

Analisis Sentimen Berbasis Aspek Kinerja Polri Menggunakan SVM dengan Pendekatan POS Tagging

I Gede Ary Suta Sanjaya¹, I Made Candiasa², Luh Joni Erawati Putri³

¹²³ Universitas Pendidikan Ganesha
Jalan Udayana No.11, Singaraja, Buleleng, Bali, Indonesia

e-mail: suta.arry@gmail.com¹, candiasaimade@undiksha.ac.id², email joni.erawati@undiksha.ac.id³

Received : April, 2024

Accepted : July, 2024

Published : August, 2024

Abstract

Various public opinions regarding POLRI's performance always appear on Twitter. Sentiment analysis is conducted to determine the trend of public opinion sentiment and the performance aspects discussed in these opinions. This research aims to identify performance aspects discussed in the tweet data and the sentiment of the tweet data based on the POLRI's performance. POS Tagging is used to classify POLRI performance aspects including "crime handling", "response's speed", "interaction with people", and "none", and SVM method to classify "positive", "negative", and "neutral" sentiments. Conditional Random Field (CRF) method and Bio-Tags technique used on POS Tagging process by labeling performance aspects-based words collection reflected to performance aspects. Sentiment classification based on POLRI performance aspects was built using SVM. 1103 tweets were obtained using Twitter API from January 2023 to October 2023 as initial datasets, which was divided into 80% training data and 20% testing data. The research results show that the sentiment classification model based on performance aspects using SVM with the POS Tagging approach obtained 81% accuracy, 83% precision, 81% recall, and 81% F1-score. Usage of another classification methods, continuity of use of initial data, and the development of the analyzed aspects can be suggested for future research development.

Keywords: Big Data, Indonesian National Police, Machine Learning, POS Tagging, Support Vector Machine

Abstrak

Beragam opini masyarakat terkait kinerja POLRI terus bermunculan pada media sosial Twitter. Analisis sentimen dilakukan untuk mengetahui kecenderungan sentimen opini masyarakat dan aspek kinerja yang dibahas pada opini tersebut. Penelitian ini bertujuan untuk mengetahui aspek kinerja yang dibahas pada data tweet dan sentimen data tweet berdasarkan aspek kinerja POLRI. POS Tagging digunakan sebagai klasifikasi aspek kinerja POLRI yang meliputi "penanganan kejahatan", "kecepatan respon", "interaksi terhadap masyarakat", dan "tidak beraspek", serta metode Support Vector Machine (SVM) sebagai proses klasifikasi sentimen "positif", "negatif", dan "netral". Proses POS Tagging dilakukan menggunakan metode Conditional Random Field (CRF), melalui teknik Bio-Tags dengan melabelkan aspek kinerja berdasarkan koleksi kata yang mencerminkan klasifikasi aspek kinerja. Pemodelan klasifikasi sentimen berdasarkan aspek kinerja POLRI dibangun dengan metode SVM. Diperoleh 1103 tweet melalui API Twitter dalam kurun waktu Januari 2023 sampai Oktober 2023 sebagai dataset awal, yang dibagi menjadi 80% data training dan 20% data testing. Hasil penelitian menunjukkan model klasifikasi sentimen berbasis aspek kinerja menggunakan SVM dengan pendekatan POS Tagging memperoleh akurasi sebesar 81%, presisi sebesar 83%, recall sebesar 81%, dan F1-score sebesar 81%. Penggunaan metode klasifikasi

machine learning lainnya, kontinuitas penggunaan data awal, dan pengembangan aspek yang dianalisis dapat menjadi saran untuk pengembangan penelitian kedepannya.

Kata Kunci: *Big Data, Machine Learning, Polri, POS Tagging, Support Vector Machine*

1. PENDAHULUAN

Kepolisian Negara Republik Indonesia (Polri) sebagai salah satu institusi yang berada di garda depan penegakan hukum harus bekerja dengan baik sehingga dapat memperoleh kepercayaan publik yang tinggi [1]. Namun, seringkali kinerja Polri menjadi sorotan masyarakat. Beberapa tahun belakangan, kepercayaan masyarakat Indonesia terhadap kinerja Polri mulai menurun seiring dengan munculnya kasus-kasus baik kasus yang *viral* maupun kasus kecil yang memperlihatkan kurangnya kinerja polisi [2]. Beragam opini masyarakat muncul pada media sosial terkait kinerja Polri. Sehingga, opini masyarakat berdasarkan aspek kinerja Polri merupakan hal yang penting untuk dianalisis sehingga dapat mengetahui bagaimana sentimen dari opini masyarakat dan apa aspek kinerja Polri yang dapat dibahas pada opini tersebut sehingga menjadi acuan peningkatan kinerja Polri kedepannya.

