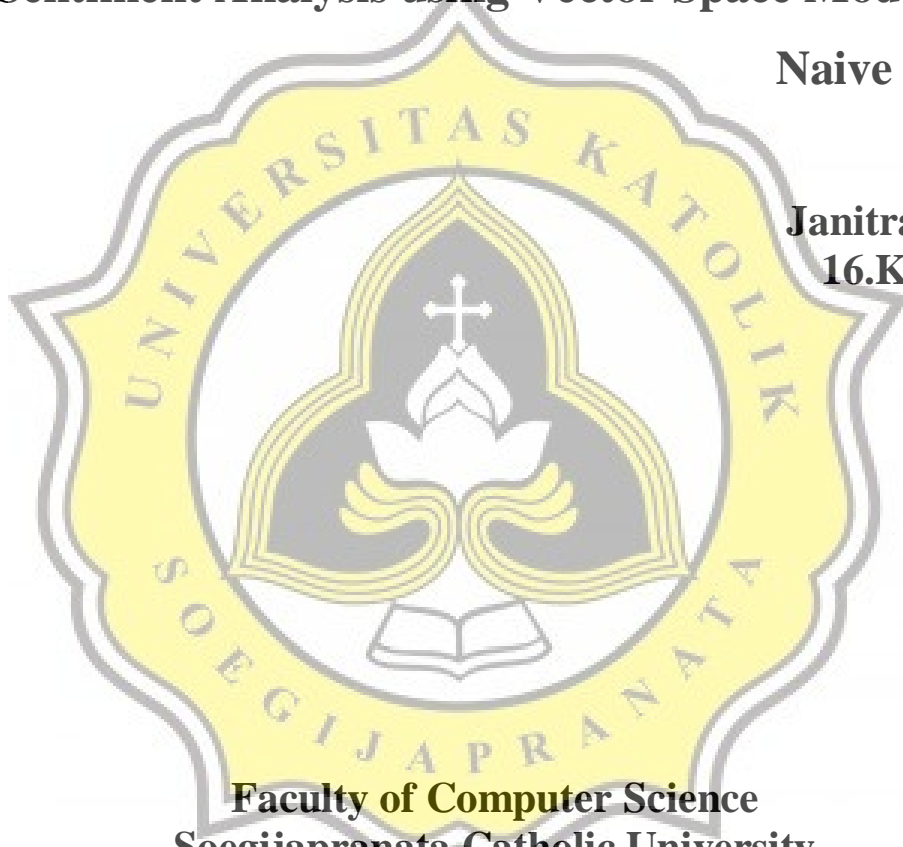




# **PROJECT REPORT**

## **Sentiment Analysis using Vector Space Model and Naive Bayes**

**Janitra Veda  
16.K1.0044**



**Faculty of Computer Science  
Soegijapranata Catholic University  
2020**



## HALAMAN PENGESAHAN

Judul Tugas Akhir: : Sentiment Analysis using Vector Space Model and Naive Bayes  
Diajukan oleh : Janitra Veda  
NIM : 16.K1.0044  
Tanggal disetujui : 19 November 2020  
Telah setuju oleh  
Pembimbing : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D  
Penguji 1 : Hironimus Leong S.Kom., M.Kom.  
Penguji 2 : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D  
Penguji 3 : Rosita Herawati S.T., M.I.T.  
Penguji 4 : Y.b. Dwi Setianto  
Ketua Program Studi : Rosita Herawati S.T., M.I.T.  
Dekan : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D

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

[sintak.unika.ac.id/skripsi/verifikasi/?id=16.K1.0044](http://sintak.unika.ac.id/skripsi/verifikasi/?id=16.K1.0044)

## HALAMAN PERNYATAAN PUBLIKASI ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan dibawah ini :

Nama : Janitra Veda  
Program Studi : Teknik Informatika  
Fakultas : Ilmu Komputer  
Jenis Karya : Skripsi

Menyetujui untuk memberikan kepada Universitas Katolik Soegijapranata Semarang Hak Bebas Royalti Noneksklusif atas karya ilmiah yang berjudul **“Sentiment Analysis using Vector Space Model and Naive Bayes”** beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif 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, 19 November 2020

Yang menyatakan



Janitra Veda

## STATEMENT OF ORIGINALITY

I, the undersigned:

