

**Perbandingan Algoritma Naïve Bayes Classifier Dan K-Nearest
Neighbors Untuk Analisis Sentimen Covid-19 Di Twitter**



OLEH

HABIBI AULIA NUR SYIFA

NPM : 18.1.03.03.0012

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PGRI KEDIRI
2023**

**Perbandingan Algoritma Naïve Bayes Classifier Dan K-Nearest
Neighbors Untuk Analisis Sentimen Covid-19 Di Twitter**

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer (S.Kom)



OLEH

HABIBI AULIA NUR SYIFA

NPM : 18.1.03.03.0012

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PGRI KEDIRI
2023**

HALAMAN PERSETUJUAN

Skripsi oleh:

HABIBI AULIA NUR SYIFA

NPM: 18.1.03.03.0012

Judul:

**Perbandingan Algoritma Naïve Bayes Classifier Dan K-Nearest
Neighbors Untuk Analisis Sentimen Covid-19 Di Twitter**

Telah disetujui untuk diajukan kepada

Panitia ujian/sidang skripsi Program

Studi Sistem Informasi

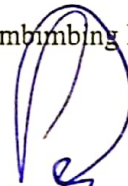
Tanggal: 23 Januari 2023

Pembimbing I



Rina Firliana, M.Kom
NIDN. 0731087703

Pembimbing II



Arie Nugroho, S.Kom, MM.
NIDN. 0712108103

HALAMAN PENGESAHAN

Skripsi oleh:

HABIBI AULIA NUR SYIFA

NPM : 18.1.03.03.0012

Judul:

**Perbandingan Algoritma Naïve Bayes Classifier Dan K-Nearest Neighbors
Untuk Analisis Sentimen Covid-19 Di Twitter**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Program Studi Sistem Informasi FT UN PGRI Kediri

Pada tanggal: 23 Januari 2023

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji:

- | | |
|---------------|------------------------------|
| 1. Ketua | : Rina Firliana, M. Kom |
| 2. Penguji I | : Anita Sari Wardani, M. Kom |
| 3. Penguji II | : Arie Nugroho, S. Kom, MM |



Mengetahui,

Dekan Fakultas Teknik



Dr. Suryo Widodo, M.Pd

NIP. 19640202 199103 1002

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Habibi Aulia Nur Syifa
Jenis Kelamin : Laki-Laki
Tempat/Tgl.Lahir : Nganjuk/ 13 Januari 2000
NPM : 18.1.03.03.0012
Fak./Jur./Prodi. : Teknik/Sistem Informasi

Menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 23 Januari 2023

Yang Menyatakan



Habibi Aulia Nur Syifa
NPM. 18.1.03.03.0012

HALAMAN MOTO DAN PERSEMBAHAN

MOTO

Ikuti Kata Hati

Persembahan:

Dengan mengucap syukur alhamdulillah, kupersembahkan hasil karya kerja kerasku untuk orang-orang yang aku sayangi:

1. Ibu dan Ayah yang tercinta yang menjadi alasan aku hidup.
2. Kakakku yang ada di surga.
3. Sahabat-sahabatku yang selalu ada saat aku membutuhkan.

ABSTRAK

Habibi Aulia Nur Syifa : Perbandingan Algoritma *Naïve Bayes Classifier* Dan *K-Nearest Neighbors* Untuk Analisis Sentimen Covid-19 Di Twitter, Skripsi, Sistem Informasi, Fakultas Teknik UNP Kediri 2023.

