ANALISIS SENTIMENT DI TWITTER TERHADAP ANIES BASWEDAN SEBAGAI BAKAL CALON PRESIDEN 2024 MENGGUNAKAN METODE K-NEAREST NEIGHBOR

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

INDRA FEBRIANSYAH
NIM. 11950115094

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2024

LEMBAR PERSETUJUAN

ANALISIS SENTIMENT DI TWITTER TERHADAP ANIES BASWEDAN SEBAGAI BAKAL CALON PRESIDEN 2024 MENGGUNAKAN METODE K-NEAREST NEIGHBOR

TUGAS AKHIR

Oleh

INDRA FEBRIANSYAH

NIM. 11950115094

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 11 Januari 2024

Pembimbing I,	Pembimbing II,
Muhammad Fikry, S.T., M.Sc.	**Yusra, S.T., M.T.**
NIP. 19801018 200710 1 002	**NIP. 19840123 201503 2 001**

LEMBAR PENGESAHAN

ANALISIS SENTIMENT DI TWITTER TERHADAP ANIES BASWEDAN SEBAGAI BAKAL CALON PRESIDEN 2024 MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Oleh

INDRA FEBRIANSYAH

NIM. 11950115094

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 11 Januari 2024

Mengesahkan,

Dekan,

Dr. Hartono, M.Pd.

NIP. 19640301 199203 1 003

Ketua Jurusan,

Iwan Iskandar, S.T., M.T.

NIP. 19821216 201503 1 003

DEWAN PENGUJI

Ketua	: Reski Mai Candra, S.T., M.Sc.
Pembimbing I	: Muhammad Fikry, S.T., M.Sc.
Pembimbing II	: Yusra, S.T., M.T.
Penguji I	: Surya Agustian, S.T., M.Kom.
Penguji II	: Eka Pandu Cynthia, S.T., M.Kom.

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Indra Febriansyah

NIM :11950115094

Tempat, Tgl. Lahir : Padang, 13 Februari 2001

Fakultas : Sains dan Teknologi

Prodi : Teknik Informatika

Judul Jurnal :

Analisis Sentiment di Twitter terhadap Anies Baswedan sebagai Bakal Calon Presiden 2024 Menggunakan Metode K-Nearest Neighbor

Menyatakan dengan sebenar-benarnya bahwa:

1. Penulisan jurnal dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu jurnal saya ini, saya nyatakan bebas dari plagiat.
4. Apabila di kemudian hari terbukti terdapat plagiat dalam penulisan jurnal saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 15 Januari 2024
Yang membuat pernyataan

Indra Febriansyah
NIM. 11950115094

G-Tech : Jurnal Teknologi Terapan
Volume 7, No. 3, Juli 2023, hal.1061-1070
E-ISSN: 2623-064x | P-ISSN: 2580-8737

Analisis Sentiment di Twitter terhadap Anies Baswedan sebagai Bakal Calon Presiden 2024 Menggunakan Metode K-Nearest Neighbor

Indra Febriansyah[1]✉, Muhammad Fikry[2], Yusra[3]
[1,2,3]Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 14-06-2023
Direvisi : 16-06-2023
Diterima : 22-06-2023

ABSTRAK

Anies Baswedan menjadi tokoh politik yang paling banyak diperbincangkan di Twitter. Banyaknya opini masyarakat di Twitter dapat dimanfaatkan untuk memperoleh gambaran positif dan negatif dengan melakukan analisis sentimen. Penelitian ini bertujuan untuk menerapkan metode K-Nearest Neighbor dalam melakukan analisis sentimen di Twitter terhadap Anies Baswedan sebagai bakal calon presiden 2024. Data *tweet* yang digunakan berjumlah 3.400 *tweets* dengan 2.130 label positif dan 1.270 label negatif. Pengujian model K-Nearest Neighbor menggunakan metode 10-*fold cross validation* dilakukan dengan dua eksperimen, yaitu pengujian tanpa seleksi fitur dan pengujian menggunakan seleksi fitur *threshold* pada nilai *Document Frequency* (DF). Metode K-Nearest Neighbor menghasilkan model dengan performa terbaik pada pengujian menggunakan seleksi fitur DF *threshold*, dengan kombinasi parameter nilai k = 14 dan DF *threshold* bernilai 12 memperoleh nilai akurasi sebesar 87,35%, presisi sebesar 87,39%, *recall* sebesar 85,3%, dan f1 score sebesar 86,13%.

Kata Kunci:
Analisis Sentimen, K-Nearest Neighbor, K-Fold Cross Validation, Bakal Calon Presiden

ABSTRACT

Anies Baswedan is the most discussed political figure on Twitter. The number of public opinions on Twitter can be utilized to obtain positive and negative images by conducting sentiment analysis. This study aims to apply the K-Nearest Neighbor method in conducting sentiment analysis on Twitter towards Anies Baswedan as a 2024 presidential candidate. The tweet data used amounted to 3,400 tweets with 2,130 positive labels and 1,270 negative labels. Testing the K-Nearest Neighbor model using the 10-fold cross validation method was carried out with two experiments, namely testing without feature selection and testing using threshold feature selection on Document Frequency (DF) value. The K-Nearest Neighbor method produces a model with the best performance in testing using DF threshold feature selection, with a parameter combination of k = 14 and DF threshold value of 12 obtaining an accuracy value of 87.35%, precision of 87.39%, recall of 85.3%, and f1 score of 86.13%.