Berdasarkan UU No 2 Tahun 2002 Pasal 13, ditegaskan bahwa Kepolisian Negara Republik Indonesia bertugas: memelihara keamanan dan ketertiban masyarakat, menegakkan hukum, dan memberikan perlindungan, pengayoman dan pelayanan kepada masyarakat [3]. Memelihara keamanan dan ketertiban masyarakat erat kaitannya dengan respon kepolisian dalam menanggapi situasi darurat atau potensial yang dapat mengancam integritas keamanan publik. Kecepatan dan efisiensi respon terhadap ancaman terhadap ketertiban masyarakat menjadi poin penting dalam menilai kinerja kepolisian. Menegakkan hukum mencakup upaya kepolisian dalam menginvestigasi, menuntaskan, dan menangani kasus kejahatan. Kapabilitas kepolisian dalam mengelola berbagai aspek hukum yang terkait dengan pelanggaran tercermin dari kinerja kepolisian dalam melakukan penanganan kejahatan. Memberikan perlindungan, pengayoman, dan pelayanan kepada masyarakat mencerminkan interaksi positif kepolisian dengan masyarakat, yang melibatkan respons terhadap kebutuhan dan kepentingan masyarakat serta upaya untuk membangun

hubungan yang positif. Beberapa tugas dan tanggung jawab utama kepolisian menjadi pertimbangan aspek kinerja yang diperhatikan dalam opini masyarakat. Berdasarkan uraian tersebut, maka berdasarkan tugas dan kewajiban Kepolisian Negara Republik Indonesia pada UU No 2 Tahun 2002 Pasal 13, maka beberapa aspek kinerja sebagai representasi tugas dari polisi yang dapat diperhatikan meliputi aspek kinerja 'penanganan kejahatan' yang merepresentasikan tugas menegakkan hukum, 'kecepatan respon' yang merepresentasikan tugas memelihara keamanan dan ketertiban masyarakat, serta 'interaksi terhadap masyarakat' yang merepresentasikan tugas memberikan perlindungan, pengayoman, dan pelayanan kepada masyarakat.

Penggunaan media sosial sebagai sarana untuk berbagi pendapat dan opini menjadi hal yang sering dilakukan oleh kalangan masyarakat. Beberapa penggunaan media sosial dalam penyampaian informasi memberikan kemudahan dalam melakukan pengumpulan data dengan jumlah yang sangat besar yang disebut dengan Big Data. Salah satu penggunaan umum media sosial sebagai sumber big data saat ini adalah Twitter [4]. Sosial media Twitter atau yang sekarang disebut 'X' tergolong banyak digunakan oleh pengguna internet di Indonesia. Berdasarkan survey GWI, media sosial yang paling banyak dipakai oleh warganet Indonesia sepanjang 2022 yaitu media sosial Twitter, yang mencapai 60,2% dari keseluruhan pengguna internet usia 16-64 tahun [5].

Keberagaman pandangan masyarakat dapat menyebabkan adanya kesulitan dalam menyaring opini dan membutuhkan waktu untuk memproses secara menyeluruh. Pemrosesan pendapat atau opini telah dikenal ini sebagai teknik analisis sentimen [6]. Analisis sentimen adalah teknik untuk proses pemahaman, ekstraksi, serta pengolahan data berupa teks yang berjalan secara otomatis sehingga memperoleh informasi sentimen yang terdapat dalam suatu opini [7]. Beberapa algoritma dalam proses *machine learning* dapat

digunakan untuk melakukan analisis sentimen. Namun berdasarkan penelitian dengan judul "Top 10 Algorithm in Data Mining", diperoleh kesimpulan bahwa Support Vector Machine (SVM) merupakan algoritma yang termasuk sepuluh metode algoritma *deep learning* terbaik untuk analisis sentimen [8]. Kelebihan dari SVM yaitu pada proses penentuan jarak yang menggunakan *support vector*, sehingga kecepatan proses komputasi menjadi lebih meningkat dan memperoleh hasil model klasifikasi yang baik, walaupun dilatih dengan himpunan data yang relatif sedikit [9].

