

TUGAS AKHIR

PERBANDINGAN KINERJA ALGORITMA *MULTINOMIAL NAÏVE BAYES (MNB)*, *MULTIVARIATE BERNOULLI* DAN *ROCCHIO ALGORITHM* DALAM KLASIFIKASI KONTEN BERITA *HOAX* BERBAHASA INDONESIA PADA MEDIA SOSIAL



HAMDHAN ASHARI

1610651083

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH JEMBER

2020

TUGAS AKHIR

PERBANDINGAN KINERJA ALGORITMA *MULTINOMIAL NAÏVE BAYES (MNB)*, *MULTIVARIATE BERNOULLI* DAN *ROCCHIO ALGORITHM* DALAM KLASIFIKASI KONTEN BERITA *HOAX* BERBAHASA INDONESIA PADA MEDIA SOSIAL

**Disusun untuk Melengkapi Tugas dan Memenuhi Syarat Kelulusan
Program Strata I Jurusan Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Jember**



HAMDHAN ASHARI

1610651083

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2020

HALAMAN PERSETUJUAN

PERBANDINGAN KINERJA ALGORITMA *MULTINOMIAL NAÏVE BAYES (MNB)*, *MULTIVARIATE BERNOULLI* DAN *ROCCHIO ALGORITHM* DALAM KLASIFIKASI KONTEN BERITA *HOAX* BERBAHASA INDONESIA PADA MEDIA SOSIAL

Oleh:

Hamdhan Ashari

1610651083

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada Sidang Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh:

Pembimbing I



Deni Arifianto, M.Kom.

NIDN. 0718068103

Pembimbing II



Habibatul Azizah Al Faruq, M.Pd.

NIDN. 0718128901

HALAMAN PENGESAHAN

PERBANDINGAN KINERJA ALGORITMA *MULTINOMIAL NAÏVE BAYES (MNB)*, *MULTIVARIATE BERNOULLI* DAN *ROCCHIO ALGORITHM* DALAM KLASIFIKASI KONTEN BERITA *HOAX* BERBAHASA INDONESIA PADA MEDIA SOSIAL

Oleh:

Hamdhan Ashari

1610651083

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada Sidang Tugas Akhir tanggal 04 Juli 2020 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui Oleh,

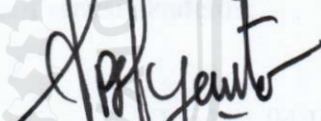
Penguji I



Agung Nilogiri, S.T., M.Kom.

NIDN. 0030037701

Pembimbing I



Deni Arifianto, M.Kom.

NIDN. 0718068103

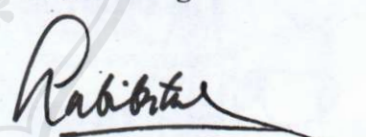
Penguji II



Ginanjar Abdurrahman, S.Si., M.Pd.

NIDN. 0714078704

Pembimbing II



Habibatul Azizah Al Faruq, M.Pd.

NIDN. 0718128901

Mengesahkan,

Dekan Fakultas Teknik

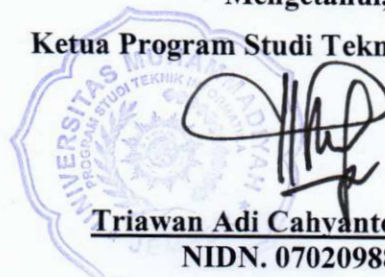


Nanang Saiful Rizal, S.T., M.T.

NIDN. 0705047806

Mengetahui,

Ketua Program Studi Teknik Informatika



Triawan Adi Cahyanto, M.Kom.

NIDN. 0702098804

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

NAMA : HAMDHAN ASHARI

NIM : 16 1065 1083

INSTITUSI : Strata-1 Program Studi Teknik Informatika Fakultas Teknik,
Universitas Muhammadiyah Jember.

Menyatakan bahwa Tugas Akhir yang berjudul “**Perbandingan Kinerja Algoritma *Multinomial Naïve Bayes (MNB)*, *Multivariate Bernoulli* Dan *Rocchio Algorithm* Dalam Klasifikasi Konten Berita *Hoax* Berbahasa Indonesia Pada Media Sosial**” bukan merupakan Tugas Akhir orang lain baik sebagian maupun keseluruhan kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak benar penulis bersedia mendapatkan sanksi akademik.

