

## Komparasi *Naïve Bayes* dan *K-NN* Dalam Analisis Sentimen di Twitter Terhadap Kemenangan Paslon 02

Alfira Fitri Nur Azizah<sup>1</sup>, Viry Puspaning Ramadhan<sup>2\*</sup>

<sup>1,2</sup> Sistem Informasi, Universitas Merdeka Malang, Jalan Terusan Dieng. 62-64 Klojen, Pisang Candi, Sukun, Kota Malang, 65146, Indonesia

**\*Email Korespondensi:**  
viry.puspaning@unmer.ac.id

### 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; *K-NN*; Komparasi; *Naïve Bayes*.

### 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:** Comparison; *K-NN*; *Naïve Bayes*; Sentiment Analysis.

## 1. Pendahuluan

Pemilu Presiden dan Wakil Presiden tahun 2024 adalah salah satu peristiwa politik yang paling dinantikan oleh seluruh masyarakat Indonesia. Pemilu merupakan implementasi nyata dari demokrasi di Indonesia, dimana warga negara memiliki peran untuk dapat ikut serta memilih pejabat publik secara langsung (Maulana, Fridawati, & Masrurroh, 2022). Hasil perhitungan suara yang dikenal sebagai *real count* pada Pemilu serentak tahun 2024 ini telah memicu berbagai macam reaksi dari berbagai kalangan masyarakat. Terdapat pendapat yang mendukung maupun yang menentang saat berjalannya proses perhitungan tersebut. Hasil Pemilu yang sejak awal menunjukkan bahwa pasangan calon nomor urut 02 unggul dari dua paslon lain menjadi perbincangan hangat di berbagai platform, termasuk Twitter.

Twitter adalah salah satu bentuk *microblogging* populer di kalangan user internet (Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani, 2023). Layanan media sosial Twitter digunakan untuk beragam tujuan dalam

berbagai bagian, seperti sebagai *platform* kampanye politik, sebagai wadah untuk melakukan aksi protes, sebagai sumber pembelajaran, dan sebagai sarana komunikasi dalam situasi darurat (Murni, Riadi, & Fadlil, 2023).

Hal diatas mendukung gagasan untuk memanfaatkan media sosial *Twitter* sebagai sumber informasi yang bisa dianalisis melalui klasifikasi sentimen. Dengan jumlah data yang besar di platform *Twitter*, data tersebut dapat disaring dan digunakan sebagai bahan penelitian yang sesuai (Wahyu Andrian, Adline Twince Tobing, Zuhdi Pane, & Kusnadi, 2023). Suatu pendekatan yang bekerja untuk mengetahui apakah opini serta pandangan yang diberikan merupakan sesuatu yang positif atau negatif dapat dijadikan sebagai acuan dalam membuat suatu keputusan, ataupun meningkatkan pelayanan merupakan definisi umum dari analisis sentimen (Mutiarra, Alkhalifi, Mayangky, & Gata, 2020). Dalam konteks penelitian ini, analisis sentimen dilakukan terhadap pandangan publik di *Twitter* mengenai hasil *real count* yang menetapkan kemenangan pasangan calon nomor urut 02 pada Pemilu 2024. Penelitian ini menerapkan algoritma probabilitas *Naive Bayes* dan *K-NN* sebagai metode klasifikasi. Pengkajian ini akan menguji dan membandingkan dua algoritma tersebut dalam upaya menganalisis sentimen opini publik di media sosial *Twitter* terhadap kemenangan paslon 02 pada Pemilu 2024.

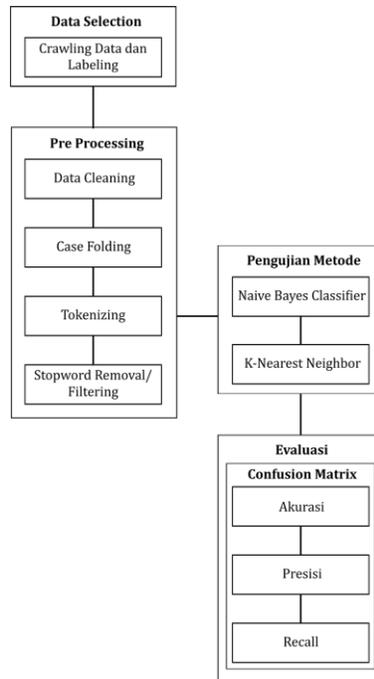
Pada penelitian sebelumnya, oleh peneliti dijadikan sebagai acuan dalam penelitian ini dengan menggunakan pendekatan yang sama namun dengan topik atau konteks yang berbeda. Diantaranya yang pertama menganalisis sentimen mengenai pemindahan Ibukota menggunakan metode *Naive Bayes* dan *K-NN* oleh (Syahril Dwi Prasetyo et al., 2023) dengan diperoleh sebanyak 2000 data yang setelah dianalisis menghasilkan nilai sebesar 82,27% untuk *Naive Bayes* dan 88,12% untuk *K-NN*. Penelitian kedua oleh (Ramadhani & Wahyudin, 2022) membahas penguraian sentimen mengenai penyuntikan *AstraZeneca* di *Twitter* dengan metodologi *Naive Bayes* dan *K-NN* menggunakan 470 data yang memperoleh akurasi pada *Naive Bayes* sebesar 88,56% dan pada *K-NN* sebesar 74,78%. Riset lainnya oleh (Syarifuddinn, 2020) dalam pembahasan sentimen umum terhadap *Covid-19* di *Twitter* dengan menerapkan strategi *Naive Bayes* dan *K-NN* dengan perolehan data sebanyak 1098 *tweets* populasi. Dari penelitian tersebut, algoritma *Naive Bayes* memperoleh akurasi senilai 63,21% sementara untuk *K-NN* sebanyak 58,10% oleh karena itu dapat disimpulkan bahwa *Naive Bayes* dianggap lebih optimal daripada *K-NN*.