Keywords **:**
Sentiment Analysis, K-Nearest Neighbor, K-Fold Cross Validation, Presidential Candidates

Corresponding Author **:**
Indra Febriansyah
Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau
Panam, Jl. HR. Soebrantas No.Km. 15, RW.15, Simpang Baru, Kota Pekanbaru, Riau
Email: 11950115094@students.uin-suska.ac.id

PENDAHULUAN

Di Indonesia, pertumbuhan internet terus meningkat setiap tahunnya. Menurut survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), jumlah orang yang menggunakan internet di Indonesia mencapai 215,63 juta pada Januari 2023, dengan tingkat penetrasi sebesar 78,19 persen dari total populasi (APJII, 2023). Berdasarkan laporan Datareportal, pada bulan Januari 2023, tercatat sebanyak 167 juta pengguna internet di Indonesia atau setara dengan 60,4 persen dari total populasi merupakan pengguna aktif media sosial (Kemp, 2023). Media sosial menjadi sarana yang memudahkan masyarakat untuk berinteraksi dan berbagi informasi secara virtual. Melalui media sosial, masyarakat semakin mudah untuk mendapatkan informasi dan memberikan pendapatnya (Gunawan dkk., 2022). Twitter merupakan media sosial yang lebih berfokus pada bersosial menggunakan teks meskipun pada versi yang baru telah mendukung format video dan foto sebagai pendukung cuitan atau *tweet* (Asro'i & Februariyanti, 2022). Laporan Datareportal pada bulan April 2023 menunjukkan Indonesia berada diurutan keenam pengguna Twitter terbanyak di seluruh dunia dengan memiliki 14,8 juta pengguna aktif (Kemp, 2023b). Hal ini menunjukkan, Twitter menjadi media sosial yang sering digunakan oleh masyarakat sebagai tempat untuk membagikan informasi dan juga mengungkapkan sebuah opini (Barros dkk., 2022). Opini tersebut dapat memuat komentar atau pendapat masyarakat yang berkaitan dengan bidang agama, politik, pendidikan, olahraga, dan lain-lain. Salah satu isu pada bidang politik yang menjadi perhatian masyarakat Indonesia adalah tokoh politik yang akan maju pada pemilihan presiden Indonesia tahun 2024.

Anies Baswedan merupakan tokoh politik yang di deklarasikan oleh partai Nasional Demokrat pada tanggal 3 Oktober 2022 sebagai bakal calon presiden Indonesia tahun 2024. Sebagai tokoh politik yang menjadi sorotan, Anies Baswedan mendapat banyak perhatian di media sosial. Berdasarkan hasil analisis yang dilakukan oleh Drone Emprit (Rahman, 2022) pada periode 3-10 Juli 2022, Anies Baswedan menjadi tokoh politik yang paling banyak di perbincangkan di media sosial dengan 105.110 *mentions*. Opini atau komentar mengenai Anies Baswedan dapat disampaikan masyarakat melalui Twitter dengan mengunggah cuitan atau *tweet*. Data *tweet* tersebut dapat dimanfaatkan untuk memperoleh gambaran positif dan negatif dengan melakukan analisis sentimen. Analisis sentimen adalah teknik yang digunakan untuk memahami data opini dan mengolah tekstual data secara otomatis untuk mengidentifikasi sentimen yang melekat pada sebuah opini (Pertiwi, 2019). Analisis sentimen dapat mengelompokkan opini untuk mengetahui sentimen positif atau sentimen negatif (Fitriyani & Hartanto, 2020). Pada penelitian ini analisis sentimen dilakukan dengan menggunakan metode K-Nearest Neighbor (K-NN). K-Nearest Neighbor (K-NN) merupakan metode sederhana yang mudah diimplementasikan dan baik dalam melakukan klasifikasi teks (Furqan dkk., 2022).

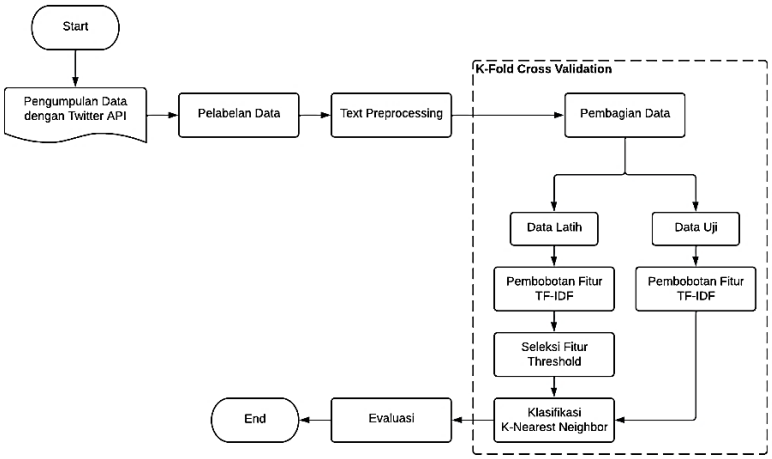
Penelitian mengenai analisis sentimen di Twitter menggunakan metode K-Nearest Neighbor telah dilakukan sebelumnya. Pada penelitian (Malik Zuhdi dkk., 2019) metode K-Nearest Neighbor digunakan dalam analisis sentimen capres Indonesia 2019 menghasilkan akurasi 83,33%. Penelitian (Amardita dkk., 2022) menggunakan metode K-Nearest Neighbor untuk analisis sentimen pada ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung menghasilkan akurasi sebesar 88,29%. Pada penelitian (Isnain dkk., 2021) menggunakan metode K-Nerest Neighbor untuk analisis sentimen masyarakat mengenai pembelajaran online menghasilkan akurasi 84,65%.

