

**ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP BUS LISTRIK MENGGUNAKAN NAIVE BAYES**

SKRIPSI

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



disusun oleh

SYARIF NURWAHID JAELANI

19.11.2861

Kepada

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

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
BUS LISTRIK MENGGUNAKAN NAÏVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

SYARIF NURWAHID JAELANI

19.11.2861

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP BUS LISTRIK MENGGUNAKAN NAIVE BAYES**

yang disusun dan diajukan oleh

Syarif Nurwahid Jaelani

19.11.2861

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Juni 2023

Dosen Pembimbing,



Ike Verawati, M.Kom
NIK. 190302237

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP BUS
LISTRIK MENGGUNAKAN NAIVE BAYES

yang disusun dan diajukan oleh

Syarif Nurwahid Jaelani

19.11.2861

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Juni 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Atik Nurmasani, S.Kom., M.Kom
NIK. 190302354

Robert Marco, M.T.
NIK. 190302228

Ike Verawati, M.Kom
NIK. 190302237



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Juni 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 : Syarif Nurwahid Jaelani
NIM : 19.11.2861

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Pengguna Twitter Terhadap Bus Listrik Menggunakan Naive Bayes

Dosen Pembimbing : IKE Verawati, M.Kom

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, 22 Juni 2023

Yang Menyatakan,



Syarif Nurwahid Jaelani

HALAMAN PERSEMBAHAN

Syukur Alhamdulillah atas segala rahmat dan hidayah-Nya serta dengan kemudahan yang diberikan oleh Allah SWT, skripsi ini peneliti persembahkan kepada:

1. Kepada Orang tua, Bapak Sugiyono, S.Pd dan Ibu Tri Umbarwati yang telah mendidik, mendoakan dan memberi dukungan yang sangat berharga untuk peneliti.
2. Ibu Ike Verawati, M.Kom selaku Dosen Pembimbing, yang telah memberikan bimbingan, pengarahan, nasihat, dukungan serta telah banyak membantu segala proses hingga skripsi ini selesai.
3. Rizqi Bio Janhefi, Feizal Reza, M.Muh.Akhyar Z, dan Bramantio Wibowo yang telah memberi semangat serta dukungan untuk peneliti.
4. Kepada teman-teman Program studi Informatika Universitas Amikom Yogyakarta yang telah memberikan dukungan yang luar biasa.
5. Kepada seluruh mahasiswa kelas Informatika 05 angkatan 2019 yang telah memberikan dukungan serta semangat kepada peneliti.

KATA PENGANTAR

Puji Syukur, Alhamdulillah atas kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya kepada kita semua sehingga kami dapat menyelesaikan skripsi yang telah diajukan sebagai salah satu syarat untuk menyelesaikan program strata satu (S1) pada program studi Informatika Universitas Amikom Yogyakarta.

Adapun penyusunan skripsi ini digunakan sebagai bukti bahwa penyusun telah menyelesaikan serta melaksanakan penelitian Skripsi. Dalam proses penyusunan laporan ini penyusun mendapatkan banyak bantuan dari berbagai pihak. Oleh karena itu kami selaku peneliti mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof Dr. M. Suyanto, M.M selaku rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta,S.Kom., M.Kom. selaku kepala Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya Duhita, M.Kom. selaku Kepala Prodi Informatika Universitas Amikom Yogyakarta.
4. Ibu Ike Verawati, M.Kom. Selaku pembimbing utama yang telah banyak membimbing serta membantu dalam menyelesaikan skripsi ini.

