

1. Information Engineering
2. Operations Engineering & Management
3. Engineering Management

**TEXT MINING TO GAIN BUSINESS, MARKET, AND COMPETITIVE  
INSIGHTS IN COURIER SERVICE INDUSTRY**

**A THESIS**

**Submitted in Partial Fulfillment of the Requirement for the Degree of  
Bachelor of Engineering in Industrial Engineering**



**EUNIKE ANNICE ARIELLA  
17 14 09215**

**INTERNATIONAL INDUSTRIAL ENGINEERING PROGRAM  
DEPARTMENT OF INDUSTRIAL ENGINEERING  
FACULTY OF INDUSTRIAL TECHNOLOGY  
UNIVERSITAS ATMA JAYA YOGYAKARTA  
2021**

## HALAMAN PENGESAHAN

Tugas Akhir Berjudul

TEXT MINING TO GAIN BUSINESS, MARKET, AND COMPETITIVE INSIGHTS IN COURIER  
SERVICE INDUSTRY

yang disusun oleh

EUNIKE ANNICE ARIELLA

171409215

dinyatakan telah memenuhi syarat pada tanggal 10 Juni 2021

		Keterangan
Dosen Pembimbing 1	: Ririn Diar Astanti, D.Eng.	Telah menyetujui
Dosen Pembimbing 2	: Ririn Diar Astanti, D.Eng.	Telah menyetujui
Tim Penguji		
Penguji 1	: Ririn Diar Astanti, D.Eng.	Telah menyetujui
Penguji 2	: Yosef Daryanto, S.T., M.Sc., Ph.D.	Telah menyetujui
Penguji 3	: The Jin Ai, D.Eng.	Telah menyetujui

Yogyakarta, 10 Juni 2021

Universitas Atma Jaya Yogyakarta

Fakultas Teknologi Industri

Dekan

ttd

Dr. A. Teguh Siswanto, M.Sc



**IDENTIFICATION PAGE**  
A THESIS ON  
**SENTIMENT ANALYSIS TO GAIN BUSINESS, MARKET, AND COMPETITIVE  
INSIGHTS IN DELIVERY SERVICE INDUSTRY**

submitted by

**Eunike Annice Ariella**

17 14 09215

was examined and approved on June 10, 2021

Faculty Supervisor,

Ririn Diar Astanti, S.T., M.MT., D.Eng., IPM

Broad of Examiners,  
Chair,

Ririn Diar Astanti, S.T., M.MT., D.Eng., IPM

Member,

Member,

Yosef Daryanto, S.T., M.Sc., Ph.D.

The Jin Ai, D.Eng.

Yogyakarta, June 10, 2021  
Universitas Atma Jaya Yogyakarta  
Faculty of Industrial Technology,  
Dean,

Dr. A. Teguh Siswanto, M.Sc

## DECLARATION OF ORIGINALITY OF RESEARCH

I certify that the research entitled "Sentiment Analysis to Gain Business, Market, And Competitive Insights in Courier Service Industry" is this thesis has not already been submitted for any other degree.

I certify that to the best of my knowledge and belief, this thesis which I wrote does not contain the works of parts of the works of other people, except those cited in the quotations and bibliography, as scientific paper should.

In addition, I certify that I understand and abide the rule stated by the Ministry of Education and Culture the Republic of Indonesia, subject to the provisions of Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 17 Tahun 2010 tentang Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi.

Signature

:



Student name

: Eunike Annice Ariella

Student ID

: 171409215

Date

: June 4, 2021

## ACKNOWLEDGEMENTS

Praises and thanks to God, the Almighty, for His grace, so that the final project report entitled "Sentiment Analysis to Gain Business, Market, And Competitive Insights in Courier Service Industry" can be compiled.

The author expresses her sincere gratitude to her supervisor, Ririn Diar Astanti, S.T., M.MT., D.Eng., IPM, for providing invaluable guidance throughout this research. In addition, many thanks are dedicated to the owner of Pos Indonesia Tentara Rakyat Yogyakarta, who supported the author during the validation phase.

The author also gives her thanks to all faculty members of the Department of Industrial Engineering UAJY for helping him in various ways. The same appreciation is extended to her UAJY friends.