Penelitian lainnya telah melakukan perbandingan metode K-Nearest Neighbor dengan metode lain dalam melakukan analisis sentimen di Twitter. Pada penelitian (Pertiwi, 2019) untuk analisis sentimen opini publik mengenai sarana dan transportasi mudik tahun 2019 menggunakan metode Naïve Bayes, Neural Network, K-Nearest Neighbor dan SVM, dengan nilai akurasi tertinggi didapatkan dengan metode K-Nearest Neighbor sebesar 90,76%. Penelitian (Rahmawati & Sukmasetya, 2022) dengan studi kasus analisis sentimen masyarakat terhadap kebijakan kominfo dengan membandingkan metode Logistic Regression, SVM, Naïve Bayes, Decision Tree, Random Forest, dan metode K-Nearest Neighbor, menghasilkan akurasi terbaik dengan menggunakan metode K-Nearest Neighbor sebesar 85,42%.

Tujuan penelitian ini adalah untuk menerapkan dan memperoleh performa metode K-Nearest Neighbor dalam melakukan analisis sentimen positif dan negatif dari opini masyarakat di Twitter terhadap Anies Baswedan sebagai bakal calon presiden 2024. Data *tweet* yang berhasil dikumpulkan dengan teknik *crawling* akan melalui tahap pelabelan data, *text preprocessing*, pembagian data latih dan data uji dengan 10-*fold cross validation*, pembobotan fitur dengan TF-IDF, seleksi fitur pada *Document Frequency* (DF), klasifikasi menggunakan metode K-Nearest Neighbor, dan evaluasi *confusion matrix*. *Confussion matrix* akan melakukan evaluasi terhadap model untuk mendapatkan nilai akurasi, presisi, recall, dan f1 score.

METODE PENELITIAN

Penelitian ini berpedoman pada tahapan yang tersusun secara sistematis agar mendapatkan hasil sesuai dengan yang diharapkan. Tahapan tersebut disusun untuk memaksimalkan penerapan metode K-Nearest Neighbor dalam melakukan analisis sentimen di Twitter terhadap Anies Baswedan sebagai bakal calon presiden 2024. Tahap-tahap penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahap Penelitian

Pengumpulan Data

Pengumpulan data merupakan tahap awal dalam melakukan proses analisis sentimen. Pada penelitian ini, pengumpulan data dilakukan dengan metode *crawling* menggunakan bahasa pemograman Python dengan memanfaatkan Twitter API. Data yang dikumpulkan adalah data *tweet* dari media sosial Twitter yang berhubungan dengan Anies Baswedan sebagai bakal calon presiden 2024 dengan kata kunci "Anies Presiden 2024", "Anies Calon Presiden 2024", "Anies Baswedan Calon Presiden 2024", dan "Anies Baswedan Presiden 2024".

Pelabelan Data

Pelabelan data dilakukan secara manual yang dilakukan oleh 5 orang sebagai anotator. Penelitian ini menggunakan 2 label yaitu positif dan negatif. Label akan ditentukan melalui hasil suara terbanyak yang diberikan oleh anotator. Data yang sudah diberi label akan disepakati dengan menghitung tingkat kesepakatan antar anotator menggunakan Fleiss Kappa. Rumus menghitung Fleiss Kappa dapat dilihat pada Persamaan 1.

$$kappa = \frac{p_\alpha - p_\varepsilon}{1 - p_\varepsilon} \quad (1)$$

p_α adalah presentase jumlah pengukuran antar anotator, sedangkan p_ε adalah presentase jumlah perubahan antar anotator. p_α dihitung dengan rumus persamaan (2) dan p_ε dihitung dengan rumus persamaan (3) berikut ini.

$$p_\alpha \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{j=1}^{k} x_{ij}^2 - m}{m(m-1)} \quad (2)$$

$$p_\varepsilon = \sum_{j=1}^{k} q_j^2 \text{ , dimana } q_j = \frac{1}{nm}\sum_{i=1}^{n} x_{ij} \quad (3)$$

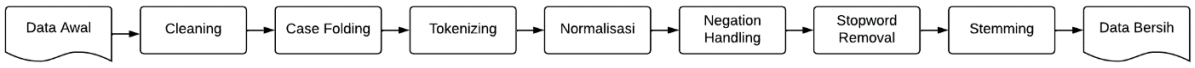
Dengan n adalah jumlah data, m adalah jumlah anotator, dan x_{ij} adalah jumlah keputusan anotator. Berdasarkan nilai kappa yang telah diperoleh, maka tingkat kesepakatan antar anotator dapat dilihat pada Tabel 1.

Tabel 1. Tingkat Kesepakatan Nilai Kappa

Indeks Kappa	*Level of Agreement*
< 0	*Poor Agreement*
0.20 – 0.00	*Slight Agreement*
0.40 – 0.21	*Fair Agreement*
0.60 – 0.41	*Moderate Agreement*
0.80 – 0.61	*Subtantial Agreement*
1.0 – 0.81	*Almost Perfect Agreement*

Text Preprocessing

Text preprocessing merupakan tahap awal dalam pemrosesan data menjadi data yang sesuai dengan kebutuhan analisis dan siap di proses ke tahap selanjutnya. *Text preprocessing* bertujuan untuk meningkatkan kualitas dari data (Furqan dkk., 2022). Tahapan *text preprocessing* pada penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. Tahapan *Preprocessing*

Pada tahapan *cleaning*, dilakukan pembersihan data dengan menghilangkan kata-kata yang tidak penting seperti URL, *username* (@), angka (0-9), *special character,* dan *white space*. Selanjutnya, pada tahapan *case folding*, dilakukan proses mengubah huruf besar menjadi huruf kecil (*lowercase*) untuk mempermudah proses selanjutnya. Setelah itu, dilakukan tahapan *tokenizing* yang merupakan proses membagi kalimat menjadi token atau kata. Tahapan normalisasi dilakukan untuk mengubah kata yang disingkat, tidak baku, dan kata yang salah pengetikan menjadi kata baku menurut ejaan bahasa Indonesia. Proses ini menggunakan kamus normalisasi yang telah dibuat oleh penulis sesuai dengan data yang digunakan.

