THE REPUBLIC OF INDONESIA DEFENSE UNIVERSITY

SENTIMENT ANALYSIS MODEL USING BI-LSTM AND LATENT DIRICHLET ALLOCATION ON YOUTUBE COMMENTS TO SUPPORT INTELLIGENCE DATA

AGUNG NUGROHO
120220405003

This Thesis was Written for the Fulfillment of the Requirement
to Earn Master's Degree in Defense Science

SCIENCE AND DEFENSE TECHNOLOGY FACULTY
CYBER DEFENSE ENGINEERING STUDY PROGRAM

BOGOR
2024

THE REPUBLIC OF INDONESIA DEFENSE UNIVERSITY

SENTIMENT ANALYSIS MODEL USING BI-LSTM AND LATENT DIRICHLET ALLOCATION ON YOUTUBE COMMENTS TO SUPPORT INTELLIGENCE DATA

AGUNG NUGROHO
120220405003

This Thesis was Written for the Fulfillment of the Requirement
to Earn Master's Degree in Defense Science

SCIENCE AND DEFENSE TECHNOLOGY FACULTY
CYBER DEFENSE ENGINEERING STUDY PROGRAM

BOGOR
2024

THESIS APPROVAL SHEET

Student Name	:	Agung Nugroho
Student Identity Number	:	120220405003
Study Program	:	Cyber Defense Engineering
Faculty	:	Science dan Defense Technology
Thesis Title	:	Sentiment Analysis Model using Bi-LSTM and Latent Dirichlet Allocation on Youtube Comments to Support Intelligence Data

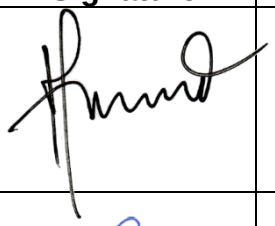
Advisor I,	Advisor II,
Dr. Ir. H. Achmad Farid Wadjdi, M.M.	Prof. Ir. Teddy Mantoro, MSc., Ph.D., SMIEEE.
Date :	Date :

Acknowledged by,
Dean of The Faculty of Defense Science and Technology,

Prof. Dr. Ir. Muhamad Asvial, M.Eng.
First Class Administrator
Date :

THESIS VERIFICATION SHEET

Student Name	:	Agung Nugroho
Student Identity Number	:	120220405003
Study Program	:	Cyber Defense Engineering
Faculty	:	Science dan Defense Technology
Thesis Title	:	Sentiment Analysis Model using Bi-LSTM and Latent Dirichlet Allocation on Youtube Comments to Support Intelligence Data

No.	Name	Signature	Date
1.	Advisor I: Dr. Ir. H. Achmad Farid Wadjdi, M.M.		26/01/2024
2.	Advisor II: Prof. Ir. Teddy Mantoro, MSc., Ph.D., SMIEEE.		26/01/2024
3.	Reviewer I: Syachrul Arief, S.Si., Ph.D.		26/01/2024
4.	Reviewer II: Dr. Yosef Prihanto, S.Si., M.Si.		26/01/2024
5.	Reviewer III: Ruby Alamsyah, M.Tr.Opsla., M.Han., MCE., CIPA., CIT., CIIQA. Kolonel Laut (P) NRP. 10342/P		26/01/2024

DECLARATION OF ORIGINALITY

I hereby declare that in this thesis there is no work or part of work that has been given to obtain a bachelor's degree at any level at a university; and to the best of my knowledge, no term, phrase, sentence, paragraph, subchapter or chapter of any work has ever been written or published; except as written in this text and mentioned in the Reference List.

If in the future it is proven that there is plagiarism in this thesis, I am willing to accept sanctions in accordance with the provisions of the applicable regulations/laws.

Bogor, January , 2024

Agung Nugroho

FOREWORDS

I would like to thank the presence of Allah SWT, the completion of the preparation of the thesis entitled Sentiment Analysis Model using Bi-LSTM and Latent Dirichlet Allocation on YouTube Comments to Support Intelligence Data can be complete.

The preparation of this thesis is intended as one of the requirements for obtaining a Master's degree in the Cyber Defense Engineering Study Program, Faculty of Science and Defense Technology, The Republic of Indonesia Defense University.