COVID-19 ialah infeksi gangguan pernapasan yang dapat menular dengan cepat. Penyakit ini muncul di Tiongkok pada bulan November 2019. Di negara Indonesia sendiri pada tanggal 10 April 2020 ditemukan sebanyak lebih dari 3000 kasus positif mengidap *COVID-19* dengan tingkat kematian sebesar 9.1%. Belum tuntasnya upaya pemutusan rantai penyebaran *COVID-19* oleh pemerintah membuat kalangan masyarakat menjadi resah, was-was dan ketakutan terhadap adanya pandemi ini. Banyak masyarakat yang ingin meluapkan aspirasinya di sosial media yang dirasa cocok sebagai tempat yang mewakili aspirasi masyarakat akan pandemi *COVID-19*. Salah satu media sosialnya yaitu *twitter*. Terdapat banyak sekali pesan teks yang dikirimkan ada pesan teks yang positif dan ada juga pesan teks yang negatif sehingga sulit untuk mengambil informasi yang selaras didalamnya karena keberagaman pesan teks yang dikirim. Salah satu cara untuk mengatasi hal tersebut ialah menggunakan analisis sentimen (pendapat). Pada penelitian ini terdapat berbagai proses antara lain pengumpulan data, preprocessing teks, pembobotan kata, Klasifikasi sentimen menggunakan algoritma *K-Nearest Neighbors* (KNN) dan *Naïve Bayes Classifiers* (NBC). Hasil yang diperoleh pada penelitian ini algoritma KNN mendapat akurasi sebesar 72.37% sedangkan algoritma NBC mendapat akurasi sebesar 67.84%. KNN menjadi algoritma klasifikasi yang paling baik untuk mengklasifikasikan sentimen (pendapat) negatif, Dimana label negatif yang diprediksi benar (TN) pada KNN lebih besar yaitu 393 dibanding NBC yang sebanyak 339. Sedangkan NBC menjadi algoritma yang paling baik untuk mengklasifikasikan sentimen (pendapat) positif, dimana label positif yang diprediksi benar (TP) pada NBC lebih besar yaitu 275 dibanding KNN yang sebanyak 262.

Kata Kunci : *COVID-19, Twitter, Analisa Sentimen, K-Nearest Neighbors, Naïve Bayes Classifiers*

KATA PENGANTAR


Puji Syukur Kami panjat kan kehadiran Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan- Nya tugas penyusunan skripsi ini dapat diselesaikan.

Skripsi dengan judul **“Perbandingan Algoritma *Naïve Bayes Classifier* Dan *K-Nearest Neighbors* Untuk Analisis Sentimen Covid-19 Di Twitter”** ini ditulis sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi UN PGRI Kediri.

Pada kesempatan ini diucapkan terimakasih dan penghargaan yang setulus-tulusnya kepada:

1. Kedua orang tua tercinta dan keluarga yang selalu mendukung penulis baik doa dan usaha selama kuliah.
2. Teman-teman yang selalu ada untuk membantu penulis saat penyusunan skripsi.
3. Bapak Arie Nugroho, S. Kom., MM dan ibu Rina Firliana, M.Kom selaku Dosen pembimbing skripsi yang selalu berusaha untuk menyempatkan waktu serta pikiran saat penulis membutuhkan bimbingan skripsi.
4. Bapak/Ibu Penguji yang telah memberikan penulis saran dan masukan untuk perbaikan dan penyelesaian skripsi ini
5. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
6. Dr. Suryo Widodo, M.Pd. selaku Dekan Fakultas Teknik Universitas Nusantara PGRI Kediri.
7. Rina Firliana, M.Kom. selaku Ketua Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri.

Kediri, 13 Januari 2023


Habibi Aulia Nur Syifa
NPM.18.1.03.03.0012

DAFTAR ISI

HALAMAN PERSETUJUAN.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN	iii
HALAMAN MOTO DAN PERSEMBAHAN	iv
ABSTRAK	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang dan Permasalahan.....	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	6
1.6. Sistematika Penulisan.....	7
BAB II LANDASAN TEORI	8
2.1 Kajian Teori.....	8
2.1.1 Text Mining	8
2.1.2 Sentimen Analysis	8
2.1.3 Preprocessing	9
2.1.4 Confusion Matrix.....	9
2.1.5 K-Fold Cross Validation.....	11
2.1.6 K-Nearest Neighbors	12
2.1.7 Eucludian Distance	12
2.1.8 Rapidminer.....	13
2.1.9 Term Frequency-Inverse Document Frequency	13
2.1.10 Naïve Bayes Classifiers	13
2.2 Kajian Penelitian Terdahulu	14

BAB III METODE PENELITIAN.....	17
3.1. Metode Penelitian.....	17
3.2. Metode Pengumpulan Data	17
3.3. Metode Analisis Data	17
3.3.1. Alur Penelitian	18
BAB IV HASIL DAN PEMBAHASAN	22
4.1. Pengumpulan Data	22
4.2. Pre-Proccesing	24
4.2.1. <i>Cleaning</i>	25
4.2.2. <i>Case Folding</i>	28
4.2.3. <i>Tokenize</i>	29
4.2.4. <i>Normalization</i>	30
4.2.5. <i>Stopword</i>	33
4.2.6. <i>Stemming</i>	35
4.3. Pembobotan Kata	37
4.4. Implementasi Algoritma.....	40
4.4.1. Implementasi Algoritma Naïve Bayes Classifier.....	40
4.4.2. Implementasi Algoritma KNN.....	42
4.5. Validasi.....	44
4.6. Evaluasi	45
4.6.1. Evaluasi Confusion Matrix Naïve Bayes	45
4.6.2. Evaluasi Confusion Matrix KNN.....	46
4.7. Perbandingan Hasil.....	47
BAB V PENUTUP.....	48
5.1. Kesimpulan.....	48
5.2. Saran	48
DAFTAR PUSTAKA	49
LAMPIRAN.....	52