Kemudian, pada tahapan *negation handling*, dilakukan penanganan terhadap kata negasi. Pada tahapan ini, jika ditemukan kata negasi seperti "jangan", "tidak", "bukan", dan "belum" pada data, maka akan dilakukan penggabungan antara kata negasi dan kata setelahnya menggunakan tanda garis bawah sehingga menjadi satu kata baru. Jika kata baru tersebut terdapat pada kamus negasi yang telah dibuat, maka kata baru tersebut akan diubah sesuai dengan kamus negasi. Setelah itu, dilakukan tahapan *stopword removal* yang merupakan proses penghapusan kata yang tidak memiliki makna atau kurang penting pada data. Proses ini menggunakan *stoplist* bahasa Indonesia pada *library* NLTK. *Natural Language Toolkit* (NLTK) merupakan library atau toolkit bahasa pemrograman Python yang menyediakan beragam komponen dan modul untuk mendukung berbagai tugas pemrosesan bahasa alami. NLTK menyediakan daftar *stop words* Bahasa Indonesia yang dapat digunakan untuk menghapus kata-kata yang terdapat pada *stop words* tersebut dari data yang digunakan, sehingga akan menyisakan data dengan kata-kata yang lebih informatif dan berfokus pada kata-kata kunci. Terakhir, dilakukan tahapan *stemming* yang merupakan proses menghapus imbuhan pada kata sehingga menjadi kata dasar sesuai dengan aturan bahasa Indonesia yang benar.

Pembagian Data

Pembagian data bertujuan untuk memisahkan data menjadi dua bagian yaitu, data latih dan data uji. Data latih akan digunakan untuk melakukan pelatihan pada model yang akan dibangun, sedangkan data uji digunakan untuk mengukur performa dari model yang telah

dibangun. Dalam penelitian ini, pembagian data dilakukan dengan metode k-*fold cross validation*. K-*fold cross validation* adalah sebuah metode yang memecah dataset menjadi dua bagian yaitu data latih dan data uji sebanyak k kelompok dimana jumlah data latih dan data uji pada tiap-tiap kelompok berjumlah sama (Daqiqil ID, 2021).

Pembobotan fitur TF-IDF

Pada penelitian ini, pembobotan fitur pada data latih dan data uji dilakukan secara terpisah. Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) bertujuan untuk memberikan bobot pada data berupa kata untuk mendapatkan data numerik yang dapat digunakan untuk melakukan tahapan klasifikasi menggunakan metode K-Nearest Neighbor. Terdapat dua perhitungan dalam mencari nilai TF-IDF, yaitu *Term Frequency* (TF) dan *Invers Document Frequency* (IDF) (Alrajak dkk., 2020). Nilai TF-IDF didapatkan dengan mengalikan nilai TF dan nilai IDF (Dharmawan dkk., 2020). TF merupakan jumlah kemunculan kata dalam setiap dokumen. Sedangkan untuk nilai IDF didapat dengan menghitung jumlah seluruh dokumen dibagi dengan nilai *Document Frequency* (DF). DF adalah jumlah dokumen yang memiliki kata. Nilai DF mengacu pada jumlah data latih.

Seleksi Fitur

Seleksi fitur bertujuan untuk menghapus fitur yang kurang relevan terhadap proses klasifikasi. Fitur yang kurang relevan ini dapat mempengaruhi performa dari model yang akan dibangun. Pada penelitian ini, dilakukan seleksi fitur berdasarkan *Document Frequency* (DF) dengan menentukan nilai ambang batas (*threshold*). Menentukan *threshold* pada *Document Frequency* (DF) mengasumsikan bahwa fitur yang jarang terdapat pada dokumen lain tidak memiliki pengaruh dalam proses klasifikasi. Jika nilai DF berada dibawah nilai *threshold* yang telah ditentukan, maka fitur tersebut akan diseleksi dan tidak digunakan dalam proses klasifikasi.

K-Nearest Neighbor

K-Nearest Neighbor adalah proses untuk mengelompokkan data ke dalam kelas-kelas yang telah ditentukan sebelumnya berdasarkan jarak terdekat atau tingkat kemiripan data tersebut dengan data set atau data latih yang ada (Malik Zuhdi dkk., 2019). Prinsip kerja algoritma dari KNN yaitu menentukan dan mencari jarak terdekat dengan nilai k *neighbor* terdekat dalam data latih dengan data yang akan diuji (Tangkelayuk & Mailoa, 2022). Nilai k digunakan untuk menyatakan jumlah tetangga terdekat yang terlibat dalam penentuan prediksi label kelas pada data uji (Gunawan dkk., 2022). Tetangga terdekat dengan jarak terkecil akan digunakan untuk menentukan label mayoritas data uji tersebut. Dalam penelitian ini, perhitungan jarak data latih dengan data uji dilakukan dengan menggunakan metode Euclidean Distance.

