

PREDICTION OF 2024 PRESIDENTIAL ELECTION USING K-NN WITH METRIC APPROACHES CHEBYSHEV AND EUCLIDEAN BASED ON TWITTER DATA INVESTIGATION

Steven Ryan Darmawan^{*1}, Muhamad Fatchan², Donny Maulana³

^{1,2,3}Informatics, Universitas Pelita Bangsa, Indonesia

Email: ¹stevenryandarmawan@mhs.pelitabangsa.ac.id, ²fatchan@pelitabangsa.ac.id,
³donny.maulana@pelitabangsa.ac.id

(Article received: January 11, 2024; Revision: January 25, 2024; Published: April 04, 2024)

Abstract

The potential difference between the popularity of presidential candidates on social media and in the general public poses a serious challenge in predicting the outcome of the 2024 presidential election. Technical constraints in collecting, cleaning and analyzing dynamic and large-scale social media data can threaten the accuracy and validity of predictions. To overcome this problem, careful steps and in-depth understanding are needed. Therefore, this study aims to predict the winner of the 2024 presidential election from the popularity of presidential candidates Anies Baswedan, Ganjar Pranowo, and Prabowo Subianto on Twitter. The K-Nearest Neighbor (K-NN) method with the Both Metric approach (Euclidean and Chebyshev) was used to analyze 51,192 tweet data through the Knowledge Discovery in Database (KDD) stage using Orange software. The evaluation results show almost the same performance, with AUC values of 0.725 for Euclidean and 0.720 for Chebyshev. The CA result was 55.6% for Euclidean and 55.4% for Chebyshev. Although F1, precision, and recall were almost the same, overall, the Euclidean metric was better. The prediction shows Prabowo Subianto as the most popular candidate on Twitter. Nonetheless, these results need to be interpreted with caution and strengthened with further analysis and additional data to get a more comprehensive conclusion. This research shows that K-NN with both metrics can provide predictions above 50%, reliable enough to be able to predict the most popular candidates on Twitter.

Keywords: 2024 Presidential Election, Chebyshev, Euclidean, K-Nearest Neighbor, Twitter.

PREDIKSI PEMENANGAN CAPRES 2024 MENGGUNAKAN K-NN DENGAN PENDEKATAN METRIK CHEBYSHEV DAN EUCLIDEAN BERBASIS INVESTIGASI DATA TWITTER

Abstrak

Perbedaan potensial antara popularitas calon presiden di media sosial dan di masyarakat umum menjadi tantangan serius dalam memprediksi hasil Pilpres 2024. Kendala teknis dalam pengumpulan, pembersihan, dan analisis data media sosial yang dinamis dan berskala besar dapat mengancam akurasi dan validitas prediksi. Untuk mengatasi masalah ini, langkah-langkah hati-hati dan pemahaman mendalam diperlukan. Oleh karena itu, penelitian ini bertujuan untuk memprediksi pemenang Pilpres 2024 dari popularitas calon presiden Anies Baswedan, Ganjar Pranowo, dan Prabowo Subianto di Twitter. Metode K-Nearest Neighbor (K-NN) dengan pendekatan Both Metric (Euclidean dan Chebyshev) digunakan untuk menganalisis 51.192 data tweet melalui tahap Knowledge Discovery in Database (KDD) menggunakan perangkat lunak Orange. Hasil evaluasi menunjukkan performa yang hampir sama, dengan nilai AUC sebesar 0.725 untuk Euclidean dan 0.720 untuk Chebyshev. Hasil CA adalah 55,6% untuk Euclidean dan 55,4% untuk Chebyshev. Meskipun F1, presisi, dan recall hampir sama, secara keseluruhan, metrik Euclidean lebih baik. Prediksi tersebut menunjukkan Prabowo Subianto sebagai kandidat paling populer di Twitter. Meskipun demikian, hasil ini perlu ditafsirkan dengan hati-hati dan diperkuat dengan analisis lebih lanjut dan data tambahan untuk mendapatkan kesimpulan yang lebih komprehensif. Penelitian ini menunjukkan bahwa K-NN dengan kedua metrik tersebut dapat memberikan prediksi di atas 50%, cukup dapat diandalkan untuk memprediksi pemenang Pilpres 2024 melalui data Twitter.

Kata kunci: 2024 Pemilihan Presiden, Chebyshev, Euclidean, K-Nearest Neighbor, Twitter.

1. PENDAHULUAN

Penelitian ini berfokus pada prediksi kemenangan calon presiden pada Pilpres 2024 di Indonesia sebagai objek penelitiannya. Pilpres merupakan momen penting dalam sistem demokrasi negara ini karena rakyat memiliki kesempatan untuk memilih pemimpinnya secara langsung. Dalam pilpres, beberapa calon presiden dari berbagai partai politik berkompetisi untuk memenangkan dukungan rakyat, baik secara langsung, maupun tidak langsung melalui media sosial seperti *Twitter* [1]. *Twitter* adalah platform media sosial yang memungkinkan individu untuk berinteraksi satu sama lain dan memperoleh informasi terkini atau berita [2]. Para peneliti sering kali menggunakan *Twitter* untuk mengambil data karena kecepatan platform ini, dengan batasan karakter maksimal 280 karakter, memungkinkan penyampaian pesan secara langsung dan singkat [3]. Oleh karena itu, penelitian dapat mengembangkan model prediktif menggunakan algoritma *K-Nearest Neighbor* (K-NN) dengan pendekatan *Both Metric* (*Chebyshev* dan *Euclidean*), untuk memperkirakan pemenang pilpres berdasarkan popularitas calon presiden di *Twitter* [4].

Di era digital saat ini, media sosial telah menjadi elemen yang tidak terpisahkan dari kehidupan sehari-hari masyarakat, termasuk dalam ranah politik [5]. Namun, terdapat potensi adanya perbedaan antara popularitas calon presiden di media sosial dengan popularitas yang sebenarnya di masyarakat luas. Hasil analisis dari media sosial khususnya *Twitter* bersifat dinamis, karena itu secara terus-menerus mengevaluasi pandangan masyarakat terhadap calon presiden tahun 2024 dapat meningkatkan akurasi penggabungan hasil penelitian yang berkaitan dengan persepsi masyarakat [6].

Selain itu, penelitian dapat menghadapi kendala teknis dalam mengumpulkan, membersihkan, dan mengolah data dari media sosial yang bersifat dinamis dan besar skala [7]. Diperlukan pengolahan data *tweet* untuk mengidentifikasi perbedaan antara respons yang bersifat positif, negatif, dan netral bagi seseorang yang ingin memahami tanggapan masyarakat di platform *Twitter* [8]. Seluruh permasalahan ini dapat berdampak pada akurasi dan validitas prediksi pemenang Pilpres 2024. Oleh karena itu, diperlukan langkah-langkah yang cermat dan pemahaman mendalam dalam mengatasi permasalahan tersebut untuk memastikan hasil prediksi yang lebih akurat dan reliabel.

