

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP  
LAYANAN PROVIDER INDIHOME MENGGUNAKAN  
ALGORITMA NAIVE BAYES**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**MUHAMMAD ENGGAR AZIZ HIBBANUARI**

**15.11.8681**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP LAYANAN  
PROVIDER INDIHOME MENGGUNAKAN ALGORITMA NAIVE  
BAYES**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**MUHAMMAD ENGGAR AZIZ HIBBANNUARI  
15.11.8681**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2023**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP  
LAYANAN PROVIDER INDIHOME MENGGUNAKAN  
ALGORITMA NAIVE BAYES**

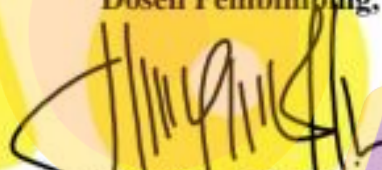
yang disusun dan diajukan oleh

**Muhammad Enggar Aziz Hibbannuari**

**15.11.8681**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 01 September 2022

Dosen Pembimbing,



Hartatik, S.T., M.Cs.

NIK. 190302282

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP**  
**LAYANAN PROVIDER INDIHOME MENGGUNAKAN**  
**ALGORITMA NAIVE BAYES**

yang disusun dan diajukan oleh

**Muhammad Enggar Aziz Hibbannuari**

**15.11.8681**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 18 Januari 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Mardhiva Hayatv, S.T., M.Kom**  
**NIK. 190302108**

**Yuli Astuti, M.Kom**  
**NIK. 190302146**

**Windha Mega Pradnya D, M.Kom**  
**NIK. 190302185**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 18 Januari 2023

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom.**  
**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

**Nama mahasiswa : Muhammad Enggar Aziz Hibbannuari**  
**NIM : 15.11.8681**

Menyatakan bahwa Skripsi dengan judul berikut:

### **Analisis Sentimen Pengguna Twitter Terhadap Layanan Provider Indihome Menggunakan Algoritma Naive Bayes**

Dosen Pembimbing : Hartatik, S.T., M.Cs.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Januari 2023

Yang Menyatakan,



Muhammad Enggar Aziz Hibbannuari

## HALAMAN PERSEMBAHAN

Puji syukur penulis haturkan kepada Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya, sehingga dapat menyelesaikan skripsi ini dengan lancar dan baik. Dengan dukungan dan semangat dari berbagai pihak.

Skripsi penulis persembahkan kepada pemberi ilmu pengetahuan, akal sehat, serta memberikan jalan mudah dengan tepat yaitu kepada Allah SWT.

Untuk kedua Orang Tua, Bapak Sudaryadi, S.ST. dan Ibu Mud'ah S.Pd.I yang selalu memberi semangat dan do'a, sehingga skripsi ini dapat selesai dengan baik.

Untuk Saudara penulis Amri Husnianto, Rahimah Rabita N. P. dan Ashill Zuhud Fadhail N. yang selalu memberi semangat dan dorongan agar menyelesaikan skripsi ini.

Untuk Ibu Hartatik, S.T., M.Cs. selaku dosen pembimbing yang telah memberikan masukan, semangat, sabar dalam membimbing dan arah kepada penulis dalam penyusunan naskah skripsi.

Untuk Almh., Ibu Armadyah Amborowati, S.Kom, M.Eng. selaku dosen dosen wali dan dosen pembimbing, semoga ilmu yang diberikan menjadi pahala yang tidak putus.

Untuk seluruh Dosen dan Civitas Akademik Universitas Amikom Yogyakarta yang telah memberi kesempatan, bantuan, pengalaman dan ilmu bagi penulis dalam perkuliahan.

Untuk Mas Fregy Damara dan Mas Wahyu yang telah memberi saran, waktu dan motivasi dalam penyusunan skripsi.