Tujuan dari perbandingan dalam penelitian ini adalah untuk mengidentifikasi metode yang paling efisien dan memiliki tingkat akurasi yang baik dalam menganalisis sentimen opini publik. Melalui adanya penelitian ini, diharapkan juga dapat diperoleh pemahaman yang lebih mendalam terkait bagaimana opini masyarakat di media sosial *Twitter* apakah cenderung positif atau negatif dalam merespon hasil *real count* terkait kemenangan pasangan calon nomor urut 02 pada Pemilu 2024.

## 2. Metode Penelitian

Berdasarkan rancangan diatas menunjukkan bahwa penelitian dilaksanakan dengan berbagai tahapan-tahapan. Dimulai dari aktivitas pemilihan data sampai tahapan akhir yaitu evaluasi model untuk mendapatkan nilai perbandingan. Adapun penjelasan dari desain penelitian diatas sebagai berikut :

Tahapan awal dari penelitian ini ialah mengumpulkan data atau melakukan seleksi data. Proses pengumpulan data dikerjakan menggunakan *Node.JS* dan *library tweet-harvest* (Ghufron, Mahabbataka Arsyada, Lukman, Haryono Putra, & Rakhmawati, 2023), yang dijalankan dengan memanfaatkan *Google Colab* untuk mengakses tweet yang relevan terkait kemenangan paslon 02. Dalam penelitian ini, menggunakan *tweet-harvest* dengan versi 2.6.1 yang *library* nya telah dimodifikasi dan diperbarui oleh si pembuat yaitu Helmi Satria.



Gambar 1. Struktur Penelitian

Dalam prosesnya metode ini memerlukan *auth token* yang diperoleh dari akun pengguna *Twitter* asli. *Node.JS* diinstal untuk menjalankan skrip *crawling* menggunakan *tweet-harvest*. *Pandas* juga diinstal untuk analisis data. Selanjutnya proses *data crawling* dilakukan dengan menggunakan beberapa kata kunci yang relevan, salah satunya “Prabowo Gibran real count” dan hanya tweet berbahasa Indonesia saja dengan kurun waktu 15 Februari 2024 – 22 April 2024. Hasil *data tweet* yang telah diambil menggunakan *tweet-harvest* dapat disimpan dalam bentuk *DataFrame pandas*, yang kemudian di ekspor ke file *CSV* atau *Excel* dengan data didapatkan sebanyak 1149 dataset. Hasil *crawling* tersebut kemudian dilabeli untuk menentukan kategori sentimen dari tweet yang diperoleh. Proses pelabelan pada *dataset* tersebut dilakukan secara manual tanpa adanya kamus khusus. Kategori label terdiri dari dua klasifikasi, yaitu positif dan negatif.

```

# Crawl Data

filename = 'Paslon02-realcount-1.csv'
search_keyword = 'Prabowo Gibran real count lang:id since:2024-02-15 until:2024-04-22'
limit = 100

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" -l {limit} --token {twitter_auth_token}
  
```

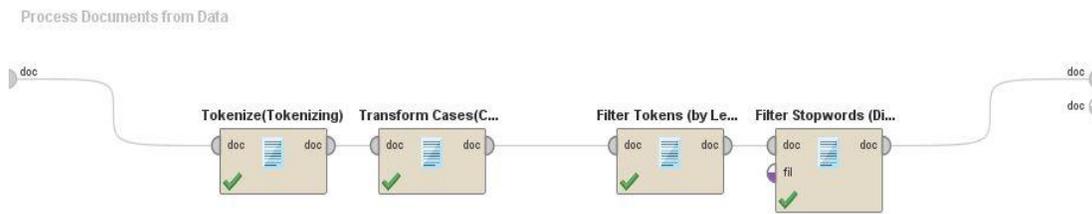
**Tweet Harvest [v2.6.0]**

Research by Helmi Satria  
Use it for Educational Purposes only!

This script uses Chromium Browser to crawl data from Twitter with your Twitter auth token.  
Please enter your Twitter auth token when prompted.

Note: Keep your access token secret! Don't share it with anyone else.  
Note: This script only runs on your local device.

Gambar 2. Proses Crawling Menggunakan Tweet-Harvest



Gambar 3. Pre-Processing

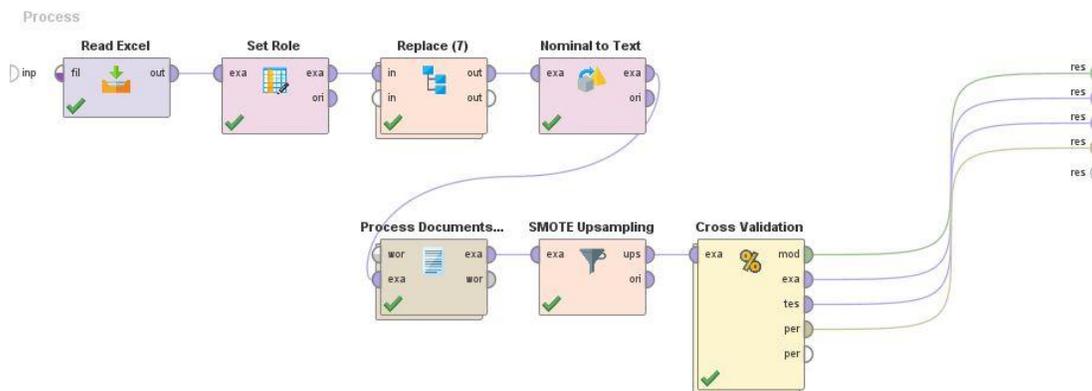
Setelah melakukan *crawling data*, maka data perlu diatur dan dibersihkan melalui tahapan *pre-processing* (Puad, Garno, & Susilo Yuda Irawan, 2023). Pada tahap ini, dilakukan proses *pre-processing* yang mencakup pembersihan data mentah dan tidak terstruktur. *Pre-processing* data merupakan langkah penting untuk memastikan bahwa *dataset* yang akan dianalisis bersih dan sesuai untuk diolah lebih lanjut. Tahapan *pre-processing* dilakukan menggunakan perangkat lunak *Rapidminer*. Tahapan yang dilakukan meliputi *Data Cleaning, Case Folding, Tokenizing, Filtering Token, dan Stopword Removal*.