The preparation of this thesis was completed thanks to the help and support from various parties, both directly and indirectly. For this reason, on this occasion the researcher would like to thank:

1. Dr. Ir. H. Achmad Farid Wadjdi, M.M. as advisor I and and Prof. Ir. Teddy Mantoro, MSc., Ph.D., SMIEEE II as advisor II for their support and guidance so far and for providing direction to researchers so that this thesis can be completed.
2. The reviewer has provided criticism and suggestions in improving this report.
3. Kolonel Laut (E) Dr. H.A. Danang Rimbawa, S.Si., M.T., M.Tr.Opsla., CEH, CSBA as Head of the Cyber Defense Engineering Study Program and all staff, lecturers and students in the Cyber Defense Engineering Study Program, as well as the entire Defense University community who have helped smooth lectures.
4. Kolonel Laut (P) Ruby Alamsyah, M.Tr.Opsla., M.Han., CIPA., CIT., CIIQA and Kolonel Laut (E) Suginta Ginting, S.Kom., MMSI., M.Tr.Hanla who has fully supported the thesis writing and lecture activities at Defense University.
5. The parties who have helped the researcher a lot during the process of collecting data and writing this thesis. Thank you for taking the time to

discuss, and for sharing your knowledge with researchers so that this scientific work can be completed.

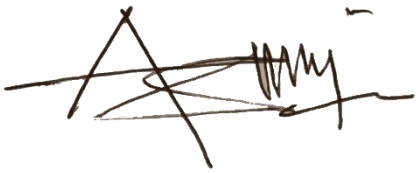
6. All my beloved family, especially father, mother, wife and children, who always pray for the smooth progress of the thesis.
7. National Cyber and Crypto Agency for its support in assignments during the lecture process and thesis work.
8. All fellow researchers whose names I cannot mention one by one due to limited space.

May Allah SWT repay the kindness of various parties for their assistance.

The researcher realizes that this thesis is still imperfect, therefore the researcher humbly hopes for constructive criticism and suggestions to support the perfection of this research.

Finally, we hope that this thesis can provide benefits to the development of defense science and be useful for stakeholders in efforts to improve national security and defense in the cyber sector.

Bogor, January , 2024



Agung Nugroho

ABSTRACT

SENTIMENT ANALYSIS MODEL USING BI-LSTM AND LATENT DIRICHLET ALLOCATION ON YOUTUBE COMMENTS TO SUPPORT INTELLIGENCE DATA

AGUNG NUGROHO

YouTube has the potential to become a propaganda medium that can disrupt national security stability. So comments on YouTube videos can be used as a medium for assessing public opinion which is formed from acts of propaganda. The presence of Artificial Intelligence technology makes it easier to assess public opinion but is hampered by accuracy and datasets. This research proposes modeling using Bi-Directional Long Short-Term Memory (Bi-LSTM) and Latent Dirichlet Allocation (LDA) using quantitative methods to assess sentiment categories and topics from comments on YouTube videos. This research aims to design a model that is novel in the form of combining a sentiment analysis model with the addition of topic category output to support intelligence data collection on social media. The research results show that the Bi-LSTM model with Word2Vec has the highest performance compared to other models, such as Logistic Regression (LR), Multinomial Naïve Bayes (MNB), K-Nearest Neighbors (KNN), and Random Forest (RF), with an average value -average accuracy, precision, recall and F1-Score reaches 98%. The results of this research contribute to the assessment of public opinion and categorization of topics currently being widely discussed by the public on YouTube as an open source of intelligence information to support defense strategies.

Keywords: Bi-LSTM, Deep Learning, LDA, Pancagatra, Youtube

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF CHARTS

LIST OF ABBREVIATIONS

AI	: Artificial Intelligence
Bi-LSTM	: Bidirectional Long Short-Term Memory
CBOW	: Continous Bag of Word
DT	: Decision Trees
Ipoleksosbudhankam	: Ideology, Politics, Economics, Socio-Culture, Defense and Security
Kemhan RI	: Ministry of Defense of the Republic of Indonesia
KNN	: k-Nearest Neighbors
LDA	: Latent Dirichlet Allocation
LR	: Logistic Regression
MNB	: Multinomial Naïve Bayes
NLP	: Natural Language Processing
OSINT	: Open Source Intelligence
RF	: Random Forest