Selain itu, metode Part-of-Speech (POS) Tagging juga dapat digunakan sebagai proses pemberian label POS atau kelas sintaktik pada tiap kata di dalam korpus [10]. Pelabelan kata pada POS Tagging dapat dilakukan dengan dua cara, yang meliputi pelabelan dengan berbasis aturan (*rule-based*), serta pelabelan dengan menggunakan probabilitas (*probability-based*) dari model yang dirancang. Teknik *Rule-based tagging* berjalan dengan proses *top-down*, yang dilakukan dengan konsultasi terhadap ahli bahasa/linguistik untuk menentukan aturan pembahasan yang digunakan sehari-hari. Sedangkan pada teknik *probability-based tagging*, proses dilakukan dengan cara *bottom-up*, yang dilakukan dengan menggunakan korpus dalam pelatihan data sehingga dapat ditentukan secara probalistik terkait tag kata terbaik di dalam suatu konteks [11]. Salah satu teknik *probability-based tagging* yang dapat digunakan adalah *Conditional Random Field* (CRF). CRF merupakan model pada *probability-based* sebagai prediksi data dengan karakteristik sekuensial. Konsep dari metode ini adalah dengan menggunakan informasi kontekstual dari label sebelumnya untuk meningkatkan jumlah informasi model serta dapat membangun prediksi yang lebih baik [12]. Apabila dibandingkan dengan metode *Hidden Markov Model* (HMM), CRF memiliki kelebihan pada sifatnya yang *conditional*, yang mana ini dapat mengurangi ketergantungan asumsi yang ditemukan pada kebutuhan metode HMM sehingga inferensi mudah dikerjakan [13].

Penelitian terkait analisis sentimen kinerja Kepolisian Negara Republik Indonesia sebelumnya sudah pernah dilakukan sebelumnya. [2] pada penelitiannya menggunakan algoritma Naïve Bayes dalam mengklasifikasi sentimen masyarakat pada

media sosial Twitter ke dalam bentuk klasifikasi sentimen positif dan negatif. Hasil penelitian diperoleh bahwa dengan nilai akurasi tertinggi diperoleh sebesar 81%. Penelitian terkait penggunaan metode POS Tagging sebelumnya pernah dilakukan, dengan berfokus pada proses POS Tagging pada berita Indonesia dengan metode CRF dan HMM. Penelitian ini memperoleh hasil bahwa metode CRF memberikan nilai akurasi yang lebih tinggi dibandingkan dengan HMM, dimana hasil POS Tagging menggunakan CRF memberikan nilai akurasi sebesar 97,68% dibandingkan dengan HMM yang memberikan nilai akurasi sebesar 96,25%. [14]. Penelitian terkait metode perbandingan metode analisis sentimen pernah dilakukan sebelumnya, dengan membandingkan metode SVM dan Naïve Bayes dalam proses analisis sentimen penggunaan kendaraan listrik di Indonesia pada Twitter. Penelitian ini memperoleh hasil bahwa metode SVM memberikan hasil akurasi yang lebih baik dibandingkan dengan Naïve Bayes, yaitu akurasi dengan metode SVM sebesar 70,83% dan metode Naïve Bayes sebesar 63,02% [15]. Penelitian terkait penggunaan SVM dalam proses analisis sentimen sebelumnya pernah dilakukan dengan melakukan perbandingan ekstraksi fitur *Bag of Words* (BoW) dan TF-IDF pada proses analisis sentimen terhadap review film pada situs IMDb. Penelitian ini memperoleh hasil bahwa analisis sentimen menggunakan SVM dengan metode ekstraksi fitur TF-IDF memberikan hasil yang lebih baik dibandingkan dengan BoW, yaitu TF-IDF sebesar 91,27% dan BoW sebesar 88,59% [16].

