ANALISIS SENTIMEN PENGGUNA INSTAGRAM PADA CALON PRESIDEN 2024 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Wahyu Surya Gemilang, Purwantoro, Carudin
Program Studi Informatika S1, Fakultas Ilmu Komputer
Universitas Singaperbangsa Karawang, Jl HS. Ronggo Waluyo Karawang, Indonesia
1910631170053@student.unsika.ac.id

ABSTRAK

Penelitian ini bertujuan untuk memahami pandangan masyarakat dengan menganalisis sentimen pengguna Instagram terkait Calon Presiden 2024 dengan menggunakan algoritma *Support Vector Machine* (SVM). Metodologi *Knowledge Discovery in Database* (KDD) digunakan untuk mengelola data komentar sebanyak 1200 dari berbagai sumber. Setelah *preprocessing*, data diseleksi menjadi 1179 komentar yang telah diberi label sentimen (positif, negatif, dan netral) digunakan untuk melatih model SVM dengan kernel linear. Evaluasi hasil dilakukan dengan menggunakan *confusion matrix*, yang menghasilkan metrik seperti akurasi, presisi, *recall*, dan *f1-score*. Pengujian dilakukan dengan empat skenario pembagian data yang berbeda (90:10, 80:20, 70:30, dan 60:40), dengan skenario ketiga (70:30) mencapai akurasi tertinggi sekitar 77% menggunakan SVM.. Penelitian ini memberikan wawasan tentang sentimen masyarakat terkait Calon Presiden 2024 di *platform* instagram, yang dapat bermanfaat bagi pemangku kepentingan seperti bawaslu dan masyarakat umum.

Kata kunci : *analisis sentimen, Calon Presiden, Support Vector Machine, Knowledge in Discovery Database*

1. PENDAHULUAN

Pada era digital ini, media sosial telah menjadi salah satu tempat atau *platform*, dimana orang dapat mengekspresikan diri dan menyampaikan pendapat mengenai berbagai topik. Instagram adalah salah satu platform media sosial yang terkenal, yang dasarnya digunakan untuk berbagi foto dan video antara pengguna. Dalam unggahan di instagram, semua orang dapat secara bebas menulis komentar untuk mengekspresikan pendapat mereka terhadap unggahan tersebut. Namun, tidak jarang masyarakat komentar yang kasar atau bahkan mengeluarkan ujaran kebencian [1].

Pemilihan presiden di Indonesia dijadwalkan akan dilaksanakan pada tahun 2024. Terkait dengan situasi ini, diskusi dan prediksi mengenai calon presiden di Indonesia telah menjadi topik yang hangat dan menarik dikalangan masyarakat Indonesia, dengan banyak orang yang mengungkapkannya melalui media sosial. Kemudian informasi disediakan dapat digunakan untuk memprediksi hasil pemilu dengan menggunakan analisis sentimen [2].

Analisis sentimen terhadap komentar pengguna instagram diperlukan untuk mengelompokkan komentar tersebut, untuk dapat mengambil berbagai bentuk termasuk tindakan penghinaan, pencemaran nama baik, penistaan, perilaku tidak pantas, provokasi, ajakan kekerasan, dan penyebaran informasi palsu. Analisis sentimen ini bertujuan untuk menganalisis opini, sentimen, dan emosi yang terdapat dalam dokumen atau data. Tugas pokok dari analisis sentimen yaitu untuk mengklasifikasikan sifat teks yang terdapat dalam kalimat atau pendapat (Murni et al., 2023).

Penelitian sebelumnya oleh [4] tentang analisis sentimen pemilihan calon Presiden tahun 2019 pada twitter menggunakan algoritma *Support Vector Machine*. Hasil pengujian terhadap 10 kombinasi skenario terbaik yaitu menggunakan TF-IDF dan *stemming* memiliki akurasi sebesar 81,58%, kemudian dengan menggunakan kombinasi *word count* dan *stemming* memiliki akurasi 77,56%. Namun, masih ada beberapa kekurangan dalam proses klasifikasi yang disebabkan oleh data yang kurang jelas.

2. TINJAUAN PUSTAKA

2.1. *Knowledge Discovery in Database (KDD)*

KDD *(Knowledge Discovery in Databases)* adalah proses yang kompleks dalam mencari serta mengidentifikasi pola *(pattern)* dalam data. Pola yang ditemukan harus benar, baru, bermanfaat, dan dapat dimengerti. KDD terkait dengan Teknik integritas dan penemuan ilmiah, interpretasi serta visualisasi pola dalam sejumlah kumpulan data [5].