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	10
Tabel 2. 2 Contoh Skenario Cross Validation.....	11
Tabel 2. 3 Kajian Penelitian Terdahulu.....	15
Tabel 2. 4 Lanjutan	16
Tabel 2. 5 Lanjutan	29
Tabel 4. 1 Contoh Datasets	23
Tabel 4. 2 Lanjutan	24
Tabel 4. 3 Contoh Hasil Filter URL.....	26
Tabel 4. 4 Contoh Hasil Filter Hashtag.....	27
Tabel 4. 5 Contoh Hasil Filter Username.....	27
Tabel 4. 6 Contoh Hasil Case Folding	28
Tabel 4. 7 Contoh Hasil Tokenize.....	30
Tabel 4. 8 Contoh Hasil Normalisasi	31
Tabel 4. 9 Contoh Hasil Stopword.....	34
Tabel 4. 10 Contoh Hasil Stemming.....	35
Tabel 4. 11 Contoh Perhitungan Manual TF-IDF Sederhana	38
Tabel 4. 12 Lanjutan	39
Tabel 4. 13 Contoh dataset perhitungan NBC	41
Tabel 4. 14 Perbandingan Hasil Confusion Matrix.....	47

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	18
Gambar 4. 1 Proporsi Datasets.....	22
Gambar 4. 2 Persiapan Data.....	23
Gambar 4. 3 Pengaturan Cleaning	25
Gambar 4. 4 Pengaturan Case Folding.....	28
Gambar 4. 5 Pengaturan Tokenize	29
Gambar 4. 6 Contoh Kamus Normalisasi	31
Gambar 4. 7 Contoh Pengaturan Proses Pertama	32
Gambar 4. 8 Rangkaian Proses Pertama	33
Gambar 4. 9 Proses Konversi Atribut Pertama	33
Gambar 4. 10 Contoh Kamus Stopword	35
Gambar 4. 11 Contoh Kamus Stemming	36
Gambar 4. 12 Rangkaian Proses Remove Duplicate	36
Gambar 4. 13 Konversi Atribut Kedua	37
Gambar 4. 14 Contoh Kata Yang Sering Muncul	38
Gambar 4. 15 Rangkaian Proses NBC	41
Gambar 4. 16 Rangkaian Proses KNN.....	43
Gambar 4. 17 Pengaturan Parameter KNN	44
Gambar 4. 18 Rangkaian Cross Validation.....	45
Gambar 4. 19 Hasil Confusion Matrix NBC.....	45
Gambar 4. 20 Hasil Confusion Matrix KNN	46

DAFTAR LAMPIRAN

Lampiran 1 Kajian Penelitian Terdahulu	52
Lampiran 2 Perhitungan Jumlah Kemunculan Kata Tiap Sentimen	54
Lampiran 3 Perhitungan bobot probabilitas sentimen untuk data latih	56
Lampiran 4 Perhitungan Vmap untuk data uji dan Penentuan kelas data D6.....	57
Lampiran 5 Dataset Penelitian	58
Lampiran 6 Isi Kamus Normalisasi.....	90
Lampiran 7 Isi Kamus Stopword	94
Lampiran 8 Isi Kamus Stemming	96

BAB I

PENDAHULUAN

1.1. Latar Belakang dan Permasalahan

Perkembangan internet yang cepat telah mengubah banyak aspek dalam komunikasi. Dengan adanya internet sebagai fasilitas yang sangat dibutuhkan semakin banyak *platform-platform* komunikasi yang berkembang termasuk *platform* media sosial. Di Indonesia pengguna media sosial mencapai 170 juta pengguna dengan 5 teratas yang paling populer yakni *youtube*, *whatsapp*, *instagram*, *facebook* dan *twitter* (Kemp, 2021).

