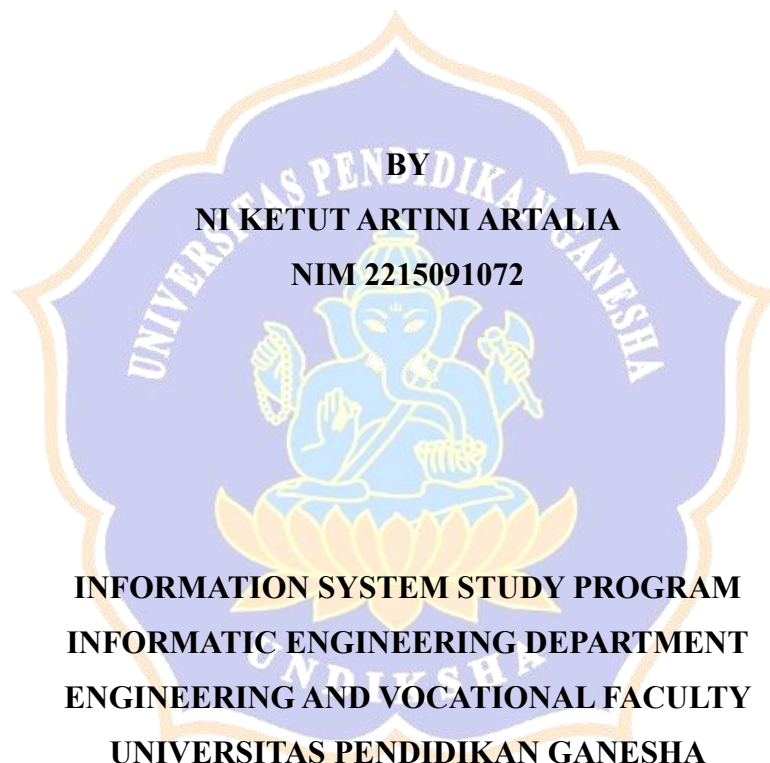


**SENTIMENT ANALYSIS OF BALI'S RAPID  
TOURISM GROWTH: COMPARING INDOBERT  
PERFORMANCE WITH INSET LEXICON AND  
MANUAL LABEL INTEGRATION**



**SINGARAJA**

**2026**

**SENTIMENT ANALYSIS OF BALI'S RAPID TOURISM  
GROWTH: COMPARING INDOBERT  
PERFORMANCE WITH INSET LEXICON AND  
MANUAL LABEL INTEGRATION**

**SKRIPSI**

**Submitted to**

**Universitas Pendidikan Ganesha**

**To fulfil one of the requirements in completing the  
Bachelor of Information Systems Study Program**



**By**

**Ni Ketut Artini Artalia**

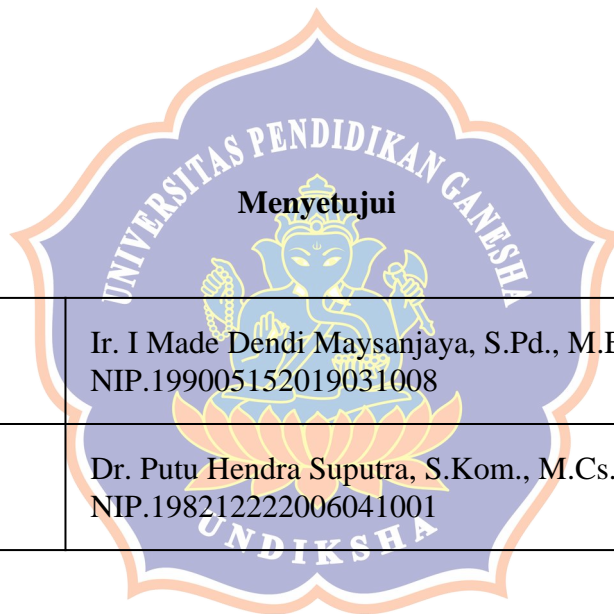
**NIM 2215091072**

**INFORMATION SYSTEM STUDY PROGRAM  
INFORMATIC ENGINEERING DEPARTMENT  
ENGINEERING AND VOCATIONAL FACULTY  
UNIVERSITAS PENDIDIKAN GANESHA**

**2026**

# SKRIPSI

## DIAJUKAN UNTUK MELENGKAPI TUGAS DAN MEMENUHI SYARAT-SYARAT UNTUK MENCAPAI GELAR SARJANA KOMPUTER



Pembimbing I	Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng. NIP.199005152019031008
Pembimbing II	Dr. Putu Hendra Suputra, S.Kom., M.Cs. NIP.198212222006041001