Untuk menangani permasalahan tersebut, teknologi analisis data dan pendekatan berbasis komputasi dapat digunakan sebagai solusi potensial yakni salah satunya data mining menggunakan algoritma K-NN dengan *Both Metric* (*Euclidean* dan *Chebyshev*) melalui *software Orange*. Teknologi ini dapat meningkatkan pemahaman kita tentang pola dukungan dan pergerakan popularitas calon presiden

dalam kampanye politik [9]. Dalam penelitian ini, algoritma *K-Nearest Neighbor* (K-NN) dipilih sebagai metode prediksi karena kelebihanannya dalam melakukan klasifikasi berdasarkan data terdekat atau dapat menyamaratakan himpunan data training yang relatif kecil [10]. Dengan pendekatan *Both Metric* (*Chebyshev* dan *Euclidean*), akan digunakan untuk mengukur jarak antara data, baik dari perbedaan maksimum maupun akar kuadrat dari jumlah kuadrat perbedaan. Kedua metrik ini akan membantu mengukur perbedaan dan kesamaan popularitas antara calon presiden melalui data *Twitter* [11].

Terdapat penelitian relevan yaitu penelitian oleh Saputra dan Primadasa pada tahun 2018 yang berjudul "Penerapan Teknik Klasifikasi untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor". Penelitian tersebut menggunakan algoritma K-NN untuk memprediksi kelulusan mahasiswa berdasarkan data Index Prestasi semester. Meskipun berfokus pada konteks pendidikan, penelitian tersebut dijadikan referensi karena terdapat kesamaan penerapan algoritma K-NN. Namun, penelitian tersebut memiliki catatan, seperti penggunaan data training yang kurang efektif dan pengaruh jumlah data *training* terhadap hasil prediksi. Oleh karena itu, adaptasi dan pertimbangan perlu dilakukan dalam penelitian tentang Pilpres 2024 di *Twitter* [12].

Selain itu, penelitian lain yang relevan telah dilakukan dalam konteks analisis politik menggunakan algoritma K-NN yaitu penelitian oleh Badrul tahun 2015 dengan judul "Prediksi Hasil Pemilu Legislatif dengan Menggunakan Algoritma *K-Nearest Neighbor*". Meskipun data yang digunakan berbeda, yaitu data KPUD bukan data dari *Twitter*, hasil penelitian menunjukkan model algoritma K-NN memiliki akurasi yang baik sebesar 81.35% dengan tingkat diagnostik *Excellent Classification*. Dengan demikian, penelitian ini dapat memberikan wawasan dan pemahaman penggunaan metode K-NN dalam konteks prediksi dan perlu melakukan adaptasi yang sesuai dengan tujuan penelitian tentang Pilpres 2024 di *Twitter* [13].

Penelitian yang relevan yakni penelitian oleh Aulia Putri, dkk (2023), yang berjudul "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir", tidak berkaitan langsung dengan pemilihan presiden, namun memiliki relevansi dalam membandingkan K-NN dengan metode lain. Dalam penelitian tersebut, dilakukan analisis, pengujian, dan evaluasi terhadap kinerja algoritma K-NN, NBC, dan SVM dalam meramalkan kelulusan mahasiswa tingkat akhir menggunakan 379 data publik. Hasilnya menunjukkan bahwa K-NN memiliki kinerja lebih tinggi dengan tingkat akurasi 87,8%, presisi 87,8%, dan recall 84%. Dalam perbandingan keseluruhan, K-NN dianggap lebih efektif dibandingkan dengan *Naive Bayes* dan *Support Vector Machine* dalam

memprediksi kelulusan mahasiswa tingkat akhir [14].

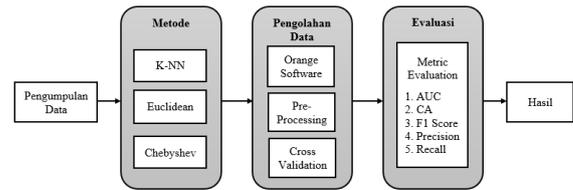
Selanjutnya, penelitian lain yang memiliki relevansi yakni dari Sudriyanto, Syahro, dan Fitriani tahun 2023 dengan judul “Perbandingan Performa Model Machine Learning Support Vector Machine, Neural Network, dan K-Nearest Neighbors dalam Prediksi Harga Saham”, tidak berkaitan langsung dengan pemilihan presiden, namun membandingkan K-NN dengan metode lain. Hasil penelitian menunjukkan bahwa model prediksi K-Nearest Neighbors (K-NN) memiliki nilai RMSE terendah, yaitu 0.037, menunjukkan akurasi tertinggi dalam memprediksi harga saham PT Astra International Tbk (ASII.JK). Meskipun Neural Network (NN) memiliki RMSE sedikit lebih tinggi (0.048), model ini tetap memberikan prediksi yang cukup akurat. Sementara itu, model Support Vector Machine (SVM) memiliki RMSE yang lebih tinggi, mencapai 0.075, menunjukkan tingkat kesalahan prediksi yang lebih tinggi dibandingkan dengan K-NN dan NN. Berdasarkan penelitian tersebut rekomendasi terbaik untuk menganalisis harga saham adalah menggunakan model K-NN [15].

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini memiliki tujuan untuk mengembangkan model prediktif yang akurat memprediksi pemenang Pilpres 2024 berdasarkan popularitas calon presiden melalui data Twitter. Dengan menggunakan algoritma K-NN serta pendekatan *Both Metric* (*Chebyshev* dan *Euclidean*), diharapkan penelitian ini dapat memberikan kontribusi penting dalam pemahaman potensi algoritma K-NN sebagai alat prediksi politik. Hasil penelitian diharapkan dapat memberikan manfaat bagi para pemangku kepentingan politik dalam merancang strategi kampanye yang lebih efektif, serta memberikan pemahaman yang lebih baik kepada masyarakat umum mengenai potensi pemenang pilpres berdasarkan popularitas calon presiden berdasarkan data di *Twitter*.

2. METODE PENELITIAN

Pada Gambar 1, tahapan penelitian dimulai dengan mengambil data dari Twitter mengenai kandidat calon presiden 2024. Metode yang digunakan adalah *K-Nearest Neighbor* (K-NN) dengan *Euclidean* dan *Chebyshev*. Kemudian melalui tahap pengolahan data dengan menggunakan *software Orange* untuk melakukan *pre-processing* guna membersihkan dan menyiapkan data sebelum dilakukan analisis lebih lanjut.

Setelah proses *pre-processing*, data akan melewati proses *cross validation* untuk menentukan data training dan data *testing* secara statistik. Tahap berikutnya adalah evaluasi hasil, digunakan untuk menghitung nilai AUC, CA, *F1 score*, *precision*, dan *recall*, yang hasilnya untuk mengevaluasi akurasi model.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Penelitian ini merupakan jenis penelitian kuantitatif, yang berfokus pada pengumpulan dan analisis data numerik dengan pendekatan matematis atau menggunakan teknik statistik [16]. Objek penelitian ini adalah analisis data berdasarkan popularitas kandidat calon presiden 2024 dari media sosial *Twitter*. Penelitian ini menggunakan metode K-NN dengan pendekatan *Both Metric* (*Chebyshev* dan *Euclidean*), data dari *Twitter* diinvestigasi untuk memprediksi calon presiden yang berpotensi menang.