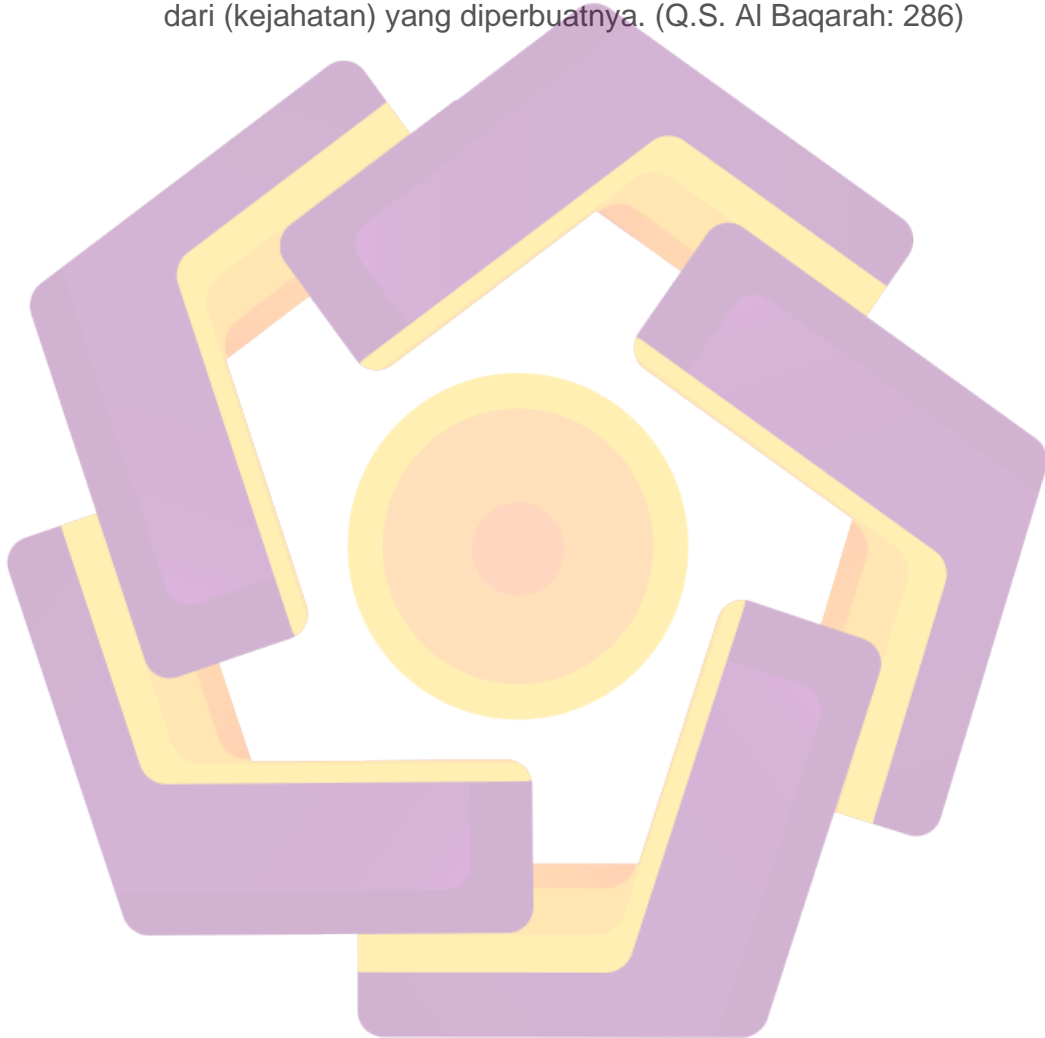
Untuk teman-teman 15-S1TI-03, yang selalu kompak dan memberikan semangat.

Untuk seluruh pihak yang mungkin tidak dapat disebutkan satu persatu, penulis haturkan terimakasih semoga ilmu dan pengalaman yang telah diberikan akan berbalik menjadi pahala dan manfaat. Penulis menyadari bahwa naskah skripsi ini memiliki kekurangan. Untuk itu penulis mengharapkan saran dan kritik dari segala aspek dalam penyusunan dan penulisan naskah skripsi ini. Penulis berharap naskah skripsi ini dapat bermanfaat bagi peneliti selanjutnya berkaitan topik serupa.

## HALAMAN MOTTO

Man Jadda Wajada

Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya. Dia mendapat (pahala) dari (kebajikan) yang dikerjakannya dan dia mendapat (siksa) dari (kejahatan) yang diperbuatnya. (Q.S. Al Baqarah: 286)



## KATA PENGANTAR

Puji dan syukur penulis haturkan kepada Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan baik. Tak lupa sholawat serta salam penulis haturkan kepada junjungan Nabi Muhammad SAW yang *insyaallah* bagi umat Muslim menjadi sauri tauladan. Penulis dapat menyelesaikan skripsi dengan judul *Analisis Sentimen Pengguna Twitter Terhadap Layanan Provider Indihome Menggunakan Algoritma Naive Bayes*.