Last but not least, her gratitude is addressed to her beloved family and friends for the love and support, who also support the author from the start until the end of this study.

Finally, the authors hope that this research can provide benefits for those who are interested.

Yogyakarta, June 4, 2021

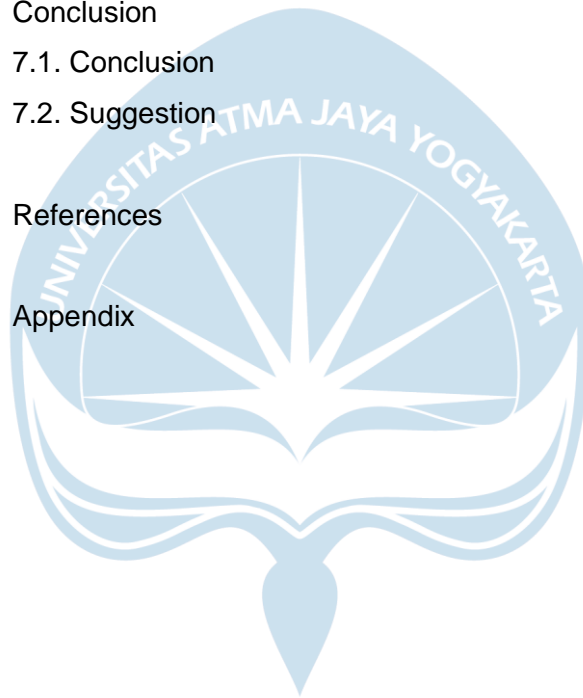
Eunike Annice Ariella

## TABLE OF CONTENT

CHAPTER	TITLE	PAGE
	Cover Page	i
	Identification Page	ii
	Declaration of Originality of Research	iv
	Acknowledgements	v
	Table of Contents	vi
	List of Tables	viii
	List of Figures	ix
	Appendix List	xi
	Abstract	xii
1	Introduction	1
	1.1. Background	1
	1.2. Problem Formulation	4
	1.3. Research Objective	4
	1.4. Scope and Limitation	5
2	Literature Review and Theoretical Background	6
	2.1. Literature Review	6
	2.2. Theoretical Background	8
3	Research Methodology	19
	3.1. Initiation Phase	22
	3.2. Diagnosis Phase	22
	3.3. Selection of Problem-Solving Method Phase	25
	3.4. Customer Feedback Collection	25
	3.5. Data Processing	25
	3.6. Result Transformation and Discussion Phase	27
4	Customer Feedback Collection	29
	4.1. Collection Method	29
	4.2. Customer Feedback Scrapping Result	32

## TABLE OF CONTENT

5	Data Processing	37
	5.1. The Text Mining Model Design	37
	5.2. Text Mining Model	39
6	Result Transformation and Discussion	87
	6.1. Critical to Quality (CTQ) Tree	87
	6.2. Discussion	94
7	Conclusion	97
	7.1. Conclusion	97
	7.2. Suggestion	98
	References	99
	Appendix	105



## LIST OF TABLES

TABLE	TITLE	PAGE
Table 2.1.	Comparison of Data Mining Method	14
Table 5.1.	Summary of Unigram Sentiment Analysis	52
Table 5.2.	Summary of Bigram Sentiment Analysis	54
Table 5.3.	Statistical Summary of Tweets Polarity Analysis	57
Table 5.4.	Result of Negative Bigram Deeper Analysis	59
Table 5.5.	Frequency of Each Negative Word and Negative Word Variety	61
Table 5.6.	Result of Positive Bigram Deeper Analysis	68
Table 5.7.	Frequency of Each Positive Word and Positive Word Variety	70





## LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 2.1.	The 5V model that currently defines Big Data	9
Figure 2.2.	Text Mining/ KDT Process	12
Figure 2.3.	Sentiment Analysis Classification Technique	13
Figure 2.4.	Lean Six Sigma Framework	18
Figure 3.1.	Flowchart of 6-Phases Research Procedure	19
Figure 3.2.	Annual Top Brand Index for Courier Service	23
Figure 4.1.	R Syntax for Scrapping Data	31
Figure 4.2.	Result from Twitter Data Scrapping	33
Figure 4.3.	R Syntax to Filter The Data Frame	34
Figure 4.4.	Filtered Twitter Data Frame from Scrapping	35
Figure 4.5.	Twitter Data Frame Summary	36
Figure 5.1.	R Syntax for Tweeter Textual Data Pre-processing	40
Figure 5.2.	Example of Non-Indonesian Language in the Filtered Data	41
Figure 5.3.	Example of External Party of Pos Indonesia in the Filtered Data	42
Figure 5.4.	Example of Duplicate Tweets in the Filtered Data	42
Figure 5.5.	Result of Cleaned and Pre-processed Data	43
Figure 5.6.	Number of Object Before and After Pre-processed	43
Figure 5.7.	R Syntax for Unigram and Bag-of-Words Text Transformation	44
Figure 5.8.	Result of Unigram – Left & Bag-of-Words – Right	45
Figure 5.9.	R Syntax for Bigram and Bag-of-Bigrams Text Transformation	45
Figure 5.10.	Result of Bigram – Left & Bag-of-2gram – Right	46
Figure 5.11.	R Syntax for Unigram and Bigram Sentiment Type Grouping	47
Figure 5.12.	R Syntax for Calculating Polarity Score	48
Figure 5.13.	R Syntax for Sentiment Analysis Pattern Identification	50
Figure 5.14.	Bar Chart of Unigram Sentiment Analysis	51
Figure 5.15.	Word Cloud of Unigram Sentiment Analysis	52

## LIST OF FIGURES

Figure 5.16.	Bar Chart of Bigram Sentiment Analysis	54
Figure 5.17.	R Syntax for Sentiment Analysis Pattern Identification	55
Figure 5.18.	Histogram of Tweets Polarity Calculation	56
Figure 5.19.	Pie Chart of Tweets Polarity Calculation	57
Figure 5.20.	Example of Deeper Analysis for Negative Word	58
Figure 5.21.	Example of Deeper Analysis for Positive Word	58
Figure 6.1.	Critical to Quality (CTQ) Tree of Agent Aspect	89
Figure 6.2.	Critical to Quality (CTQ) Tree of Delivery Aspect	90
Figure 6.3.	Critical to Quality (CTQ) Tree of Product Aspect	91
Figure 6.4.	Critical to Quality (CTQ) Tree of Product Aspect	92
Figure 6.5.	Critical to Quality (CTQ) Tree of Tracking System Aspect	93



## LIST OF APPENDIXES

APPENDIX	TITLE	PAGE
Appendix 1:	Categorization of Negative Sentiment Classification	105
Appendix 2:	Categorization of Positive Sentiment Classification	106
Appendix 3:	Confusion Matrix of Sentiment Prediction Model	107
Appendix 4:	Performance Measurement of Sentiment Prediction Model	108



## ABSTRACT

The increasing use of e-commerce has led to the emergence of many couriers' service brands. The increased number of courier services has caused fierce market competition that affects experienced courier brands, POS Indonesia. To maintain the position and win the competition as one of Indonesia's top courier service brands, it is important always to evaluate the company's performance. If the company cannot meet the customers' needs, wants, and expectations, the percentage of market coverage will decrease, which will result in the company being unable to compete.

Analyzing feedback/reviews from customers through social media can be utilized to compare POS Indonesia's performance with other brands, which can later become valuable insights for improvement. The conventional method such as interview, survey, a focus group was preferred as a method to collect data in the past. However, with technological and data developments, text mining methods can be used to collect data effectively and efficiently.

This study aims to use text mining sentiment analysis to process secondary unstructured data from Twitter into valuable insights that contributed to the process to satisfy the customers. The secondary unstructured data was processed through Knowledge Discovery in Textual Database (KDT), resulting in complex and uncategorized Voice of Customers (VOCs). The result then transformed into Critical to Quality (CTQ) Trees to create comprehensive insight that the general reader can easily understand.

The result of text mining stated that there are five aspects that still unsatisfactory to the customer: delivery procedure, customer service, agent, tracking system, and product. The results are already reflecting the real customer feedback that enters the overall system of POS Indonesia.

**Keywords: Courier Service Industry, Customer Feedback, Improvement, Sentiment Analysis, Social Media Data.**