



PROJECT REPORT
SENTIMENT ANALYSIS ON HOMESCHOOLING
TOPIC USING LONG SHORT TERM MEMORY AND
SUPPORT VECTOR MACHINE

JULIUS ALVIN WIDJAYA
18.K1.0020

Faculty of Computer Science
Soegijapranata Catholic University
2022

HALAMAN PENGESAHAN



Judul Tugas Akhir : Sentiment Analysis On Homeschooling Topic Using Long Short Term Memory And Support Vector Machine

Diajukan oleh : Julius Alvin Widjaya

NIM : 18.K1.0020

Tanggal disetujui : 15 Juli 2022

Telah setuju oleh

Pembimbing : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D

Penguji 1 : Yonathan Purbo Santosa S.Kom., M.Sc

Penguji 2 : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D

Penguji 3 : Hironimus Leong S.Kom., M.Kom.

Penguji 4 : Rosita Herawati S.T., M.I.T.

Penguji 5 : Y.b. Dwi Setianto S.T., M.Cs.

Penguji 6 : Yulianto Tejo Putranto S.T., M.T.

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=18.K1.0020

HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan dibawah ini :

Nama : Julius Alvin Widjaya

Program Studi : Teknik Informatika

Fakultas : Ilmu Komputer

Jenis Karya : Skripsi

Menyetujui untuk memberikan kepada Universitas Katolik Soegijapranata Semarang Hak Bebas Royalti Noneklusif atas karya ilmiah yang berjudul “Sentiment Analysis On Homeschooling Topic Using Long Short Term Memory And Support Vector Machine” beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneklusif ini Universitas Katolik Soegijapranata berhak menyimpan, mengalihkan media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir ini selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 15 Juli 2022

Yang menyatakan



Julius Alvin Widjaya

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : JULIUS ALVIN WIDJAYA

ID : 18.K1.0020

declare that this work, titled “Sentiment Analysis On Homeschooling Topic Using Long Short Term Memory And Support Vector Machine”, 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, 15 Juli 2022



Julius Alvin Widjaya

18.K1.0020

ACKNOWLEDGMENT

First, I would like to thank God, because of His blessings and grace, I was able to complete this project well and smoothly. On this occasion, I would like to thank everyone who has provided input, support, and motivation for the entire process of making this project.

I would like to thank my supervisor, R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D. who has guided me while I made this project to completion.

I also want to thank my parents, family, friends, and friends for giving me advice and encouragement to complete my project and study at Soegijapranata Catholic University.

Finally, I hope that everyone who is involved in this project that I have created, both those who are written above and those who are not, will always be blessed by God wherever they are.

Semarang, 15 Juli 2022



Julius Alvin Widjaya

ABSTRACT

Social media is a medium that can be used to share information. In addition, social media is also a suitable place to see the conditions that exist in society. Besides being useful as a medium for sharing information, social media is also often used to convey public opinion or sentiment on a topic being discussed. To find out sentiment or opinion on a topic, this research uses the Sentiment Analysis method.

Sentiment Analysis can be classified well using several models such as Deep Learning and Supervised Learning. The Deep Learning model that will be used in this research is the Long Short Term Memory method, and the Supervised Learning model that will be used in this case is the Support Vector Machine method. The data used is topic data about English Homeschooling taken from Twitter. In this study, the Long Short Term Memory method will be compared with the Support Vector Machine method based on the results of accuracy, precision, recall and F1-Score obtained.

Based on the test results using the Long Short-Term Memory and Support Vector Machine methods, it is found that the Long Short Term Memory method is superior with an Accuracy value of 83.8%, Precision 88.2%, Recall 92.3%, and F1-Score 90.2% compared to the results of the Support Vector Machine method with an accuracy value of 71%, Precision 83.9%, Recall 79.6%, and F1-Score 81.7%.