Setelah tahapan *pre-processing*, akan diuraikan pengimplementasian pendekatan untuk analisis sentimen opini publik terhadap kemenangan paslon 02 pada Pemilu 2024 memakai perangkat lunak statistik *Rapidminer*. Algoritma *Naïve Bayes Classifier* dan *K-Nearest Neighbor* diterapkan dalam pengujian ini. Kemudian hasil dari pengujian kedua metode tersebut akan dilakukan komparasi atau perbandingan untuk mengidentifikasi metode manakah yang menghasilkan nilai akurasi tinggi hingga bisa dikatakan lebih efektif dalam melakukan analisis sentimen opini publik di *Twitter*. Pada tahapan pengujian metode dalam penelitian ini, prosesnya meliputi persiapan data, penyeimbangan data, pembagian dataset menjadi data latih dan data pengujian, implementasi kedua teknik, hingga evaluasi yang merupakan hasil akhir pengujian (Nishom, 2019).

Pada Langkah evaluasi, kinerja *Naïve Bayes* serta *K-Nearest Neighbor* dievaluasi. *Confusion Matrix* dipakai dalam mengevaluasi kinerja sistem tersebut dengan mengukur antara lain *accuracy* untuk mengukur seberapa mirip nilai yang diharapkan dengan nilai aktual, *presisi* untuk melakukan evaluasi akurasi prediksi positif model dengan berdasarkan prediksi positifnya, dan *recall* untuk mengevaluasi jumlah nilai positif sebenarnya yang diidentifikasi sebagai *true positive* oleh model melalui klasifikasi.

### 3. Hasil

Pada pembahasan analisis sentimen ini, proses utama analisis meliputi beberapa tahapan yang akan dibahas meliputi pengambilan dan pelabelan data, *pre-processing*, pengujian kedua metode, serta evaluasi yang merupakan hasil akhir pengujian. Berikut proses keseluruhan analisis pada gambar 4.



Gambar 4. Proses Utama Analisis Sentimen

Proses pengumpulan data atau *data crawling* dilakukan menggunakan *tweet-harvest* yang dijalankan dengan memanfaatkan *Google Colab* untuk mengakses *tweet* yang relevan terkait kemenangan paslon 02 pada Pemilu 2024. Hasil *crawling* didapatkan sebanyak 1149 data *tweet* berdasarkan beberapa kata kunci relevan yang telah digabungkan dalam satu file *excel* dan hanya *tweet* berbahasa Indonesia saja dengan kurun waktu 15 Februari 2024 – 22 April 2024 disajikan pada gambar dibawah ini.