2.2. Instagram

Dalam Instagram, pengguna dapat berinteraksi dengan memberikan tanda suka atau komentar pada gambar yang diunggah, serta saling berkirim pesan berupa teks, gambar atau video. Terkait dengan komentar Instagram, pengguna dapat membatasi komentar yang dianggap tidak pantas atau tidak diperlukan pada gambar yang diunggah. Fitur lainnya adalah *Top Comment*, yang menampilkan komentar yang paling banyak disukai oleh pengguna sehingga muncul di bagian atas daftar komentar [6]

2.3. *Support Vector Machine*

Support Vector Machine adalah sistem pembelajaran mesin yang menggunakan fungsi-fungsi linear dalam fitur dengan dimensi yang tinggi. SVM dilatih dengan algortima pembelajaran yang didasarkan pada teori optimasi. Keakuratan model yang dihasilkan oleh SVM sangat tergantung pada parameter dan fungsi kernel yang digunakan. SVM dibagi menjadi dua jenis yaitu linear dan non-linear.

SVM linear digunakan pada data yang dapat dipisahkan secara linear dengan menggunakan hyperlane dan soft margin. Sedangkan SVM non-linear menggunakan kernel trick untuk mentransformasi data ke dalam ruang dimensi yang lebih tinggi [7].

2.4. *Confusion Matrix*

Confusion Matrix adalah sebuah alat yang digunakan untuk mengukur kinerja suatu metode prediksi dengan cara mengitung tingkat kebenaran dalam proses klasifikasi. Akurasi prediksi dapat dihitung dengan menggunakan *Confusion Matrix*. Jumlah data prediksii dari system dan jumlah data prediksi dari pakar sebagai variabel masukan *Confusion Matrix*. Kemudian, setiap system memberikan masing-masing metode *fuzzy* nilai akurasi, presisi, *recall* dan *F-1 Score* [8].

2.5. Analisis Sentimen

Dalam bidang *data mining*, analisis sentimen adalah subbidang ilmu yang memiliki kegunaan dalam menganlisis, memproses, dan mengekstrak data teks dari berbagai sumber seperti layanan, produk, individu, organisasi, peristiwa, atau topik tertentu. Tujuan utamanya adalah untuk memperoleh informasi dari sekumpulan data yang ada. Analisis sentimen merupakan sebuah area penelitian yang relatif baru dalam pemrosesan bahasa alami (NLP) yang bertujuan untuk mengidentifikasi unsur subjektif dalam teks serta mengklasifikasikan sentimen yang terkandung dalam opini tersebut [9].

2.6. *Text Mining*

Text mining merupakan bagian dari *data mining* yang terlibat dalam penggunaan alat anaisis untuk menggali pengetahuan dari sejumlah dokumen dari waktu ke waktu. Konsep *text mining* sering digunakan untuk mengelompokkan dokumen-dokumen berdasarkan topiknya dalam proses yang disebut klasifikasi. Proses klasifikasi ini bertujuan untuk membantu pengguna memahami dan menganalisis konten dokumen secara lebih efisien [10].

3. METODE

Dalam penelitian ini, digunakan metode KDD (*Knowledge Discovery in Database)*, sebuah pendekatan yang digunakan dalam riset untuk menghasilkan pengetahuan baru dari data yang sudah ada. Metode ini melibatkan 5 tahap, yaitu:

1. *Data Selection*
   Data komentar dari instagram Calon Presiden 2024 diambil melalui *exportcomment.com* dan disimpan dalam format xlsx.
2. *Preprocessing*
   Data diseleksi melalui tahap *cleaning, case folding, tokenizing, stemming, stopword removal*, dan *normalization.*
3. *Transformation*
   Data diubah menjadi represntatif numerik menggunakan metode TF-IDF setelah sentimen dilabeli dengan kamus *lexicon*.
4. *Data Mining*
   Klasifikasi sentimen dilakukan dengan algoritma *Support Vector Machine* (SVM) dengan *kernel linear*.
5. *Evaluation*
   Performa model dievaluasi dengan akurasi, presisi, *recall*, dan *f1-score* dalam berbagai skenario pembagian data.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini adalah melakukan analisis sentimen pengguna instagram terhadap calon presiden 2024 dengan menggunakan algoritma *Support Vector Machine*. Data dikumpulkan selama periode waktu tertentu untuk memastikan representasi yang baik dari komentar yang analisis. Berikut proses penelitiannya:

4.1. Data Selection

Data komentar Instagram yang digunakan dalam penelitian in diambil dari postingan akun kompas.com, detik.com, sindonews, dan katadatacoid yang terkait dengan topik calon presiden 2024 yaitu Prabowo, Ganjar Pranowo, dan Anies Baswedan menggunakan *website scraper exportcomments.com*. Dapat dilihat pada Gambar 1 dataset ini terdiri dari beberapa atribut, diantaranya yaitu *username, date, text.*

	Username	Date	Text
0	sandro.oles	2023-07-04 06:35:35	Gampang ,u usul aja ahok,bukan susah cari calo...
1	2martono	2023-07-04 06:53:45	@bandung_64 ganjar bagus kendaraan politiknya ...
2	ahmadi.arifin	2023-07-04 06:57:56	@irvanharnadi liat aja taun taun mendatang wkwk
3	rifqy_kingking	2023-07-04 07:10:44	bapak prabowo harus mencontoh buk megawati ,ba...
4	zhafarian	2023-07-04 07:19:17	@brothers_reptiless moon maap ini gimana cerit...
...	...	...	...
1195	rigobertoerwin	2023-07-24 13:58:39	Nol besar ente
1196	rigobertoerwin	2023-07-24 14:00:07	@moms_al_and_zikra kadrunwati
1197	pambudi318	2023-07-24 14:15:01	Makin kesini ini org ngomong tuh bikin enek🤮
1198	pambudi318	2023-07-24 14:15:49	@anisty531 terbaik versi apa tolong kasih tau
1199	ahmadmuladi	2023-07-24 14:17:32	@anisty531 musibah buat Indonesia bukan Anies ...

1200 rows × 4 columns

Gambar 1. Contoh dataset hasil *crawling*

4.2. *Cleaning*

Tahap pembersihan (*cleaning*) merupakan proses eliminasi *url*, angka, tanda baca, dan simbol-simbol yang tidak relevan dalam proses klasifikasi. Contoh hasil dari langkah pembersihan ini dapat dilihat pada Gambar 2.

Text	clean_pengguna	cleaning
Gampang ,u usul aja ahok,bukan susah cari calo...	None	Gampang u usul aja ahok bukan susah cari calon...
@bandung_64 ganjar bagus kendaraan politiknya ...	ganjar bagus kendaraan politiknya banteng mer...	ganjar bagus kendaraan politiknya banteng mer...
@irvanharnadi liat aja taun taun mendatang wkwk	liat aja taun taun mendatang wkwk	liat aja taun taun mendatang wkwk
bapak prabowo harus mencontoh buk megawati ,ba...	None	bapak prabowo harus mencontoh buk megawati bag...
@brothers_reptiless moon maap ini gimana cerit...	moon maap ini gimana ceritanya putin wkwk	reptiless moon maap ini gimana ceritanya putin...

Gambar 2. Contoh hasil tahap *cleaning*

4.3. *Case Folding*

Proses *case folding* adalah langkah untuk mengonversi seluruh karakter dalam teks menjadi huruf kecil. Tujuannya adalah untuk memastikan bahwa data yang akan diolah memiliki karakter yang seragam, sehingga memudahkan langkah *preprocessing* berikutnya, yaitu *tokenizing*. Contoh hasil dari tahap *case folding* dapat dilihat pada Gambar 3.

cleaning	case_folding
Gampang u usul aja ahok bukan susah cari calon...	gampang u usul aja ahok bukan susah cari calon...
ganjar bagus kendaraan politiknya banteng mer...	ganjar bagus kendaraan politiknya banteng mer...
liat aja taun taun mendatang wkwk	liat aja taun taun mendatang wkwk
bapak prabowo harus mencontoh buk megawati bag...	bapak prabowo harus mencontoh buk megawati bag...
reptiless moon maap ini gimana ceritanya putin...	reptiless moon maap ini gimana ceritanya putin...
...	...
Nol besar ente	nol besar ente