Twitter merupakan salah satu media sosial yang populer digunakan sebagai wadah untuk komunikasi di masyarakat. Dengan menggunakan *Twitter* pengguna dapat mengirim dan menerima pesan secara langsung kepada pengguna lain. Di Indonesia *Twitter* mendapatkan peringkat ke-5 sebagai media sosial yang paling banyak digunakan. Menjadikan *Twitter* sebagai salah satu *platform* untuk beropini dan menyampaikan pendapat masyarakat mengenai berbagai isu dan perbincangan di Indonesia pada masa pandemi *COVID-19* (Fikri et al. 2020)

COVID-19 atau dapat disebut coronavirus 19 ialah infeksi gangguan pernapasan yang dapat menular dengan cepat. Penyakit *COVID-19* muncul di daerah Wuhan, Tiongkok pada bulan november 2019 (Shereen et al. 2020). Organisasi kesehatan dunia atau World Health Organization selanjutnya memberikan penamaan jenis corona virus baru ini sejak munculnya kejadian

luar biasa di wuhan Cina pada desember 2019 yaitu *Severe Acute Respiratory Syndrome Coronavirus2* atau bisa disebut (SARS-COV2) yang menjadi penyebab dari *COVID-19*. Virus *COVID-19* sendiri masuk ke negara Indonesia dari awal maret sampai saat ini dan telah diperkirakan berlanjut sampai dengan waktu yang tidak dapat ditentukan (Lestari et al, 2021).

Di negara Indonesia sendiri pada tanggal 10 april 2020 ditemukan sebanyak lebih dari 3000 kasus positif mengidap *COVID-19*, 282 orang sembuh dan 306 orang meninggal dunia dengan tingkat kematian sebesar 9.1% (Kartika Sari et al., 2021). Belum tuntasnya upaya pemutusan rantai penyebaran *COVID-19* oleh pemerintah membuat kalangan masyarakat menjadi resah, was-was dan ketakutan terhadap adanya pandemi ini (Mahfud et al. 2020). Banyak masyarakat yang ingin meluapkan aspirasinya di sosial media yang dirasa cocok sebagai tempat yang mewakili aspirasi masyarakat akan pandemi *COVID-19*. Salah satu media sosialnya yaitu *twitter*, dimana pada *twitter* dimungkinkan pengguna akun mengirimkan hingga 140 karakter pada pesan teksnya (Handayani et al. 2021). Terdapat banyak sekali pesan teks yang dikirimkan ada pesan teks yang positif dan ada juga pesan teks yang negatif sehingga sulit untuk mengambil informasi yang selaras didalamnya karena keberagaman pesan teks yang dikirim. Salah satu cara untuk mengatasi hal tersebut ialah menggunakan analisis sentimen(pendapat).

Analisis sentimen(pendapat) masuk kedalam bidang diantara beberapa bidang penelitian seperti *Natural Language Processing*, *Machine Learning* dan *Data Mining*. Analisis sentimen atau opini merupakan pemisahan sentimen dari

isi sebuah kalimat atau teks. Analisis sentimen(pendapat) dilakukan untuk mengelompokkan kalimat atau sentimen berdasarkan pola sifat didalamnya. Pola sifat ini merupakan sentimen yang memiliki aspek positif, negatif atau netral (Dwi Septiana et al. 2021). Studi yang melakukan penelitian sebelumnya tentang analisis sentimen(pendapat) di *twitter* terkait dengan COVID-19 (Handayani et al. 2021). Penelitian tersebut menggunakan kamus sastra dalam preprocessingnya, menggunakan algoritma *Naïve Bayes Classifiers* dan mendapatkan hasil akurasi sebesar 78% .Penelitian lain terkait dengan COVID-19 (Astari et al. 2020). Penelitian tersebut menggunakan data private berjumlah 796 *tweet* serta algoritma *Naïve Bayes Classifiers* dan mendapat akurasi sebesar 67%.(Dwi Septiana et al. 2021) dalam penelitiannya menggunakan 3 tahap *preprocessing* yaitu *case folding*, *tokenize* dan *stopword*, menggunakan algoritma *naïve bayes classifiers* dan didapat hasil akurasi sebesar 69%.

