

SKRIPSI

**STUDI KOMPARATIF ALGORITMA *NAÏVE BAYES CLASSIFIER* DAN *K-NEAREST NEIGHBOR* UNTUK ANALISIS SENTIMEN OPINI PUBLIK DI *TWITTER* MENGENAI KEMENANGAN PASLON 02
(Studi Pada Real Count Pemilu 2024)**



Oleh :

Alfira Fitri Nur Azizah

NIM: 20083000097

**PROGRAM S1 SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA
MALANG
2024**

**STUDI KOMPARATIF ALGORITMA *NAÏVE BAYES*
CLASSIFIER DAN *K-NEAREST NEIGHBOR* UNTUK
ANALISIS SENTIMEN OPINI PUBLIK DI *TWITTER*
MENGENAI KEMENANGAN PASLON 02
(Studi Pada Real Count Pemilu 2024)**



SKRIPSI

Diajukan Sebagai Salah Satu Persyaratan
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :

Alfira Fitri Nur Azizah

NIM: 20083000097

**PROGRAM S1 SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA
MALANG
2024**

MOTTO

“Maka sesungguhnya bersama kesulitan pasti ada kemudahan. Maka apabila engkau telah selesai (dari suatu urusan), tetaplah bekerja keras (untuk urusan yang lain). Dan hanya kepada Tuhanmulah engkau berharap.”

(Q.S Al-Insyirah : 6-7)

“Selalu ada harga dalam setiap proses. Nikmati saja lelah-lelah itu. Lebarkan lagi rasa sabar itu. Semua yang kau investasikan untuk menjadikan dirimu serupa yang kau impikan, mungkin tidak akan selalu berjalan lancar. Tapi gelombang-gelombang itu yang bisa kau ceritakan.”

(Boy Candra)

“Orang tua dirumah menanti kepulanganmu dengan hasil yang membanggakan, jangan kecewakan mereka. Simpan keluhmu, sebab letihmu tak sebanding dengan perjuangan mereka menghidupimu.”

(Ika df)

LEMBAR PERSETUJUAN SKRIPSI

Nama : Alfira Fitri Nur Azizah
NIM : 20083000097
Program Studi : Sistem Informasi
Bidang Kajian : Data Mining
Judul : Studi Komparatif Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor Untuk Analisis Sentimen Opini Publik di Twitter Mengenai Kemenangan Paslon 02 (Studi Pada Real Count Pemilu 2024)

Malang, 9 Juli 2024

DISETUJUI DAN DITERIMA

Ketua Program Studi
S1 Sistem Informasi



Luthfi Irdana, S.Pd., M.Pd.
NIDN. 0711059203

Dosen Pembimbing



Viry Puspaning Ramadhan, A.Md.Li., S.Kom., M.Kom.
NIDN. 0725019602

Dekan

Fakultas Teknologi Informasi



Dr. Mardiana Ardarwati, SE., M.Si.
NIDN. 0716037601

LEMBAR PENGESAHAN SKRIPSI

Dipersiapkan dan disusun oleh

Nama : Alfira Fitri Nur Azizah
NIM : 20083000097
Judul : Studi Komparatif Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor Untuk Analisis Sentimen Opini Publik di Twitter Mengenai Kemenangan Paslon 02 (Studi Pada Real Count Pemilu 2024)

Telah dipertahankan di depan Dosen Penguji pada,

Hari : Rabu
Tanggal : 17 Juli 2024
Tempat : Fakultas Teknologi Informasi
Universitas Merdeka Malang

Susunan Penguji

Ketua Penguji



Rahmatina Hidayati, S.Kom., MT.
NIDN. 0720028902

Sekretaris Penguji



Viry Puspaning Ramadhan, A.Md.Li., S.Kom., M.Kom.
NIDN. 0725019602

Anggota Penguji



Firnanda Al-Islama Achyunda Putra, S.Kom., M.Kom.
NIDN. 0706129401

Skrripsi ini Telah Diterima Sebagai Salah Satu Persyaratan untuk
Memperoleh Gelar Sarjana Sistem Informasi

Malang, 29 Juli 2024

Dekan Fakultas Teknologi Informasi
Universitas Merdeka Malang



Dr. Mardiana Andarwati, SE., M.Si.
NIDN. 0716037601

SURAT PERNYATAAN ORISINALITAS SKRIPSI

Saya yang bertandatangan di bawah ini:

Nama : Alfira Fitri Nur Azizah
NPM : 20083000097
Program Studi : S1 Sistem Informasi
Judul : Studi Komparatif Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor Untuk Analisis Sentimen Opini Publik di Twitter Mengenai Kemenangan Paslon 02 (Studi Pada Real Count Pemilu 2024)
Fakultas : Teknologi Informasi
Perguruan Tinggi : Universitas Merdeka Malang

Menyatakan dengan sesungguhnya bahwa skripsi saya yang berjudul :

“Studi Komparatif Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor Untuk Analisis Sentimen Opini Publik di Twitter Mengenai Kemenangan Paslon 02 (Studi Pada Real Count Pemilu 2024)”

adalah benar-benar hasil karya saya dan bukan karya orang lain (plagiasi/jiplakan) serta tidak didasarkan pada data palsu, baik sebagian maupun keseluruhan kecuali dalam bentuk kutipan daftar pustaka.