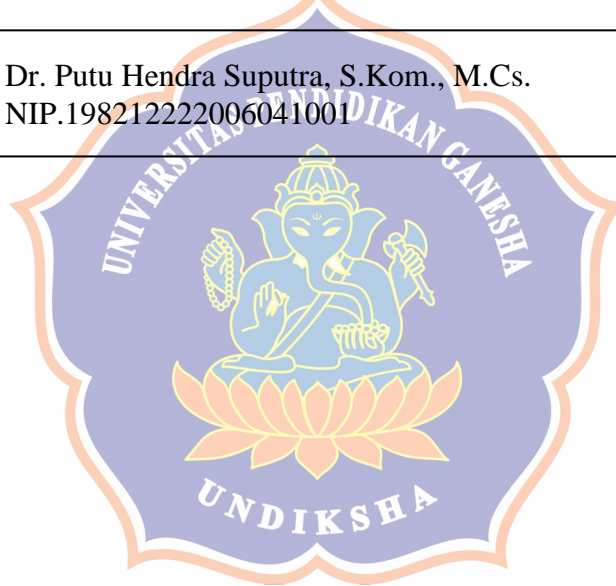


- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

Skripsi oleh NI KETUT ARTINI ARTALIA ini  
telah dipertahankan di depan dewan penguji  
Pada tanggal 17 April 2026

### Dewan Penguji

Ketua	Ir. Gede Surya Mahendra, S.Pd., M.Kom. NIP.199003132022031009
Anggota	Ir. Kadek Teguh Dermawan, S.Pd., M.Kom. NIP.199406132025061003
Anggota	Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng. NIP.199005152019031008
Anggota	Dr. Putu Hendra Suputra, S.Kom., M.Cs. NIP.198212222006041001



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSRé - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSRé
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan  
Universitas Pendidikan Ganesha  
guna memenuhi syarat-syarat untuk mencapai gelar Sarjana Komputer

### Menyetujui

Ketua Ujian	Ir. Made Windu Antara Kesiman, S.T., M.Sc., Ph.D. NIP.198211112008121001
Sekretaris Ujian	Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng. NIP.199005152019031008



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

## STATEMENT

I hereby declare that the thesis entitled “*Sentiment Analysis of Bali’s Rapid Tourism Growth: Comparing IndoBERT Performance with InSet Lexicon and Manual Label Integration*”, along with all its contents, is entirely my own original work. I confirm that I have not engaged in any form of plagiarism or improper citation that violates established academic ethical standards. I fully understand and accept the consequences, including any sanctions that may be imposed upon me, should any violation of academic integrity be discovered in this work in the future, or if any claims arise regarding the originality of this thesis.

Singaraja, 24 April 2026

Declared by,



Ni Ketut Artini Artalia

NIM. 2215091072

## **DEDICATION**

The author would like to express the highest praise and deepest gratitude to  
**Ida Sang Hyang Widhi Wasa**  
for His blessings and grace, which have enabled the author to complete this  
thesis successfully through the ease and guidance bestowed.

This thesis is dedicated to:

### **BELOVED PARENTS**

(I Made Karna & Ni Ketut Muspa)

For their endless prayers, unconditional love, unwavering support, and  
guidance, and for always striving to provide everything needed throughout the  
author's educational journey, enabling the author to successfully achieve a  
bachelor's degree.

### **BELOVED SIBLINGS**

(Beda Arta, Adi Arta, Tri Arta, Dewi Kartika, Lisdyanti)

For their constant support, guidance, and unwavering belief in the author  
throughout the educational journey.