Keterbaruan penelitian ini dari penelitian sebelumnya adalah penggunaan tahap *preprocessing* yang lebih banyak serta pembuatan kamus *stopword* dan *stemming* sendiri yang didasarkan pada dataset yang dimiliki serta penggunaan data publik. Penulis menggunakan *rapidminer* sebagai alat bantu menganalisis data. Dalam pengumpulan datanya penulis menggunakan dataset publik yang diambil dari website *kaggle* yang terdiri dari *tweet* bersentimen positif dan negatif (Prastyo et al. 2020). Data kemudian dilakukan proses pembobotan kata untuk mengetahui pengaruh suatu kata dalam dokumen. Teknik yang digunakan yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF), teknik pembobotan kata dari suatu dokumen menggunakan *case folding* tokenisasi,

normalisasi, *stopwords*, dan *stemming*. Teknik ini digunakan karena cukup mudah dipelajari dan cukup mudah dalam penerapan untuk permasalahan keakuratan suatu dokumen (Rofiqi et al. 2019). Data yang telah diproses menggunakan TF-IDF kemudian dilakukan pengklasifikan untuk mendapat sentimen positif dan negatif. Metode yang digunakan yaitu *naive bayes classifier*(NBC) ditambah *k-nearest neighbors*(KNN) sebagai perbandingan metode. Metode *naive bayes classifier* sering digunakan dalam teknik klasifikasi karena mudah dalam proses *training* dan klasifikasi serta memiliki keakuratan tinggi dalam mengklasifikasikan dokumen dengan banyak (Wijaya et al. 2021).

Maka dari itu, berdasarkan uraian diatas penulis mengusulkan untuk membuat sebuah penelitian berjudul Perbandingan Algoritma *Naive Bayes Classifier* Dan *K-Nearest Neighbors* Untuk Analisis Sentimen *Covid-19* Di *Twitter* dan diharapkan penelitian ini dapat membantu dalam mengetahui kecenderungan opini masyarakat terkait *COVID-19*.

1.2. Rumusan Masalah

1. Seberapa tinggi akurasi yang didapat dari penerapan analisis sentimen(pendapat) menggunakan algoritma NBC dan KNN pada dataset kumpulan opini pengguna twitter tentang *COVID-19* yang dimiliki?
2. Algoritma mana antara NBC atau KNN yang paling baik untuk mengklasifikasikan sentimen(pendapat) berdasarkan hasil analisis sentimen(pendapat) pada dataset kumpulan opini pengguna twitter tentang *COVID-19* yang dimiliki ke dalam sentimen(pendapat) positif dan negatif?

1.3. Batasan Masalah

1. Penelitian dilakukan dengan menggunakan data publik tweet terkait dengan *COVID-19* dikirim pada 23 Maret 2020 sampai 14 Mei 2020 yang didapat dari website *kaggle* .
2. Penelitian ini hanya meneliti topik tentang *COVID-19* di Indonesia pada media sosial *twitter*.
3. Penelitian ini tidak meliputi kata dan kalimat yang cara penulisannya tidak umum seperti emotikon.
4. Penelitian ini meneliti pesan teks (*tweet*) berbahasa Indonesia
5. Klasifikasi sentimen dibagi menjadi dua kategori sentimen yaitu negatif dan positif.
6. *Preprocessing*, implementasi algoritma dan pengujian menggunakan *Rapidminer*.
7. Klasifikasi menggunakan algoritma NBC dan algoritma KNN sebagai perbandingan
8. Stemming dilakukan secara manual berdasarkan temuan kata pada dataset

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan yang ingin dicapai dalam penelitian ini sebagai berikut:

1. Untuk mengklasifikasikan opini pengguna twitter tentang *COVID-19* berdasarkan hasil dari analisis sentimen(pendapat) pada dataset kumpulan opini pengguna twitter tentang *COVID-19* ke dalam sentimen(pendapat) positif dan negatif.
2. Ingin melakukan perbandingan algoritma dengan menerapkan algoritma NBC dan algoritma KNN di dalam analisis sentimen(pendapat) pada dataset kumpulan opini pengguna twitter tentang *COVID-19* yang dimiliki.

1.5. Manfaat Penelitian

Adapun manfaat dari penelitian ini sebagai berikut:

1. Pembaca dan penulis mendapatkan informasi tentang pandangan para pengguna twitter terkait *COVID-19* berdasarkan hasil analisis sentimen(pendapat) pada dataset yang dimiliki berupa kumpulan pesan teks opini pengguna twitter tentang *COVID-19*, apakah cenderung positif atau negatif.
2. Pembaca dan penulis dapat mengetahui informasi tentang perbedaan hasil confusion matrix antara implementasi algoritma NBC dan algoritma KNN berdasarkan dari hasil analisis sentimen(pendapat) pesan teks opini pengguna twitter tentang *COVID-19* pada dataset yang dimiliki berupa kumpulan pesan teks opini pengguna twitter tentang *COVID-19*.