Jember, 04 Juli 2020

Yang Menyatakan,



Hamdhan Ashari

NIM. 16 1065 1083

UNGKAPAN TERIMA KASIH

Puji syukur kepada Allah S.W.T atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir ini. Atas segala upaya, bimbingan, dan arahan dari semua pihak, penulis mengucapkan terimakasih sebesar- besarnya kepada:

1. Allah SWT berkat segala ridho, rahmat dan hidayah-Nya penulis dapat menyelesaikan segala urusan dalam menyusun laporan Tugas Akhir dan diberikan kesempatan mendapatkan gelar Sarjana Komputer.
2. Bapak Nanang Saiful Rizal, S.T., M.T selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
3. Bapak Triawan Adi Cahyanto, M.Kom selaku Kepala Program Studi Teknik Informatika di Universitas Muhammadiyah Jember.
4. Bapak Deni Arifianto M.Kom selaku dosen pembimbing 1 tugas akhir yang telah memberikan bimbingan dari awal sampai akhir dengan penuh kesabaran sehingga tugas akhir ini dapat terselesaikan.
5. Ibu Habibatul Azizah Al Faruq M.Pd selaku dosen pembimbing 2 tugas akhir yang telah memberikan bimbingan dari awal sampai akhir dengan penuh kesabaran sehingga tugas akhir ini dapat terselesaikan.
6. Bapak Agung Nilogiri, S.T., M.Kom selaku dosen penguji 1 yang telah memberikan kritik, saran dan masukan yang membangun dalam penelitian ini.
7. Bapak Ginanjar Abdurrahman, S.Si., M.Pd selaku dosen penguji 2 yang telah memberikan kritik, saran dan masukan yang membangun dalam penelitian ini.
8. Kedua orang tua saya Imhan dan Wiwik Murtiwinarsih, serta kakak perempuan saya Heni Rahmawati S.P terimakasih atas *support* dan doanya dalam penyusunan laporan tugas akhir ini.
9. Teman-teman seperjuangan yang senasib dan setanah air, baik “HIMAJO, KELUARGA JARKOM, TELUR CLUB” terimakasih atas doa dan dukungannya selama ini, dan terimakasih atas kebersamaannya.

MOTTO

“Barang siapa keluar untuk mencari ilmu maka dia berada di jalan Allah SWT”

(HR. Turmudzi)

**“Waktu itu bagaikan pedang, jika kamu tidak memanfaatkannya untuk memotong,
maka ia akan memotongmu”**

(HR. Muslim)

**“Barang siapa yang memberi kemudharatan kepada seorang muslim, maka Allah
memberi kemudharatan kepadanya, barang siapa yang merepotkan (menyusahkan)
seorang muslim maka Allah akan menyusahkan dia”**

(HR. Abu Dawud)

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya”

(QS. Al-Baqarah: 286)

**“Barang siapa yang bersungguh-sungguh, sesungguhnya kesungguhan tersebut untuk
kebaikan dirinya sendiri”**

(QS. Al-Ankabut: 6)

KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT yang telah ridho, rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir dengan judul **“Perbandingan Kinerja Algoritma *Multinomial Naïve Bayes (MNB)*, *Multivariate Bernoulli* dan *Rocchio Algorithm* Dalam Klasifikasi Konten Berita *Hoax* Berbahasa Indonesia Pada Media Sosial”**. Adapun tujuan penulisan laporan tugas akhir ini adalah untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan Sarjana pada Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember.

Penulis mengucapkan banyak terima kasih kepada bapak dan ibu pembimbing, serta bapak penguji yang telah memberikan bimbingan, kritik dan saran yang membangun sehingga tugas akhir ini dapat terselesaikan.

Penulis menyadari bahwa penyusunan skripsi ini masih banyak kekurangan dan jauh dari kata sempurna maka dari itu penulis sangat mengharapkan kritik dan saran yang membangun demi kesempurnaan tugas akhir ini. Semoga tulisan ini dapat bermanfaat tidak saja untuk penulis tetapi bagi pembaca.