Penulis haturkan terimakasih kepada seluruh pihak yang terlibat baik secara langsung atau tidak dalam penyusunan skripsi ini khususnya kepada:

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku Ketua Program Studi Informatika.
4. Ibu Hartatik, S.T., M.Cs. selaku Dosen Pembimbing yang telah sabar dan memberi arahan serta masukan kepada penulis dalam penyusunan skripsi ini.
5. Almh., Ibu Armadyah Amborowati, S.Kom, M.Eng. selaku dosen dosen wali dan dosen pembimbing, semoga ilmu yang diberikan menjadi pahala yang tidak putus.
6. Dosen Penguji yang telah memberi masukan dan saran kepada penulis sehingga dapat menyusun skripsi dengan lebih baik.
7. Seluruh Dosen dan Civitas Akademik Universitas Amikom Yogyakarta yang telah memberi kesempatan, bantuan, pengalaman dan ilmu bagi penulis dalam perkuliahan.
8. Kedua orang tua penulis, Bapak Sudaryadi, S.ST. dan Ibu Mud'ah S.Pd.I yang selalu memberi semangat dan do'a, sehingga skripsi ini dapat selesai dengan baik.
9. Saudara penulis Amri Husnianto, Rahimah Rabita N. P. dan Ashill Zuhud Fadhail N. yang selalu memberi semangat dan dorongan agar menyelesaikan skripsi ini.



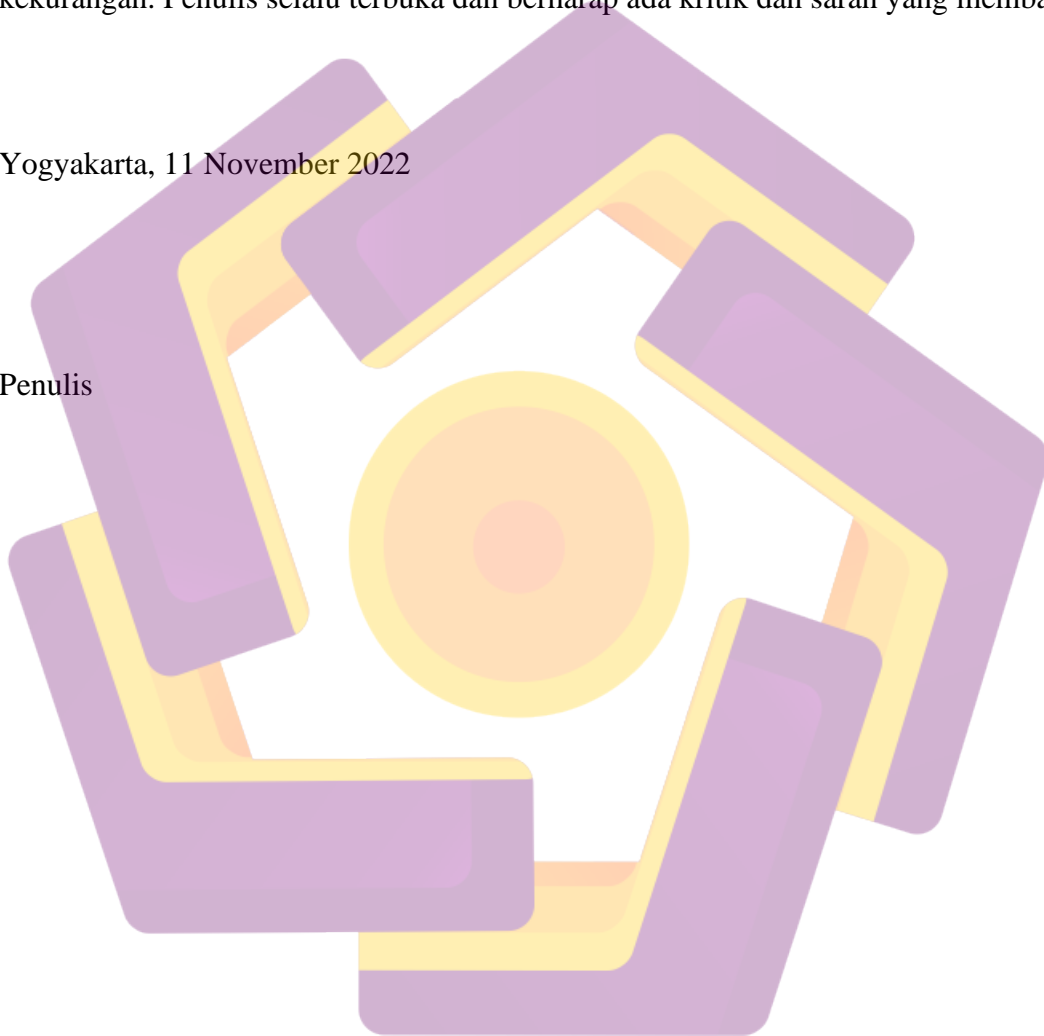
10. Untuk Mas Fregy Damara dan Mas Wahyu yang telah memberi saran, waktu dan motivasi dalam penyusunan skripsi.