Keyword: Sentiment Analysis, Long Short Term Memory, Support Vector Machine, Homeschooling, Twitter

TABLE OF CONTENTS

COVER	i
APPROVAL AND RATIFICATION PAGE (Heading plain)	Error! Bookmark not defined.
DECLARATION OF AUTHORSHIP	iii
ACKNOWLEDGMENT	v
ABSTRACT (Abstract Title)	vi
TABLE OF CONTENTS	vii
LIST OF FIGURE	x
LIST OF TABLE	xi
CHAPTER 1 INTRODUCTION.....	1
1.1. Background.....	1
1.2. Problem Formulation.....	2
1.3. Scope.....	2
1.4. Objective.....	3
CHAPTER 2 LITERATURE STUDY.....	4
CHAPTER 3 RESEARCH METHODOLOGY.....	8
3.1 Artificial Neural Network.....	8
3.2 Reccurental Neural Network.....	8
3.3. Activation Function.....	9
3.4. Adam Optimization.....	10
3.5. Binary Cross Entropy.....	10
3.6. Batch Size dan Epoch.....	11
3.7. Natural Language Processing (NLP)	11
3.8. Sentiment Analysis.....	11
3.9. Machine Learning (ML)	11

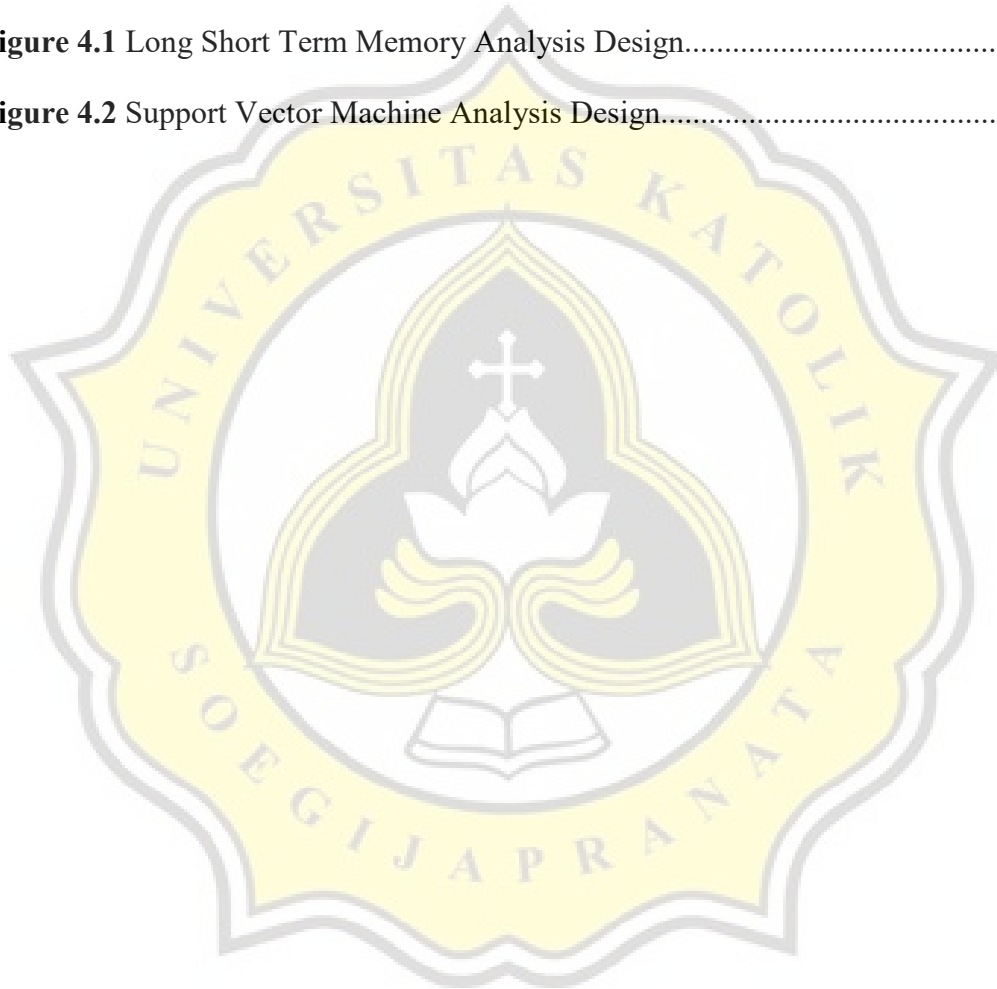
3.9.1	Supervised Learning.....	12
3.9.2	Unsupervised Learning.....	12
3.10.	Count-Vectorizer.....	12
3.11.	Data source.....	13
3.12.	Labeling.....	14
3.13.	Preprocessing.....	15
3.13.1.	Cleansing.....	15
3.13.2.	Case Folding.....	16
3.13.3.	Stopwords Removal.....	16
3.13.4.	Tokenization.....	16
3.14.	Training Data and Testing Data.....	17
3.15.	Word Embedding.....	17
3.15.1.	Word Index.....	18
3.15.2.	Padding / Pad Sequences.....	18
3.16.	Long Short Term Memory (LSTM)	19
3.17.	Fully Connected Layer.....	22
3.18.	Support Vector Machine (SVM)	22
3.19.	Parameter Regularization.....	23
3.20.	Confusion Matrix.....	23
CHAPTER 4 ANALYSIS AND DESIGN.....		25
4.1.	Analysis.....	25
4.2.	Design.....	26
CHAPTER 5 IMPLEMENTATION AND RESULT.....		27
5.1.	Implementation.....	27
5.2.	Result.....	30
CHAPTER 6 CONCLUSION.....		37

