

**PERBANDINGAN ALGORITMA NAÏVE BAYES DAN DECISION TREE
PADA SENTIMEN ANALISIS**

SKRIPSI



disusun oleh

I Putu Wibina Karsa Gumi

17.11.1614

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

**PERBANDINGAN ALGORITMA NAÏVE BAYES DAN DECISION TREE
PADA SENTIMEN ANALISIS**

SKRIPSI

Diajukan untuk memenuhi salah satu persyaratan memperoleh gelar Sarjana Komputer
pada Program Studi Informatika



Disusun Oleh

I PUTU WIBINA KARSA GUMI

17.11.1614

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

PERSETUJUAN

SKRIPSI

**PERBANDINGAN ALGORITMA NAÏVE BAYES DAN DECISION TREE PADA
SENTIMEN ANALISIS**

yang dipersiapkan dan disusun oleh

I Putu Wibina Karsa Gumi

17.11.1614

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 2 Juli 2022

Dosen Pembimbing,

Hartatik, S.T., M.Cs.

NIK. 190302232

PENGESAHAN

SKRIPSI

**PERBANDINGAN ALGORITMA NAÏVE BAYES DAN DECISION TREE PADA
SENTIMEN ANALISIS**

yang dipersiapkan dan disusun oleh

I Putu Wibina Karsa Gumi

17.11.1614

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Juli 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Yuli Astuti, M.Kom
NIK. 190302146

Irma Rofni Wulandari, S.Pd., M.Eng
NIK. 190302329

Hartatik, S.T., M.Cs.
NIK. 190302232

Skripsi ini telah diterima sebagai salah satu persyaratan
Untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Juli 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanid Al Fatta, S.Kom., M.Kom
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi Pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar Pustaka

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah tanggung jawab saya pribadi

Yogyakarta, 2 Juli 2022



I Putu Wibina Karsa Gumi

NIM. 17.11.1614

MOTTO

“No man is hurt but by himself.”

- Diogenes

“I'd rather be optimistic and wrong than pessimistic and right.”

- Elon Musk

“There is only one way to happiness and that is to cease worrying about things which are beyond the power of our will.”

- Epictetus

“The only true wisdom is in knowing you know nothing.”

- Socrates

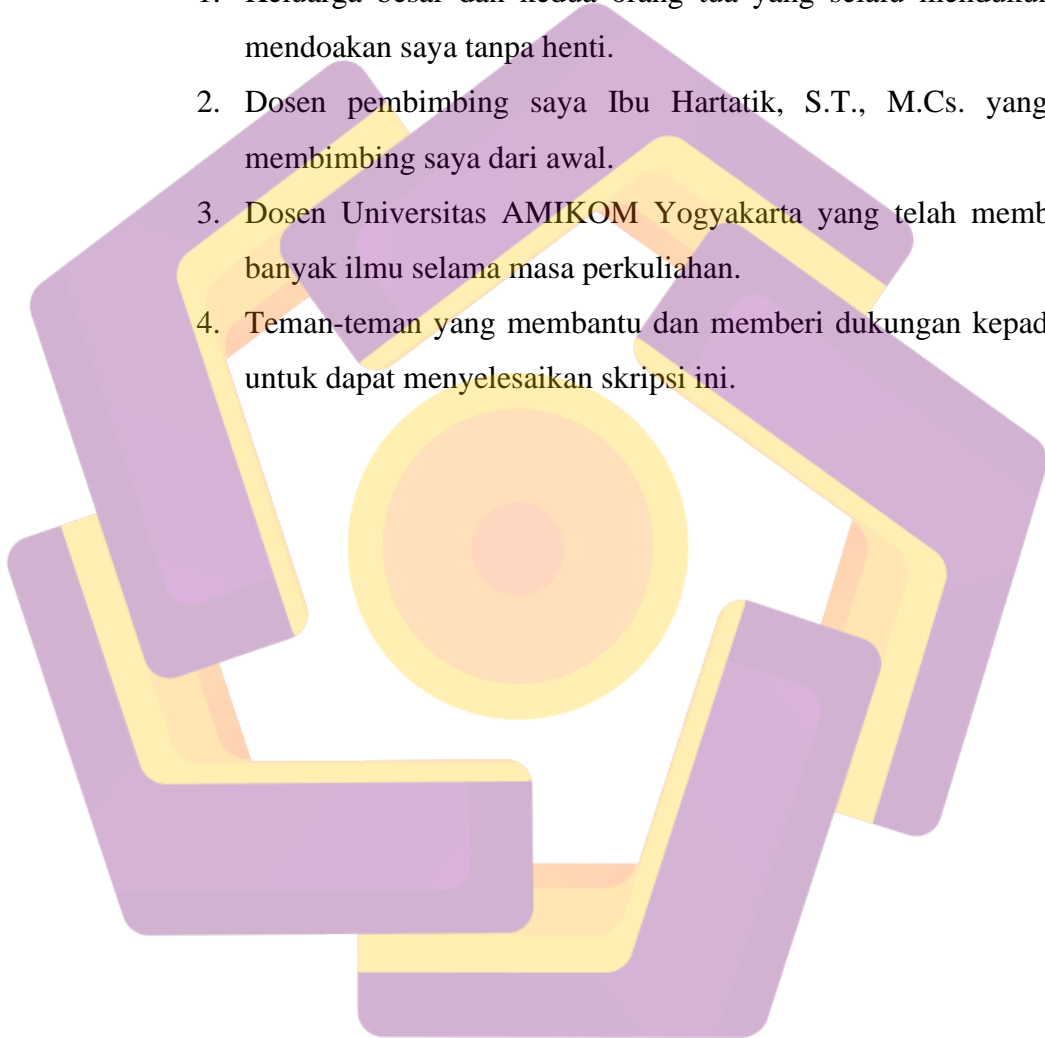
“First say to yourself what you would be and then do what you have to do.”