11. Untuk teman-teman 15-S1TI-03, yang selalu kompak dan memberikan semangat.

Penulis menyadari dalam penyusunan naskah skripsi ini tentu masih memiliki kekurangan. Penulis selalu terbuka dan berharap ada kritik dan saran yang membangun.

Yogyakarta, 11 November 2022

Penulis



## DAFTAR ISI

HALAMAN JUDUL .....	ii
HALAMAN PERSETUJUAN .....	iii
HALAMAN PENGESAHAN .....	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	v
HALAMAN PERSEMBAHAN .....	vi
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN.....	xv
INTISARI .....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian.....	2
1.5 Manfaat Penelitian .....	3
1.6 Sistematika Penulisan .....	3
BAB II TINJAUAN PUSTAKA .....	4
2.1 Studi Literatur.....	4
2.2 Dasar Teori .....	10
2.2.1 Text Mining.....	10
2.2.2 Indihome.....	10
2.2.3 Natural Language Processing.....	10
2.2.4 Vectorizer .....	11
2.2.5 Scikit Learn .....	11
2.2.6 Python.....	12
2.2.7 Natural Language ToolKit .....	12
2.2.8 Sentiment Analysis.....	13
2.2.9 Classification.....	13
2.2.10 Tweet Crawler .....	14
2.2.11 Training Data.....	14
2.2.12 Testing Data.....	14
2.2.13 Pre-Processing .....	15
2.2.13.1 Cleaning Data .....	15

2.2.13.2 Case Folding .....	16
2.2.13.3 Tokenization .....	16
2.2.13.4 Stopwords .....	16
2.2.13.5 Stemming .....	16
2.2.14 Feature Transform .....	17
2.2.15 TF-IDF .....	17
2.2.16 Train Split Test .....	18
2.2.17 Naïve Bayes Classifier .....	19
2.2.18 Cross Validatiton .....	21
<b>BAB III METODE PENELITIAN .....</b>	<b>24</b>
3.1 Objek Penelitian .....	24
3.2 Metode Pengumpulan Data .....	24
3.2.1 Studi Pustaka .....	24
3.2.2 Interview .....	24
3.3 Alur Penelitian .....	24
3.3.1 Training .....	26
3.3.1.1 Mulai .....	26
3.3.1.2 Preprocessing .....	26
3.3.1.3 Feature Transform .....	27
3.3.1.3.1 TF-IDF .....	27
3.3.1.3.1 Vectorizer .....	27
3.3.1.4 Model Training .....	28
3.3.1.4.1 Train Test Split .....	28
3.3.1.4.2 Validation Process .....	28
3.3.1.4.3 Cross Validation .....	28
3.3.1.4.4 Training Process .....	28
3.3.1.4.5 Model Training Naïve Bayes .....	28
3.3.1.4.6 Training Cross Validation .....	28
3.3.1.4.7 Confusion Matrix .....	28
3.3.1.4.8 Model .....	29
3.3.2 Prediction .....	29
3.3.2.1 Get Tweet Data .....	29
3.3.2.1.1 Input .....	29
3.3.2.1.2 Tweet Crawling .....	29
3.3.2.1.3 Hasil Tweet .....	29
3.3.2.1.4 Prediction .....	29
3.3.2.1.4.1 Preprocessing .....	29
3.3.2.1.4.2 Clean Tweet .....	29
3.3.2.1.4.3 Prediction Tweet .....	30
3.3.2.1.4.3.1 Feature Transform .....	30
3.3.2.1.4.3.2 Vector of Tweet .....	30
3.3.2.1.4.3.3 Model Prediction .....	30
3.3.2.1.4.3.4 Tweet Sentiment Probability .....	30
3.4 Alat dan Bahan Penelitian .....	30
3.4.1 Data Penelitian .....	30
3.4.2 Alat atau Instrumen .....	31
3.4.3 Bahan Penelitian .....	32
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>34</b>