Penelitian ini menggunakan data sekunder jenis dokumenter, dalam bentuk arsip yang diperoleh menggunakan bahasa pemrograman *Python* dan alat *Snsrape* untuk mengambil data dari *Twitter*. *Snsrape* adalah sebuah *library Python* yang memungkinkan pengambilan data tweet secara langsung melalui API *Twitter*. Data yang dikumpulkan mencakup 51.192 *tweet* di *Twitter*, termasuk jumlah *like*, *retweet*, dan *reply* yang diambil dalam periode tertentu dari Januari 2021 hingga Juli 2023 untuk menggambarkan tren popularitas calon presiden menjelang Pilpres 2024.

Kata kunci seperti "Ganjar Presiden," "Prabowo Presiden," "Anies Presiden," "Calon Presiden 2024," "Capres2024," dan "Presiden2024" digunakan sebagai parameter pencarian. Selain itu, peneliti juga mengutip sumber-sumber dari beberapa jurnal dan hasil penelitian relevan. Dengan menggunakan alat *Snsrape* dan bahasa pemrograman *Python*, pengumpulan data dilakukan secara efisien dan dapat diandalkan untuk memperoleh data tweet terkait popularitas calon presiden 2024. *Snsrape* memungkinkan peneliti untuk mengatur output data dalam format yang sesuai, seperti *file CSV*, untuk memfasilitasi analisis lebih lanjut.

2.2. Metode yang Digunakan

2.2.1. K-Nearest Neighbor (K-NN)

Penelitian dengan metode *K-Nearest Neighbor* (K-NN) menggunakan algoritma yang terarah, di mana *instance query* yang baru diklasifikasikan berdasarkan mayoritas label kelas pada tetangga terdekatnya. Dalam konteks prediksi pemenangan Capres, K-NN akan digunakan untuk mengklasifikasikan calon berdasarkan kesamaan fitur dengan data *tweet* yang dianalisis. *Euclidean* dan *Chebyshev* menjadi perhitungan jarak yang populer digunakan dalam K-NN [17].

2.2.2. Euclidean

Jarak *Euclidean* adalah metrik yang digunakan untuk mengukur jarak antara dua titik dalam ruang multidimensi [18]. Rumus jarak *Euclidean* adalah sebagai berikut.

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_{ik} - y_{jk})^2} \quad (1)$$

Keterangan:

- $d_{(x,y)}$ = Jarak antara objek x ke objek y
- x_{ik} = Rata-rata nilai objek ke- i berdasarkan atribut ke- k
- y_{jk} = Rata-rata nilai objek ke- j berdasarkan atribut ke- k
- i = Banyaknya objek
- m = Banyaknya atribut

2.2.3. Chebyshev

Chebyshev adalah metode yang digunakan untuk mengukur jarak antara dua objek berdasarkan nilai absolut dari selisih koordinat mereka. Rumus jarak *Chebyshev* mengikuti definisi sektor maksimum, di mana jarak antara dua sektor adalah perbedaan terbesar di antara mereka sepanjang setiap dimensi koordinat [19]. Rumusnya sebagai berikut.

$$d(x, y) = \sum_{i=1}^n |x_{ik} - y_{jk}| \quad (2)$$

Keterangan:

- $d_{(x,y)}$ = Jarak antara objek x ke objek y
- x_{ik} = Rata-rata nilai objek ke i berdasarkan atribut ke- k
- y_{jk} = Rata-rata nilai objek ke j berdasarkan atribut ke- k
- i = Banyaknya objek
- n = Banyaknya atribut

2.3. Pengolahan Data

2.3.1. Orange Tools

Penelitian ini menggunakan *Orange* dengan mengimplementasikan antarmuka yang memungkinkan pengguna untuk membuat alur kerja dengan menghubungkan *widget* yang telah ditentukan, digunakan untuk analisis data yang memadukan kemudahan pemrograman visual dengan fleksibilitas pemrograman *Python* [20].

2.3.2. Pre-Processing

Pre-processing merujuk pada berbagai jenis proses yang dilakukan pada data mentah untuk mempersiapkan langkah prosedur berikutnya. Tujuan dari *pre-processing* untuk mentransformasi data ke dalam suatu format yang lebih mudah dan efektif untuk kebutuhan pengguna. Pembersihan data dalam proses *pre-processing* mencakup penghilangan data yang tidak lengkap (*incomplete*), penghilangan data yang mengandung *noise* (*noisy*), dan penanganan data yang tidak konsisten (*inconsistent*) [21].

Dimensionality Reduction merupakan teknik dalam pre-processing untuk meningkatkan efisiensi analisis data, sehingga membantu mempertahankan informasi penting dalam dataset, terutama pada data besar dengan banyak atribut [22]. Jadi, dalam proses pre-processing melibatkan penghilangan *ID user*, *URL tweet*, *quote*, dan tanda baca yang terdapat dalam teks tweet. Selain itu, *tweet* yang tidak relevan juga dihapus.

2.3.3. Cross Validation

Penelitian ini menggunakan *cross-validation* untuk membagi data menjadi subset latih dan uji guna mengevaluasi model atau algoritma. Metode ini melibatkan pertukaran secara bergantian antara subset latih dan uji selama proses pembagian dataset, di mana subset latih digunakan untuk pelatihan model, dan subset uji untuk pengujian kinerjanya [23].

Salah satu teknik dari *cross validation* adalah *k-fold*, di mana data dibagi menjadi K bagian dengan ukuran yang sama. Misalnya, penggunaan *k-fold=5* dalam *cross-validation* bertujuan untuk mengurangi bias pada data. Proses pelatihan dan pengujian dilakukan sebanyak K kali. Ilustrasi alur kerja *cross validation* dapat dilihat pada Gambar 2 berikut ini [24].

Iteration 1	Test	Train	Train	Train	Train
Iteration 2	Train	Test	Train	Train	Train
Iteration 3	Train	Train	Test	Train	Train
Iteration 4	Train	Train	Train	Test	Train
Iteration 5	Train	Train	Train	Train	Test

Gambar 2. Contoh Simulasi *Cross Validation*

2.4. Evaluasi

Penelitian ini menggunakan beberapa metrik evaluasi seperti *AUC*, *CA*, *F1*, *Precision*, dan *Recall*. *Classification Accuracy* atau *CA* mengukur kemampuan model dalam memprediksi keseluruhan dataset dengan benar. *Precision* menilai akurasi prediksi positif, sementara *Recall* mengukur proporsi data positif yang berhasil diprediksi. *F1 Score* digunakan untuk membandingkan keseimbangan antara *Recall* dan *Precision*, memberikan gambaran keseluruhan tentang performa model. Metrik *AUC* mengukur kualitas keseluruhan model klasifikasi berdasarkan kurva *ROC*, dengan nilai yang lebih besar menunjukkan kemampuan model yang lebih baik dalam membedakan antara kelas-kelas [25].