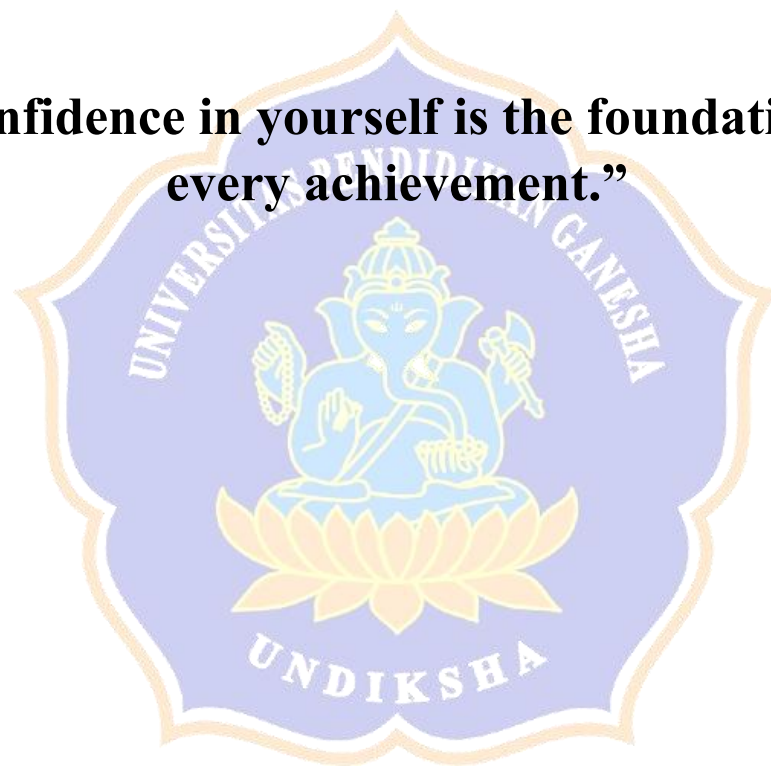
### **BELOVED NIECES AND NEPHEWS**

(Vanya, Hanah, Zanya, Vibi)

For the joy, laughter, and happiness they bring, which have been a source of  
strength and encouragement throughout this journey.

**MOTTO**

**“Confidence in yourself is the foundation of every achievement.”**



## ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to Tuhan Yang Maha Esa for His blessings and grace, which have enabled the author to complete this thesis entitled **“Sentiment Analysis of Bali’s Rapid Tourism Growth: Comparing IndoBERT Performance with InSet Lexicon and Manual Label Integration.”** This thesis is submitted in partial fulfillment of the requirements for obtaining a Bachelor’s degree in Computer Science at Universitas Pendidikan Ganesha.

The completion of this thesis would not have been possible without the invaluable assistance, support, and encouragement both moral and material from various parties. Therefore, on this occasion, the author would like to express sincere gratitude to the following:

1. Mr Prof. Dr. Kadek Rihendra Dantes, S.T., M.T., as the Dean of the Faculty of Engineering and Vocational the for the motivation and facilities provided that have enabled the author to complete the study.
2. Mr Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng., as the Head of the Information Systems Study Program as well as the First Supervisor, for providing guidance, direction, advice, and motivation to the author throughout the preparation of this thesis.
3. Mr Dr. Putu Hendra Suputra, S.Kom., M.Cs., as the Head of the Department of Informatics Engineering as well as the Second Supervisor, for the invaluable guidance, direction, advice, and motivation provided to the author throughout the preparation of this thesis.
4. Mr Ir. Gede Surya Mahendra, S.Pd., M.Kom., as the First Examiner, for the valuable suggestions, guidance, direction, and motivation provided to the author throughout the preparation of this thesis.
5. Mr Ir. Kadek Teguh Dermawan, S.Pd., M.Kom., as the Second Examiner, for the invaluable suggestions, guidance, direction, and motivation provided to the author throughout the preparation of this thesis, as well as for the extensive knowledge shared with the author during the course of study.

6. All lecturers of the Information Systems Study Program, for their dedication in educating and imparting knowledge to the author throughout the academic journey.
7. Mrs Komang Sri Ayu Sudarmini, A.Md.Par.,S.Pd., Mr I Dewa Gede Pujana, S.Pd., and Mr I Made Adi Arta who have kindly assisted the author in completing the research data required for the completion of this thesis.
8. Beloved parents, siblings, and nieces and nephews, for their endless support, prayers, motivation, and unconditional love given to the author.
9. Fellow companions in this journey (Utta, Dian, Nabila, April, and Utik), for their constant encouragement, motivation, happiness, and support throughout the academic journey.
10. Fellow members of Sifors IKI 22, for their encouragement and motivation to strive together in completing our studies.
11. In particular, the author would like to express gratitude to herself for all the perseverance, dedication, and resilience shown in completing this thesis. Thank you for not giving up and for continuing to believe in your own abilities throughout this challenging journey.