Name : Janitra Veda

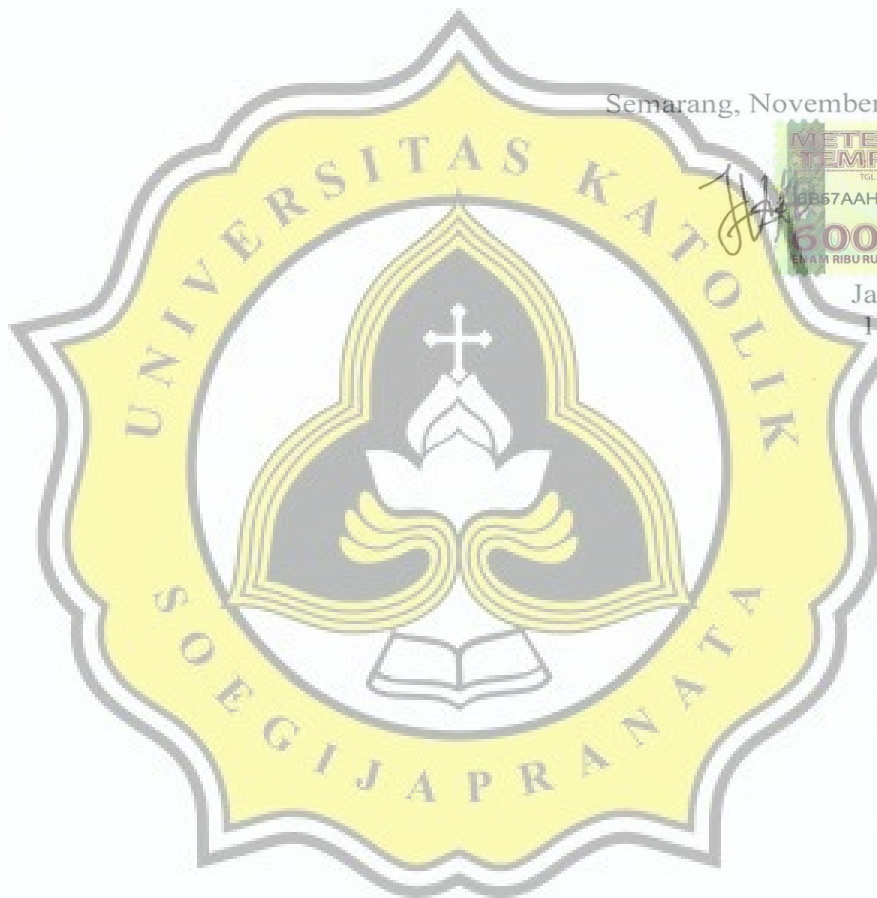
ID : 16.K1.0044

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 plagiarized or copied other works, I am ready to accept a sanction.

Semarang, November, 19, 2020



Janitra Veda  
16.K1.0044



## ACKNOWLEDGEMENTS

Praise the author to God, because with his grace and grace, the writer can complete the thesis entitled "Sentiment Analysis using Vector Space Model and Naive Bayes". This thesis is intended to fulfill one of the exam requirements in order to obtain a Bachelor of Computer (S.Kom) degree in the Informatics Engineering study program at the Computer Science faculty of Soegijapranata Catholic University. During the writing of this thesis, the author received a lot of help and support so that he could complete this thesis. Therefore, the author would like to thank profusely to:

1. R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D as a supervisor at Unika Soegijapranata Semarang majoring in informatics engineering.
2. Informatics engineering lecturers at Unika Soegijapranata Semarang who have provided knowledge, guidance, and advices during my studies.
3. My big family that give me motivation and support to complete the study period at Unika Soegijapranata.
4. All my friends who have helped me in any way, thank you for providing support and friendship that helped me through all this.

The author realizes that this thesis is far from perfect because of the limited knowledge and experience possessed. Therefore, all constructive criticism and suggestions will be accepted by the author with pleasure. The author hopes that this thesis can be useful for all parties who need it.

Semarang, 19 November 2020



Janitra Veda

## ABSTRACT

*With the progress of times like this, social media is an important thing in life. Freedom of opinion is also very large on social media, like on the Twitter, Facebook, Instagram and many others. Therefore, this opinion can be analyzed by obtaining data from the public.*

*This project aims to analyze the sentiments that exist on Twitter with social distancing hashtags and compare which algorithm is better to use. This project compares 2 algorithms, namely the Vector Space Model and the Naive Bayes.*