conversation_id	created_at	favorite_count	full_text	id_str	image_url
1.75944E+18	Mon Feb 19 04:49:17 +0000 2024	7347	Update Rekap Nasional Real Count. Sabtu 19 Februari 2024 11:45 WIB Anies-Muhammad 35 23	1.75944E+18	https://pbs.twimg.com/media/GGrjBEWakAA7E9s...
1.75989E+18	Tue Feb 20 10:37:57 +0000 2024	26851	Real Count KPU Total suara Masuk 72 37% Sumber CNN TV 01. Anies - Muhaimin Menang Tel	1.75989E+18	https://pbs.twimg.com/media/GGxvZa8AAH126.jp...
1.77047E+18	Wed Mar 20 15:09:16 +0000 2024	178	SAH KPU menetapkan Prabowo & Gibran sebagai pemenang pada Pemilu 2024 berdasar	1.7705E+18	https://pbs.twimg.com/media/GIH28jiaAACFD.jpg
1.76852E+18	Fri Mar 15 06:05:41 +0000 2024	202	Banjir Ucapan Selamat datang dari para Pemimpin dari belahan penjuru dunia atas kemenan	1.7685E+18	https://pbs.twimg.com/ext_tw_video_thumb/1768...
1.76819E+18	Thu Mar 14 07:59:29 +0000 2024	67	KPU telah menyelesaikan Real Count Rekapitulasi pleno di 21 Provinsi secara terbuka dan trar	1.7682E+18	https://pbs.twimg.com/ext_tw_video_thumb/1768...
1.77046E+18	Wed Mar 20 14:44:04 +0000 2024	2473	Komis Pemilihan Umum (KPU) RI telah menetapkan hasil real count Pemilu 2024 pada Rabu (	1.7705E+18	https://pbs.twimg.com/media/GIHwovbAAAsyPj.jpg
1.77017E+18	Tue Mar 19 19:28:29 +0000 2024	16	Update Hasil Real Count Pilpres 2024 Selasa 19 Maret: Anies-Imin Prabowo-Gibran Ganjar-Mi	1.7702E+18	https://pbs.twimg.com/media/GIHd1z1asAAH70.jpg
1.76387E+18	Sat Mar 02 10:21:46 +0000 2024	25	Suara Prabowo-Gibran semakin tak terkejar oleh paslon 01 dan 03. Yang sono ngarep menan	1.7639E+18	https://pbs.twimg.com/media/GH4bKwaAAADfG...
1.76488E+18	Tue Mar 05 05:01:35 +0000 2024	2	Update real count sementara Pilpres 2024 di Sumbagsel Lor! â€¦ Perhitungan Pilpres di websit	1.7649E+18	https://pbs.twimg.com/media/GH17qBQAAMK0S...
1.7628E+18	Wed Feb 28 11:04:13 +0000 2024	5	Survey LSI menemukan Mayoritas Publik percaya & menerima hasil real count KPU. Prab	1.7628E+18	https://pbs.twimg.com/media/GH1eDLbaAAAS5m...
1.77028E+18	Wed Mar 20 02:38:34 +0000 2024	12	Update Hasil Real Count Pilpres 2024 Selasa 19 Maret: Anies-Imin Prabowo-Gibran Ganjar-Mi	1.7703E+18	https://pbs.twimg.com/media/GID17egBQAAMK0S...
1.76741E+18	Tue Mar 12 04:22:57 +0000 2024	13	Real Count https://t.co/XW3PpE309 Nyatakan Prabowo-Gibran Pemenang Pilpres 2024 #Pri	1.7674E+18	https://pbs.twimg.com/media/GIc1XAsAAwV6e.jpg
1.77061E+18	Thu Mar 11 00:48:45 +0000 2024	138	Sebagai pendukung Ganjar saya mengucapkan selamat atas kemenangan pasangan Prabowo	1.7706E+18	https://pbs.twimg.com/media/GIH7kpuasAAhLH5.jp...
1.76777E+18	Wed Mar 13 04:25:47 +0000 2024	42	REAL COUNT VERSI GUE SENDIRI. SUMBER C1 @KPU C1 @KawalPemilu_org dan @detikcom	1.7678E+18	https://pbs.twimg.com/media/GIHgITMb0AAh16.jpg
1.76289E+18	Tue Mar 28 17:25:00 +0000 2024	0	Jasa Foto video acara kudukaan pemakaman kremasi JABODETABEK dan sekitarnya. WA : 087	1.7629E+18	https://pbs.twimg.com/media/GHbwhvhsAA0RoV...
1.7649E+18	Tue Mar 05 06:44:42 +0000 2024	1	Prabowo Subianto Mengingatkan Agar Keunggulan Sementara Dirinya Dan Mas Gibran Dalar	1.7649E+18	https://pbs.twimg.com/ext_tw_video_thumb/1764...
1.76715E+18	Mon Mar 11 11:21:28 +0000 2024	6288	Kami mengucapkan selamat kepada @prabowo dan @gibran_tweet telah memenangkan Pilj	1.7671E+18	
1.7676E+18	Tue Mar 12 17:12:16 +0000 2024	0	Real Count Kawalpemilu Nyatakan Prabowo-Gibran Menang Pilpres 2024 https://t.co/YDcM0	1.7676E+18	https://pbs.twimg.com/media/GIFGwYnYbKAAzRBX...
1.76817E+18	Wed Mar 14 06:48:11 +0000 2024	0	Yang hangat real count https://t.co/QX4ZYLSUJ ucapan selamat kepada @prabowo & amp;	1.7682E+18	https://pbs.twimg.com/media/GIHkFvAIAAX8f9.jp...
1.76281E+18	Wed Feb 28 11:56:26 +0000 2024	116	Paslon 02 Prabowo-Gibran sementara memperoleh sebanyak 2.945.256 suara atau 58 75 per	1.7628E+18	https://pbs.twimg.com/media/GHbBarSawAAJW7...
1.76341E+18	Fri Mar 01 03:44:20 +0000 2024	35	Real Count KPU 77 90%: Prabowo-Gibran 58 83% AMIN 24 49% Ganjar-Mahfud 16 68% https://t.co	1.7634E+18	
1.76786E+18	Wed Mar 13 10:36:09 +0000 2024	263	Real Count https://t.co/Fpk0J0fN Nyatakan Pasangan Prabowo-Gibran Menang https://t.co	1.7679E+18	
1.76783E+18	Wed Mar 13 08:21:00 +0000 2024	4	TPN Ganjar Pranowo-Mahfud MD mengontarai kemenangan Prabowo-Gibran versi real cou	1.7678E+18	https://pbs.twimg.com/ext_tw_video_thumb/1767...
1.76668E+18	Sun Mar 10 04:20:00 +0000 2024	2	Sudah nyata sekali hasil dari Quick Count Merupakan bukti ilmiah dari kemenangan Prabowo (	1.7667E+18	https://pbs.twimg.com/media/GISBPH4XI AESMzr.jj...

Gambar 5. Hasil Crawling Data

Setelah hasil *crawling data* didapatkan, selanjutnya dilakukan pelabelan data untuk menetapkan nilai sentimen dari data *tweet* yang diperoleh (Juanita, 2020). Proses labeling pada data penelitian ini dilakukan secara manual tanpa adanya kamus khusus. Pembagian label terdiri dari dua klasifikasi, yaitu positif dan negatif. Dari hasil pelabelan data *tweet* secara manual pada gambar diatas, didapatkan kategori positif sebanyak 897 label dan kategori negatif sebanyak 252 label.

Tahapan selanjutnya yaitu *pre-processing*. Fase pengolahan data ini bertujuan untuk mempermudah analisis pada tahap selanjutnya (Supriyanto, Alita, & Isnain, 2023). Pada tahapan ini, proses pra-pemrosesan dilakukan pada perangkat lunak *rapidminer*. Berikut adalah langkah-langkah *pre-processing* yang dilakukan dalam penelitian ini:

Proses pembersihan data adalah dengan penghapusan karakter tertentu, tanda baca, dan *URL* tambahan yang tidak relevan atau tidak diperlukan dari setiap *tweet* karena biasanya tidak memberikan kontribusi signifikan terhadap analisis sentimen.