The author fully realizes that this thesis is still far from perfect due to the limitations of the author's abilities. Therefore, for the improvement of this thesis, the author sincerely welcomes any constructive criticism and suggestions from various parties. The author hopes that this thesis will be beneficial and useful for all, particularly for the development of education in Bali

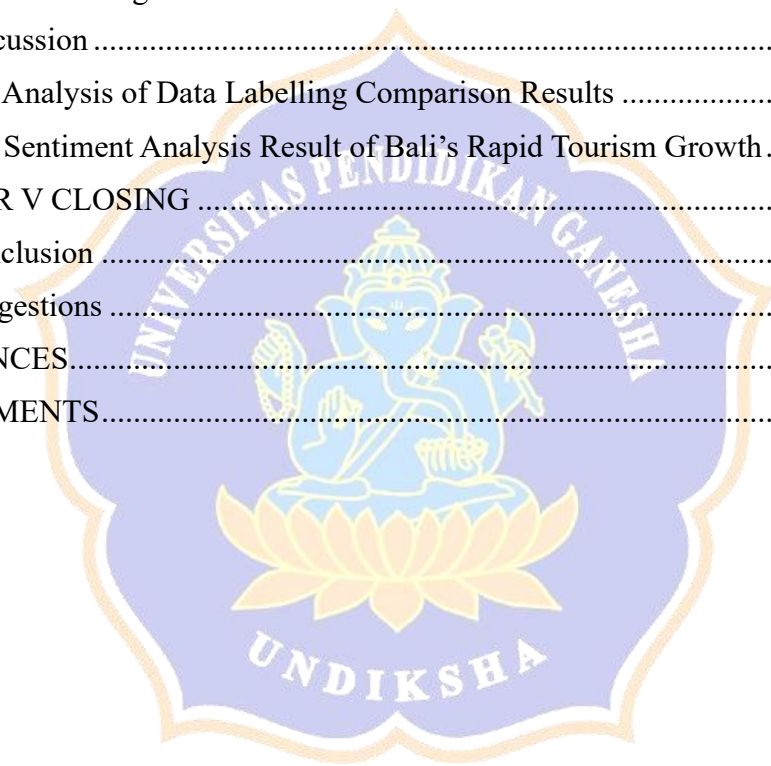
Singaraja, 24 April 2026

Author

## LIST OF CONTENT

STATEMENT.....	iii
DEDICATION .....	iv
MOTTO.....	v
ABSTRACT.....	vi
ABSTRAK .....	vii
ACKNOWLEDGEMENTS .....	viii
LIST OF CONTENT.....	x
LIST OF TABLES.....	xii
LIST OF FIGURES.....	xiii
LIST OF ATTACHMENTS .....	xiv
CHAPTER I INTRODUCTION.....	1
1.1 Background .....	1
1.2 Problem Identification.....	4
1.3 Problem Statement .....	5
1.4 Research Objective.....	5
1.5 Scope of Research .....	6
1.6 Research Significance .....	6
CHAPTER II LITERATURE REVIEW .....	8
2.1 State of The Art .....	8
2.2 Theoretical Basis.....	18
2.2.1 Rapid Tourism Growth.....	18
2.2.2 TikTok .....	18
2.2.3 Data Mining .....	20
2.2.4 Deep Learning.....	22
2.2.5 Sentiment Analysis.....	23
2.2.6 Scraping Data .....	25
2.2.7 Preprocessing Data.....	25
2.2.8 InSet Lexicon .....	26
2.2.9 IndoBERT .....	27
2.2.10 K-Fold Cross Validation.....	29
2.2.11 Confusion Matrix .....	29
CHAPTER III RESEARCH METHODOLOGY .....	45
3.1 Research Type .....	45

3.2	Research scheme .....	45
CHAPTER IV RESULT AND DISCUSSION.....		57
4.1	Result .....	57
4.1.1	Data Collecting.....	57
4.1.2	Manual Labelling Data.....	59
4.1.3	Data Preprocessing.....	61
4.1.4	Automatic Labelling Data (Inset Lexicon).....	63
4.1.5	Data Augmentation.....	65
4.1.6	Modelling dan Evaluation .....	67
4.1.8	Visualizing.....	83
4.2	Discussion .....	86
4.2.1	Analysis of Data Labelling Comparison Results .....	86
4.2.2	Sentiment Analysis Result of Bali's Rapid Tourism Growth.....	88
CHAPTER V CLOSING .....		90
5.1	Conclusion .....	90
5.2	Suggestions .....	91
REFERENCES.....		93
ATTACHMENTS.....		98