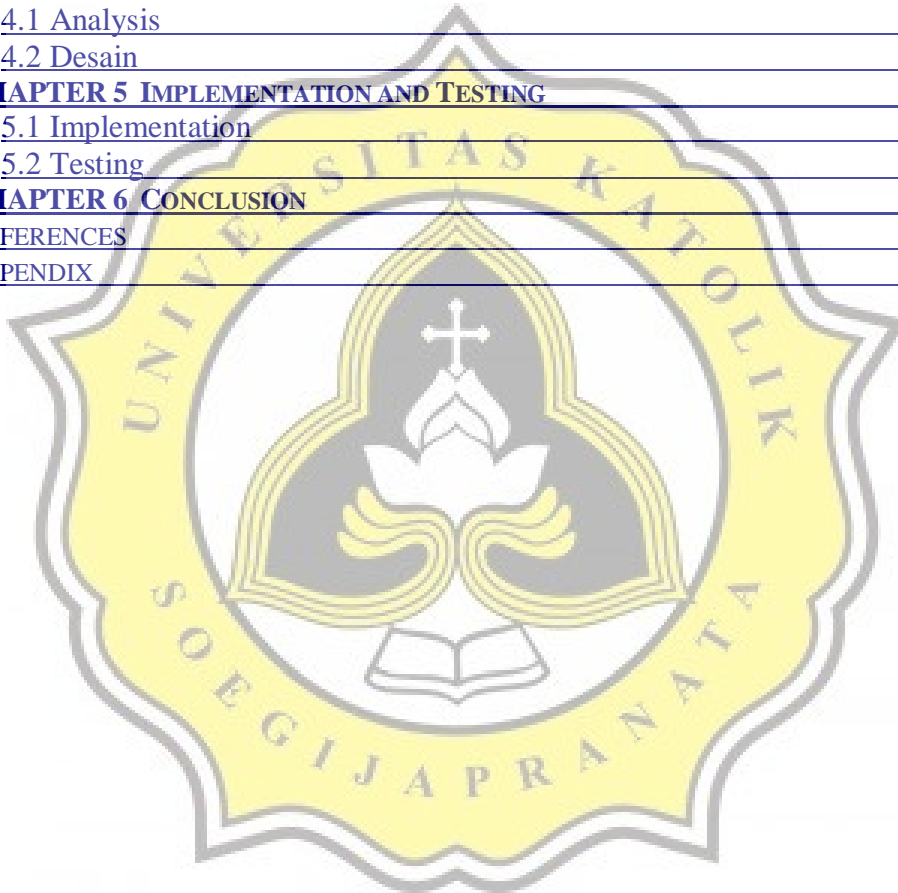
*The final results obtained are in the form of performance results from each algorithm using different test schemes. To see maximum results.*

*Keyword: Vector Space Model, Naive Bayes, Sentiment Analysis*



## TABLE OF CONTENTS

Cover	i
APPROVAL AND RATIFICATION PAGE	ii
STATEMENT OF ORIGINALITY	iii
ABSTRACT	iv
TABLE OF CONTENTS	v
ILLUSTRATION INDEX	vi
INDEX OF TABLES	vii
<b>CHAPTER 1 INTRODUCTION</b>	<b>1</b>
1.1 Background	1
1.2 Problem Formulation	1
1.3 Scope	1
1.4 Objective	2
<b>CHAPTER 2 LITERATURE STUDY</b>	<b>3</b>
<b>CHAPTER 3 RESEARCH METHODOLOGY</b>	<b>5</b>
<b>CHAPTER 4 ANALYSIS AND DESIGN</b>	<b>6</b>
4.1 Analysis	6
4.2 Desain	7
<b>CHAPTER 5 IMPLEMENTATION AND TESTING</b>	<b>8</b>
5.1 Implementation	8
5.2 Testing	8
<b>CHAPTER 6 CONCLUSION</b>	<b>9</b>
REFERENCES	
APPENDIX	A



## ILLUSTRATION INDEX

<u>Illustration 3.3: TF-IDF</u>	20
<u>Illustration 3.4.1: Data Train</u>	21
<u>Illustration 3.4.2: Data Test</u>	22
<u>Illustration 3.5: Vector Space Model</u>	22
<u>Illustration 4.1.1: Akurasi, Presisi, Recall</u>	24
<u>Illustration 4.1.2: F1-Score</u>	24
<u>Illustration 5.1: Flowchart VSM &amp; Naive Bayes</u>	30





## INDEX OF TABLES

Table 4.2.1: VSM Data	25
Table 4.2.2: VSM Analysis Data 150-50	26
Table 4.2.3: VSM Analysis Data 100-98	26
Table 4.2.4: VSM Analysis Data 50-148	26
Table 4.2.5: VSM Calculation Data 150-50	26
Table 4.2.6: VSM Calculation Data 100-98	26
Table 4.2.7: VSM Calculation Data 50-148	27
Table 4.3.1.: Naive Bayes Data	28
Table 4.3.2: Naive Bayes Analysis Data 150-50	28
Table 4.3.3: Naive Bayes Analysis Data 100-98	28
Table 4.3.4: Naive Bayes Analysis Data 50-148	28
Table 4.3.5: Naive Bayes Calculation Data 150-50	29
Table 4.3.6: Naive Bayes Calculation Data 100-98	29
Table 4.3.7: Naive Bayes Calculation Data 50-148	29