6.1. Conclusion.....	37
6.2. Suggestion.....	37
REFERENCES.....	38
APPENDIX.....	a



LIST OF FIGURE

Figure 3.1 Architecture ANN.....	8
Figure 3.2 Architecture RNN.....	9
Figure 3.3 Long Short-Term Memory Network Structure.....	19
Figure 3.4 Best Hyperplane and Maximum Margin.....	22
Figure 4.1 Long Short Term Memory Analysis Design.....	26
Figure 4.2 Support Vector Machine Analysis Design.....	26



LIST OF TABLE

Table 3.1. Public Opinion Web Crawling Results.....	13
Table 3.2. Labeling results on the LSTM method.....	14
Table 3.3. Labeling results on the SVM method.....	15
Table 3.4. Preprocessing Data Cleansing.....	15
Table 3.5. Preprocessing Case Folding.....	16
Table 3.6. Preprocessing Stopwords Removal.....	16
Table 3.7. Preprocessing Tokenization.....	17
Table 3.8. Split Training Data and Testing Data.....	17
Table 3.9. Confusion Matrix.....	23
Table 5.1. LSTM Train Results.....	31
Table 5.2. LSTM Confusion Matrix Evaluation Results.....	32
Table 5.3. LSTM Test Results.....	32
Table 5.4. LSTM All Test Results.....	33
Table 5.5. SVM Confusion Matrix Evaluation Results.....	35
Table 5.6. SVM Test Results.....	35
Table 5.7. SVM All Test Results.....	36
Table 5.8. SVM Undersampling Method Confusion Matrix Evaluation Results.....	37
Table 5.9. SVM Undersampling Method Test Results.....	37
Table 5.10. SVM Undersampling Method All Test Results.....	38
Table 5.11. SVM Kernel All Test Result.....	39
Table 5.12. SVM Polynomial Kernel Confussion Matrix Evaluation Results.....	39
Table 5.13. SVM Polynomial Kernel Test Results.....	40
Table 5.14. SVM RBF Kernel Confussion Matrix Evaluation Results.....	40

Table 5.15. SVM RBF Kernel Test Results.....40

Table 5.16. SVM Sigmoid Kernel Confussion Matrix Evaluation Results.....41

Table 5.17. SVM Sigmoid Kernel Test Results.....41