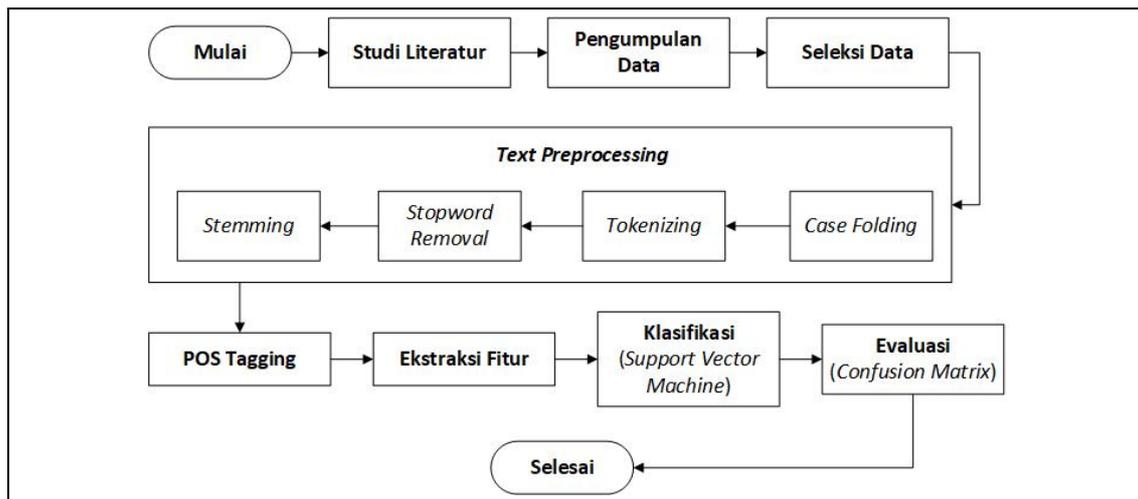
Penelitian mengenai klasifikasi aspek kinerja Polri dan klasifikasi sentimen berdasarkan aspek kinerja Polri pada opini masyarakat di Twitter belum pernah dilakukan sebelumnya. Berdasarkan hal tersebut, dilakukan penelitian terkait analisis sentimen masyarakat berbasis aspek kinerja Polri menggunakan SVM dengan pendekatan POS Tagging. Penelitian ini bertujuan untuk mengetahui performansi dari model klasifikasi sentimen berbasis aspek kinerja Polri menggunakan SVM dengan pendekatan POS Tagging, yang berfokus pada penentuan aspek kinerja menggunakan POS Tagging serta klasifikasi sentimen berbasis aspek kinerja yang menggunakan SVM, dimana penelitian ini belum pernah dilakukan sebelumnya. Klasifikasi aspek Polri meliputi 'penanganan kejahatan', 'kecepatan respon',

‘interaksi terhadap masyarakat’, dan ‘tidak beraspek’ serta klasifikasi sentimen yang dikelompokkan menjadi kelas sentimen ‘positif’, ‘negatif’, dan ‘netral’.

2. METODE PENELITIAN

Metode penelitian penting untuk dijabarkan sehingga alur penelitian menjadi lebih

sistematis dan memperoleh hasil yang sesuai dengan tujuan penelitian [17]. Diagram alur penelitian terkait proses analisis sentimen berbasis aspek kinerja melalui media Twitter menggunakan metode SVM dan pendekatan POS Tagging ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

2.1 Pengumpulan Data

Tahap pengumpulan data yang dilakukan pada penelitian ini adalah dengan melakukan *crawling* data *tweet* dari Twitter menggunakan Twitter API melalui *tools* Tweet Harvest, yang nantinya hasil dari proses *crawling* akan disimpan ke dalam file dengan format CSV. Data yang diambil adalah *tweet* dari pengguna Twitter dalam *range* waktu dari Januari 2023 sampai Oktober 2023 yang membahas mengenai Polri dengan kata kunci yang digunakan meliputi ‘polri’, ‘polisi’, ‘#polri’, dan ‘#polisi’.