Evaluasi

Evaluasi digunakan untuk mengukur performa model yang telah dibangun pada proses pelatihan data latih menggunakan metode K-Nearest Neighbor. Evaluasi dilakukan dengan menggunakan metode *Confusion Matrix*. *Confusion matrix* merupakan metode yang cocok digunakan untuk permasalahan klasifikasi (Daqiqil ID, 2021). Penelitian ini menggunakan Confusion matrix dengan dua kelas yaitu positif dan negatif yang disajikan pada Tabel 2.

Tabel 2. Confusion Matrix

	Predicted Positive	*Predicted Negative*
Actual Positive	*True Positive (TP)*	*False Negative (FN)*
Actual Negative	*False Positive (FP)*	*True Negative (TN)*

HASIL DAN PEMBAHASAN

Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data *tweet* berbahasa Indonesia yang mengandung opini atau sentimen mengenai Anies Baswedan sebagai bakal calon presiden 2024. Data dikumpulkan pada rentang waktu bulan Oktober sampai dengan November 2022. Jumlah

data yang berhasil dikumpulkan pada proses *crawling* sebanyak 5355 *tweets*. Pada data *tweet* tersebut masih terdapat kalimat yang bukan opini atau sentimen, sehingga harus dilakukan seleksi data terlebih dulu agar data siap digunakan untuk analisis sentimen. Jumlah data *tweet* setelah dilakukan proses seleksi sebanyak 3400 *tweets*.

Pelabelan Data

Pada tahapan ini, pelabelan data *tweet* dilakukan secara manual oleh 5 anotator. Label untuk setiap data di tentukan berdasarkan suara terbanyak yang diberikan oleh setiap anotator, seperti contoh pelabelan data pada Tabel 3.

Tabel 3. Pelabelan Data

No	*Tweet*	Anotator					Label
		1	2	3	4	5	
1	@susipudjiastuti Pengen cepet" Anies jadi presiden, biar hal lawak gini gak muncul	Negatif	Positif	Negatif	Positif	Positif	**Positif**
2	@RomitsuT Sudahlah anies tidak akan jadi presiden	Negatif	Negatif	Negatif	Negatif	Negatif	**Negatif**
3	@JarnasABW Siap berjuang untuk pak Anies presiden RI 2024..	Positif	Positif	Positif	Positif	Positif	**Positif**
4	@jansen_jsp saya yakin Anies gak akan pernah jadi presiden kecuali bermimpi	Negatif	Negatif	Negatif	Negatif	Negatif	**Negatif**
5	@tatakujiyati Mantap, Buk! Pak Anies emang lebih cocok merawat rumah, daripada jadi presiden. ????????	Positif	Negatif	Positif	Negatif	Positif	**Positif**

Pelabelan pada data tweet menghasilkan label positif sebanyak 2130 dan label negatif sebanyak 1270. Untuk menentukan apakah label tersebut sudah dapat digunakan dalam proses klasifikasi, maka harus di hitung terlebih dahulu tingkat kesepakatan antar anotator dengan menggunakan Fleiss Kappa. Hasil dari perhitungan Fleiss Kappa yaitu 0,68 dengan kategori *subtantial agreement*. *Subtantial agreement* berarti data yang sudah di label berada pada tingkat kesepakatan yang kuat dari kelima anotator. Adapun visualisasi data *tweet* pada setiap kelas label positif dan kelas label negatif dapat dilihat pada Gambar 3.

Gambar 3. ***Word Cloud Data Tweet***

Pada Gambar 3 disajikan visualisasi data *tweet* pada setiap kelas label positif dan kelas label negatif. Kata-kata yang paling penting dan sering muncul pada data ditunjukkan dengan ukuran font yang lebih besar.

Text Preprocessing

Text preprocessing bertujuan untuk memproses data mentah yang diperoleh dari hasil crawling menjadi data yang siap untuk di analisis. Hasil dari tahap *text preprocessing* dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Text Preprocessing*

Preprocessing	Sebelum	Sesudah
Cleaning	**@jansen_jsp** saya yakin Anies gak akan pernah jadi presiden kecuali bermimpi	saya yakin Anies gak akan pernah jadi presiden kecuali bermimpi
Case folding	saya yakin Anies gak akan pernah jadi presiden kecuali bermimpi	saya yakin anies gak akan pernah jadi presiden kecuali bermimpi
Tokenizing	saya yakin anies gak akan pernah jadi presiden kecuali bermimpi	[saya, yakin, anies, gak, akan, pernah, jadi, presiden, kecuali, bermimpi]
Normalisasi	[saya, yakin, anies, **gak**, akan, pernah, jadi, presiden, kecuali, bermimpi]	[saya, yakin, anies, **tidak**, akan, pernah, jadi, presiden, kecuali, bermimpi]
Negation Handling	[saya, yakin, anies, **tidak**, **akan**, pernah, jadi, presiden, kecuali, bermimpi]	[saya, yakin, anies, **mustahil**, pernah, jadi, presiden, kecuali, bermimpi]
Stopword removal	[**saya**, **yakin**, anies, mustahil, **pernah**, **jadi**, presiden, kecuali, bermimpi]	[anies, mustahil, presiden, kecuali, bermimpi]
Stemming	[anies, mustahil, presiden, kecuali, **bermimpi**]	anies mustahil presiden kecuali **mimpi**

K-Nearest Neighbor (K-NN)

