Iteration on Two-Label Classification..... | 48 |
| Table 4. 8 Classification Report on Two-Label Classification.....      | 50 |



## LIST OF FIGURES

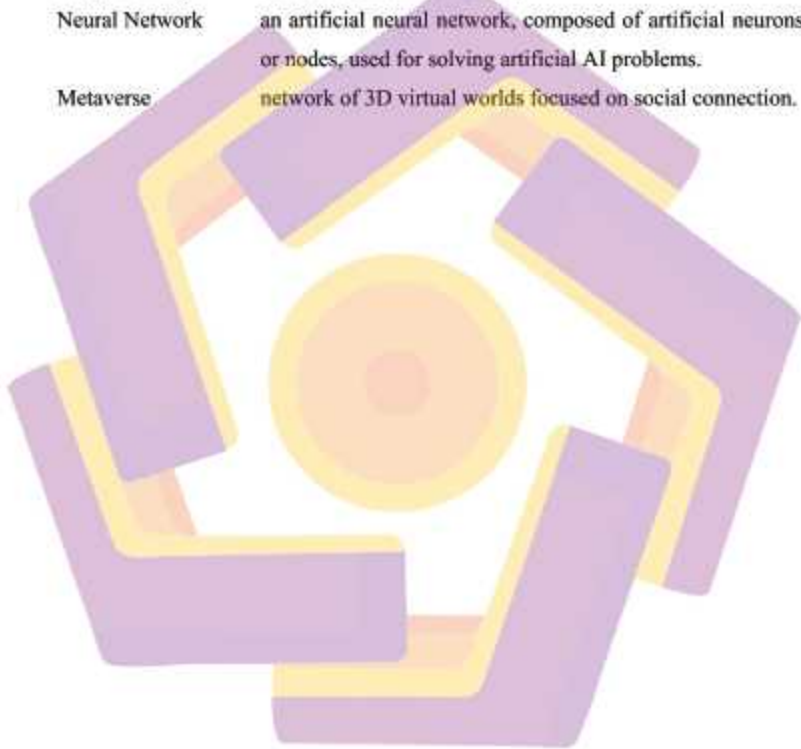
|   |    |
|---|----|
| Figure 2. 1 RNN Formula .....   | 15 |
| Figure 2. 2 Image Representation of Recurrent Neural Network .....                          | 15 |
| Figure 2. 3 Image Representation of Unrolled Recurrent Neural Network .....                 | 16 |
| Figure 2. 4 Image Representation of LSTM and its formula .....                              | 17 |
| Figure 2. 5 Softmax Function .....  | 18 |
| Figure 3. 1 Research Stages.....  | 22 |
| Figure 3. 2 Research Pipeline.....  | 24 |
| Figure 3. 3 Results of 50 YouTube videos from the Youtube Data.....                         | 26 |
| Figure 3. 4 Results of all the comments (without replies) with specific.....                | 28 |
| Figure 4. 1 All the required libraries to perform a sentiment analysis on RNN and LSTM..... | 31 |
| Figure 4. 2 Dataset Implementation.....   | 32 |
| Figure 4. 3 Total Dataset Sentiment Count of Positives, Neutrals, and Negatives.....        | 32 |
| Figure 4. 4 Plot Labeled Sentiment.....   | 33 |
| Figure 4. 5 Pre-Processing Method.....  | 34 |
| Figure 4. 6 Word Clouds for Negative Sentiment from YouTube Comments .....                  | 35 |
| Figure 4. 7 Word Clouds for Positive Sentiment from YouTube Comments .....                  | 35 |
| Figure 4. 8 Word Clouds for Positive Sentiment from YouTube Comments .....                  | 35 |
| Figure 4. 9 Tokenizer on processed comments.....  | 36 |
| Figure 4. 10 Data Splitting.....  | 36 |
| Figure 4. 11 Data Count after being split.....  | 36 |
| Figure 4. 12 LSTM Model Summary.....  | 37 |
| Figure 4. 13 Early Stopping to prevent overfitting.....                                     | 38 |
| Figure 4. 14 RNN/LSTM Model Training with KFold Cross-Validation .....                      | 39 |
| Figure 4. 15 Testing the previously trained model .....                                     | 44 |
| Figure 4. 16 Confusion Matrix.....  | 45 |
| Figure 4. 17 Learning Curve of Accuracy .....   | 46 |
| Figure 4. 18 Learning Curve of Loss .....   | 46 |
| Figure 4. 19 Two-Label Classification.....  | 48 |

|  |    |
|--|----|
| Figure 4. 20 Two-Label Confusion Matrix .....                            | 49 |
| Figure 4. 21 Learning Curve of Accuracy on Two-Label Classification..... | 50 |
| Figure 4. 22 Learning Curve of Loss on Two-Label Classification .....    | 50 |



## GLOSSARY

|                  |   |
|------------------|---|
| Machine Learning | a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. |
| Deep Learning    | a subset of machine learning.   |
| Neural Network   | an artificial neural network, composed of artificial neurons or nodes, used for solving artificial AI problems.   |
| Metaverse        | network of 3D virtual worlds focused on social connection.  |





## ABSTRACT

The Metaverse is a universe composed of computer-generated spaces. It is a next-generation platform for social interaction, digital entertainment, and commerce, much like the internet, and is powered by the blockchain. It is a decentralized platform where anyone can create and experience unique digital spaces which allow users to create and experience immersive 3D worlds and applications. It's the next evolution in online social interaction that is still currently in development with some uncertainties. Therefore, this research becomes interesting and important to certain parties who want to know the pros and cons of what Metaverse does, how it affects people, and how people feel about it. In this study, the author proposes to conduct a sentiment analysis using a Recurrent Neural Network (RNN) to analyze the responses of certain individual opinions from YouTube commentary on the Metaverse to predict and classify them as someone that reacts positively, negatively, or neutrally towards the Metaverse and testing it with F1-Score. The result of this study shows that Recurrent Neural Network Long Short-Term Memory Sentiment Analysis on Metaverse-related opinion produced an F1-Score of 66% with a precision of 66% and recall at 67% and an F1-Score of 80% with a precision of 80% and recall at 79% for RNN/LSTM Two-Label Classification.

**Keywords:** The Metaverse, Neural Network, Recurrent Neural Network, Long Short-Term Memory, YouTube, Sentiment Analysis