Gambar 3. Contoh hasil tahap *case folding*

4.4. *Tokenizing*

Pada proses *tokenizing* dilakukan untuk memisahkan teks menjadi bagian-bagian yang lebih terstruktur sehingga mempermudah proses analisis selanjutnya, seperti penghapusan karakter khusus, menghitung frekuensi kata, atau membangun representasi vektor dari teks. Berikut contoh hasil dari tahapan *tokenizing* dapat dilihat pada Gambar 4.

case_folding	tokenizing
gampang u usul aja ahok bukan susah cari calon...	[gampang, u, usul, aja, ahok, bukan, susah, ca...
ganjar bagus kendaraan politiknya banteng mer...	[ganjar, bagus, kendaraan, politiknya, banteng...
liat aja taun taun mendatang wkwk	[liat, aja, taun, taun, mendatang, wkwk]
bapak prabowo harus mencontoh buk megawati bag...	[bapak, prabowo, harus, mencontoh, buk, megawa...
reptiless moon maap ini gimana ceritanya putin...	[reptiless, moon, maap, ini, gimana, ceritanya...

Gambar 4. Contoh hasil tahap *tokenizing*

4.5. *Stemming*

Pada tahap *stemming* dilakukan untuk mengubah kata-kata dalam teks menjadi bentuk dasar dengan cara memotong akhiran atau awalan dari kata-kata tersebut sehingga kata-kata yang memiliki makna atau arti yang sama tetapi memiliki variasi bentuk dapat dianggap setara. Pada Gambar 5 berikut contoh hasil tahapan *stemming*.

tokenizing	stemming
[gampang, u, usul, aja, ahok, bukan, susah, ca...	[gampang, u, usul, aja, ahok, bukan, susah, ca...
[ganjar, bagus, kendaraan, politiknya, banteng...	[ganjar, bagus, kendara, politik, banteng, mer...
[liat, aja, taun, taun, mendatang, wkwk]	[liat, aja, taun, taun, datang, wkwk]
[bapak, prabowo, harus, mencontoh, buk, megawa...	[bapak, prabowo, harus, contoh, buk, megawati,...
[reptiless, moon, maap, ini, gimana, ceritanya...	[reptiless, moon, maap, ini, gimana, cerita, p...

Gambar 5. Tahap *stemming*

4.6. *Stopword Removal*

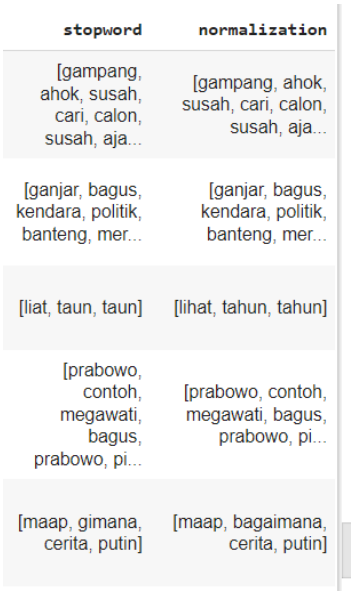
Selanjutnya akan dilakukan tahap *stopword removal* dengan proses untuk meningkatkan efisiensi analisis teks dengan mengurangi jumlah kata yang diproses dan meningkatkan kualitas analisis dengan menghilangkan kata yang tidak relevan. Berikut contoh hasil dari tahapan *stopword removal* dapat dilihat pada Gambar 6.

stemming	stopword
[gampang, u, usul, aja, ahok, bukan, susah, ca...	[gampang, ahok, susah, cari, calon, susah, aja...
[ganjar, bagus, kendara, politik, banteng, mer...	[ganjar, bagus, kendara, politik, banteng, mer...
[liat, aja, taun, taun, datang, wkwk]	[liat, taun, taun]
[bapak, prabowo, harus, contoh, buk, megawati,...	[prabowo, contoh, megawati, bagus, prabowo, pi...
[reptiless, moon, maap, ini, gimana, cerita, p...	[maap, gimana, cerita, putin]

Gambar 6. Contoh hasil tahap *stopword removal*

4.7. *Normalization*

Dalam tahap *normalizationi* atau normalisasi, digunakan kamus KBBI untuk melakukan koreksi terhadap kata-kata yang tidak baku menjadi kata baku. Tujuan dari proses ini adalah untuk mengubah kata-kata tersebut menjadi bentuk yang lebih standar dan sesuai dengan entri Kamus Besar Bahasa Indonesia (KBBI). Contoh hasil dari langkah *normalization* dapat dilihat dalam Gambar 7.