- Epictetus

PERSEMBAHAN

Puja dan puji syukur saya panjatkan dalam doa kepada Tuhan Yang Maha Esa karenanyalah saya memiliki kekuatan untuk menyelesaikan skripsi ini dengan baik. Skripsi ini saya persembahkan kepada semua pihak yang terlibat langsung maupun tidak langsung, yaitu untuk :

1. Keluarga besar dan kedua orang tua yang selalu mendukung dan mendoakan saya tanpa henti.
2. Dosen pembimbing saya Ibu Hartatik, S.T., M.Cs. yang telah membimbing saya dari awal.
3. Dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama masa perkuliahan.
4. Teman-teman yang membantu dan memberi dukungan kepada saya untuk dapat menyelesaikan skripsi ini.



KATA PENGANTAR

Puji syukur saya selaku penulis panjatkan kepada Tuhan Yang Maha Esa, karena atas karunia dan rahmatnya saya dapat menyelesaikan skripsi ini. Skripsi dengan judul “PERBANDINGAN ALGORITMA NAÏVE BAYES DAN DECISION TREE PADA SENTIMEN ANALISIS” disusun sebagai salah satu syarat untuk menyelesaikan program sarjana pada Universitas AMIKOM Yogyakarta.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan, maka dari itu penulis menerima berbagai kritik dan saran dari semua pihak. Semoga skripsi yang sederhana ini bisa bermanfaat bagi penulis dan pembaca. Apabila terdapat kesalahan saya memohon maaf, terima kasih.

Yogyakarta 2 Juli 2022



I Putu Wibina Karsa Gumi

Daftar Isi

JUDUL	i
PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN	iv
MOTTO	v
PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	2
1.5 Batasan Masalah	3
1.6 Sistematika Penulisan	3
BAB II	4
LANDASAN TEORI	4
2.1 Tinjauan Pustaka.....	4
2.2 Sentimen Analisis	7
2.3 Twitter dan API	7
2.3.1 Crawling Data	8
2.4 Text Mining	8
2.5 Text Preprocessing.....	9

2.5.1	Cleaning	9
2.5.2	Case Folding	9
2.5.3	Stemming	10
2.5.4	Filtering.....	10
2.5.5	Tokenizing	11
2.6	Klasifikasi	11
2.6.1	Supervised Learning	11
2.6.2	Naïve Bayes	12
2.6.3	Decision Tree.....	14
2.7	TF-IDF.....	15
2.8	Confusion Matrix.....	15
2.8.1	Accuracy	16
2.8.2	Precision.....	16
2.8.3	Recall	16
2.8.4	F1 Score.....	16
BAB III	17
METODE PENELITIAN	17
3.1	Perancangan dan Alur Penelitian	17
3.2	Studi Pustaka.....	18
3.3	Pengambilan Data	19
3.4	Preprocessing dan Labeling	20
3.5	Perancangan Sistem	20
3.6	Analisa Hasil.....	21
3.7	Kesimpulan	21
3.8	Alat Penelitian.....	21
3.8.1	Perangkat Keras	21

