

**PERBANDINGAN AKURASI ALGORITMA MULTINOMIAL,
BERNOULLI DAN GAUSSIAN NAÏVE BAYES UNTUK
SENTIMEN ANALISIS HACKER BJORKA PADA SOSIAL
MEDIA TWITTER**

SKRIPSI



Oleh :

BAGAS CAKRA WIRADANA

18081010100

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2023

**PERBANDINGAN AKURASI ALGORITMA
MULTINOMIAL, BERNOULLI DAN GAUSSIAN NAÏVE BAYES
UNTUK SENTIMEN ANALISIS HACKER BJORKA PADA
SOSIAL MEDIA TWITTER**

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh Gelar

Sarjana Komputer Program Studi Informatika



Oleh :

BAGAS CAKRA WIRADANA

18081010100

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2023

LEMBAR PENGESAHAN

Judul : PERBANDINGAN AKURASI ALGORITMA
MULTINOMIAL, BERNOULLI DAN GAUSSIAN NAÏVE
BAYES UNTUK SENTIMEN ANALISIS HACKER BJORKA
PADA SOSIAL MEDIA TWITTER.
Oleh : BAGAS CAKRA WIRADANA
NPM : 18081010100

Telah Diseminarkan Dalam Ujian Skripsi Pada :
Hari Kamis Tanggal 26 Januari 2023

Mengetahui

Dosen Pembimbing

Dosen Penguji

1.

1.

Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

Yisti Vita Via, S.ST, M.Kom

NIP. 19860425 2021212 001

2.

2.

Wahyu S. J. Saputra, S.Kom., M.Kom.

NIP. 19860825 2021211 003

Hendra Maulana, S.Kom, M.Kom

NPT. 201198 31 223248

Menyetujui

Dekan

Koordinator Program Studi

Fakultas Ilmu Komputer

Informatika

Dr. Novirina Hendrasarie, S.T, M.T

NIP. 19681126 199403 2 001

Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa program studi Informatika UPN “Veteran” Jawa Timur, yang bertanda tangan dibawah ini :

Nama : Bagas Cakra Wiradana

NPM : 18081010100

Menyatakan bahwa judul skripsi yang saya ajukan dan kerjakan yang berjudul :

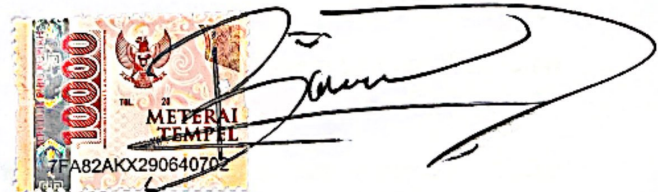
**“PERBANDINGAN AKURASI ALGORITMA MULTINOMIAL,
BERNOULLI, DAN GAUSSIAN NAÏVE BAYES UNTUK SENTIMEN
ANALISIS HACKER BJORKA PADA SOSIAL MEDIA TWITTER”**

Bukan merupakan plagiat dari skripsi/tugas akhir/penelitian orang lain dan juga bukan merupakan produk atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam daftar pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun di institusi pendidikan lainnya.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka saya siap menerima segala konsekuensinya.

Surabaya, 26 Januari 2023

Penulis,

A handwritten signature in black ink is written over a 1000 Rupiah postage stamp. The stamp features the Garuda Pancasila emblem and the text '1000', 'REPUBLIK INDONESIA', and 'MEPERAI TEMPEL'. The serial number '7FA82AKX290640702' is visible at the bottom of the stamp.

BAGAS CAKRA WIRADANA

NPM. 18081010100

PERBANDINGAN AKURASI MULTINOMIAL, BERNOULLI, DAN GAUSSIAN NAÏVE BAYES UNTUK SENTIMEN ANALISIS HACKER BJORKA PADA SOSIAL MEDIA TWITTER

Nama Mahasiswa : **Bagas Cakra Wiradana**

NPM : **18081010100**

Program Studi : **Informatika**

Dosen Pembimbing : **Fetty Tri Anggraeny, S.Kom, M.Kom**

**Wahyu Syaifullah Jauharis Saputra, S.Kom.,
M.Kom.**

ABSTRAK

Pada era teknologi yang semakin maju ini penggunaan sosial media telah digemari oleh semua kalangan, yang bertujuan untuk mendapatkan informasi, mengeluarkan hak suara dengan berpendapat, bersosialisasi dan lain-lain. Di sosial media banyak terdapat informasi yang mengandung opini, pendapat ataupun tanggapan, hal ini sering menyebabkan terjadinya ketimpangan antara informasi yang sedang dibicarakan dan informasi yang diinterpretasikan serta banyaknya penyebaran informasi yang tidak benar adanya sehingga tidak benar-benar bisa dipahami sentiment yang ingin diutarakan maupun disampaikan. Maka dari itu penelitian ini bertujuan untuk mengetahui akurasi dari percobaan pembelajaran mesin dengan menerapkan model Multinomial Naïve bayes dan Gaussian Naïve Bayes yang akan melakukan analisis terhadap sentiment mengenai tweet dengan topik Bjorka. Penelitian ini dilakukan dengan menerapkan kedua model Naïve Bayes tersebut dengan menggunakan kurang lebih 1500 data tweet. Data yang dimiliki dibagi dengan perbandingan 90% data training, dan 10% data testing. Percobaan menggunakan Multinomial Naïve Bayes menghasilkan akurasi sebesar 73%, sedangkan percobaan menggunakan Gaussian Naïve Bayes menghasilkan akurasi sebesar 55%.