4.1	Data Cleaning .....	34
4.2	Case Folding .....	35
4.3	Tokenization.....	36
4.4	Stopwords.....	37
4.5	Stemming .....	39
4.6	Feature Transform .....	40
4.6.1	TF-IDF.....	40
4.7	Pembagian Data Training dan Data Testing .....	45
4.8	Naïve Bayes Classifier .....	45
4.8.1	Likelihood .....	45
4.8.2	Prior Data.....	47
4.8.3	Posterior Data .....	47
4.8.4	Prediksi Kelas .....	48
4.8.4.1	Frekuensi Tiap Kata.....	48
4.8.4.2	Menghitung Skor Kelas Positif.....	49
4.8.4.3	Menghitung Skor Kelas Negatif.....	50
4.9	Cross Validation.....	50
4.9.1	Hasil Pengujian KFold 1.....	52
4.9.2	Hasil Pengujian KFold 2.....	52
4.9.3	Hasil Pengujian KFold 3.....	53
4.9.4	Hasil Pengujian KFold 4.....	53
4.9.5	Hasil Pengujian KFold 5.....	54
4.10	Confusion Matrix .....	55
4.10.1	Hasil Pengujian Confusion Matrix .....	56
4.11	Perbandingan Dengan Penelitian Sebelumnya .....	59
4.12	Implementasi Pengumpulan Data .....	59
4.13	Implementasi Sistem.....	60
4.13.1	Konfigurasi Server.....	60
4.13.2	Aplikasi Server .....	60
4.14	Implementasi Penyajian .....	61
4.14.1	Tampilan Home .....	61
4.14.2	Train Model .....	61
4.14.3	Model Performance .....	62
4.14.4	Sentiment Analysis .....	62
4.14.4.1	Keyword dan Total Tweet .....	62
4.14.4.2	Memilih Model Training .....	63
4.14.4.3	Tweet Type .....	63
4.14.4.4	Tweet Filters .....	63
4.14.4.5	Input Analyze Tweet .....	63
4.14.4.6	Tweet Insight .....	64
4.15	Hasil Klasifikasi Sentimen Analysis .....	65
4.15.1	Hasil Uji Perhitungan Confusion Matrix.....	68
BAB V PENUTUP .....		71
5.1	Kesimpulan.....	71
5.2	Saran .....	71
REFERENSI .....		73
LAMPIRAN.....		77