Jember, 04 Juli 2020

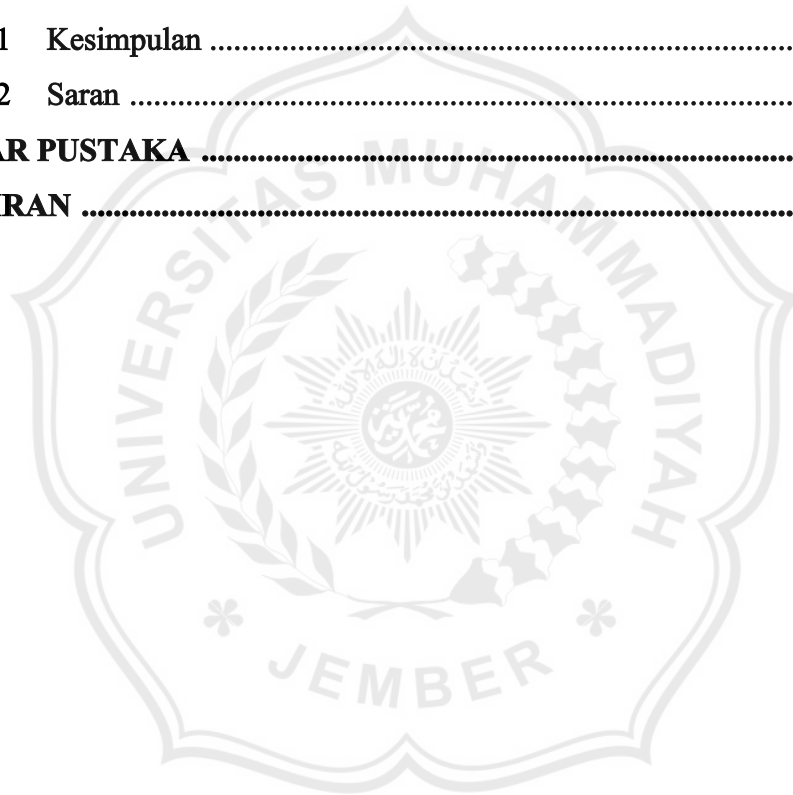
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
ABSTRAK	v
ABSTRACT	vi
UNGKAPAN TERIMA KASIH	vii
MOTTO	viii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR GRAFIK	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	6
2.1 Media Sosial	6
2.2 <i>Text Mining</i>	7
2.3 <i>Text Processing</i>	7
2.4 Pembobotan TF-IDF	8
2.5 <i>Data Mining</i>	9
2.6 <i>Multinomial Naïve Bayes</i>	10
2.7 <i>Rocchio Algorithm</i>	12
2.8 <i>Multivariate Bernoulli</i>	13
2.9 <i>Jupyter Notebook</i>	14

2.10	<i>K-Fold Cross Validation</i>	14
2.11	Penelitian Terdahulu	14
2.11.1	Rasywir dan Purwarianti, 2015	15
2.11.2	Afriza dan Adisantoso, 2018	15
2.11.3	Adhi, Saptono dan Anggrainingsih, 2017	15
2.11.4	Rahman, Wiranto dan Doewes, 2017	16
BAB III METODE PENELITIAN		17
3.1	Rancangan Penelitian	17
3.2	Pengumpulan Data	17
3.3	<i>Text Normalization</i>	20
3.4	<i>Text Processing</i>	21
3.5	Pembobotan	21
3.6	Klasifikasi	24
3.6.1	<i>Multinomial Naïve Bayes (MNB)</i>	24
3.6.2	<i>Rocchio Algorithm</i>	26
3.6.3	<i>Multivariate Bernoulli</i>	30
3.7	Validasi dan Evaluasi	33
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Pengumpulan Data	34
4.2	<i>Pre-processing Data</i>	35
4.2.1	<i>Case Folding</i>	35
4.2.2	<i>Tokenizing</i>	37
4.2.3	<i>Filtering</i>	39
4.2.4	<i>Stemming</i>	41
4.3	Pembobotan Kata	42
4.4	Hasil Klasifikasi	42
4.4.1	Hasil Klasifikasi <i>Multinomial Naïve Bayes (MNB), Bernoulli dan Rocchio</i> pada <i>Fold K = 2</i>	43
4.4.2	Hasil Klasifikasi <i>Multinomial Naïve Bayes (MNB), Bernoulli dan Rocchio</i> pada <i>Fold K = 4</i>	45

4.4.3 Hasil Klasifikasi <i>Multinomial Naïve Bayes (MNB)</i> , <i>Bernoulli</i> dan <i>Rocchio</i> pada <i>Fold K = 5</i>	48
4.4.4 Hasil Klasifikasi <i>Multinomial Naïve Bayes (MNB)</i> , <i>Bernoulli</i> dan <i>Rocchio</i> pada <i>Fold K = 8</i>	51
4.4.5 Hasil Klasifikasi <i>Multinomial Naïve Bayes (MNB)</i> , <i>Bernoulli</i> dan <i>Rocchio</i> pada <i>Fold K = 10</i>	54
4.5 Pengujian Menggunakan Validasi Data	62
BAB V SARAN DAN KESIMPULAN	63
5.1 Kesimpulan	63
5.2 Saran	64
DAFTAR PUSTAKA	65
LAMPIRAN	68



DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	17
Gambar 3.2 Halaman utama <i>turnbackhoax.id</i>	18
Gambar 3.3 Tampilan salah satu berita <i>hoax</i>	19
Gambar 3.4 Tampilan narasi dan Penjelasan Berita	20
Gambar 3.5 Proses <i>Pre-processing</i>	21



DAFTAR TABEL

Tabel 3.1	Contoh Data untuk Pembobotan	21
Tabel 3.2	Hasil Perhitungan <i>Term Frequency (TF)</i>	21
Tabel 3.3	Hasil Perhitungan <i>Inverse Document Frequency (IDF)</i>	22
Tabel 3.4	Hasil Perhitungan Pembobotan <i>TF-IDF</i>	23
Tabel 3.5	Nilai Probabilitas <i>term</i> dokumen uji	25
Tabel 3.6	Hasil Perhitungan nilai <i>centroid</i>	26
Tabel 3.7	Hasil Perhitungan vektor <i>centroid</i> \times vektor dokumen uji	27
Tabel 3.8	Hasil Perhitungan panjang vektor <i>centroid</i> pada seluruh <i>term</i>	28
Tabel 3.9	Hasil Perhitungan panjang vektor dokumen uji	29
Tabel 3.10	Nilai Probabilitas <i>term</i> dokumen uji	31
Tabel 3.11	Nilai satu di kurang probabilitas <i>term</i> dokumen uji	32
Tabel 4.1	Beberapa Data Konten Berita	34
Tabel 4.2	Proses <i>case folding</i>	35
Tabel 4.3	Proses <i>tokenizing</i>	37
Tabel 4.4	Proses <i>filtering</i>	39
Tabel 4.5	Proses <i>stemming</i>	41
Tabel 4.6	Hasil Pembobotan Data Konten Berita	42
Tabel 4.7	<i>Confusion Matrix</i> pada algoritma <i>MNB</i> $K=2$	43
Tabel 4.8	<i>Confusion Matrix</i> pada algoritma <i>Bernoulli</i> $K=2$	43
Tabel 4.9	<i>Confusion Matrix</i> pada algoritma <i>Rocchio</i> $K=2$	44
Tabel 4.10	<i>Confusion Matrix</i> pada algoritma <i>MNB</i> $K=4$	45
Tabel 4.11	<i>Confusion Matrix</i> pada algoritma <i>Bernoulli</i> $K=4$	46
Tabel 4.12	<i>Confusion Matrix</i> pada algoritma <i>Rocchio</i> $K=4$	47
Tabel 4.13	<i>Confusion Matrix</i> pada algoritma <i>MNB</i> $K=5$	48
Tabel 4.14	<i>Confusion Matrix</i> pada algoritma <i>Bernoulli</i> $K=5$	49
Tabel 4.15	<i>Confusion Matrix</i> pada algoritma <i>Rocchio</i> $K=5$	50
Tabel 4.16	<i>Confusion Matrix</i> pada algoritma <i>MNB</i> $K=8$	51
Tabel 4.17	<i>Confusion Matrix</i> pada algoritma <i>Bernoulli</i> $K=8$	52
Tabel 4.18	<i>Confusion Matrix</i> pada algoritma <i>Rocchio</i> $K=8$	53

Tabel 4.19	<i>Confusion Matrix</i> pada algoritma <i>MNB</i> $K=10$	54
Tabel 4.20	<i>Confusion Matrix</i> pada algoritma <i>Bernoulli</i> $K=10$	55
Tabel 4.21	<i>Confusion Matrix</i> pada algoritma <i>Rocchio</i> $K=10$	56
Tabel 4.22	Hasil Akurasi, Presisi dan <i>recall</i> keseluruhan pada <i>MNB</i>	58
Tabel 4.23	Hasil Akurasi, Presisi dan <i>recall</i> keseluruhan pada <i>Bernoulli</i>	59
Tabel 4.24	Hasil Akurasi, Presisi dan <i>recall</i> keseluruhan pada <i>Rocchio</i>	60
Tabel 4.25	Hasil Prediksi Data Baru	62



DAFTAR GRAFIK

Grafik 4.1 Hasil Akurasi, Presisi dan <i>Recall</i> pada <i>Fold K</i> = 2	45
Grafik 4.2 Hasil Akurasi, Presisi dan <i>Recall</i> pada <i>Fold K</i> = 4	48
Grafik 4.3 Hasil Akurasi, Presisi dan <i>Recall</i> pada <i>Fold K</i> = 5	51
Grafik 4.4 Hasil Akurasi, Presisi dan <i>Recall</i> pada <i>Fold K</i> = 8	54
Grafik 4.1 Hasil Akurasi, Presisi dan <i>Recall</i> pada <i>Fold K</i> = 10	57