Tabel 1. Hasil Data Cleaning

Sebelum Data Cleaning	Setelah Data Cleaning
Kami mengucapkan selamat kepada @prabowo dan @gibran_tweet telah memenangkan Pilpres satu putaran saja dengan kira-kira 58% suara (hasil real count 82% foto TPS di https://t.co/dTC6Wm2Bxu). Selamat menyambut hilal Ramadhan dan menunaikan ibadah di bulan suci ini. Semoga berkah.	Kami mengucapkan selamat kepada dan telah memenangkan Pilpres satu putaran saja dengan kira-kira 58 suara hasil real count 82 foto TPS di Selamat menyambut hilal Ramadhan dan menunaikan ibadah di bulan suci ini. Semoga berkah.

Langkah selanjutnya yaitu *case folding* atau normalisasi teks, dalam hal ini dilakukan perubahan semua karakter atau teks menjadi huruf kecil untuk menghindari perbedaan antara huruf besar dan huruf kecil (Supriyanto et al., 2023). Tabel dibawah ini menampilkan hasil daripada *case folding*.

Tabel 2. Hasil Case Folding

Sebelum Case Folding	Setelah Case Folding
Kami mengucapkan selamat kepada dan telah memenangkan Pilpres satu putaran saja dengan kira-kira 58 suara hasil real count 82 foto TPS di Selamat menyambut hilal Ramadhan dan menunaikan ibadah di bulan suci ini. Semoga berkah.	kami mengucapkan selamat kepada dan telah memenangkan pilpres satu putaran saja dengan kira-kira 58 suara hasil real count 82 foto tps di selamat menyambut hilal ramadhan dan menunaikan ibadah di bulan suci ini. semoga berkah.

Proses selanjutnya adalah *tokenizing* atau tokenisasi. Dalam hal ini dilakukan pembagian kalimat menjadi kata-kata terpecah. Dengan prosedur ini, kalimat bisa diurai menjadi kata terpisah, memungkinkan analisis lebih mendalam terhadap tingkatan kata.

Tabel 3. Hasil Tokenizing

Sebelum Tokenizing	Setelah Tokenizing
kami mengucapkan selamat kepada dan telah memenangkan pilpres satu putaran saja dengan kira-kira 58 suara hasil real count 82 foto tps di selamat menyambut hilal ramadhan dan menunaikan ibadah di bulan suci ini. semoga berkah.	kami mengucapkan selamat kepada telah memenangkan pilpres satu putaran saja dengan kira kira suara hasil real count foto selamat menyambut hilal ramadhan menunaikan ibadah bulan suci semoga berkah

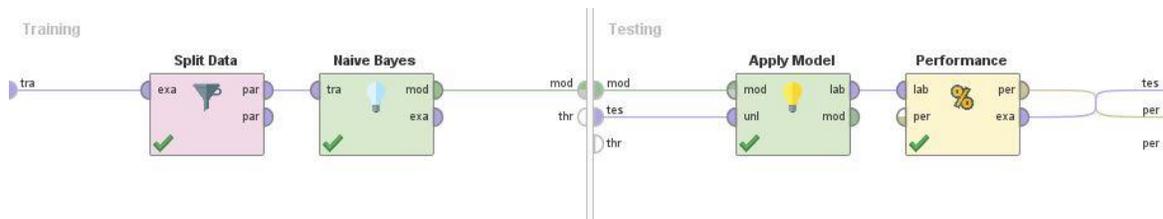
Untuk meningkatkan kualitas data teks, ditambahkan operator *Filter Tokens by Length* setelah *tokenizing*. Operator ini berfungsi menghilangkan *token* yang terlalu pendek atau terlalu panjang. Dengan menyaring *token* berdasarkan panjangnya, data menjadi lebih bersih dan relevan. Dalam penelitian ini, panjang minimum *token* ditetapkan 4 karakter dan maksimum 25 karakter.

Stopword Removal merupakan salah satu tahapan penting dalam pre-processing yang merupakan untuk menyederhanakan kata, mengurangi dimensi, dan penekanan pada kata-kata yang relevan untuk analisis (Supriyanto et al., 2023). Dalam tahapan ini, terdapat stopword list yang di dapatkan dari website Kaggle. Dengan menghapus kata-kata umum yang tidak relevan, kita bisa meningkatkan kualitas analisis teks serta kinerja model machine learning.

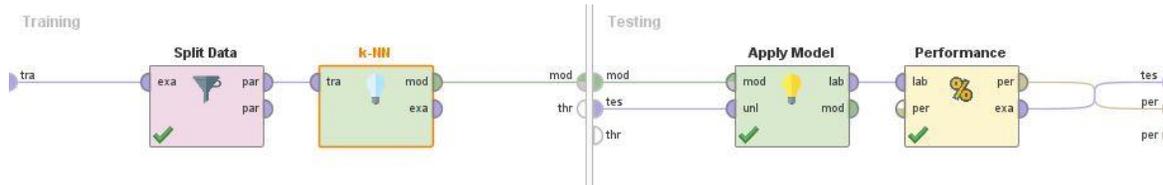
Tabel 4. Hasil Dari Stopword Removal

Sebelum Stopword Removal	Sesudah Stopword Removal
kami mengucapkan selamat kepada telah memenangkan pilpres satu putaran saja dengan kira kira suara hasil real count foto selamat menyambut hilal ramadhan menunaikan ibadah bulan suci semoga berkah	selamat memenangkan pilpres putaran suara hasil real count foto selamat menyambut hilal ramadhan menunaikan ibadah suci semoga berkah