2.4.1. AUC (Area Under Curve)

AUC merupakan metrik umum yang digunakan untuk membandingkan kinerja prediksi dari model pengklasifikasi. Rentang nilai *AUC* adalah antara 0 hingga 1, dan semakin mendekati nilai 1 mengindikasikan bahwa model tersebut mampu melakukan klasifikasi dengan baik [26]. Cara

menghitungnya dapat dijelaskan melalui rumus berikut.

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \quad (3)$$

2.4.2. CA (Classification Accuracy)

Akurasi adalah persentase nilai yang berasal dari total data uji yang berhasil diidentifikasi secara benar [27].

Meskipun nilai akurasi tinggi, tidak selalu dapat dianggap sebagai indikator keberhasilan prediksi model, terutama ketika terdapat ketidakseimbangan kelas. Oleh karena itu, penting untuk mengamati *Recall* dan *Precision* kelas sebagai faktor kritis dalam mengevaluasi kinerja ketika melakukan klasifikasi pada data yang tidak memiliki ketidakseimbangan kelas [28]. Rumus *Accuracy* ditunjukkan oleh rumus berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Keterangan :

- TP = True Positif
- TN = True Negatif
- FP = False Positif
- FN = False Negatif

2.4.3. F1 Score

F1-Score merupakan suatu ukuran yang membandingkan *precision* dan *recall*, yang digunakan untuk mengevaluasi ketepatan pengujian data. [29]. Rumus *F1-Score* dapat dilihat di bawah ini.

$$F1\ Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (5)$$

2.4.4. Precision

Merupakan perbandingan antara jumlah prediksi positif yang benar (*True Positif*) dengan keseluruhan hasil prediksi negatif yang benar (*True Negatif*) [30].

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

2.4.5. Recall

Nilai *recall* menggambarkan tingkat keberhasilan dalam mengidentifikasi informasi dengan benar, baik itu positif atau negatif, pada data tertentu [31].

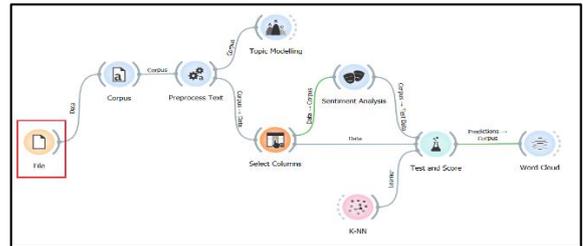
$$Recall = \frac{TP}{TP + FN} \quad (7)$$

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data Twitter

Dalam *Orange*, terdapat berbagai *widget* yang merupakan simbol secara visual yang dapat diatur dan dihubungkan satu sama lain untuk membentuk alur kerja analisis data, serta memungkinkan *Orange* untuk mengelola data teks atau gambar [2]. Alur

kerja proses pengumpulan data dapat dilihat pada Gambar 2 berikut ini.



Gambar 2. Widget Proses Pengumpulan Data di Orange

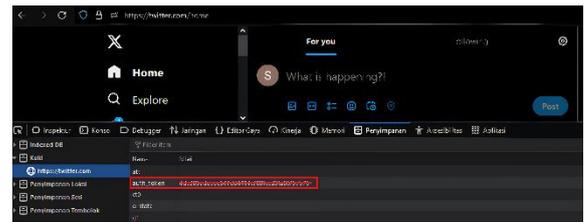
Data Capres 2024 dari Twitter dikumpulkan dengan kata kunci seperti "Ganjar Presiden," "Prabowo Presiden," dan "Anies Presiden." Data mencakup jumlah *tweet*, *like*, *retweet*, *reply*, dan *quote* untuk setiap calon. Pengumpulan data menggunakan modul *Python*, khususnya *Sns scrape*, dengan langkah-langkah crawling sebagai berikut.

- a. Menginstal *Sns scrape* dan *Pandas* menggunakan *Google Colabs*, seperti diperlihatkan pada Gambar 3.

```
[ ] !pip install pandas
!curl -sL https://deb.nodesource.com/setup_18.x | sudo -E bash -
!sudo apt-get install -y nodejs
!npm install -g npm@9.8.0
```

Gambar 3. Proses Instal *Sns scrape* dan *Pandas*

- b. Pada halaman awal *Twitter*, klik kanan untuk membuka *inspection*, kemudian akses ke menu *Storage*, pada bagian *Cookies*, terdapat link *Twitter* yang berisi *auth_token*. Salin *auth_token* tersebut untuk dimasukkan ke perintah *Sns scrape*. Dapat dilihat pada Gambar 4 berikut ini.



Gambar 4. Mengakses *Auth Token* Akun *Twitter*

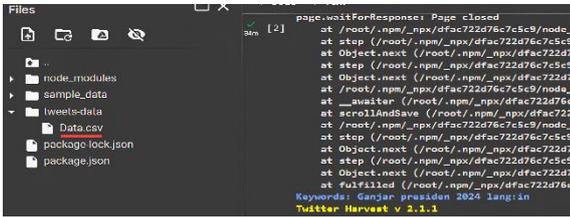
- c. Gunakan *Sns scrape* untuk mengambil *tweet* berbahasa Indonesia yang mencakup "Calon Presiden 2024" dengan memasukkan *auth_token* *Twitter* pribadi. Ambil maksimal 1.000 *tweet*, seperti yang ditunjukkan pada Gambar 5.

```
!python3 crawl.py
filename = 'Data.csv'
search_keyword = 'Ganjar presiden 2024 lang:in'
limit = 1000
!npmx yes tweet harvest@latest -o "[filename]" -s "[search_keyword]" -l [limit]

Welcome to the Twitter Crawler
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.
? What's your Twitter auth token? 
```

Gambar 5. Perintah *Sns scrape* untuk Mengambil Data *Twitter*

- d. Setelah pencarian selesai, pada Gambar 6 terlihat bahwa data dengan format .csv akan muncul pada folder *tweets-data*. Kemudian simpan file data csv tersebut untuk digunakan.



Gambar 6. Data Akses CSV

- e. Berikut ini adalah Tabel 1 yang berisi hasil data yang telah berhasil diambil melalui proses *crawling*, yang menjadi data mentah sebelum diseleksi.