1.6. Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penelitian.

BAB II LANDASAN TEORI

Bab ini berisi teori-teori yang digunakan untuk memahami permasalahan yang dibahas pada penelitian ini yaitu teori yang berkaitan tentang *text mining*, sentimen analisis, pembobotan kata, *NBC*, *KNN* dan penelitian terdahulu.

BAB III METODE PENELITIAN

Bab ini berisi tentang tahapan yang akan dilakukan pada penelitian dimana tahapan dideskripsikan dalam bentuk diagram alir serta masing-masing tahapan berisi penjelasan rinci yang terkait.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang proses yang terdapat pada tahap penelitian yang telah dibuat serta hasil temuan-temuan yang didapat saat penelitian.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dari semua uraian pada bab-bab sebelumnya serta saran yang diharapkan berguna untuk pengembangan selanjutnya.

DAFTAR PUSTAKA

- Astari, N. M. A. J., Dewa Gede Hendra Divayana, & Gede Indrawan. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1), 27–29. <https://doi.org/10.30864/jsi.v15i1.332>
- Dwi Septiana, R., Budi Susanto, A., & Tukiyat, T. (2021). Analisis Sentimen Vaksinasi Covid-19 Pada Twitter Menggunakan Naive Bayes Classifier Dengan Feature Selection Chi-Squared Statistic Dan Particle Swarm Optimization. 49–56.
- Emzir. (2010). *Metodologi penelitian pendidikan kualitatif & kualitatif*. Jakarta: Raja Grafindo Persada.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook : advanced approaches in analyzing unstructured data*. Cambridge University Press.
- Fikri, M. I., Sabrila, T. S., Azhar, Y., & Malang, U. M. (2020). Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, 10(02), 71–76.
- Hadianti, S., Yosep Tember, F., Mandiri, N., Raya, J., No, J., Melayu, C., & Timur, J. (n.d.). ANALISIS SENTIMENT COVID-19 DI TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SVM. *Jurnal Teknologi Informasi*, 6(1). Retrieved from www.Kaggle.com.
- Handayani, E. T., & Sulistiyawati, A. (2021). ANALISIS SENTIMEN RESPON MASYARAKAT TERHADAP KABAR HARIAN COVID-19 PADA TWITTER KEMENTERIAN KESEHATAN DENGAN METODE KLASIFIKASI NAIVE BAYES. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(3), 32–37. Retrieved from <http://jim.teknokrat.ac.id/index.php/JTISI>
- Kartika, E., & Gondohanindijo, J. (2020). Rancang Bangun Model Sentimen Analisis Review Produk Pada Toko Online Menggunakan Naive Bayes. *Seminar Nasional Hasil Penelitian Dan Pengabdian Kepada Masyarakat*, 201–212.
- Kartika Sari, E., Ria EMD, B., Karina Putri, M., Eka Rosita, M., Studi, P. S., & Tinggi Ilmu Kesehatan Akbidyo, S. (2021). *PANGAN FUNGSIONAL SEBAGAI ALTERNATIF PENUNJANG IMUN DI MASA PANDEMI*.
- Kemdikbud. (2016, October 28). KBBi Daring. Retrieved December 11, 2022, from <https://kbbi.kemdikbud.go.id/>
- Kemp, S. (2021). DIGITAL 2021: INDONESIA. Retrieved December 27, 2021, from DATA REPORTAL website: <https://datareportal.com/reports/digital-2021-indonesia>
- Krisandi, N., Helmi, B., & Prihandono, I. (2013). *ALGORITMA k-NEAREST NEIGHBOR DALAM KLASIFIKASI DATA HASIL*