Pada tahapan ini dilakukan pengujian pada kedua metode dalam hal ini *Naïve Bayes* dan *K-Nearest Neighbor*. Sebelum proses uji dilakukan, dataset dipisahkan menjadi data pelatihan dan data pengujian. Dalam tahapnya, operator *split data* ditambahkan kedalam proses *Rapidminer*. Operator *split data* dikonfigurasi untuk menetapkan rasio pembagian data. Pada penelitian ini, peneliti menggunakan rasio 90:10 (90% data training, 10% data testing). Peneliti juga menerapkan beberapa pembagian rasio sebagai perbandingan yaitu 80:20 (80% data training, 20% data testing), 70:20 (70% data training, 20% data testing), dan 60:40 (60% data training, 40% data testing). Parameter Cross-Validation digunakan dalam proses pengujian. Didalam *Rapidminer*, parameter Cross-Validation berfungsi untuk menentukan atau menilai akurasi suatu metode.



Gambar 6. Proses Uji Metode Naive Bayes



Gambar 7. Proses Uji Metode K-NN

#### 4. Pembahasan

Tahapan terakhir dari proses penelitian ini adalah evaluasi model dari hasil pengujian yang dilakukan pada teknik *Naive Bayes* dan *K-Nearest Neighbor*. Evaluasi hasil dari kedua pendekatan ini melibatkan penggunaan *confusion matrix* yang memberikan nilai akurasi, presisi, dan *recall* dari data uji (Khotimah & Utami, 2022). Untuk menghasilkan nilai yang optimal, penelitian ini menerapkan *k-fold Cross Validation* dengan nilai  $k=10$  (Palepa, Pratiwi, & ..., 2024). Evaluasi model ini penting untuk memahami seberapa akurat model dapat memprediksi sentimen. Berikut penjelasan dari hasil evaluasi masing-masing algoritma :

accuracy: 87.35% +/- 1.81% (micro average: 87.35%)

	true Positif	true Negatif	class precision
pred. Positif	686	16	97.72%
pred. Negatif	211	881	80.68%
class recall	76.48%	98.22%	

Gambar 8. Confusion Matrix Pada Metode Naive Bayes

Berdasarkan gambar diatas, menunjukkan akurasi yang dihasilkan dari model *Naive Bayes Classifier* adalah 87.35% dengan *margin error* +/- 1.81% (*micro average: 87.35%*). Hasil evaluasi menunjukkan bahwa algoritma *Naive Bayes* mampu mengklasifikasikan data secara memuaskan (Mufidah, Winarno, Alzami, Udayanti, & Sani, 2022). Rincian hasil evaluasi diatas antara lain *True Positif* (TP) sebanyak 686 merupakan besar *tweet* positif yang berhasil diprediksi dengan benar, *True Negatif* (TN) sebanyak 881 merupakan besar *tweet* negatif yang berhasil diprediksi dengan benar. Model salah memprediksi 16 *tweet* negatif sebagai positif (*False Positif* (FP)), *False Negatif* (FN) sebanyak 211 merupakan *tweet* positif yang diprediksi sebagai negatif. Selanjutnya *Precision* untuk kelas positif mencapai 97.72% yang menampilkan persentase *tweet* yang sebenarnya positif dari semua *tweet* yang diprediksi positif oleh model, sedangkan *Precision* untuk kelas negatif mencapai 80.68% menunjukkan persentase *tweet* yang memang sebenarnya negatif dari semua *tweet* yang diprediksi negatif oleh model.

Terakhir adalah *Recall* untuk kelas positif senilai 76.48% yang artinya persentase *tweet* positif yang berhasil diprediksi benar dari seluruh *tweet* positif sebenarnya dan untuk kelas negatif senilai 98.22%, menunjukkan persentase *tweet* negatif yang berhasil diprediksi dengan benar dari seluruh *tweet* negatif sebenarnya. Hasil Evaluasi menampilkan bahwa algoritma *Naive Bayes* memiliki kemampuan yang baik dalam mengidentifikasi kelas negatif (*Recall: 98.22%*) dibandingkan dengan kelas positif (*Recall: 76.48%*). Dalam hal ini *Precision*

positif juga sangat tinggi (*Precision*: 97.72%), mengindikasikan bahwa prediksi positif yang dibuat oleh algoritma ini sebagian besar tepat atau benar.

accuracy: 69.68% +/- 3.14% (micro average: 69.68%)

	true Positif	true Negatif	class precision
pred. Positif	360	7	98.09%
pred. Negatif	537	890	62.37%
class recall	40.13%	99.22%	

Gambar 9. Confusion Matrix Pada K-Nearest Neighbor

Berdasarkan gambar diatas, menunjukkan akurasi yang dihasilkan dari model *K-Nearest Neighbor* adalah 69.68% dengan *margin error* +/- 3.14% (*micro average*: 69.68%). Evaluasi ini menunjukkan bahwa algoritma *K-Nearest Neighbor* memiliki performa lebih rendah dibandingkan dengan algoritma *Naïve Bayes* (Ramadhani & Wahyudin, 2022). Rincian hasil evaluasi diatas antara lain *True Positif* (TP) sebanyak 360 merupakan besar *tweet* positif yang berhasil diprediksi dengan benar, *True Negatif* (TN) sebanyak 890 merupakan besar *tweet* negatif yang berhasil diprediksi dengan benar. Model salah memprediksi 7 *tweet* negatif sebagai positif (*False Positif* (FP)), *False Negatif* (FN) sebanyak 537 merupakan *tweet* positif yang diprediksi sebagai negatif. Selanjutnya *Precision* untuk kelas positif mencapai 98.09% yang menunjukkan persentase *tweet* yang benar-benar positif dari semua *tweet* yang diprediksi positif oleh model, sedangkan *Precision* untuk kelas negatif mencapai 62.37% menunjukkan persentase *tweet* yang memang sebenarnya negatif dari semua *tweet* yang diprediksi negatif oleh model.