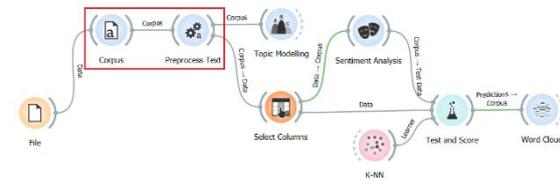
Tabel 1. *Crawling Data Results*

No.	Date	URL Tweet	Teks Tweet	Quotes	Reply	Re-tweet	Like	User-name
1	06/07/2023	https://twitter.com/indo_pejuang1680324014081	“Dukung ganjar jadi presiden 2024. #GanjarCapres”	0	0	0	0	indo_pejuang
2	08/07/2023	https://twitter.com/kenikir2269768087408640	“Saya juga maunya 2024 Pak @prabowo presiden ke-8 RI....”	0	1	5	17	Kenikir27
3	09/07/2023	https://twitter.com/TauhidSHA95080960	“Anies Baswedan in sya Allah RI1 2024”	0	0	0	0	TauhidSHA
...
51.190	7/07/2023	https://twitter.com/MarbotUdien348645405085696	“Ganjar Pranowo dinilai memiliki daya tarik tersendiri bagi kalangan emak-emak...”	2	1	13	32	MarbotUdien

51.191	11/07/2023	https://twitter.com/bersama_nies/status/1678737497554321408	“yang kita punya HANYA NAMA BAIK, jadi jaga nama baik itu baik- baik” ~ Anies Baswedan ~” “Yusril Ihza Mahendra Bakal Beri Dukungan kepada Prabowo Subianto di Pilpres 2024? ...”	0	0	0	0	bersama_nies
51.192	12/07/2023	https://twitter.com/is_pelssy993312752238593	“Anies Baswedan in sya Allah RI1 2024”	0	1	0	1	is_pelssy

3.2. Pre-Processing Data

Pada Gambar 7 menunjukkan alur kerja *pre-processing* yang dibuat berdasarkan *widget* di Orange.



Gambar 7. *Widget Pre-Processing* di Orange

Setelah mengumpulkan data, langkah selanjutnya adalah melakukan *pre-processing* untuk mengubah format data agar mudah diklasifikasikan [33]. Ini melibatkan penghapusan *URL*, *quote*, dan tanda baca dalam teks *tweet*. *Pre-processing* penting untuk memastikan kebersihan, relevansi, dan kesiapan data untuk analisis lebih lanjut. Tabel 2 menunjukkan hasil *pre-processing* data.

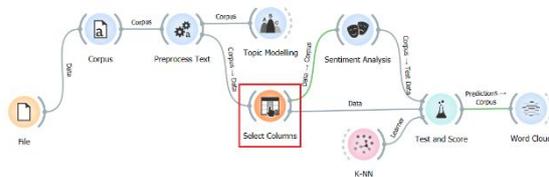
Tabel 2. Hasil Data Setelah *Pre-Processing*

No.	Date	Teks Tweet	Reply	Re-tweet	Like	User-name
1	06/07/2023	“Dukung ganjar jadi presiden 2024. #GanjarCapres”	0	0	0	indo_pejuang
2	08/07/2023	“Saya juga maunya 2024 Pak @prabowo presiden ke-8 RI....”	1	5	17	Kenikir27
3	09/07/2023	“Anies Baswedan in sya Allah RI1 2024”	0	0	0	Tauhid_SHA

51.190	7/07/2023	“Ganjar Pranowo dinilai memiliki daya tarik tersendiri bagi kalangan emak-emak...”	1	13	32	Marbot_Udien
51.191	11/07/2023	“yang kita punya HANYA NAMA BAIK, jadi jaga nama baik itu baik-baik” ~ Anies Baswedan ~”	0	0	0	bersam anies
51.192	12/07/2023	“Yusril Ihza Mahendra Bakal Beri Dukungan kepada Prabowo Subianto di Pilpres 2024? ...”	1	0	1	is_pelsy

3.3. Ekstraksi Fitur

Pada Gambar 8 menunjukkan alur kerja transformasi atau ekstraksi fitur dengan menggunakan widget “Select Columns” yang dibuat di Orange.



Gambar 8. Widget Select Columns K-NN di Orange

Peneliti melakukan ekstraksi fitur dari data yang telah dipersiapkan. Fitur-fitur yang diambil meliputi jumlah *tweet*, *like*, *retweet*, dan *reply*, untuk tiap calon presiden dan menjadi atribut dalam model K-NN untuk memprediksi kemenangan Capres 2024 berdasarkan data *Twitter*. Hasil ekstraksi fitur terhadap data calon presiden dapat terlihat pada Tabel 3 berikut ini.

Tabel 3. Hasil Ekstraksi Fitur Data Calon Presiden 2024

Calon Presiden	Tahun	Jumlah Tweet	Jumlah Like	Jumlah Retweet	Jumlah Reply
Anies	2021	5.576	192.744	38.039	25.689
	2022	5.782	456.430	106.688	50.073
	2023	4.962	190.922	51.694	24.272
Ganjar	2021	7.504	156.571	31.142	18.227
	2022	5.821	144.699	64.146	23.548
	2023	4.360	85.859	20.532	12.217
Prabowo	2021	5.971	150.510	29.136	46.194
	2022	6.303	172.197	38.076	50.171
	2023	4.913	137.824	46.626	41.662
Jumlah		51.192	1.687.756	426.079	292.053

Peneliti menggunakan empat fitur *Twitter* (*tweet*, *like*, *retweet*, *reply*) dari setiap calon presiden untuk memprediksi pemenang Capres 2024 dengan

model K-NN. Anies menonjol dengan jumlah *retweet* dan *like* yang lebih tinggi. Evaluasi dilakukan dengan metode K-NN dan pendekatan *Both Metric* (*Chebyshev* dan *Euclidean*) menggunakan *Orange*.

3.4. Perhitungan K-NN dan Evaluasi Kinerja Berdasarkan Both Metric (Euclidean dan Chebyshev)

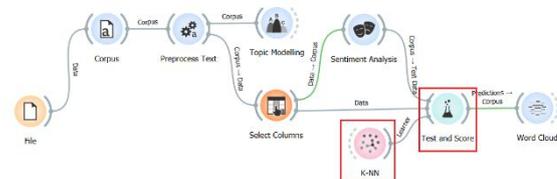
Setelah menghitung *Euclidean* dan *Chebyshev*, peneliti memilih 5 data *training* terdekat ($K=5$). Urutan pemilihan data dapat dilihat pada Tabel 4.

Tabel 4. Hasil Perhitungan *Euclidean* dan *Chebyshev*

No.	Hasil Perhitungan Euclidean	Hasil Perhitungan Chebyshev
1.	Anies 2021 yaitu 13.862,30	Anies 2021 yaitu 13.655
2.	Prabowo 2021 yaitu 22.101,87	Prabowo 2021 yaitu 17.490
3.	Prabowo 2022 yaitu 36.454,64	Prabowo 2022 yaitu 34.373
4.	Ganjar 2021 yaitu 71.816,05	Ganjar 2022 yaitu 58.840
5.	Ganjar 2022 yaitu 74.127,27	Ganjar 2021 yaitu 70.712

Selanjutnya, melakukan *voting* untuk menentukan klasifikasi calon presiden berdasarkan mayoritas dari K data *training* terdekat. Menggunakan jarak *Euclidean* dan *Chebyshev* berdasarkan hasil perhitungan, mayoritas adalah "Prabowo" karena ada 2 hasil yang menjadi jarak terdekat. Jadi, berdasarkan K-NN dengan $K=5$ dan menggunakan jarak *Euclidean* atau *Chebyshev*, klasifikasi calon presiden untuk tahun 2024 adalah "Prabowo" berdasarkan data *testing* (2023), dan data *training* (2021 dan 2022).