- PRODUKSI KELAPA SAWIT PADA PT. MINAMAS KECAMATAN PARINDU*. 02(1), 50.
- Kurniawan, A., & Adinugroho, S. (2019). *Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features* (Vol. 3). Retrieved from <http://j-ptiik.ub.ac.id>
- Kurniawan, R., & Apriliani, A. (2020). Analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari Twitter berbasis web scraper. *Jurnal Instek (Informatika Sains Dan Teknologi)*, 5(1), 67–75.
- Lestari, M. P., Witarsyah, D. J., & Hamami, F. (n.d.). *PERAMALAN PERTAMBAHAN PASIEN COVID-19 MENGGUNAKAN SUPPORT VECTOR REGRESSION FORECASTING GROWTH OF COVID-19 PATIENTS USING SUPPORT VECTOR REGRESSION*.
- Mahfud, I., & Gumantan, A. (2020). Survey Of Student Anxiety Levels During The Covid-19 Pandemic. *Jp.Jok (Jurnal Pendidikan Jasmani, Olahraga Dan Kesehatan)*, 4(1), 86–97. <https://doi.org/10.33503/jp.jok.v4i1.1103>
- Muhammad, Z., Ramadhani, R., & Rizqifaluthi, H. (2018). Process Mining Akademik Sekolah Menggunakan RapidMiner. *MATICS*, 10(2), 39. <https://doi.org/10.18860/mat.v10i2.5745>
- Nabhan, A. A., Rahayudi, B., & Ratnawati, D. E. (2021). *Klasifikasi Tweets Masyarakat yang Membicarakan Layanan GoFood dan GoRide pada GoJek Dimedia Sosial Twitter Selama Masa Kenormalan Baru (New Normal) dengan Metode Naïve Bayes* (Vol. 5). Retrieved from <http://j-ptiik.ub.ac.id>
- Pitria, P. (2014). Analisis sentimen pengguna twitter pada akun resmi samsung indonesia dengan menggunakan naïve bayes . *Doctoral Dissertation*.
- Prastyo, P. H., Sumi, A. S., Dian, A. W., & Permanasari, A. E. (2020). Tweets Responding to the Indonesian Government’s Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel. *Journal of Information Systems Engineering and Business Intelligence*, 6(2), 112. <https://doi.org/10.20473/jisebi.6.2.112-122>
- Prof. Dr. Sugiyono. (2013). *METODE PENELITIAN KUANTITATIF, KUALITATIF DAN R & D*. ALFABETA.
- Putu Gita Naraswati, N., Cindy Rosmilda, D., Desinta, D., Khairi, F., Damaiyanti, R., & Nooraeni, R. (n.d.). *SISTEMASI: Jurnal Sistem Informasi Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification*. Retrieved from <http://sistemasi.ftik.unisi.ac.id>
- Rafiqa Cahyani. (2019). *ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH PUBLIK PESERTA PILPRES 2019* (SKRIPSI). UIN SUNAN AMPEL SURABAYA, Surabaya.
- Ramadhanta, E., Razaq, M., Jacob, D. W., & Hamami, F. (2021). Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran

- Online Selama Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Perbandingan Algoritma Klasifikasi. *E-Proceeding of Engineering*, 8(5), 9000–9006.
- Rivki, M., & Bachtiar, A. M. (2017). IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR DALAM PENGKLASIFIKASIAN FOLLOWER TWITTER YANG MENGGUNAKAN BAHASA INDONESIA. *Jurnal Sistem Informasi*, 13(1), 31. <https://doi.org/10.21609/jsi.v13i1.500>
- Rofiqi, M. A., Fauzan, Abd. C., Agustin, A. P., & Saputra, A. A. (2019). Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 58–64. <https://doi.org/10.28926/ilkomnika.v1i2.18>
- Saad, S. E., & Yang, J. (2019). Twitter Sentiment Analysis Based on Ordinal Regression. *IEEE Access*, 7, 163677–163685. <https://doi.org/10.1109/ACCESS.2019.2952127>
- Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. Retrieved from <https://t.co/9Wl0aWpfD5>
- Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020, July 1). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, Vol. 24, pp. 91–98. Elsevier B.V. <https://doi.org/10.1016/j.jare.2020.03.005>
- Syarifuddin, M. (2020). ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN KNN. *INTI Nusa Mandiri*, 15(1), 23–28. <https://doi.org/10.33480/inti.v15i1.1347>
- Tala, F. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*.
- Wijaya, T. N., Indriati, R., & Muzaki, M. N. (2021). Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter. *Jambura Journal of Electrical and Electronics Engineering*, 2(2), 78–83.
- Zainal A. Hasibuan, Ph. D. (2007). *METODOLOGI PENELITIAN PADA BIDANG ILMU KOMPUTER DAN TEKNOLOGI INFORMASI*.