Terakhir adalah *Recall* untuk kelas positif senilai 40.13% yang artinya persentase *tweet* positif yang berhasil diprediksi benar dari seluruh *tweet* positif sebenarnya dan untuk kelas negatif senilai 99.22%, menunjukkan persentase *tweet* negatif yang berhasil diprediksi dengan benar dari seluruh *tweet* negatif sebenarnya. Untuk hasil pengujian menggunakan beberapa perbandingan split data pada kedua algoritma yang juga dilakukan oleh peneliti dapat diperhatikan pada tabel dibawah.

Tabel 5. Hasil Dengan Beberapa Perbandingan Split data

	Naïve Bayes			K-Nearest neighbor		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
70:30	86.57%	95.18%	77.03%	67.12%	98.42%	34.78%
60:40	86.73%	93.76%	78.71%	65.83%	98.63%	32.11%
80:20	86.68%	96.08%	76.48%	68.56%	98.54%	37.68%

Gambar diatas memperlihatkan hasil perbandingan kinerja algoritma *Naïve Bayes* dan K-Nearest Neighbor (K-NN) pada tiga skenario pembagian data yakni : 70:30, 60:40, dan 80:20. Dari tabel tersebut, terlihat bahwa *Naïve Bayes* secara konsisten mencapai akurasi lebih tinggi di semua skenario dibandingkan dengan K-NN, dengan nilai akurasi sekitar 86% sedangkan K-NN berkisar 65-68%. Selain itu, *Naïve Bayes* juga menunjukkan nilai precision dan recall yang lebih seimbang. Meskipun precision K-NN sangat tinggi (diatas 98%), namun recall nya sangat rendah (sekitar 32-37%). Ini menunjukkan bahwa *Naïve Bayes* lebih andal dalam mengidentifikasi *tweet* yang relevan dengan sentimen tertentu secara keseluruhan, sementara K-NN cenderung lebih selektif dan sering gagal.

Evaluasi memanfaatkan confusion matrix diatas mengindikasikan bahwa algoritma *Naïve Bayes* menunjukkan kinerja yang unggul dalam menganalisis pandangan publik terhadap kemenangan paslon 02 pada Pemilu 2024. Dengan tingkat akurasi yang cukup tinggi, model ini memperlihatkan kemampuan yang lebih baik dalam memprediksi data negatif dibandingkan data positif, yang ditunjukkan oleh nilai *Recall* dan *Precision* yang lebih

tinggi untuk kelas negatif. Hasil tersebut mendukung bahwa *Naïve Bayes* merupakan algoritma yang efektif untuk analisis sentimen dalam konteks penelitian ini (Ramadhani & Wahyudin, 2022).

Sementara untuk hasil evaluasi menggunakan *confusion matrix* diatas pada algoritma *K-Nearest Neighbor* menunjukkan algoritma tersebut memiliki performa yang lebih rendah jika dikomparasikan bersama *Naïve Bayes* dalam melakukan analisis sentimen opini publik di *Twitter* terhadap kemenangan paslon 02 pada Pemilu 2024. Hasil pengujian menunjukkan bahwa meskipun *K-NN* memiliki *Precision* yang sangat tinggi untuk kelas positif, *Recall* yang rendah untuk kelas positif dan *Precision* yang rendah untuk kelas negatif menyoroti keterbatasan algoritma ini dalam mendeteksi *tweet* positif dan negatif secara konsisten.

Terkait kecenderungan sentimen masyarakat berdasarkan hasil evaluasi menggunakan algoritma *Naïve Bayes* serta *K-Nearest Neighbor*, dapat ditentukan sentimen publik terhadap kemenangan paslon 02 pada Pemilu 2024. Untuk menentukan apakah sentimen publik lebih ke positif atau negatif, kita dapat menghitung persentase masing-masing kelas berdasarkan prediksi dari algoritma *Naïve Bayes* yang memiliki akurasi lebih tinggi. Dari total *dataset* sebesar 1794 (total dari *Naïve Bayes confusion matrix*), sentimen positif terdiri dari 702 (TP + FP = 686 + 16) jika dihitung  $(702 / 1794) * 100 = 39.12\%$ , sedangkan sentimen negatif terdiri dari 1092 (TN + FN = 881 + 211 = 1092) jika dihitung  $(1092 / 1794) * 100 = 60.88\%$ . Dari analisis tersebut dapat disimpulkan bahwa sentimen publik terhadap kemenangan paslon 02 lebih banyak yang negatif dibandingkan yang positif. Sentimen positif adalah 39.12%, sedangkan sentimen negatif sebesar 60.88%. ini menunjukkan bahwa meskipun ada dukungan positif terhadap kemenangan paslon 02, mayoritas opini publik di *Twitter* cenderung negatif.