Pada Gambar 9 menunjukkan alur kerja evaluasi kinerja model di *Orange* dengan menggunakan widget “K-NN” dan “Test and Score”.



Gambar 9. Widget K-NN dan Test and Score untuk Evaluasi Kinerja Model di Orange

Widget "Test and Score" adalah alat untuk mengevaluasi kinerja model atau algoritma dalam analisis data [34]. Fungsinya adalah membagi dataset menjadi data *training* dan data *testing*, kemudian menguji seberapa baik model mampu memprediksi data yang belum pernah dilihat sebelumnya. Prosesnya dimulai dengan memasukkan data ke dalam widget, memilih model yang akan dievaluasi, dan membagi data sesuai proporsi. Model K-NN dilatih dengan data *training* dan digunakan untuk memprediksi data *testing*.

Hasil prediksi dibandingkan dengan nilai aktual dari data *testing* untuk mengukur AUC, CA, F1 Score, Precision, dan Recall. Output dari widget ini adalah nilai kinerja model K-NN pada data *testing*.

Widget "K-NN" adalah bagian dari kategori Classification di Orange yang membangun model K-NN untuk klasifikasi data. Pengguna dapat mengatur parameter K yang menentukan jumlah tetangga terdekat yang digunakan. K-NN digunakan untuk melakukan klasifikasi terhadap objek baru berdasarkan tetangga terdekatnya. Kelas yang paling sering muncul dan paling dekat di antara tetangga-tetangga tersebut akan menjadi hasil klasifikasi [35].

Dengan kombinasi kedua widget ini, pengguna dapat membangun dan menguji model K-NN secara sistematis. Evaluasi kinerja membantu memahami sejauh mana model K-NN dapat berkinerja dengan baik dan memilih parameter K yang optimal untuk menghasilkan model yang akurat dan dapat diandalkan.

Berdasarkan perhitungan menggunakan Orange, hasilnya dibagi menjadi dua dalam Both Metric (Euclidean dan Chebyshev) adalah sebagai berikut.

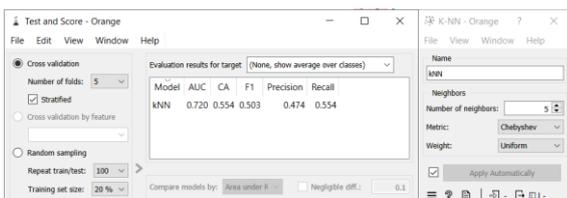
a. Evaluasi Kinerja dari Hasil Euclidean Berdasarkan Perhitungan di Orange



Gambar 10. Widget Test and Score Hasil Evaluasi Euclidean di Orange

Data dari 51.192 *tweet* diproses dengan aplikasi Orange dan 5-fold cross validation. Model K-NN dengan jarak Euclidean menunjukkan performa yang menjanjikan untuk memprediksi calon presiden 2024, dengan AUC sebesar 0.725. Tingkat akurasi keseluruhan model (CA) adalah 0.556, F1 Score sebesar 0.510, Precision 0.482, dan Recall 0.556. Meskipun model memberikan kontribusi penting, terdapat potensi peningkatan kinerja dengan pertimbangan metode K-NN lain atau pengoptimalan fitur-fitur dan algoritma yang lebih kompleks.

b. Evaluasi Kinerja dari Hasil Chebyshev Berdasarkan Perhitungan di Orange



Gambar 11. Widget Test and Score Hasil Evaluasi Chebyshev di Orange

Selain Euclidean, peneliti menggunakan metrik jarak Chebyshev dalam metode K-NN untuk

mengklasifikasikan data *tweet* calon presiden 2024 dari Twitter. Data dibagi menjadi 5 fold menggunakan cross validation untuk evaluasi yang lebih akurat. Hasil penelitian pada Gambar 11 menunjukkan AUC sebesar 0.720, CA sebesar 0.554, dan F1 Score sebesar 0.503, menandakan kemampuan model dalam membedakan kelas positif dan negatif serta keseimbangan Precision dan Recall. Precision sebesar 0.474 menunjukkan kemampuan model mengklasifikasikan calon presiden yang sebenarnya, sedangkan Recall sebesar 0.554 menunjukkan kemampuan model mengidentifikasi calon presiden yang sebenarnya dari total calon presiden.

Eksperimen dengan metrik Euclidean menunjukkan hasil sedikit lebih baik, tetapi kedua metode memberikan hasil yang hampir serupa dan dapat diandalkan. Penggunaan metrik jarak Chebyshev memberikan keuntungan dalam menangani data tidak linier. Secara keseluruhan, metode K-NN dengan metrik Chebyshev cukup menjadi pilihan yang tepat untuk klasifikasi data *tweet* calon presiden 2024 dari Twitter dalam penelitian tersebut.

4. DISKUSI

Peneliti menganalisis data Twitter sebanyak 51.192 data terkait calon presiden 2024 (Anies Baswedan, Ganjar Pranowo, Prabowo Subianto) menggunakan metode K-NN dengan Euclidean dan Chebyshev sebagai metrik jarak. Empat fitur yang dievaluasi: "Tweet", "Like", "Retweet", "Reply". Kinerja model dievaluasi dengan metrik AUC, CA, F1 Score, Precision, dan Recall dihitung berdasarkan hasil 5 fold cross-validation, sehingga memberikan pandangan holistik tentang kinerja model K-NN dalam berbagai aspek. Hasilnya dapat terlihat dari Tabel 5 berikut ini.

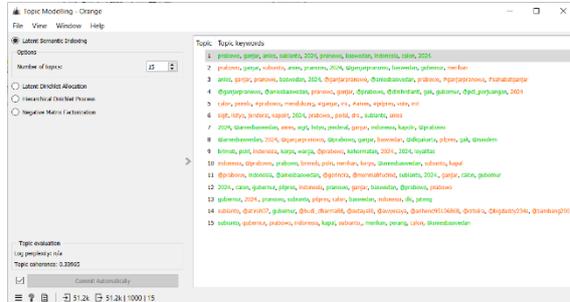
Tabel 5. Perbandingan Hasil Kinerja Metrik Euclidean dan Chebyshev

Kinerja Metrik	Jarak Metrik Euclidean	Jarak Metrik Chebyshev
AUC	0.725	0.720
CA	0.556	0.554
F1-Score	0.510	0.503
Precision	0.482	0.474
Recall	0.556	0.554

Model K-NN dengan metrik Euclidean dan Chebyshev memiliki performa hampir mendekati dalam memprediksi calon presiden berdasarkan data Twitter, dengan nilai AUC, CA, F1 Score, Precision, dan Recall yang sedikit berbeda tetapi memberikan hasil klasifikasi yang dapat diandalkan. Penelitian ini berkontribusi dalam analisis data Twitter tentang calon presiden 2024, memberikan wawasan bagi pemangku kepentingan, dan menyarankan potensi peningkatan kinerja melalui eksplorasi metode K-

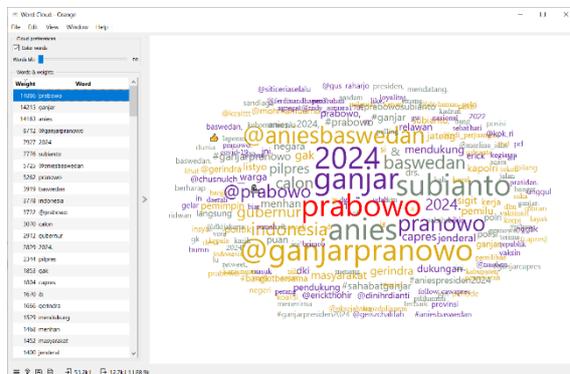
NN lainnya atau penggunaan algoritma *machine learning* yang lebih kompleks di masa depan.