Kata kunci: *Sentimen analisis, tweet, Multinomial Naïve Bayes, Gaussian Naïve Bayes*

KATA PENGANTAR

Dengan mengucapkan puji syukur kehadirat Allah SWT yang telah memberikan petunjuk, pertolongan serta seluruh nikmat-Nya kepada kita, dengan izin dan ridho-Nya lah penulis dapat menyelesaikan laporan skripsi ini yang berjudul “Perbandingan Akurasi Algoritma Multinomial, Bernoulli, dan Gaussian Naïve Bayes Untuk Sentimen Analisis Hacker Bjorka Pada Sosial Media Twitter”

Banyak dukungan dan bantuan yang didapatkan selama melakukan penelitian hingga akhirnya mampu menyelesaikan penulisan laporan skripsi ini. Dengan rasa hormat, penulis mengucapkan terima kasih kepada seluruh pihak terkait yang turut membantu dan terlibat dalam penyusunan laporan ini dari awal hingga akhir.

Penulis menyadari mungkin masih ada kekurangan pada laporan skripsi ini mengingat keterbatasan pengetahuan dan kemampuan. Oleh karena itu, penulis sangat mengharapkan masukan, kritik, dan saran yang dapat membangun untuk mengembangkan laporan ini.

Surabaya, 26 Januari 2023

Penyusun

UCAPAN TERIMA KASIH

Dalam penyusunan dan penulisan skripsi ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari beberapa pihak. Oleh karena itu dalam kesempatan ini penulis menyampaikan terima kasih kepada Allah SWT yang selalu ada dan memberi kelancaran, kemudahan, kekuatan dan sumber dari segala pencerahan.

Tidak lupa penulis berterima kasih kepada yang terhormat :

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Dr. Novirina Hendrasarie, S.T, M.T. Selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Fetty Tri Anggraeny, S.Kom, M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Fetty Tri Anggraeny, S.Kom, M.Kom. selaku dosen pembimbing I dan Bapak Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.. selaku dosen pembimbing II, yang telah sabar, ikhlas meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, motivasi, arahan, dan saran-saran yang sangat berharga kepada penulis selama penyusunan skripsi ini.
5. Staff dan dosen Teknik Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmu, arahan, dan pengalamannya selama perkuliahan.
6. Kedua orang tua penulis, Wahyu Setya Andhika, S.H, M.H dan Etty Hartiningsih, S.Sos. yang selalu memberi semangat, motivasi, dan mendoakan untuk kebaikan, keberhasilan maupun kesuksesan penulis.
7. Nadia Ristya Dewi, Helna Freecenta, dan Taufiqur Rohman, yang telah sabar berjuang bersama dalam suka maupun duka dengan saling menguatkan dan saling memberikan dukungan selama perkuliahan di Teknik Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
8. Teman-teman Teknik Informatika angkatan 2018, yang sudah bersedia

membantu, mengisi, dan menjadi bagian dalam cerita hidup dan perkuliahan penulis.

9. Semua pihak yang tidak dapat penulis ucapkan yang juga turut mendoakan dan menyemangati penulis dalam menyelesaikan skripsi ini.

Terima kasih atas segala bantuan, dukungan dan doa-doanya semoga Allah SWT memberikan balasan yang setimpal kepada semua pihak yang telah membantu penulis sehingga penulis mampu dan berhasil menyelesaikan penelitian dan proses penyusunan laporan penelitian skripsi dengan baik dan tepat.

Surabaya, 26 Januari 2023

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	i
ABSTRAK.....	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH.....	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah	3
1.4. Tujuan	4
1.5. Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Pendahulu.....	5
2.2 Analisis Sentimen	6
2.3 Twitter.....	6
2.4 Text Mining	7
2.5 Text Preprocessing	7
2.5.1 Case Folding	8
2.5.2 Tokenizing	8
2.5.3 Stopword Removal.....	8
2.5.4 Stemming	9
2.6 Ekstraksi Fitur.....	9
2.6.1 TF-IDF	10
2.7 Naïve Bayes Classifier.....	11
2.7.1 Multinomial Naïve Bayes	12
2.7.2 Bernoulli Naïve Bayes	13
2.7.3 Gaussian Naïve Bayes.....	13
2.8 Confusion Matrix	13
2.8.1 Accuracy	15
2.8.2 Precision.....	15