Demikian pernyataan ini saya buat dengan sebenar-benarnya dan apabila ternyata di dalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiat, maka saya bersedia mendapatkan sanksi akademik berupa pencabutan gelar akademik sesuai dengan peraturan perundang-undangan yang berlaku.

Malang, 2 Agustus 2024

Yang membuat pernyataan,



Alfira Fitri Nur Azizah

ABSTRAK

Pemilu Presiden dan Wakil Presiden tahun 2024 menjadi peristiwa politik paling dinanti oleh masyarakat Indonesia. Hasil perhitungan suara atau *real count* Pemilu 2024 telah menimbulkan beragam reaksi baik pro maupun kontra di berbagai *platform* sosial media khususnya *Twitter*, karena keunggulan paslon nomor urut 02. Penelitian ini memanfaatkan *Twitter* sebagai sumber data untuk interpretasi opini. Metode *Naïve Bayes* serta *K-NN* dipilih dalam studi ini yang kemudian diuji dan dibandingkan. Hasil penelitian menyajikan hasil *Naïve Bayes* dengan tingkat ketepatan sebesar 87.35% +/- 1.81% (micro average: 87.35%), sedangkan besaran akurasi yang didapatkan algoritma *K-NN* sebesar 69.68% +/- 3.14% (micro average: 69.68%) dengan perbandingan split data 90:10. Hasil analisis tersebut menunjukkan bahwa *Naïve Bayes* lebih unggul dari *K-Nearest Neighbor*.

Kata Kunci : Analisis Sentimen, *Naïve Bayes*, *K-NN*, Komparasi

ABSTRACT

The 2024 Presidential and Vice Presidential Election stands out as a highly awaited political event by the people of Indonesia. The vote counts result, or real count, of the 2024 election have sparked a variety of reactions, both supportive and opposing, especially on social media platforms like Twitter, due to the lead of candidate pair number 02. This study utilizes Twitter as a data source for opinion interpretation. The Naïve Bayes and K-NN were chosen in this study, and their performances are tested and compared. The research results present Naïve Bayes with an accuracy rate of 87.35% +/- 1.81% (micro average: 87.35%), while K-NN algorithm achieved an accuracy rate of 69.68% +/- 3.14% (micro average: 69.68%) using a data partition ratio of 90:10. The analysis results indicate that Naïve Bayes is more effective than K-Nearest Neighbor.

Keywords : Sentiment Analysis, Naïve Bayes, K-NN, Comparison

LEMBAR PERSEMBAHAN

Dengan segala kerendahan hati dan rasa syukur yang mendalam, skripsi ini saya persembahkan kepada :

1. Allah SWT sebagai bentuk rasa syukur saya karena telah memberikan nikmat karunia pertolongan yang tiada henti hingga saat ini.
2. Kedua orang tua yang dengan tulus serta penuh keikhlasan, mencurahkan segala kasih sayang dan cintanya, serta yang senantiasa mendoakan dan memberikan semangat juga dukungan dengan sepenuh hati.
3. Tak lupa dipersembahkan kepada diri sendiri, sebuah persembahan istimewa. Terimakasih telah bertahan sejauh ini dan tidak pernah berhenti berusaha serta berdoa untuk menyelesaikan skripsi ini.