Pengujian ini bertujuan untuk mendapatkan performa dari algoritma K-NN dalam melakukan analisis sentimen menggunakan data *tweet* yang berkaitan dengan Anies Baswedan sebagai bakal calon presiden 2024. Pengujian dilakukan dengan beberapa eksperimen, yaitu pengujian model K-NN tanpa seleksi fitur dan pengujian model K-NN menggunakan seleksi fitur dengan beberapa nilai *threshold* pada *Document Frequency* (DF). Data *tweet* yang digunakan dalam proses pengujian sebanyak 3.400 *tweets* yang dibagi menjadi data latih dan data uji. Pembagian data latih dan data uji dilakukan dengan menggunakan 10-*fold cross validation* untuk membagi data yang setara dengan rasio 90:10. Pengujian akan dilakukan dengan beberapa parameter nilai k yang berbeda, yaitu k = 2, k = 4, k = 6, k = 8, k = 10, k = 12, k = 14, k = 16, k = 18, dan k = 20. Nilai k yang menghasilkan nilai akurasi tertinggi dipilih sebagai model terbaik dalam penerapan metode K-NN. Pengujian pertama terhadap model K-Nearest neighbor dengan 10-*fold cross validation* dilakukan tanpa menggunakan seleksi fitur. Hasil pengujian ini, dapat dilihat pada Tabel 5.

Tabel 5. Hasil Pengujian Model K-NN Tanpa Seleksi Fitur

Fold	Akurasi (%) Setiap Nilai K									
	2	4	6	8	10	12	14	16	18	20
1	77,94	80,88	79,41	79,71	80,29	84,71	85,29	85,71	83,24	82,94
2	71,18	72,06	74,70	74,12	75,88	76,47	76,18	75,88	77,65	77,35
3	70,59	75,29	76,47	75,59	77,65	77,35	76,76	78,82	77,35	76,76
4	69,71	73,82	74,71	77,06	74,71	75,88	78,24	78,24	78,82	78,24
5	71,47	75,29	75,88	77,06	77,35	77,06	76,47	77,35	76,18	75,88
6	71,76	75,59	77,94	79,12	78,24	79,41	80,88	81,18	82,35	81,76
7	70	73,24	74,29	75	77,94	76,47	77,65	76,47	77,35	77,35
8	71,47	73,53	76,76	75,59	76,47	75	74,41	75,88	76,47	77,65
9	70,59	75,59	75,29	75,88	76,18	76,47	76,76	75,59	78,24	77,94
10	67,06	73,53	75,29	76,47	77,65	77,94	77,94	77,65	77,65	77,35

Pada Tabel 5 menunjukkan bahwa performa terbaik model K-Nearest Neighbor tanpa menggunakan seleksi fitur, diperoleh pada *fold* pertama dengan parameter nilai k = 14, nilai akurasi mencapai 85,29%.

Pengujian kedua, dilakukan dengan menggunakan seleksi fitur beberapa nilai *threshold* pada *Document Frequency* (DF). Nilai *threshold* yang digunakan adalah 4, 8, 12, 16, dan 20. Pada setiap pengujian perubahan nilai *threshold*, dilakukan juga perubahan pada parameter nilai k sehingga diperoleh kombinasi nilai *threshold* dan parameter nilai k yang menghasilkan performa terbaik. Pada pengujian ini, kombinasi nilai *threshold* dan parameter *k* yang menghasilkan

performa terbaik terdapat pada $k = 14$. Hasil dari pengujian model K-Nearest Neighbor dengan 10-*fold cross validation* menggunakan seleksi fitur nilai *threshold* pada *Document Frequency* (DF) dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengujian Model K-NN dengan Seleksi Fitur *Threshold*

Nilai K	Akurasi (%) Setiap Nilai DF *Threshold*				
	4	8	12	16	20
2	74,41	74,12	74,12	73,82	72,65
4	77,65	80,29	81,47	78,82	77,94
6	81,18	82,06	82,06	81,18	81,76
8	81,18	81,47	83,82	81,47	79,12
10	81,18	82,65	84,71	80,88	82,65
12	82,06	82,35	83,53	81,76	81,47
14	82,35	83,82	87,35	82,65	82,35
16	83,82	84,71	85,88	82,35	82,06
18	82,94	84,71	85,88	85	81,18
20	83,82	83,82	86,76	86,47	80,56

Dari Tabel 6 dapat dilihat bahwa performa terbaik ditunjukkan pada *fold* pertama dengan kombinasi nilai DF *threshold* = 12 dan parameter nilai k =14, mencapai akurasi sebesar 87,35%.

Evaluasi

Evaluasi *Confusion Matrix* dilakukan untuk mengukur kinerja model K-Nearest Neighbor dalam klasifikasi sentimen. Hasil pengukuran kinerja model pada setiap pengujian dapat dilihat pada Tabel 7.