Yogyakarta, 22 Juni 2023

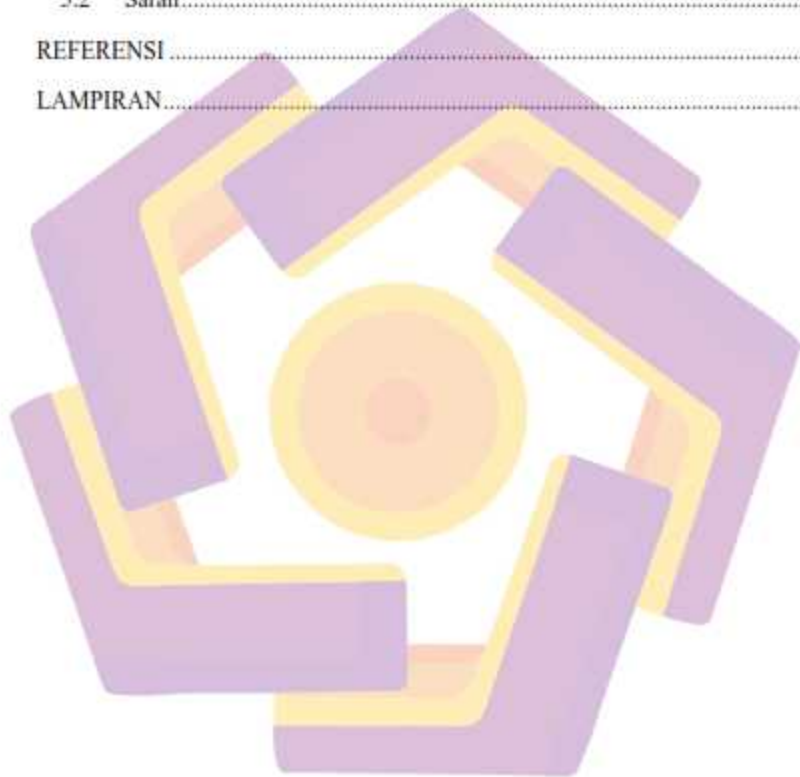
Syarif Nurwahid jaelani

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
INTISARI	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	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 Bus Listrik.....	10

2.2.2	Text Mining	10
2.2.3	Text Preprocessing	10
2.2.4	Analisis Sentimen	11
2.2.5	Twitter	12
2.2.7	Jupyter Notebook	12
2.2.8	TF-IDF (Term Frequency Inverse Document Frequency).....	12
2.2.9	Metode K-Fold Cross Validation	13
2.2.10	Hyperparameter.....	14
2.2.11	Grid Search	14
2.2.12	Naive Bayes Classifier.....	14
2.2.13	Multinomial Naive Bayes	15
2.2.14	Lexicon Based.....	16
2.2.15	Confusion Matrix.....	17
BAB III METODE PENELITIAN		19
3.1	Alur Penelitian.....	19
3.2	Alat dan Bahan.....	22
3.2.1	Alat/Perangkat.....	22
3.2.2	Data Penelitian	23
3.2.3	Instrumen Pengumpulan Data.....	27
3.2.4	Multinomial Naive Bayes	27
BAB IV HASIL DAN PEMBAHASAN		29
4.1	Pengumpulan Data	29
4.2	Preprocessing Data	30
4.3	Pelabelan Lexicon Indonesia.....	34
4.4	Pembobotan TF-IDF	36

4.5	Klasifikasi dengan Metode Multinomial Naive Bayes	38
4.6	Evaluasi	44
BAB V PENUTUP		50
5.1	Kesimpulan.....	50
5.2	Saran.....	50
REFERENSI		51
LAMPIRAN.....		54



DAFTAR TABEL

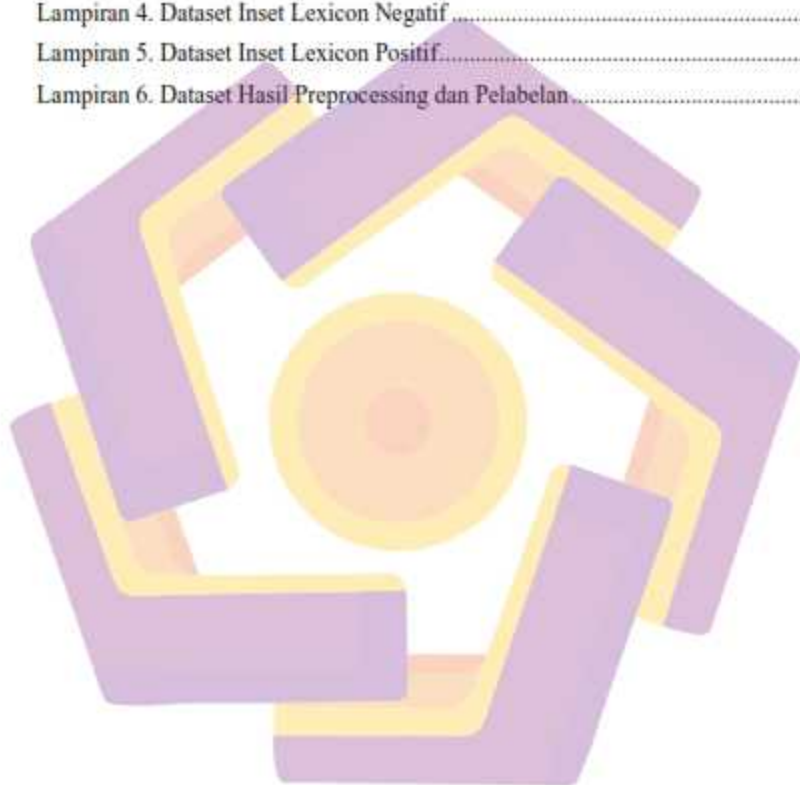
Tabel 2.1. Keaslian Penelitian	10
Tabel 2.2. Confusion Matrix	16
Tabel 3.1. Sampel Data Tweet	22
Tabel 3.2. Sampel dataset Colloquial Indonesian Lexicon	24
Tabel 3.3. Sampel dataset lexicon positif	25
Tabel 3.4. Sampel dataset lexicon negatif	25
Tabel 3.5. Sampel dataset emoji to text Indonesia	25
Tabel 4.1. Konversi Emoji	29
Tabel 4.2. Case Folding	29
Tabel 4.3. Cleaning	30
Tabel 4.4. Tokenizing	30
Tabel 4.5. Normalisasi	31
Tabel 4.6. Stopword removal	31
Tabel 4.7. Stemming	32
Tabel 4.8. Perhitungan Inset lexicon	33
Tabel 4.9. Hasil Pelabelan	33
Tabel 4.10. Dokumen Sampel	34
Tabel 4.11. Pembobotan Tf-Idf	35
Tabel 4.12. Dokumen Latih	36
Tabel 4.13. Term Frequency Data Latih	36
Tabel 4.14. Probabilitas Data Latih	38
Tabel 4.15. Contoh dokumen Data Uji	39
Tabel 4.16. Perbandingan hasil klasifikasi	42
Tabel 4.17. Evaluasi Multinomial Naive Bayes	44
Tabel 4.18. Evaluasi Multinomial NB + Hyperparameter	45

DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian	18
Gambar 4.1. Script scraping.....	28
Gambar 4.2. Dataset hasil scraping.....	28
Gambar 4.3. Script Inset Lexicon base	32
Gambar 4.4. Distribusi Pelabelan Inset.....	34
Gambar 4.5. Script Pembobotan Kata.....	35
Gambar 4.6. Hasil Pembobotan Tf-IDF.....	36
Gambar 4.7. Script Model Klasifikasi	40
Gambar 4.8. Hasil Prediksi Model.....	41
Gambar 4.9. Hasil Prediksi Model dengan Optimasi.....	41
Gambar 4.10. Confusion Matrix Naive Bayes.....	42
Gambar 4.11. Confusion Matrix Model + Hyperparameter	43
Gambar 4.12. Perbandingan Akurasi.....	49

DAFTAR LAMPIRAN

Lampiran 1. Dataset hasil scrapping	54
Lampiran 2. Dataset Kamus_Alay	54
Lampiran 3. Dataset Emoji to Text	55
Lampiran 4. Dataset Inset Lexicon Negatif	55
Lampiran 5. Dataset Inset Lexicon Positif	55
Lampiran 6. Dataset Hasil Preprocessing dan Pelabelan	56



INTISARI

Analisis setntimen twitter merupakan salah satu metode dalam mengidentifikasi dan mengklasifikasi pendapat kedalam sentimen positif atau negatif pada tweet. Salah satu topik yang dibahas pada media sosial twitter yang mendapat berbagai pendapat yang pro dan kontra adalah bus listrik. Naïve bayes dapat digunakan untuk mengklasifikasikan sentimen serta mudah untuk diimplementasikan dan memiliki tingkat akurasi yang cukup tinggi. Tujuan dari penelitian ini adalah mengimplementasikan naïve bayes untuk menkalsifikasikan sentimen mengenai bus listrik apakah sentimennya mengarah ke positif atau negatif. Beberapa tahapan yang dilakukan yaitu pengambilan data, preprocessing, pelabelan lexicon, pembobotan kata, pembagian data latih dan uji, klasifikasi naïve bayes, dan valusasi confusion matrix. Hasil dari tahapan tersebut dari 4 kali percobaan rasio pembagian data yang berbeda diperoleh sentimen terbanyak pada sentimen positif yang mencapai 77,31% dari data latih pada rasio pembagian data 6:4 dan akurasi yang diperoleh setelah dilakukan optimalisasi sebesar 78%.

Kata kunci: Analisis Sentimen, Naive Bayes, Bus Listrik.



ABSTRACT

Twitter sentiment analysis is a method of identifying and classifying opinions into positive or negative sentiments on tweets. One of the topics discussed on social media Twitter which received various opinions, the pros and cons, is the electric bus. Naïve Bayes can be used to classify sentiments and is easy to implement and has a fairly high level of accuracy. The purpose of this study is to implement naïve bayes to classify sentiments about electric buses whether the sentiments are positive or negative. Several stages were carried out, namely data collection, preprocessing, lexicon labeling, word weighting, training and test data distribution, naïve Bayes classification, and confusion matrix valuation. The results of these stages from 4 trials of different data sharing ratios obtained the most positive sentiment which reached 77.31% of the training data at a data sharing ratio of 6:4 and the accuracy obtained after optimization was 78%.

Keyword: *Sentiment Analysis, Naïve Bayes, Electric Bus.*