Hasil penelitian prediksi pemenang pilpres 2024 dapat terlihat dalam *widget Topic Modeling* dan *Word Cloud*, yang tersedia di *Orange* berikut ini.



Gambar 12. *Widget Topic Modeling*

Dalam Gambar 12, dapat terlihat 15 topik kata yang sering muncul dalam data *tweet* yang diambil, di topik nomor 1 kata yang paling banyak muncul adalah nama-nama calon kandidat pilpres 2024, kemudian pada topik nomor 2 didominasi oleh Prabowo, pada topik nomor 3 didominasi oleh Ganjar, dan pada topik nomor 4 didominasi oleh Anies. Kata yang sering muncul juga dapat divisualisasikan dalam *Word Cloud* pada Gambar 13 berikut ini.



Gambar 13. *Widget Word Cloud*

Dalam visualisasi *Word Cloud*, kata "Prabowo" merupakan kata yang paling banyak disebut dengan total sebanyak 14.886 kemunculan, diikuti oleh kata "Ganjar" dengan 14.215 kemunculan, dan kata "Anies" dengan 14.163 kemunculan.

Perlu ditegaskan bahwa ketiga nama calon kandidat presiden muncul secara berurutan dalam visualisasi *Word Cloud*. Hal ini menunjukkan bahwa Prabowo merupakan kandidat calon presiden 2024 yang paling populer dan sering disebut dalam percakapan di *Twitter* berdasarkan data yang telah dianalisis.

Berdasarkan hasil analisis dan perhitungan menggunakan metode K-NN dengan metrik *Both Metric (Euclidean dan Chebyshev)*, prediksi untuk pemenang calon presiden 2024 berdasarkan data *Twitter* adalah Prabowo Subianto. Model K-NN

mengidentifikasi Prabowo Subianto sebagai calon yang diprediksi menjadi pemenang berdasarkan atribut dan data *Twitter* yang telah diberikan.

Penelitian ini dihubungkan dengan penelitian relevan oleh Saputra dan Primadasa (2018) tentang prediksi kelulusan mahasiswa menggunakan K-NN. Meskipun berfokus pada konteks pendidikan, penelitian tersebut memberikan referensi yang berguna terkait penerapan algoritma K-NN. Penelitian lain oleh Badrul (2015) dalam konteks prediksi hasil pemilu legislatif menggunakan K-NN dengan data KPUD juga menjadi relevan. Meskipun data yang digunakan berbeda, hasilnya menunjukkan akurasi yang baik, sehingga memberikan wawasan dan pemahaman yang dapat diadaptasi dalam penelitian Pilpres 2024 di *Twitter*. Perlu diperhatikan bahwa hasil prediksi tentang pemenang Pilpres 2024 berdasarkan *Twitter* sebaiknya diinterpretasikan hati-hati dan diperkuat dengan analisis lebih lanjut serta data dari berbagai sumber untuk keakuratan dan komprehensifitas yang lebih baik.

Namun, perlu diingat bahwa hasil prediksi ini hanya didasarkan pada data *Twitter* yang diolah dengan metode K-NN dan menggunakan metrik evaluasi *Both Metric (Euclidean dan Chebyshev)*. Hasil prediksi ini sebaiknya diinterpretasikan dengan hati-hati dan perlu diperkuat dengan analisis lebih lanjut serta data dari berbagai sumber lainnya untuk menghasilkan kesimpulan yang lebih komprehensif dan akurat mengenai calon presiden 2024. Meskipun demikian, hasil ini memberikan wawasan yang menarik dalam memahami dinamika percakapan tentang calon presiden di *platform* media sosial seperti *Twitter*.

5. KESIMPULAN

Kesimpulan penelitian ini bahwa metode K-NN dengan *Both Metric (Euclidean dan Chebyshev)* berhasil digunakan untuk memprediksi pemenang Pilpres 2024 berdasarkan popularitas di *Twitter*. Kedua metode menunjukkan bahwa Prabowo Subianto adalah calon yang paling populer. Metode K-NN dengan *Both Metric* ini memberikan tingkat keakuratan di atas 50%, dengan AUC 72,5% dan CA 55,6% untuk *Euclidean*, serta AUC 72% dan CA 55,4% untuk *Chebyshev*. Namun, secara keseluruhan hasil metrik *Euclidean* lebih baik dibanding *Chebyshev*. Dengan hasil ini, maka perlu diinterpretasikan dengan hati-hati dan memerlukan analisis lebih lanjut serta penambahan data dari berbagai sumber untuk kesimpulan yang lebih komprehensif dan akurat mengenai pemenang Pilpres 2024.

Dalam pengembangan penelitian ini, disarankan untuk menggabungkan data dari berbagai sumber media sosial dan platform lainnya guna mendapatkan gambaran yang lebih komprehensif tentang popularitas calon presiden secara keseluruhan. Selain itu, perlu dilakukan studi lebih

lanjut dengan mempertimbangkan penggunaan metode K-NN yang menggunakan metrik jarak lainnya, serta eksplorasi penggunaan algoritma machine learning yang lebih kompleks agar dapat meningkatkan performa prediksi.