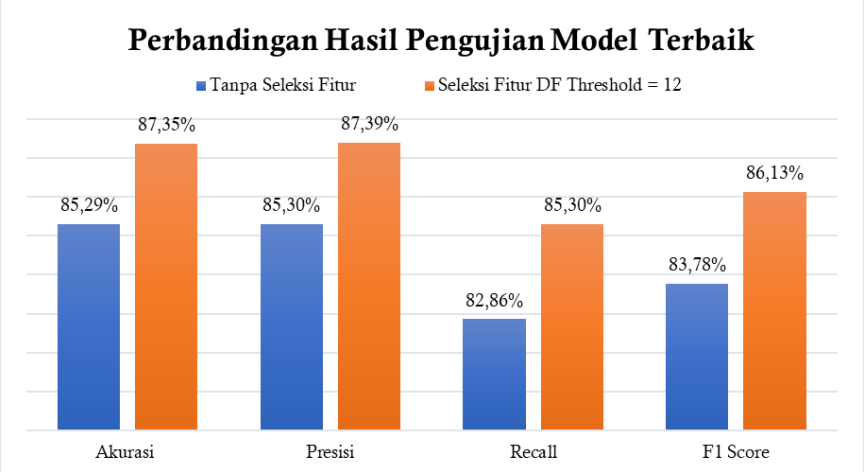
Tabel 7. *Confusion Matrix* Model Terbaik

Evaluasi		K-NN Tanpa Seleksi Fitur		K-NN dengan Seleksi Fitur DF *Threshold* = 12	
		Predicted			
		Positive	*Negative*	*Positive*	*Negative*
Actual	***Positive***	197	16	199	14
	Negative	34	93	29	98

Dari hasil evaluasi pengujian yang di tampilkan pada Tabel 7, dapat diperoleh performa model dengan menghitung nilai akurasi, presisi, *recall*, dan f1 score pada model terbaik setiap eksperimen. Pengujian eksperimen pertama yang dilakukan tanpa menggunakan seleksi fitur, berhasil memprediksi 231 label positif dan 109 label negatif. Mendapatkan model terbaik pada parameter nilai k = 14 dengan nilai akurasi 85,29%, presisi 85,3%, *recall* 82,86%, dan f1 score 83,78%. Sedangkan eksperimen kedua, dilakukan dengan menggunakan seleksi fitur beberapa nilai *threshold* pada *Document Frequency* (DF). Model berhasil memprediksi 228 label positif dan 112 label negatif. Model terbaik terdapat pada kombinasi parameter nilai k = 14 dengan *threshold* = 12, menghasilkan nilai akurasi mencapai 87,35%, presisi 87,39%, *recall* 85,3%, dan f1 score 86,13%.

Perbandingan Model

Berdasarkan hasil pengujian pada model K-Nearest Neighbor dengan 10-*fold cross validation*, dilakukan perbandingan hasil model terbaik pada setiap eksperimen untuk mengidentifikasi model dengan performa terbaik. Hasil perbandingan kedua model terbaik pada setiap eksperimen disajikan pada Gambar 4.



Gambar 4. Hasil Perbandingan Model Terbaik

Gambar 4 menyajikan perbandingan antara pengujiaan tanpa seleksi fitur dan pengujian menggunakan seleksi fitur DF *Threshold*. Perbandingan kedua model menunjukkan eksperimen dengan menggunakan seleksi fitur DF *threshold* memperoleh model yang lebih baik. Performa terbaik diperoleh pada kombinasi nilai *threshold* = 12 dan parameter nilai k = 14, menghasilkan nilai akurasi, persisi, *recall* dan f1 score secara berturut turut yaitu 87,35%, 87,39%, 85,3%, dan 86,13%. Hal ini menunjukkan bahwa penggunaan seleksi fitur nilai *threshold* pada *Document Frequency* (DF) dapat meningkatkan performa dari model K-Nearest Neighbor dalam klasifikasi sentimen dengan 10-*fold cross validation*.

KESIMPULAN DAN SARAN

Hasil penelitian ini menunjukkan bahwa metode K-Nearest Neighbor berhasil diterapkan dalam analisis sentimen di Twitter terhadap Anies Baswedan sebagai bakal calon presiden 2024. Model K-Nearest Neighbor dengan performa terbaik terdapat pada *fold* pertama dengan kombinasi parameter nilai k = 14 dan nilai *threshold* = 12 pada *Document Frequency* (DF), dengan akurasi mencapai 87,35%, presisi 87,39%, *recall* 85,3%, dan f1 score 86,13%. Hasil pengujian menunjukkan bahwa penggunaan seleksi fitur *threshold* pada *Document Frequency* (DF) dapat meningkatkan performa model K-Nearest Neighbor. Penelitian selanjutnya dapat menerapkan seleksi fitur lain seperti Chi-Square, Particle Swarm Optimization (PSO), atau teknik lainnya untuk meningkatkan kinerja model.

REFERENSI

Alrajak, M. S., Ernawati, I., & Nurlaili, I. (2020). Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada Twitter dengan Algoritma K-Nearest Neighbor (K-NN). *Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, 110–122.

Amardita, R. S., Adiwijaya, A., & Purbolaksono, M. D. (2022). Analisis Sentimen terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN. *Jurnal Riset Komputer*, *9*(1), 62–68. https://doi.org/10.30865/jurikom.v9i1.3793

APJII, T. (2023, Maret 10). *Survei APJII Pengguna Internet di Indonesia Tembus 215 Juta Orang*. Asosiasi Penyelenggara Jasa Internet Indonesia. https://apjii.or.id/berita/d/survei-apjii-pengguna-internet-di-indonesia-tembus-215-juta-orang

Asro'i, A. & Februariyanti, H. (2022). Analisis Sentimen Pengguna Twitter terhadap Perpanjangan PPKM Menggunakan Metode K-Nearest Neighbor. *Jurnal Khatulistiwa Informatika*, *10*(1), 17–24. https://doi.org/https://doi.org/10.31294/jki.v10i1.12624

Daqiqil ID, I. (2021). *Machine Learning : Teori, Studi Kasus dan Implementasi Menggunakan Python* (Edisi 1). UR PRESS.