3.8.2	Perangkat Lunak	21
BAB IV	22
IMPLEMENTASI DAN ANALISA HASIL	22
4.1	Perancangan Alur Desain Sistem.....	22
4.2	Pengumpulan Data.....	22
4.3	Preprocessing	24
4.3.1	Cleaning	24
4.3.2	Case Folding	26
4.3.3	Stemming	26
4.3.4	Stopword Removal	27
4.3.5	Tokenizing	28
4.4	Labeling Data.....	30
4.5	TF-IDF	30
4.6	Klasifikasi	31
4.6.1	Naïve Bayes	31
4.6.2	Decision Tree.....	33
4.7	Uji Coba Sistem.....	39
4.8	Analisa Hasil.....	39
4.8.1	Hasil Prediksi Sentimen dengan Naïve Bayes	39
4.8.2	Hasil Prediksi Sentimen dengan Decision Tree.....	42
4.8.3	Perbandingan Naïve Bayes dan Decision Tree.....	46
BAB V	48
PENUTUP	48
5.1	Kesimpulan	48
5.2	Saran	48
DAFTAR PUSTAKA	49

Daftar Tabel

Tabel 2.1 Tabel Perbandingan Penelitian	5
Tabel 2.2 Contoh Case Folding	9
Tabel 2.3 Contoh Stemming	10
Tabel 2.4 Contoh Filtering	10
Tabel 2.5 Contoh Tokenizing	11
Tabel 4.1 Cleaning	25
Tabel 4.2 Case Folding	26
Tabel 4.3 Stemming	27
Tabel 4.4 Stopword Removal	28
Tabel 4.5 Tokenizing	29
Tabel 4.6 Labeling Data	30
Tabel 4.7 Hasil Prediksi Naïve Bayes	40
Tabel 4.8 Keterangan Tabel Confusion Matrix	40
Tabel 4.9 Hasil Prediksi Decision Tree	43
Tabel 4.10 Keterangan Tabel Confusion Matrix	43
Tabel 4.11 Perbandingan Hasil Evaluasi	46

Daftar Gambar

Gambar 2.1 Alur Analisis Klasifikasi Naïve Bayes	13
Gambar 2.2 Decision Tree.....	14
Gambar 3.1 Alur Penelitian	18
Gambar 3.2 Alur Preprocessing	20
Gambar 4.1 Rancangan Alur Sistem	22
Gambar 4.2 Alur Preprocessing	24
Gambar 4.3 Akurasi Naïve Bayes	40
Gambar 4.4 Precision Naïve Bayes	41
Gambar 4.5 Recall Naïve Bayes.....	42
Gambar 4.6 F1 Score Naïve Bayes	42
Gambar 4.7 Akurasi Decision Tree.....	44
Gambar 4.8 Precision Decision Tree.....	44
Gambar 4.9 Recall Decision Tree.....	45
Gambar 4.10 F1 Score Decision Tree.....	45

INTISARI

Perkembangan teknologi informasi dalam beberapa tahun terakhir sangat pesat dimana hal tersebut mendorong penggunaan internet dan pertukaran informasi. Beragam sosial media digunakan masyarakat untuk membagi opini mereka atau sekadar berinteraksi dengan orang lain. Analisis Sentimen adalah cabang dari Natural Language Processing (NLP) yang dapat menyaring dan mengategorikan opini masyarakat pada sosial media.

Penelitian ini memanfaatkan data dari twitter untuk membandingkan dua algoritma klasifikasi yaitu Naïve Bayes Classifier dan Decision Tree. Parameter perbandingan adalah nilai akurasi, presicion, recall, dan f1 score dengan bantuan Confusion Matrix.

Percobaan penelitian dilakukan dengan membagi data menjadi dua skenario dimana skenario 1 dengan 200 data dan skenario 2 dengan 800 data. Data masing-masing scenario kemudian dibagi lagi menjadi 70% data latih dan 30% data uji dari total 1000 data. Hasil dari percobaan menunjukkan bahwa Naïve Bayes memiliki akurasi tertinggi dengan 85% dibandingkan Decision Tree dengan 78% pada skenario kedua.

Kata-kunci: Analisis Sentimen, Naïve Bayes, Decision Tree, Twitter.

ABSTRACT

The development of information technology in recent year is really fast which pushing the usage of internet. People use social media to share their opinion or just as simple as interacting with each other. Sentiment analysis is a branch of Natural Language Processing (NLP) that can filter and categorizing people opinion on social media.

This study uses twitter data to compare two algorithm that is Naïve Bayes Classifier and Decision Tree. The comparison parameter is the value of accuracy, precision, recall, and f1 score with the help of Confusion Matrix.

This study divided the data into two scenarios where the first one with 200 data and the second scenario with 800 data. Data from each scenario then divided again into 70% training data and 30% test data from the total of 1000 data. The result show that Naïve Bayes Classifier have much higher accuracy with 85% compared to Decision Tree with 78% on the second scenario.

Keywords: *Sentiment Analysis, Naïve Bayes, Decision Tree, Twitter.*