stopword	normalization
[gampang, ahok, susah, cari, calon, susah, aja...	[gampang, ahok, susah, cari, calon, susah, aja...
[ganjar, bagus, kendara, politik, banteng, mer...	[ganjar, bagus, kendara, politik, banteng, mer...
[liat, taun, taun]	[lihat, tahun, tahun]
[prabowo, contoh, megawati, bagus, prabowo, pi...	[prabowo, contoh, megawati, bagus, prabowo, pi...
[maap, gimana, cerita, putin]	[maap, bagaimana, cerita, putin]

Gambar 7. Contoh hasil tahap *normalization*

4.8. *Transformation*

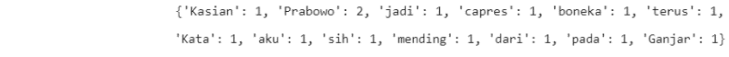
Sebelum melakukan tahap *transformation*, data dilakukan proses labelisasi, yang dilakukan dengan memanfaatkan kamus *lexicon* dari penelitian sebelumnya yang dilakukan oleh [11]. Proses penentuan label pada suatu data komentar didasarkan pada akumulasi nilai dari seluruh kata yang terdapat dalam komentar tersebut. Jika nilai akumulasi suatu komentar adalah >0, maka komentar tersebut akan diberi label positif. Sebaliknya, jika nilai akumulasi <0, maka komentar tersebut akan diberi label negatif, dan apabila nilai sentimen 0 maka kalimat sentimen tersebut masuk ke dalam kelas netral. Hasil dari proses labelisasi data komentar dapat dilihat pada Tabel 1.

Tabel 1. Hasil labelisasi data komentar

Kelas Sentimen	Jumlah
Positif	132
Negatif	162
Netral	885
Total	**1179**

Jumlah data komentar setelah dilakukan proses preprocessing yaitu 1179. Hasil dari labelisasi sentimen dengan kamus *lexicon* berhasil mengelompokkan sebanyak 1179 ke dalam 3 kelas yaitu 132 untuk data kelas positif, data ke dalam kelas negatif 162 dan data kelas netral 885.

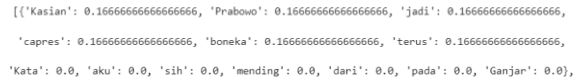
Selanjutnya pada tahap *transformation*, kata-kata yang telah melalui proses *preprocessing*, dilakukan pembobotan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk mengubah setiap kata menjadi representatif numerik. Pada tahap awal, perhitungan dilakukan untuk menghitung berapa kali setiap kata muncul dalam setiap komentar. Contoh hasil perhitungan jumlah kata pada setiap komentar dapat dilihat dalam Gambar 8.



```
{'Kasian': 1, 'Prabowo': 2, 'jadi': 1, 'capres': 1, 'boneka': 1, 'terus': 1,
'Kata': 1, 'aku': 1, 'sih': 1, 'mending': 1, 'dari': 1, 'pada': 1, 'Ganjar': 1}
```

Gambar 8. Contoh jumlah kata pada komentar

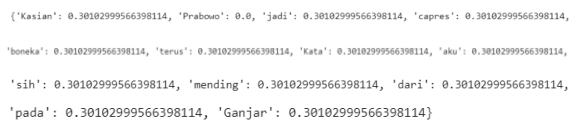
Dibawah ini adalah contoh hasil dari nilai *Term Frequency* (TF) yang bisa dilihat dalam Gambar 9.



```
[{'Kasian': 0.16666666666666666, 'Prabowo': 0.16666666666666666, 'jadi': 0.16666666666666666,
 'capres': 0.16666666666666666, 'boneka': 0.16666666666666666, 'terus': 0.16666666666666666,
'Kata': 0.0, 'aku': 0.0, 'sih': 0.0, 'mending': 0.0, 'dari': 0.0, 'pada': 0.0, 'Ganjar': 0.0},
```