## DAFTAR TABEL

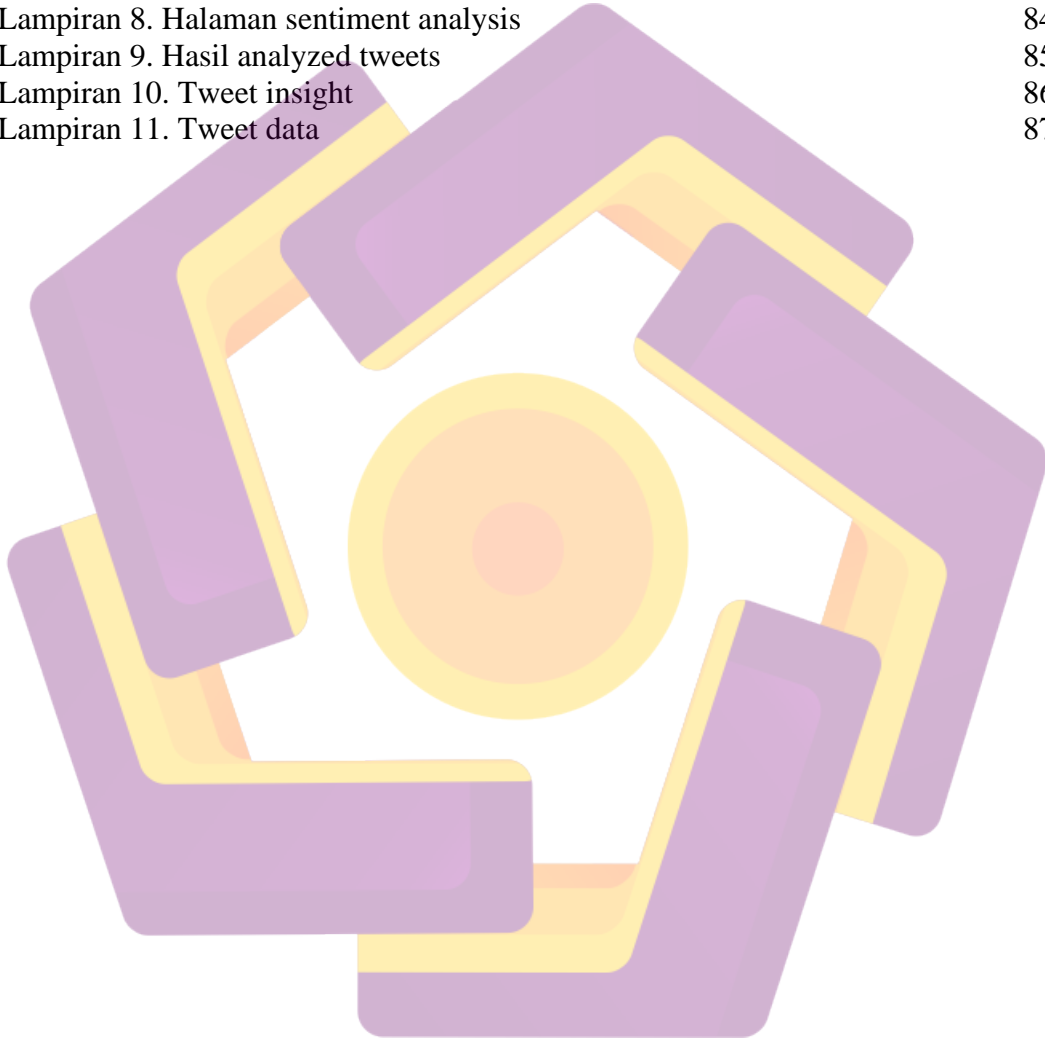
Tabel 2.1. Keaslian Penelitian	6
Tabel 2.2. Confusion Matrix	22
Tabel 3.1. Tabel Alur Penelitian	25
Tabel 3.2. Sampel Data Penelitian	31
Tabel 4.1. Data Awal Atau Data Mentah	34
Tabel 4.2. Data Setelah Data Cleaning	35
Tabel 4.3. Hasil Case Folding	35
Tabel 4.4. Data Case Folding Dan Tokenization	36
Tabel 4.5. Data Token Dan Hasil Stopwords	38
Tabel 4.6. Data Stopwords Dan Hasil Stemming	39
Tabel 4.7. Term Frequency	40
Tabel 4.8. Term Frequency - Inverse Document Frequency	42
Tabel 4.9. Term Frequency - Inverse Document Frequency Positive	43
Tabel 4.10. Term Frequency - Inverse Document Frequency Negative	44
Tabel 4.11. Pembagian Data Training Dan Data Testing	45
Tabel 4.12. Dataset Baru Untuk Prediksi	48
Tabel 4.13. Kfold 1	51
Tabel 4.14. Kfold 2	51
Tabel 4.15. Kfold 3	51
Tabel 4.16. Kfold 4	51
Tabel 4.17. Kfold 5	51