Dhaefinawan, L. R., Arwani, I., & Ratnawati, D. E. (2020). Analisis Sentimen pada Sosial Media Twitter terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *4*(3), 959–965.

Farras, I., Mahdiana, D., & Rahajoe, A. D. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Analisis Sentimen Ulasan SiCepat Ekspres pada Twitter. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi*, 441–448. https://senafti.budiluhur.ac.id/index.php

Fitrisani, N. K., & Hartanto, A. D. (2020). Analisis Sentimen terhadap Tokoh Publik Menggunakan Support Vector Machine. *Media Informasi Analisa dan Sistem*, *5*(1), 8–12. https://doi.org/https://doi.org/10.54367/means.v5i1.615

Furqan, M., Sriani, & Mayang Sari, S. (2022). Analisis Sentimen Menggunakan K-Nearest Neighbor terhadap New Normal Masa Covid-19 Di Indonesia. *Techno.com*, *21*(1), 52–61. https://doi.org/https://doi.org/10.33633/tc.v21i1.5446

Gunawan, R., Septiadi, R., Apri Wenando, F., Mukhtar, H., & Syahril. (2022). K-Nearest Neighbor (KNN) untuk Menganalisis Sentimen terhadap Kebijakan Merdeka Belajar Kampus Merdeka pada Komentar Twitter. *Jurnal CoSciTech (Computer Science and Information Technology)*, *3*(2), 152–158. https://doi.org/10.37859/coscitech.v3i2.3841

Ismain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(2), 121–130. https://doi.org/10.22146/ijccs.65176

Kemp, S. (2023a, Februari 9). *Digital 2023: Indonesia*. DATAREPORTAL. https://datareportal.com/reports/digital-2023-indonesia?

Kemp, S. (2023b, Mei 11). *Twitter Users, Stats, Data & Trends*. DATAREPORTAL. https://datareportal.com/essential-twitter-stats?

Malik Zuhdi, A., Utami, E., & Raharjo, S. (2019). Analisis Sentiment Twitter terhadap Capres Indonesia 2019 dengan Metode K-NN. *Jurnal INFORMA Politeknik Indonusa Surakarta*, *5*(2), 2442–7942. https://doi.org/https://doi.org/10.46808/informa.v5i2.73

Pertiwi, M. W. (2019). Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, KNN dan SVM. *INTI NUSA MANDIRI*, *14*(1), 27–32.

Rahman, A. (2022, Juli 14). *Analisis Popularitas dan Favorabilitas Beberapa Tokoh Politik*. Drone Emprit. https://pers.droneemprit.id/analisis-popularitas-dan-favorabilitas-beberapa-tokoh-politik/

Rahmawati, C., & Sukmasetya, P. (2022). Sentimen Analisis Opini Masyarakat Terhadap Kebijakan Kominfo atas Pemblokiran Situs non-PSE pada Media Sosial Twitter. *Jurnal Riset Komputer*, *9*(5), 1393–1400. https://doi.org/10.30865/jurikom.v9i5.4950

Tangkelayuk, A., & Mailoa, E. (2022). Klasifikasi Kualitas Air Menggunakan Metode KNN, Naive Bayes Dan Decision Tree. *Jurnal Teknik Informatika dan Sistem Informasi*, *9*(2), 1109–1119. https://doi.org/https://doi.org/10.35957/jatisi.v9i2.2048

G-TECH
JURNAL TEKNOLOGI TERAPAN
E-ISSN : 2623-064X
P-ISSN: 2580-8737
Penerbit :
Fakultas Sains dan Teknologi
Universitas Islam Raden Rahmat Malang
jurnal.gtech@gmail.com +62 856-4986-9498

Malang, 22 Juni 2023

Nomor : 07.30/LoA/G-TECH/F.SAINTEK/VI/2023
Perihal : **Letter of Acceptance (LoA)**

Author Yth.
Indra Febriansyah, Muhammad Fikry, & Yusra
Universitas Islam Negeri Sultan Syarif Kasim Riau**,** Indonesia

Salam Hormat,

Berdasarkan artikel Bapak/Ibu yang telah diserahkan ke redaksi **G-Tech: Jurnal Teknologi Terapan** dengan judul :

Analisis Sentiment di Twitter terhadap Anies Baswedan sebagai Bakal Calon Presiden 2024 Menggunakan Metode K-Nearest Neighbor

bersama ini kami sampaikan bahwa hasil penilaian mitra bestari dan keputusan dewan redaksi, artikel Bapak/Ibu dinyatakan **Diterima (*Accepted*)** untuk dimuat di Jurnal kami pada Edisi Vol. 7 No. 3 Juli 2023. Setelah Anda menerima surat ini, kami informasikan juga bahwa biaya publikasi G-Tech : Jurnal Teknologi Terapan sebesar untuk mendukung biaya penyebaran akses terbuka, pengelolaan, pengeditan naskah serta manajemen jurnal dan publikasi secara umum.
Anda dapat menstransfer biaya tersebut ke Rekening **Bank Mandiri** berikut:

Nomor Rekening :
Nama Rekening :

Mohon konfirmasi pembayaran Anda melalui WA dengan melampirkan bukti transfer bank. Artikel Anda akan dipublikasikan setelah pembayaran dikonfirmasi.
Terima kasih telah mempercayai G-Tech: Jurnal Teknologi Terapan sebagai sarana untuk publikasi Anda.
Hormat Kami,

Pimpinan Redaksi

Dr. Mojibur Rohman, M.Pd