Gambar 9. Hasil dari *term frequency*

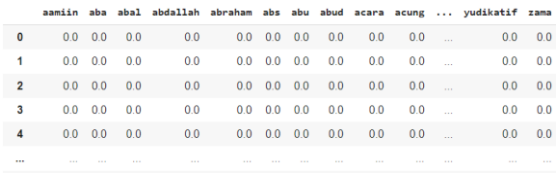
Dibawah ini adalah contoh hasil dari nilai *Inverse Document Frequency* (IDF) yang bisa dilihat dalam Gambar 10.



```
{'Kasian': 0.30102999566398114, 'Prabowo': 0.0, 'jadi': 0.30102999566398114, 'capres': 0.30102999566398114,
'boneka': 0.30102999566398114, 'terus': 0.30102999566398114, 'Kata': 0.30102999566398114, 'aku': 0.30102999566398114,
'sih': 0.30102999566398114, 'mending': 0.30102999566398114, 'dari': 0.30102999566398114,
'pada': 0.30102999566398114, 'Ganjar': 0.30102999566398114}
```

Gambar 10. Contoh hasil idf

Kemudian dibawah ini contoh hasil dari nilai *Term Frequency-Inverse Document Frequency* (TF-IDF) yang bisa dilihat dalam Gambar 11.



	aamiin	aba	abal	abdallah	abraham	abs	abu	abud	acara	acung	...	yudikatif	zama
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...

Gambar 11. Hasil tf-idf

4.9. *Data Mining*

Setelah melewati tahap *preprocessing* dan *transformation*, langkah berikutnya adalah membagi data menjadi dua dataset, yakni data pelatihan dan data pengujian. Pembagian ini dilakukan dalam empat perbandingan yang berbeda, yaitu 90:10, 80:20, 70:30, dan 60:40. Proses ini dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dengan *kernel linear* yang telah dilakukan, sebagaimana ditampilkan dalam Tabel 2.

Tabel 2: Hasil akurasi 4 skenario

Skenario	Akurasi (%)
Skenario 1 (90 : 10)	71
Skenario 2 (80 : 20)	73
Skenario 3 (70 : 30)	77
Skenario 4 (60 : 40)	76

4.10. *Evaluation*

Tahap evaluasi ini dilakukan pengujian data menggunakan Confusion Matrix. Pada tahap ini, proses pengujian dilakukan terhadap hasil klasifikasi yang dihasilkan oleh Support Vector Machine dengan menggunakan kernel linear. Hasil pengujian model melibatkan beberapa metrik, termasuk akurasi, presisi, *recall*, dan *f1-score*. Dibawah ini adalah hasil evaluasi yang dihitung untuk setiap skenario model yang telah diuji:

1. Skenario (90:10)

Berikut adalah hasil dari pengujian yang telah dilakukan pada klasifikasi komentar pengguna

instagram dengan menggunakan algoritma *Support Vector Machine* (SVM) yang menggunakan *kernel linear*, dengan jumlah data pelatihan sebanyak 1061 dan data pengujian sebanyak 118. Berikut hasilnya dapat dilihat dalam Gambar 12.

Confusion Matrix:

	precision	recall	f1-score	support
negative	0.33	0.05	0.09	19
neutral	0.74	0.98	0.84	84
positive	0.25	0.07	0.11	15
accuracy			0.71	118
macro avg	0.44	0.37	0.35	118
weighted avg	0.61	0.71	0.63	118

Accuracy: 0.711864406779661
Precision: 0.6113337914185372
Recall: 0.711864406779661
F1-score: 0.6267150334054884

Gambar 12. Hasil Uji skenario 1 SVM *kernel linear*

Dari Gambar 12, dapat disimpulkan bahwa dengan menggunakan perbandingan data pelatihan sebesar 90% dan data pengujian sebesar 10%, hasil klasifikasi mencapai tingkat akurasi sebesar 71%, presisi 61%, *recall* 71%, dan *f1-score* 63%.

2. Skenario 2 (80:20)

Dibawah ini adalah hasil pengujian yang telah dilakukan untuk mengklasifikasikan komentar pengguna instagram menggunakan algoritma *Support Vector Machine* (SVM) dengan *kernel linear*. Dalam pengujian ini, terdapat 943 data latihan dan 236 data pengujian. Detailnya bisa dilihat pada Gambar 13.

Classification Report:

	precision	recall	f1-score	support
negative	0.40	0.10	0.16	41
neutral	0.75	0.97	0.85	171
positive	0.50	0.12	0.20	24
accuracy			0.73	236
macro avg	0.55	0.40	0.40	236
weighted avg	0.67	0.73	0.66	236

Accuracy: 0.7330508474576272
Precision: 0.6670647149460709
Recall: 0.7330508474576272
F1-score: 0.6628317944716575

Gambar 13 Hasil pengujian skenario 2 SVM *kernel linear*

Dari data yang terdapat pada Gambar 13, terlihat bahwa hasil klasifikasi dengan menggunakan perbandingan data pelatihan sebanyak 80% dan data pengujian sebanyak 20% menghasilkan akurasi sekitar 73%, presisi sekitar 73%, *recall* sekitar 73% dan *f1-score* sekitar 66%.