## DAFTAR GAMBAR

Gambar 2.1. K-Fold Cross Validation	21
Gambar 3.1. Tabel Alur Penelitian	25
Gambar 4.1. Hasil K-Fold 1	52
Gambar 4.2. Hasil K-Fold 2	53
Gambar 4.3. Hasil K-Fold 3	53
Gambar 4.4. Hasil K-Fold 4	54
Gambar 4.5. Hasil K-Fold 5	54
Gambar 4.6. Hasil K-Fold	55
Gambar 4.7. Hasil Confusion Matrix Training Data	55
Gambar 4.8. Hasil Confusion Matrix Validation Data	56
Gambar 4.9. Accuracy	57
Gambar 4.10. Recall	57
Gambar 4.11. Precision	58
Gambar 4.12. F1 Score	58
Gambar 4.13. Script Crawling Tweet	59
Gambar 4.14. Hasil Data Raw	60
Gambar 4.15. Konfigurasi Api Key	60
Gambar 4.16. Flask App Server	60
Gambar 4.17. Tampilan Home	61
Gambar 4.18. Train Model	61
Gambar 4.19. Dataset Format	62
Gambar 4.20. List Model Performance	62
Gambar 4.21. Set Keyword Dan Total Data	62
Gambar 4.22. Memilih Model Training	63
Gambar 4.23. Status Sentiment Analysis	64
Gambar 4.24. Tweet Insight	64
Gambar 4.25. Average Model Probability Value	65
Gambar 4.26. Total Data Uji	65
Gambar 4.27. Pengujian Kfold Ke-1	66
Gambar 4.28. Pengujian Kfold Ke-2	66
Gambar 4.29. Pengujian Kfold Ke-3	67
Gambar 4.30. Pengujian Kfold Ke-4	67
Gambar 4.31. Pengujian Kfold Ke-5	68
Gambar 4.32. Hasil Confusion Matrix Training Data	68
Gambar 4.33. Hasil Confusion Matrix Validation Data	69
Gambar 4.34. Hasil Uji Confusion Matrix	69

## DAFTAR LAMPIRAN

Lampiran 1. Source code scrape tweet	77
Lampiran 2. Tabel alur penelitian training process	78
Lampiran 3. Tabel alur penelitian prediction process	79
Lampiran 4. Halaman home sentiment analysis	80
Lampiran 5. Halaman training model	81
Lampiran 6. Halaman model performance	82
Lampiran 7. Halaman hasil performance	83
Lampiran 8. Halaman sentiment analysis	84
Lampiran 9. Hasil analyzed tweets	85
Lampiran 10. Tweet insight	86
Lampiran 11. Tweet data	87



## INTISARI

Indihome merupakan salah satu penyedia layanan internet (ISP) yang ada di Indonesia yang jangkauan layanan sudah tercakup ke seluruh wilayah Indonesia. Sebuah penyedia layanan seperti Indihome pasti memiliki keluhan atau aduan tentang kepuasan atau kelayakan menggunakan Indihome, melalui data pengguna Twitter untuk mendapatkan akurasi nilai positif dan negatif terhadap layanan Indihome.

Menggunakan data tweet pada Twitter merupakan memanfaatkan layanan yang disediakan untuk keluhan ataupun aduan terhadap Indihome. Pengambilan data tweet pada Twitter menggunakan script Bahasa Python. Data tersebut di klasifikasi menggunakan Naive Bayes Classifier.

Naive Bayes Classifier merupakan algoritma yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh Ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa mendatang berdasarkan data pada masa sebelumnya. Analisis Sentimen memiliki tahapan yaitu crawling, pre-processing, pembobotan kata, pembentukan model dan klasifikasi sentimen. Pada pembobotan kata digunakan metode TF - IDF. Data yang ada akan diklasifikasikan ke dalam 2 kelas, yaitu positif dan negatif. Data tersebut akan diuji melalui confusion matrix dengan parameter akurasi, recall dan precision.

**Kata kunci:** Naïve Bayes, Analisis Sentimen, Klasifikasi, Indihome, Confusion Matrix.



## ABSTRACT

*Indihome is one of the internet service providers (ISPs) in Indonesia whose service coverage has been covered throughout Indonesia. A service provider like Indihome must have a complaint or complaint about satisfaction or eligibility to use Indihome, through Twitter user data to get an accuracy of positive, negative or neutral values against Indihome services.*

*Using tweet data on Twitter is taking advantage of services provided for complaints or complaints against Indihome. Retrieving tweet data on Twitter using Python script. The data was classicized using the Naive Bayes Classifier.*

*Naive Bayes Classifier is an algorithm that utilizes probability calculations and statistics proposed by British Scientist Thomas Bayes, namely predicting future probabilities based on data in the past. Sentiment Analysis has stages namely crawling, pre-processing, word weighting, model formation and sentiment classification. In the weighting of the word used the TF-IDF method. The existing data will be classified into 2 classes, namely positive, and negative. The data will be tested through a confusion matrix with parameters of accuracy, recall and precision.*

**Keyword:** *Naive Bayes, Sentiment Analysis, Classification, Indihome, Confusion Matrix*