2.2 Seleksi Data

Data *tweet* yang telah dikumpulkan selanjutnya akan masuk ke tahap seleksi data. Dilakukan pelabelan sentimen secara manual pada data awal yang dalam pelabelannya melibatkan ahli bahasa yang merupakan salah seorang guru Bahasa Indonesia di salah satu sekolah negeri di Bali. Proses pelabelan data dilakukan sebelum *text preprocessing*, dikarenakan pelabelan dilakukan secara manual sehingga pada saat proses pelabelan, *dataset* harus dalam kondisi utuh. Hal ini bertujuan untuk menghindari tidak

validnya proses pelabelan karena hilangnya makna dari suatu *tweet* apabila dilakukan *text preprocessing*.

2.3 Text Preprocessing

Setelah data dilakukan pelabelan manual, tahap selanjutnya adalah dengan melakukan *text preprocessing*, yaitu dengan serangkaian pemrosesan teks pada *tweet* sehingga nantinya teks pada *tweet* dapat dianalisa dengan mudah pada tahapan analisis sentimen berikutnya. Adapun, tahapan *text processing* yang dilakukan meliputi:

- 1) *Case Folding*, yaitu tahap mengkonversi keseluruhan teks pada *tweet* menjadi bentuk standar dengan proses perubahan karakter dari besar menjadi kecil.
- 2) *Tokenizing*, yaitu proses memotong-motong kalimat menjadi string input berdasarkan tiap kata penyusunnya. Ini bertujuan untuk membagi teks menjadi unit yang tepat yang disebut token.
- 3) *Stopword Removal*, yaitu tahap menghapus beberapa kata yang tidak memiliki pengaruh pada kalimat

- 4) *Stemming*, yaitu tahap mengubah setiap kata pada kalimat menjadi kata dasar.

2.4 POS Tagging

Tahapan POS Tagging dilakukan dengan menggunakan teknik Bio-Tagging (*Beginning, Inside, Outside*) pada POS Tagging dengan menentukan koleksi kata yang merepresentasikan aspek kinerja aspek kinerja 'penanganan kejahatan', 'kecepatan respon', 'interaksi terhadap masyarakat', dan 'tidak beraspek'.

Proses POS Tagging dilakukan dengan metode *Conditional Random Fields* (CRF). CRF adalah model probabilitas yang digunakan sebagai proses pengenalan pola berurutan, seperti POS Tagging[18]. Persamaan umum mengenai perhitungan CRF dapat dilakukan dengan formula berikut[19].

$$P(y|x) = \exp(wx, y) \sum_y \exp(w\psi(x, y)) \quad (1)$$

Nilai y merupakan label data dan x adalah input data, sedangkan ψ adalah fungsi fitur dan w adalah bobot. Proses pada metode CRF mempertimbangkan ketergantungan kontekstual antar token, yang memungkinkan pengambilan keputusan yang lebih cerdas dan kontekstual dalam proses tagging aspek kinerja Polri.

2.5 Ekstraksi Fitur

Tahapan Ekstraksi Fitur merupakan proses untuk mengubah data teks menjadi format yang dapat diolah oleh algoritma pembelajaran mesin. Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan pada penelitian ini untuk mengekstraksi fitur dari setiap dataset. Proses TF-IDF terdiri dari beberapa tahapan sebagai berikut.

- 1) *Term Frequency* (TF), yaitu mengukur frekuensi kemunculan kata dalam suatu dokumen. Kata yang muncul lebih sering memiliki bobot TF yang lebih tinggi.
- 2) *Inverse Document Frequency* (IDF), yaitu mengukur seberapa unik sebuah kata dengan mempertimbangkan frekuensi kemunculan kata pada dokumen. Kata yang memiliki tingkat frekuensi kemunculan kecil pada korpus teks akan memiliki bobot IDF yang lebih tinggi.

Metode TF-IDF dilakukan dengan menghitung bobot (W) pada setiap data *tweet* terhadap kata kunci, yang perhitungannya dapat dinotasikan pada formula berikut [20].