## 5. Kesimpulan

Berdasarkan dari hasil penelitian yang telah dilakukan sebelumnya, dapat diperoleh beberapa Kesimpulan. Dalam studi ini, algoritma *Naïve Bayes Classifier* menunjukkan nilai keakuratan yang lebih unggul dibandingkan dengan *K-Nearest Neighbor*. Tingkat ketepatan yang didapatkan *Naïve Bayes* sebesar 87.35% dengan deviasi +/- 1.81% (*micro average*: 87.35%), sementara akurasi yang didapatkan *K-NN* hanya mencapai 69.68% dengan deviasi +/-3.14% (*micro average*: 69.68%). Dalam penelitian ini menunjukkan bahwa sentimen opini publik di *Twitter* cenderung negatif daripada positif dalam merespon kemenangan paslon 02 pada Pemilu 2024. Dalam penelitian ini penggunaan rasio pembagian data 90:10 terbukti efektif dalam memberikan hasil yang konsisten untuk kedua algoritma. Walaupun rasio lain juga diuji pada penelitian ini, rasio 90:10 memberikan hasil yang paling optimal. Dalam penelitian ini menerangkan bahwa media sosial khususnya *Twitter*, memiliki peranan yang penting dalam mencerminkan opini publik terhadap hasil Pemilu 2024, menegaskan signifikansi *platform* ini dalam konteks politik. Dalam penelitian ini langkah-langkah *preprocessing* seperti *case folding*, *tokenizing*, *filtering tokens by length*, dan *stopword removal* efektif dalam meningkatkan kualitas data untuk analisis sentimen, sehingga berkontribusi pada hasil yang lebih akurat.

Berdasarkan dari hasil penelitian diatas, saran yang dapat disampaikan oleh peneliti untuk penelitian mendatang. Penggunaan *dataset* yang lebih besar dan beragam dapat meningkatkan kemampuan generalisasi model. Hal tersebut cukup membantu dalam memperoleh hasil yang lebih representatif dan akurat. Untuk lebih memperbaiki kualitas data teks sebelum dianalisis, penelitian selanjutnya dapat mengembangkan tahapan *preprocessing* lebih lanjut seperti menggunakan teknik *stemming* atau *lemmatization*. Selain menerapkan pendekatan *Naïve Bayes Classifier* dan *K-Nearest Neighbor*, disarankan untuk mengeksplorasi algoritma lain seperti *Support Vector Machine (SVM)*, *Random Forest*, atau *Deep Learning* untuk membandingkan kinerja dan efektivitasnya dalam analisis sentimen.

## Referensi

Ghufon, M. R., Mahabbataka Arsyada, M. F., Lukman, M. R., Haryono Putra, Y. A., & Rakhmawati, N. A. (2023). Analisis Sentimen Pengguna Twitter Terhadap Pemilu 2024 Berbasis Model XLM-T. *J-Intech*, 11(2), 307–315. <https://doi.org/10.32664/j-intech.v11i2.1013>

- Juanita, S. (2020). Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. *Jurnal Media Informatika Budidarma*, 4(3), 552. <https://doi.org/10.30865/mib.v4i3.2140>
- Khotimah, A. C., & Utami, E. (2022). Comparison Naive Bayes Classifier, K-Nearest Neighbor, and Support Vector Machine in the classification of individual on twitter account. *Jurnal Teknik Informatika (JUTIF)*, 3(3), 673–680.
- Maulana, G. I., Fridawati, P. I., & Masruroh, H. (2022). Pemilu 2024: Pesta demokrasi akbar pembangkit kondisi ekonomi pasca pandemi. *Jurnal Integrasi Dan Harmoni Inovatif Ilmu-Ilmu Sosial (JIHI3S)*, 2(8), 751–757. <https://doi.org/10.17977/um063v2i8p751-757>
- Mufidah, F. S., Winarno, S., Alzami, F., Udayanti, E. D., & Sani, R. R. (2022). Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter Dengan Algoritma Naive Bayes Classifier. *JOINS (Journal of Information System)*, 7(1), 14–25. <https://doi.org/10.33633/joins.v7i1.5883>
- Murni, M., Riadi, I., & Fadlil, A. (2023). Analisis Sentimen HateSpeech pada Pengguna Layanan Twitter dengan Metode Naive Bayes Classifier (NBC). *JURIKOM (Jurnal Riset Komputer)*, 10(2), 566. <https://doi.org/10.30865/jurikom.v10i2.5984>
- Mutiara, T. A., Alkhalifi, Y., Mayangky, N. A., & Gata, W. (2020). Analisis Sentimen Opini Publik Mengenai Larangan Mudik pada Twitter Menggunakan Naive Bayes. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, Vol. 6, p. 85. <https://doi.org/10.24014/coreit.v6i2.9727>
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>
- Palepa, M. J., Pratiwi, N., & ... (2024). Analisis Sentimen Masyarakat Tentang Pengaruh Politik Identitas Pada Pemilu 2024 Terhadap Toleransi Beragama Menggunakan Metode K-Nearest Neighbor. *JUPI (Jurnal ..., 9(1), 389–401.*
- Puad, S., Garno, G., & Susilo Yuda Irawan, A. (2023). Analisis Sentimen Masyarakat Pada Twitter Terhadap Pemilihan Umum 2024 Menggunakan Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1560–1566. <https://doi.org/10.36040/jati.v7i3.6920>
- Ramadhani, S. H., & Wahyudin, M. I. (2022). Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naive Bayes dan K-NN. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(4), 526–534. <https://doi.org/10.35870/jtik.v6i4.530>
- Supriyanto, J., Alita, D., & Isnain, A. R. (2023). Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 4(1), 74–80. <https://doi.org/10.33365/jatika.v4i1.2468>
- Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN. *Jurnal KomtekInfo*, 10, 1–7. <https://doi.org/10.35134/komtekinfo.v10i1.330>
- Syarifuddin, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn. *INTI Nusa Mandiri*, 15(1), 23–28. <https://doi.org/10.33480/inti.v15i1.1347>
- Wahyu Andrian, B., Adline Twince Tobing, F., Zuhdi Pane, I., & Kusnadi, A. (2023). Implementation of Naive Bayes Algorithm in Sentiment Analysis of Twitter Social Media Users Regarding Their Interest to Pay the Tax. *International Journal of Science, Technology & Management*, 4(6), 1733–1742. <https://doi.org/10.46729/ijstm.v4i6.1015>