3. Skenario 3 (70:30)

Dibawah ini terdapat hasil dari pengujian yang telah dilakukan dalam mengklasifikan komentar pengguna instagram dengan memanfaatkan algoritma *Support Vector Machine* (SVM) yang menggunakan *kernel linear*. Dalam pengujian ini, terdapat 825 data pelatihan dan 354 data pengujian.

Confusion Matrix:

	precision	recall	f1-score	support
negative	0.38	0.06	0.10	52
neutral	0.78	0.99	0.87	269
positive	0.50	0.12	0.20	33
accuracy			0.77	354
macro avg	0.55	0.39	0.39	354
weighted avg	0.70	0.77	0.70	354

Accuracy: 0.768361581920904
Precision: 0.6974643131748738
Recall: 0.768361581920904
F1-score: 0.6963713726155215

Gambar 14 Hasil pengujian skenario 3 SVM *kernel linear*

Dari Gambar 14, terlihat bahwa hasil klasifikasi dengan menggunakan perbandingan data pelatihan sebanyak 70% dan data pengujian sebanyak 30% menghasilkan akurasi sebesar 77%, presisi 70%, *recall* 77%, dan *f1-score* 70%.

4. Skenario 4 (60:40)

Dibawah ini terdapat hasil pengujian yang telah dilakukan dalam mengklasifikasikan komentar pengguna instagram dengan menggunakan algoritma *Support Vector Machine* (SVM) yang menerapkan *kernel linear*. Dalam pengujian ini, terdapat 707 data pelatihan dan 472 data pengujian. Detailnya bisa dilihat pada Gambar 15.

Confusion Matrix:

	precision	recall	f1-score	support
negative	0.75	0.04	0.08	71
neutral	0.76	1.00	0.86	355
positive	0.33	0.02	0.04	46
accuracy			0.76	472
macro avg	0.61	0.35	0.33	472
weighted avg	0.72	0.76	0.67	472

Accuracy: 0.7584745762711864
Precision: 0.7178843174776744
Recall: 0.7584745762711864
F1-score: 0.665402004538897

Gambar 15 Hasil pengujian skenario 4 SVM *kernel linear*

Berdasarkan Gambar 15 dapat dilihat bahwa hasil dari klasifikasi dengan menggunakan presentase 70% data training dan 30% data testing, didapatkan hasil dengan nilai accuracy 76%, precision 72%, recall 76% dan f1-score 67%.

4.11. *WORDCLOUD*

Pada penelitiani ini, pengujian dilakukan dengan membagi data ke dalam empat skenario yang berbeda, yakni dengan perbandingan 90:10, 80:20, 70:30, dan 60:40. Hasil dari pengujian ini menunjukkan bahwa pada skenario ketiga, akurasi mencapai nilai tertinggi yaitu 77% dengan menggunakan *kernel linear*. Sebagai ilustrasi tambahan, contoh hasil dari pencarian kata yang sering muncul dapat dilihat dalam Gambar 16.

Gambar 16 *Wordcloud* topik calon presiden

Berdasarkan Gambar 16 menunjukkan Wordcloud yang memperlihatkan kata-kata yang sering muncul dalam dataset. Kata yang sering muncul diantaranya, yaitu "ganjar", "tidak", "prabowo", dan "anis". Ukuran kata dalam Wordcloud menggambarkan seberapa sering kata tersebut muncul dalam komentar masyarakat pada postingan instagram. Hal ini menunjukkan masyarakat Indonesia banyak menanggapi tentang calon presiden tersebut.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian yang telah dilakukan, beberapa kesimpulan yang dapat ditarik adalah sebagai berikut: Dalam penelitian ini, komentar-komentar pengguna instagram yang berkaitan dengan Calon Presiden 2024, seperti Ganjar Pranowo, Praboowo Subianto, dan Anis Baswedan. Metode yang digunakan adalah algoritma *Support Vector Machine* dan metodologi *Knowledge Discovery in Database*. Data komentar yang diperoleh sebanyak 1200 dengan melalui web scraping exportcomments.com dan melalui proses Preprocessing berhasil menghasilkan 1179 data yang digunakan dan dibagi menjadi 3 kelas, yaitu kelas positif sebanyak 132 data, kelas negatif sebanyak 162 data dan kelas netral sebanyak 885 data. Penggunaan algoritma Support Vector Machine dengan kernel linear dalam tahap data mining mampu mengklasifikasikan komentar dengan cukup baik. Pengujian ini menggunakan 4 Skenario yaitu 90:10, 80:20, 70:30, dan 60:40. Dengan menggunakan confusion matrix untuk melakukan pengujian skor, berdasarkan hasilnya untuk skenario 3 yaitu 70:30, didapatkan akurasi tertinggi sebesar 77%, serta nilai presisi, recall, dan f1-score yang signifikan.