$$W_{dt} = TF_{dt} * IDF_{dt} \quad (2)$$

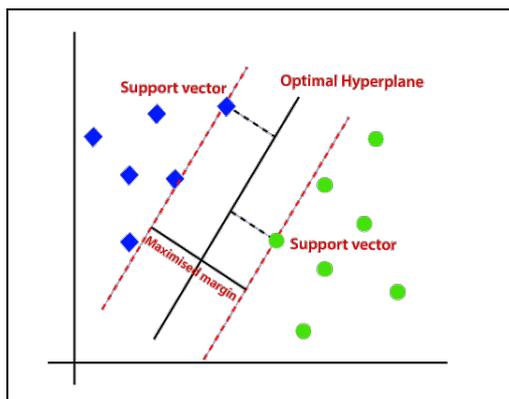
W_{dt} merupakan bobot dokumen ke- d terhadap kata ke- t , TF_{dt} merupakan banyaknya kata yang dicari pada sebuah dokumen, IDF_{dt} merupakan *Inversed Document Frequency* ($\log(N/df)$) dimana N adalah total dokumen dan df adalah banyaknya dokumen yang mengandung kata yang dicari.

Implementasi TF-IDF akan menghasilkan vektor dari fitur untuk setiap dataset. Vektor ini mencakup bobot TF-IDF untuk setiap kata dalam *dataset*, menciptakan representasi numerik yang mencerminkan karakteristik dan signifikansi setiap kata pada proses analisis sentimen.

2.6 Analisis Sentimen

Setelah dilakukan ekstraksi fitur, tahapan selanjutnya adalah dengan melakukan proses analisis sentimen. Tahap proses analisis sentimen dilakukan dengan metode *Support Vector Machine* (SVM). SVM tergolong metode *machine learning* populer yang digunakan untuk melakukan proses *data mining* [21]. *Dataset* yang digunakan pada proses analisis sentimen diambil dari tahapan *Text Preprocessing* sebelumnya, dengan pembagian 80% *dataset training* dan 20% *dataset testing*.

Secara sederhana, prinsip dasar SVM yaitu *linear classifier*, dengan mem-plot setiap *dataset* ke dalam ruang n -dimensi, (n adalah sejumlah fitur yang dimiliki), dan memisahkan *dataset* menjadi dua atau lebih kelas melalui garis pemisah yang disebut dengan *hyperplane*. Ilustrasi dari pemetaan *dataset* dengan *support vector machine* dapat dilihat pada Gambar 2.



Gambar 2. Ilustrasi Sederhana Klasifikasi Support Vector Machine
[Sumber: Pathak dkk. (2022) [22]]

Gambar 2.1 merupakan ilustrasi dari klasifikasi dengan algoritma SVM. Garis *hyperlane* terbaik adalah garis yang berada di antara kedua *support vector* dengan jarak margin maksimal antara garis dan *support vector* [9]. Persamaan umum mengenai perhitungan SVM dilakukan dengan formula berikut[23].

$$f(x) = W^T \cdot x + b \quad (3)$$

Proses analisis sentimen pada penelitian ini melibatkan dua fitur yang terdiri dari data teks pada *tweet* dan pelabelan/ tag aspek kinerja yang telah dilakukan pada tahap POS Tagging, sehingga fitur dapat digambarkan sebagai berikut.

$$x = \begin{bmatrix} \text{Tweet1} & \text{Aspek Kinerja} \\ \text{Tweet2} & \text{Aspek Kinerja} \\ \text{Tweet3} & \text{Aspek Kinerja} \\ \dots & \dots \end{bmatrix}$$

Berdasarkan hal ini, ‘aspek kinerja’ terdiri dari ‘penanganan kejahatan’, ‘kecepatan respon’, ‘interaksi dengan masyarakat’, dan ‘tidak berespek’. Melalui dua fitur dalam proses analisis sentimen, maka notasi persamaan (3) dapat diubah menjadi sebagai berikut.

$$f(x) = w_{tweet} \cdot x_{tweet} + w_{aspek} \cdot x_{aspek} + b \quad (4)$$

Nilai x_{tweet} diperoleh dari pembobotan nilai *tweet* berdasarkan bobot setiap kata yang membentuk *tweet* tersebut. Nilai x_{aspek} diasumsikan dengan nilai 1 untuk label yang berespek ‘penanganan kejahatan’, ‘kecepatan

respon’, dan ‘interaksi terhadap masyarakat’ dan nilai 0 untuk label ‘tidak berespek’.