UCAPAN TERIMAKASIH

Penyusunan skripsi ini tidak terlepas dari dukungan dan semangat dari berbagai pihak baik yang terlibat langsung maupun tidak langsung dalam penulisan skripsi ini, yang telah memberikan doa, dukungan, bimbingan, dan bantuan baik moril maupun materil kepada penulis. Untuk itu penulis ingin mengucapkan rasa hormat dan terimakasih sebesar-besarnya kepada :

1. Universitas Merdeka Malang, almamater tercinta yang telah memberikan saya kesempatan untuk menimba ilmu dan berkembang selama masa studi. Terimakasih atas dukungan fasilitas serta lingkungan akademik yang kondusif.
2. Ibu Dosen Pembimbing yang telah memberikan bimbingan, arahan, serta dukungan selama proses penyusunan skripsi ini.
3. Bapak dan Ibu Dosen serta Karyawan Program Studi S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang. Terimakasih atas ilmu, pengalaman, dan bantuan yang telah diberikan selama masa studi saya di kampus ini.
4. Kedua orang tua tercinta, Bapak dan Mama, yang selalu menjadi sumber inspirasi, motivasi, dan kekuatan dengan doa-doa dan dukungan tiada henti.
5. Adik-adikku tersayang, yang selalu memberikan semangat dan keceriaan.
6. Sahabat-sahabat tercinta, yang selalu ada disaat suka dan duka. Memberikan dukungan, tawa, dan kebersamaan yang sangat berarti.
7. Teman-teman di kampus, terimakasih atas kebersamaan, dukungan, dan kerjasama selama masa studi hingga penyusunan skripsi ini.
8. Semua pihak yang telah membantu, yang tidak bisa saya sebutkan satu per satu. Terimakasih atas kontribusi, bantuan, serta doa-doanya.
9. Untuk diri saya sendiri. Terimakasih kepada diri saya sendiri Alfira Fitri Nur Azizah. Saya bangga pada diri saya sendiri, mari bekerjasama untuk lebih berkembang lagi menjadi pribadi yang lebih baik dari hari ke hari.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi dengan judul: **“Studi Komparatif Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor Untuk Analisis Sentimen Opini Publik di Twitter Mengenai Kemenangan Paslon 02 (Studi Pada Real Count Pemilu 2024)”**. Penulis menyadari bahwa skripsi ini tidak akan terwujud tanpa bantuan serta dukungan dari beberapa pihak, sehingga pada kesempatan ini dengan segala hormat penulis menyampaikan rasa terimakasih yang tidak terhingga kepada orang-orang yang telah memberikan bantuan maupun masukan sehingga dapat terselesaikannya skripsi ini. Oleh karena itu, pada kesempatan ini penulis menyampaikan terimakasih kepada :

1. Ibu Dr. Mardiana Andarwati, S.E., M.Si. selaku Dekan Fakultas Teknologi Informasi Universitas Merdeka Malang.
2. Ibu Luthfi Indana, S.Pd., M.Pd. selaku Ketua Program Studi S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
3. Ibu Viry Puspaning Ramadhan, A.Md,Li., S.Kom., M.Kom. selaku dosen pembimbing yang dengan sangat luar biasa telah memberi bimbingan, arahan, serta motivasi dalam menyelesaikan skripsi ini mulai penyusunan proposal hingga selesai.
4. Bapak dan Ibu Dosen Program Studi S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
5. Bapak dan Ibu Karyawan S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
6. Kepada kepala keluarga saya, Bapak tercinta M. Agung Prihatin, yang dengan kasih sayang dan doa selalu menyertai setiap langkahku. Kepada Mama Sri Wahyuningsih tersayang, yang selalu memberikan cinta, dukungan, dan semangat yang tak pernah pudar. Serta kedua adikku tercinta Aulia Dewi Puspitasari dan M. Ikhsan Abdillah yang selalu memberi keceriaan dan semangat baru setiap harinya.

7. Tak lupa, terimakasih yang mendalam penulis sampaikan kepada sahabat-sahabat yang tersayang dan tercinta Amirhotul Imsiyyah, Agung Nur Dwianto, dan Achmad Wachizzudin. Terutama kepada Amirhotul Imsiyyah yang selalu mendukung dan menyemangati penulis tanpa henti serta menjadi pendengar yang siap mendengarkan segala keluh kesah penulis tanpa menghakimi. Kehadiran kalian memberikan semangat dan kekuatan yang luar biasa sepanjang penulisan skripsi ini.
8. Segenap teman baik di kelas 'E' dan 'Data 2' Sistem Informasi angkatan 2024 yang telah berperan banyak memberikan pengalaman dan pembelajaran selama di bangku kuliah. Terimakasih telah membuat kehidupan perkuliahan terasa begitu cepat dan penuh kebahagiaan, *see you on top guys*.
9. Seluruh pihak yang berperan secara langsung maupun tidak langsung yang tidak dapat disebutkan satu per satu.
10. *Last but not least*, penulis ucapkan terimakasih serta apresiasi besar kepada pemilik NIM 20083000097 atas nama diri sendiri Alfira Fitri Nur Azizah. Terimakasih atas tanggung jawab yang luar biasa dalam menyelesaikan skripsi ini hingga tuntas.