Berdasarkan hasil penelitian ini, teradapat beberapa saran yang dapat dikembangkan lagi pada penelitian selanjutnya diantaranya, yaitu: Dapat dipertimbangkan untuk menggunakan sumber data yang lebih beragam, termasuk komentar dari media sosial lainnya. Lakukan optimasi lebih lanjut terhadap parameter SVM dan kernel yang digunakan. Untuk penelitiannya selanjutnya dapat dipertimbangkan penggunaan metode seleksi fitur berbasis information gain untuk meningkatkan tingkat akurasi hasil analisis.

DAFTAR PUSTAKA

[1] M. Kurnia Maulidina and E. Itje Sela, "ANALISIS SENTIMEN KOMENTAR WARGANET TERHADAP POSTINGAN INSTAGRAM MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DAN TF-IDF (Studi Kasus: Instagram Gubernur Jawa Barat Ridwan Kamil)," 2020. Accessed: Jul. 17, 2023. [Online]. Available: http://eprints.uty.ac.id/id/eprint/6332

[2] W. Budiharto and M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," *J Big Data*, vol. 5, no. 1, Dec. 2018, doi: 10.1186/s40537-018-0164-1.

[3] M. Murni, I. Riadi, and A. Fadlil, "Analisis Sentimen HateSpeech pada Pengguna Layanan Twitter dengan Metode Naïve Bayes Classifier (NBC)," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 566, Apr. 2023, doi: 10.30865/jurikom.v10i2.5984.

[4] R. Valentini Siwabessy Peristiwari, A. Herdiani, and A. Romadhony, "Analisis Sentimen Masyarakat Terhadap Hasil Kerja Petahana Dalam Kaitan Dengan Pemilihan Presiden tahun 2019 Pada Sosial Media Twitter Menggunakan Support Vector Machine (SVM)," 2019. Accessed: Jul. 11, 2023. [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/viewFile/9839/9700

[5] T. I. Hermanto and M. A. Sunandar, "ANALISIS DATA SEBARAN PENYAKIT MENGGUNAKAN ALGORITMA DENSITY BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE," 2020. doi: https://doi.org/10.33084/jsakti.v3i1.1775.

[6] V. A. Flores, L. Jasa, and L. Linawati, "Analisis Sentimen untuk Mengetahui Kelemahan dan Kelebihan Pesaing Bisnis Rumah Makan Berdasarkan Komentar Positif dan Negatif di Instagram," *Majalah Ilmiah Teknologi Elektro*, vol. 19, no. 1, p. 49, Oct. 2020, doi: 10.24843/mite.2020.v19i01.p07.

[7] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, "SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM," *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: https://t.co/NfhnfMjtXw

[8] H. E. W. Candana, I. G. A. Gunadi, and D. G. H. Divayana, "PERBANDINGAN FUZZY TSUKAMOTO, MAMDANI DAN SUGENO DALAM PENENTUAN HARI BAIK PERNIKAHAN BERDASARKAN WARIGA MENGGUNAKAN CONFUSION MATRIX," *Jurnal Ilmu Komputer Indonesia (JIK)*, vol. 6, no. 2, pp. 14–22, 2021.

[9] V. Kevin Sitanayah Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization)," 2020. [Online]. Available: www.tripadvisor.com,

[10] C. F. Hasri and D. Alita, "PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN TERHADAP DAMPAK VIRUS CORONA DI TWITTER," *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 3, no. 2, pp. 145–160, 2022, [Online]. Available: http://jim.teknokrat.ac.id/index.php/informatika

[11] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, Feb. 2021, doi: 10.34148/teknika.v10i1.311