2.7 Evaluasi

Setelah proses klasifikasi sentimen, selanjutnya dilakukan tahap evaluasi. *Confusion matrix* digunakan untuk memperoleh nilai performansi dari penggunaan metode *Support Vector Machine*. *Confusion matrix* adalah sebuah matriks berukuran 2x2 sebagai representasi hasil klasifikasi berupa nilai biner dari sebuah data. Dilakukan perhitungan untuk mengukur performa klasifikasi, di antaranya akurasi, presisi, *recall*, dan *F-measure* [24]. Adapun hasil yang diperoleh meliputi nilai akurasi (*accuracy*), nilai presisi (*precision*), nilai *recall*, dan nilai *f1-score*. Perhitungan masing-masing *confusion matrix* dapat dilakukan dengan formula sebagai berikut.

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \quad (5)$$

$$Precision = \frac{Tp}{Tp+Fp} \quad (6)$$

$$Recall = \frac{Tp}{Tp+Fn} \quad (7)$$

$$F = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (8)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data *tweet* dilakukan melalui teknik *crawling* menggunakan *tools* Tweet Harvest yang dilakukan dengan sekali proses pengambilan data dengan range data *tweet* dari Januari 2023 sampai Oktober 2023. Proses pengumpulan data meliputi pengambilan data *tweet* menggunakan kata kunci yang berkaitan dengan Polri. Hasil dari proses *crawling* yaitu diperoleh data sekitar 1103 *tweet* yang akan disimpan ke dalam file dengan format CSV. Adapun hasil proses *crawling* dapat dilihat pada Gambar 3 dan Gambar 4.



Gambar 3. Total Data Diperoleh

	created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str	conversation_id_str	username	tweet_url
0	Mon Feb 27 23:38:51+0000 2023	1,63E+18	Malik Ohoiwer: Penetapan Tersangka Irjen Ferdy...	0	0	0	0	in	1,62E+18	1,63E+18	Nonstopid	https://twitter.com/Nonstopid/status/163035178...
1	Mon Feb 27 22:43:10+0000 2023	1,63E+18	Apel Pagi Awal Kinerja Polri Yang Presisi." #D...	0	0	4	7	in	8,50E+17	1,63E+18	humasreswaropen	https://twitter.com/humasreswaropen/status/163...
2	Mon Feb 27 16:23:07+0000 2023	1,63E+18	@chimmykuu sampai detik ini saya masih kurang ...	0	0	0	0	in	1,47E+18	1,63E+18	jokower_sejati	https://twitter.com/jokower_sejati/status/16301...
3	Mon Feb 27 13:00:47+0000 2023	1,63E+18	Membuka langsung Rakernis Gabungan Pusedokkes, ...	3	42	35	266	in	1,43E+18	1,63E+18	ListyoSigIP	https://twitter.com/ListyoSigIP/status/163019...
4	Mon Feb 27 12:59:22+0000 2023	1,63E+18	Kapolda Aceh Irjen Pol. Drs. Ahmad Haydar, S. ...	1	0	0	3	in	1,25E+18	1,63E+18	bidhumasaceh	https://twitter.com/bidhumasaceh/status/163019...
5	Mon Feb 27 10:10:38+0000 2023	1,63E+18	@kailixia @anisizzor @andikjy @jujuprek @tamriz...	0	0	0	0	in	5,22985093	1,63E+18	sayangxoloe	https://twitter.com/sayangxoloe/status/16301...
6	Mon Feb 27 09:16:56+0000 2023	1,63E+18	Ihwadsda Polda Kaltara melaksanakan Audit Kiner...	0	0	0	1	in	1,60E+18	1,63E+18	polda_kaltara	https://twitter.com/polda_kaltara/status/1630...
7	Mon Feb 27 08:33:06+0000 2023	1,63E+18	@CNNIndonesia Tenaga medis polri dan TNI, beke...	0	0	0	1	in	1,60E+18	1,63E+18	syafrialdasi38	https://twitter.com/syafrialdasi38/status/163...
8	Mon Feb 27 07:22:05+0000 2023	1,63E+18	@chimmykuu Beneran jelek banget dah kinerja po...	0	0	0	0	in	1,19E+18	1,63E+18	kojen_116	https://twitter.com/kojen_116/status/163010597...
9	Mon Feb 27 07:06:15+0000 2023	1,63E+18	Lemkapi Puji Kapolri soal Mutasi Pejabat Polri...	0	0	0	0	in	1,63E+18	1,63E+18	Suar4Merakyat	https://twitter.com/Suar4Merakyat/status/16301...