DAFTAR PUSTAKA

- [1] Pemerintah Republik Indonesia, "Undang-Undang Republik Indonesia Nomor 42 Tahun 2008 tentang Pemilihan Umum Presiden dan Wakil Presiden," Jakarta: Sekretariat Negara, 2008. Tersedia: https://www.dpr.go.id/dokjdi/document/uu/UU_2008_42.pdf. [Diakses pada 30 Juni 2023].
- [2] Hartanto, Anggit Dwi, et al. "Job Seeker Profile Classification of Twitter Data Using The Naïve Bayes Classifier Algorithm Based on The DISC Method," 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE, 2019.
- [3] E. Utami, "Comparison Naïve Bayes Classifier, K-Nearest Neighbor And Support Vector Machine In The Classification Of Individual On Twitter Account," *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no.3, 673-680, 2022.
- [4] H. Hamiruddin, "Survei dan Konsultan Politik: Membangun Popularitas dan Elektabilitas Politik," *Al-Din: Jurnal Dakwah dan Sosial Keagamaan*, vol. 7, no. 1, 2021.
- [5] N. Faradis, N. A. Al Fauzah, and M. I. Al Anshori, "Media Sosial dan Persepsi Publik: Analisis Strategi Kampanye Digital Calon Presiden Indonesia 2024," In Prosiding Seminar Nasional Ilmu Ilmu Sosial (SNIIS), vol. 2, pp. 643-652, 2023.
- [6] A. D. Akmal, I. Permana, H. Fajri, and Y. Yuliarti, "Opini Masyarakat Twitter terhadap Kandidat Bakal Calon Presiden Republik Indonesia Tahun 2024," *Jurnal Manajemen dan Ilmu Administrasi Publik (JMIAP)*, vol. 4, no. 4, 287-295, 2022.
- [7] W. Winarsih, and I. Irwansyah, "Proteksi Privasi Big Data Dalam Media Sosial," *Jurnal Audience: Jurnal Ilmu Komunikasi*, vol. 3, no. 1, 1-33, 2020.
- [8] D. Faroek, M. Yusuf, and G. Syatauw, "Sentiment Analysis of the Popularity of Parties Supporting the 2024 Presidential Candidates on Twitter Using the Naive Bayes Classifier Algorithm," *Antivirus: Jurnal Ilmiah Teknik Informatika*, vol. 17, no. 2, 216-227, 2023.
- [9] M. Syamsurrijal, A. Nurmandi, H. Jubba, M. Hidayati, T. Baharuddin, and Z. Qodir, "Prediction Candidates and Political Parties in the Presidential Election 2024 in Indonesia Based on Twitter," *ResearchSquare*, pp. 1-9, 2021.
- [10] R. D. Liklikwatil, E. Noersasongko, and C. Supriyanto, "Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Memprediksi Harga Komoditi Karet," *E-JURNAL JUSITI: Jurnal Sistem Informasi dan Teknologi Informasi*, vol. 7, no. 2, 172-182, 2018.
- [11] M. Yuliana, "Perbandingan Penskalaan Dimensi Ganda Non-Metrik dengan Menggunakan Tiga Jarak (Euclidean, Chebyshev, dan Minkowski)," Skripsi, Universitas Brawijaya, Malang, 2015. (Kutipan pada Pendahuluan)
- [12] A. Y. Saputra and Y. Primadasa, "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Techno. Com*, vol. 17, no. 4, pp. 395-403, 2018.
- [13] M. Badrul, "Prediksi Hasil Pemilu Legislatif Dengan Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Pilar Nusa Mandiri*, vol. 11, no. 2, pp. 152-160, 2015.
- [14] A. Putri, C. S. Hardiana, E. Novfuja, F. T. P. Siregar, R. Rahmadden, Y. Fatma, and R. Wahyuni, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, 20-26, 2023.
- [15] S. Sudriyanto, F. Syahro, and N. Fitriani, "Perbandingan Performa Model Machine Learning Support Vector Machine, Neural Network, dan K-Nearest Neighbors dalam Prediksi Harga Saham," *Jurnal Advanced Research Informatika*, vol. 2, no. 1, 13-21, 2023.
- [16] Sugiyono, "Metodologi Penelitian Kuantitatif, Kualitatif, dan R&D," Bandung: ALFABETA, 2019.
- [17] D. Yanosma, A. Johar, and K. Anggriani, "Implementasi Metode K-Nearest Neighbor (KNN) dan Simple Additive Weighting (SAW) dalam Pengambilan Keputusan Seleksi Anggota PASKIBRAKA," *Rekursif: Jurnal Informatika*, vol. 4, no. 2, 2016.
- [18] M. Yuliana, "Perbandingan Penskalaan Dimensi Ganda Non-Metrik dengan Menggunakan Tiga Jarak (Euclidean, Chebyshev, dan Minkowski)," Skripsi, Universitas Brawijaya, Malang, 2015. (Kutipan pada Metode/Perancangan)
- [19] I. M. K. Karo, A. Khosuri, J. S. I. Septory, and D. P. Supandi, "Pengaruh Metode Pengukuran Jarak pada Algoritma k-NN untuk Klasifikasi Kebakaran Hutan dan Lahan," *Jurnal Media Informatika Budidarma*, vol. 6, no. 2, pp. 1174-1182, 2022.

- [20] Wikipedia, "Orange (Software)," Wikipedia, 2023. [Online]. Tersedia: Orange (software) - Wikipedia. [Diakses pada 8 Juli 2023].
- [21] P. Meilina, "Penerapan data mining dengan metode klasifikasi menggunakan decision tree dan regresi," *Jurnal Teknologi*, vol. 7, no. 1, pp. 11-20, 2015.
- [22] H.N. Rahmah and R. Sulistiyowati, "Penggunaan Software Orange Data Mining pada Implementasi Text Mining dalam Analisis Sentimen Netizen di Twitter Terhadap Kelangkaan Minyak Goreng," *Sigma-Mu*, vol. 14, no. 2, pp. 1-11, 2022.
- [23] M. D. F. Tino, H. Hasanah and T. D. Santosa, "Perbandingan Algoritma Support Vector Machines (SVM) dan Neural Network untuk Klasifikasi Penyakit Jantung." *Infotech Journal*, 9(1), 232-235, 2023.
- [24] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 81-86, 2020.
- [25] V. R. Sari, F. Firdausi, and Y. Azhar, "Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 4, no. 2, pp. 1-9, 2020.
- [26] A. W. Hutomo, "Perbandingan Kinerja AUC dan G-Means pada Machine Learning Berbasis Seleksi Fitur Algoritma Genetika untuk Prediksi Cacat Perangkat Lunak," Bachelor's Thesis, Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah, Jakarta, 2020.
- [27] Y. I. Kurniawan, A. Fatikasari, M. L. Hidayat, & M. Waluyo, "Prediction For Cooperative Credit Eligibility Using Data Mining Classification With C4. 5 Algorithm," *Jurnal Teknik Informatika (JUTIF)*, vol. 2, no. 2, 67-74, 2021.
- [28] W. Yusuf, R. Witri, & C. Juliane, "Model Prediksi Penjualan Jenis Produk Tekstil Menggunakan Algoritma K-Nearest Neighbor (K-NN)," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 7, no. 1, 2022.
- [29] D. D. Putri, G. F. Nama, & W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, 2022.
- [30] M. Maulidah, W. Gata, R. Aulianita, & C. I. Agustyaningrum, "Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku. E-Bisnis," *Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 13, no. 2, 89-96, 2020.
- [31] M. Azhari, Z. Situmorang, & R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4. 5, Random Forest, SVM dan Naive Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, 640-651, 2021.
- [32] R. R. Waliansyah and C. Fitriyah, "Muttaqin, F. A., & Bachtiar, A. M. (2016). Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak 'Dodo Kids Browser,'" *Jurnal Edukasi dan Penelitian Informatika (Jepin)*, vol. 5, no. 2, 2019.
- [33] F. A. Muttaqin and A. M. Bachtiar, "Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak Dodo Kids Browser" *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, pp. 1-8, 2016.
- [34] A. Winantu and C. Khatimah, "Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbor Dalam Memprediksi Prestasi Siswa," *INTEK: Jurnal Informatika dan Teknologi Informasi*, vol. 6, no. 1, pp. 58-64, 2023.
- [35] I. Yolanda and H. Fahmi, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT. Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 3, no. 3, pp. 9-15, 2020.