## LIST OF TABLES

Table 2. 1 State of the Art.....	8
Table 2. 2 Labelling Using InSet Lexicon .....	27
Table 2. 3 Confusion Matrix General Structure .....	30
Table 2. 4 Confusion Matrix Multiclass Structure .....	31
Table 3. 1 Keyword and Hashtag for Scraping Data.....	47
Table 3. 2 Example of TikTok Comments Dataset.....	47
Table 3. 3 Example of Data Cleansing Process.....	49
Table 3. 4 Example Translation Pipeline Process .....	50
Table 3. 5 Example of Stemming Process.....	51
Table 3. 6 Example of Preprocessing Result.....	51
Table 3. 7 Example of InSet Lexicon Labelling .....	52
Table 3. 8 Example of Confusion Matrix Multiclass Result.....	54
Table 4. 1 Result of TikTok Comments Data Collection.....	58
Table 4. 2 Example of Comments That do not Meet the Criteria .....	58
Table 4. 3 Manual Data Labelling Result.....	59
Table 4. 4 Cleansing Data Process Result.....	61
Table 4. 5 Translation Pipeline Result.....	61
Table 4. 6 Post Translation Cleansing Result.....	62
Table 4. 7 Normalization Result.....	62
Table 4. 8 Stemming Result .....	63
Table 4. 9 InSet Lexicon Labelling Result.....	64
Table 4. 10 Google Colaboratory Specification.....	67
Table 4. 11 Classification Method Testing Scenario .....	67
Table 4. 12 Cross Validation Result Scenario 1 .....	69
Table 4. 13 Final Classification Report Scenario 1 .....	70
Table 4. 14 Cross Validation Result Scenario 2 .....	71
Table 4. 15 Final Classificaation Report Scenario 2 .....	72
Table 4. 16 Cross Validation Scenario 3 .....	73
Table 4. 17 Final Classification Report Scenario 3.....	74
Table 4. 18 Cross Validation Scenario 4 .....	75
Table 4. 19 Final Classification Report Scenario 4.....	76
Table 4. 20 Evaluation Summary .....	77
Table 4. 21 Cross Validation POC Scenario.....	81
Table 4. 22 Final Classification Report POC Scenario .....	83

## LIST OF FIGURES

Figure 1.1 Number of Tourist Visits to Bali in 2019-2023 .....	1
Figure 2. 1 Illustration of Pre-Training and Fine-tuned IndoBERT .....	28
Figure 3. 1 Research Scheme .....	46
Figure 3. 2 Data Preprocessing Stages .....	48
Figure 3. 3 IndoBERT Modelling .....	53
Figure 4. 1 Rating and Usage Statistic of the Scraper Agent .....	57
Figure 4. 2 Number of Each Sentiment Class Manual Labelling Classes .....	60
Figure 4. 3 Distribution of InSet Lexicon Labelling Result .....	65
Figure 4. 4 Manual Labelling Class Distribution.....	66
Figure 4. 5 InSet Labelling Class Distribution.....	67
Figure 4. 6 Accuracy Graph per Fold IndoBERT + Manual Labelling.....	69
Figure 4. 7 Confusion Matrix IndoBERT + Manual Labelling.....	70
Figure 4. 8 Accuracy Graph per Fold IndoBERT + InSet Labelling .....	71
Figure 4. 9 Confusion Matrix IndoBERT + InSet Labelling .....	72
Figure 4. 10 Accuracy Graph per Fold IndoBERT + Manual Labelling with ROS .....	73
Figure 4. 11 Confusion Matrix IndoBERT + Manual Labelling with ROS .....	74
Figure 4. 12 Accuracy Graph per Fold IndoBERT + InSet Labelling with ROS..	75
Figure 4. 13 Confusion Matrix IndoBERT + InSet Labelling with ROS.....	76
Figure 4. 14 Evaluation Summary Chart.....	78
Figure 4. 15 Class Distribution (Dataset without Slang Word).....	80
Figure 4. 16 Class Distribution After ROS (Dataset without Slang Word).....	81
Figure 4. 17 Accuracy Graph per Fold POC Scenario .....	82
Figure 4. 18 Confusion Matrix POC Scenario .....	82
Figure 4. 19 Home Page Visualization .....	85
Figure 4. 20 Result Page Visualization .....	86

## LIST OF ATTACHMENTS

Attachment 1. Profile .....	98
Attachment 2. Cleansing Data Code .....	99
Attachment 2. Translation Pipeline Code .....	99
Attachment 3. Post Translation Cleansing Code.....	100
Attachment 4. Normalization Code .....	101
Attachment 5. Stemming Code .....	101
Attachment 6. InSet Lexicon Labelling Code.....	102
Attachment 7. IndoBERT Modelling Code.....	103
Attachment 5. Data Request Letter SMP Negeri 1 Sukawati .....	110
Attachment 6. Data Request Letter SMK Negeri 1 Tampaksiring.....	112
Attachment 7. Data Request Letter Café Gabriel.....	114
Attachment 8. Data Annotation Statement Letter .....	115
Attachment 9. Documentation with annotators.....	118