Akhirnya, dengan segala kerendahan hati peneliti menyadari keterbatasan penelitian ini karena keterbatasan peneliti. Oleh karena itu, peneliti mengharapkan kritik dan saran demi kesempurnaan penelitian ini. Semoga karya yang sederhana ini dapat memberikan manfaat bagi para pembaca dan umumnya bagi mahasiswa Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.

Malang, 2 Agustus 2024

Penulis



Alfira Fitri Nur Azizah

DAFTAR ISI

MOTTO.....	ii
LEMBAR PERSETUJUAN SKRIPSI.....	iii
LEMBAR PENGESAHAN SKRIPSI.....	iv
SURAT PERNYATAAN ORISINALITAS SKRIPSI	v
ABSTRAK	vi
ABSTRACT.....	vii
LEMBAR PERSEMBAHAN	viii
UCAPAN TERIMAKASIH.....	ix
KATA PENGANTAR	x
DAFTAR ISI.....	xii
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
DAFTAR RIWAYAT HIDUP	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	3
1.5 Batasan Penelitian.....	3
BAB II KAJIAN PUSTAKA	4
2.1 Landasan Teori.....	4
2.2 Penelitian Terdahulu.....	10
2.3 Hipotesis	13
BAB III METODE PENELITIAN	14
3.1 Desain Penelitian	14
3.2 Lokasi Penelitian.....	16
3.3 Populasi dan Sampel.....	16
3.4 Teknik Pengumpulan Data.....	17
3.5 Teknik Analisis Data	19
BAB IV HASIL DAN PEMBAHASAN	22

4.1 Hasil Penelitian	22
4.2 Pembahasan	43
BAB V KESIMPULAN DAN SARAN	45
5.1 Kesimpulan.....	45
5.2 Saran.....	45
DAFTAR PUSTAKA.....	46
LAMPIRAN.....	49

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	10
Tabel 4.1 Hasil Data Cleaning	26
Tabel 4.2 Hasil Case Folding	27
Tabel 4.3 Hasil Stopword Removal	30

DAFTAR GAMBAR

Gambar 1. Desain penelitian	14
Gambar 2. Proses Pengumpulan Data	17
Gambar 3. Alur Metode Naïve Bayes Classifier	19
Gambar 4. Alur Metode K-Nearest Neighbor	20
Gambar 5. Import Library	22
Gambar 6. Proses Crawling Data Twitter	23
Gambar 7. Menyimpan Hasil Crawling Data Tweet	23
Gambar 8. Hasil Crawling Data Tweet	24
Gambar 9. Hasil Pelabelan Manual	24
Gambar 10. Proses Preprocessing Data	25
Gambar 11. Proses Data Cleaning	25
Gambar 12. Tampilan Hasil Case Folding Pada Rapidminer	27
Gambar 13. Hasil Tokenizing Pada Rapidminer	28
Gambar 14. Parameter Filter Tokens by Length	29
Gambar 15. Parameter Filter Stopwords (Dictionary)	29
Gambar 16. Proses Utama Analisis Sentimen	30
Gambar 17. Split Dataset	31
Gambar 18. Jumlah Dataset Awal	32
Gambar 19. Jumlah Dataset Setelah SMOTE	32
Gambar 20. Proses Uji Metode Naïve Bayes Classifier	33
Gambar 21. Proses Uji Metode K-NN	33
Gambar 22. Confusion Matrix Naïve Bayes	34
Gambar 23. Hasil Perbandingan 80:20	35
Gambar 24. Hasil Perbandingan 70:30	36
Gambar 25. Hasil Perbandingan 60:40	36
Gambar 26. Confusion Matrix K-NN	37
Gambar 27. Hasil Perbandingan 80:20	38
Gambar 28. Hasil Perbandingan 70:30	39
Gambar 29. Hasil Perbandingan 60:40	39
Gambar 30. Hasil <i>Naïve Bayes</i> Bulan Februari	40

Gambar 31. Hasil <i>Naïve Bayes</i> Bulan Maret	40
Gambar 32. Hasil <i>Naïve Bayes</i> Bulan April.....	40
Gambar 33. Hasil <i>K-Nearest Neighbor</i> Bulan Februari.....	42
Gambar 34. Hasil <i>K-Nearest Neighbor</i> Bulan Maret.....	42
Gambar 35. Hasil <i>K-Nearest Neighbor</i> Bulan April.....	42

DAFTAR LAMPIRAN

Lampiran 1 Output Rapidminer	49
Lampiran 2 Form Revisi Seminar Hasil.....	51
Lampiran 3 Keterangan Pemeriksaan Kemiripan	53
Lampiran 4 Hasil Cek Turnitin	54