2.8.3	Recal	15
2.8.4	F1-Score.....	15
BAB III METODOLOGI.....		16
3.1	Tahapan pengerjaan tugas akhir.....	16
3.2	Studi Literatur	16
3.3	Analisa dan Desain	17
3.4	Akuisisi Data dan Labelling.....	19
3.5	Preprocessing dan Ekstraksi Fitur.....	23
3.6	Klasifikasi Data.....	31
3.7	Pengujian Sistem dengan Confusion Matrix.....	31
BAB IV HASIL DAN PEMBAHASAN		32
1.1	Lingkungan Penelitian	32
1.1.1	Spesifikasi Perangkat keras.....	32
1.1.2	Spesifikasi Perangkat Lunak.....	33
1.2	Proses pengambilan data.....	33
1.3	Preprocessing data	39
1.4	Ekstraksi Fitur.....	55
1.4.1	TF-IDF	55
1.5	Split train dan test	57
1.6	Proses Klasifikasi.....	59
1.7	Proses Pengujian	60
BAB V PENUTUPAN.....		67
5.1	Kesimpulan	67
5.2	Saran	67
DAFTAR PUSTAKA		69

DAFTAR TABEL

Tabel 2.1 Case Folding	8
Tabel 2.2 Tokenizing	8
Tabel 2.3 Stopword Removal	9
Tabel 2.4 Stemming	9
Tabel 2.5 Multiclass Confusion Matrix	14
Tabel 3.1 Tahapan Pengerjaan Tugas Akhir	16
Tabel 3.2 Labelling pada tweet	22
Tabel 4.1 Spesifikasi perangkat keras yang digunakan	32
Tabel 4.2 Spesifikasi perangkat lunak yang digunakan	33
Tabel 4.3 Confusion matrix Multinomial Naïve Bayes	63
Tabel 4.4 Confusion matrix Bernoulli Naïve Bayes	63
Tabel 4.5 Confusion matrix Gaussian Naïve Bayes	64
Tabel 4.6 Classification Report Multinomial Naïve Bayes	64
Tabel 4.7 Classification Report Bernoulli Naïve Bayes	65
Tabel 4.8 Classification Report Gaussian Naïve Bayes	65

DAFTAR GAMBAR

Gambar 2.1 Tahapan Text Preprocessing	8
Gambar 3.1 Flowchart Sistem.....	17
Gambar 3.2 Tahapan Akuisisi Data dan Labelling.....	19
Gambar 3.3 Grafik tren dari Google tren pencarian keyword “Bjorka”	20
Gambar 3.4 Crawling Tweet	21
Gambar 3.5 Labelling.....	23
Gambar 3.6 Pemisahan data text untuk tahap preprocessing	24
Gambar 3.7 Tahapan Preprocessing.....	24
Gambar 3.8 Tweet berdasarkan kelas	25
Gambar 3.9 Flowchart proses Case Folding.....	25
Gambar 3. 10 Flowchart alur proses Tokenizing.....	26
Gambar 3.11 Flowchart alur proses Stopword Removal	27
Gambar 3.12 Flowchart alur proses Stemming.....	28
Gambar 3.13 File csv berisi tweet setelah preprocessing	29
Gambar 3.14 Tahapan Ekstraksi fitur TF-IDF.....	30
Gambar 3.15 Tahapan klasifikasi.....	31
Gambar 4.1 API token twitter developer.....	34
Gambar 4.2 File csv hasil crawling data	36
Gambar 4.3 Proses pelabelan data tweet.....	37
Gambar 4.4 Jumlah data Sentimen yang dibagi tiga kelas.....	38
Gambar 4.5 Bar chart data sentiment.....	39
Gambar 4.6 Data yang akan diproses.....	41
Gambar 4.7 data setelah melalui proses Case Folding.....	43
Gambar 4.8 Data setelah melalui proses tokenizing	44
Gambar 4.9 Data setelah melalui proses Stopword removal	46
Gambar 4.10 Data setelah melalui proses Stemming	48
Gambar 4.11 Data hasil stemming.....	49
Gambar 4.12 Data setelah proses remove punctuation	50
Gambar 4.13 Data bersih dan sentimen	51
Gambar 4.14 file csv berisi data setelah preprocessing	52
Gambar 4. 15 Load data file preprocessing	54
Gambar 4.16 konversi ke label polaritas.....	55
Gambar 4.17 vektor TF-IDF dari data text	56
Gambar 4.18 numpy array integer data text	57
Gambar 4.19 Data yang digunakan untuk training	58
Gambar 4.20 Data yang digunakan untuk testing	58
Gambar 4.21 Classification Report Multinomial Naïve Bayes.....	61
Gambar 4.22 Classification Report Bernoulli Naïve Bayes.....	61
Gambar 4.23 Classification Report Gaussian Naïve Bayes	61
Gambar 4.24 Accuracy Multinomial Naïve Bayes	62
Gambar 4.25 Accuracy Bernoulli Naïve Bayes	62
Gambar 4.26 Accuracy Gaussian Naïve Bayes.....	62