Gambar 4. Hasil Proses Pengumpulan Data

3.2 Seleksi Data

Setelah data *tweet* berhasil dikumpulkan, selanjutnya adalah tahap seleksi data. Tahap ini melakukan proses pelabelan sentimen pada setiap data *tweet* dengan melibatkan ahli bahasa yang merupakan guru Bahasa Indonesia. Label sentimen yang akan diberikan pada keseluruhan *dataset*, meliputi label positif, negatif, dan netral. Data hasil dari proses pelabelan sentimen secara manual akan disimpan pada file CSV dengan nama *dataset_labelled.csv*. Adapun, contoh hasil pelabelan sentimen manual yang telah dilakukan dan distribusi label sentimen dari keseluruhan data berturut-turut dapat dilihat pada Gambar 5 dan Gambar 6.

```
data = pd.read_csv("dataset_labelled.csv", sep=';')
data[['full_text', 'polarity_2']].head(5)
```

	full_text	polarity_2
0	Malik Ohoiwer: Penetapan Tersangka Irjen Ferdy...	positive
1	Apel Pagi Awal Kinerja Polri Yang Presisi." #D...	neutral
2	@chimmykuu sampai detik ini saya masih kurang ...	negative
3	Membuka langsung Rakernis Gabungan Pusedokkes, ...	positive
4	Kapolda Aceh Irjen Pol. Drs. Ahmad Haydar, S. ...	positive

Gambar 5. Contoh Data yang Sudah Dilabelkan Sentimen Manual

```
print(data['polarity_2'].value_counts())
```

polarity_2	
negative	404
positive	377
neutral	322
Name: count, dtype: int64	

Gambar 6. Jumlah Data pada Setiap Label Sentimen

Terlihat pada Gambar 5 dan Gambar 6 bahwa beberapa *dataset* diberikan pelabelan sentimen yang dilakukan secara manual, yang meliputi pelabelan sentimen positif, negatif, dan netral.

3.3 Text Preprocessing

Setelah dilakukan proses pelabelan sentimen secara manual, tahap selanjutnya adalah melakukan *text preprocessing*. Tahap ini bertujuan untuk mengolah dan membersihkan *dataset* sehingga dapat digunakan pada proses klasifikasi aspek kinerja dan klasifikasi sentimen. Proses *text preprocessing* dilakukan melalui beberapa tahapan, yang akan dijabarkan sebagai berikut.

a. Cleaning dan Casefolding

Tahap *cleaning* merupakan proses membersihkan *dataset* dari karakter atau kata yang tidak diperlukan seperti angka, tanda baca, spasi pada awal dan akhir kalimat, *username*, *url*, *single char*, serta melakukan modifikasi dengan mengganti karakter *html* dengan tanda petik dan mengganti *line* baru dengan spasi. Selain itu, pada tahap ini juga dilakukan proses *case folding* yang merupakan proses mengkonversi keseluruhan teks menjadi huruf kecil (*lowercase*). Contoh hasil proses *cleaning* dan *casefolding* yang telah dilakukan dapat dilihat pada Gambar 7.

	full_text	text_clean
0	Malik Ohoiwer: Penetapan Tersangka Irjen Ferdy...	malik ohoiwer penetapan tersangka irjen ferdy ...
1	Apel Pagi Awal Kinerja Polri Yang Presisi." #D...	apel pagi awal kinerja polri yang presisi divi...
2	@chimmykuu sampai detik ini saya masih kurang ...	chimmykuu sampai detik ini saya masih kurang p...
3	Membuka langsung Rakernis Gabungan Pusedokkes, ...	membuka langsung rakernis gabungan pusedokkes p...

Gambar 7. Hasil Proses *Cleaning* dan *Casefolding*

b. Tokenizing

Setelah proses *cleaning* dan *case folding* dilakukan, selanjutnya adalah melakukan *tokenizing* dengan memotong kalimat *tweet* menjadi string input berdasarkan tiap kata penyusunnya. Contoh hasil *tokenizing* yang telah dilakukan dapat dilihat pada Gambar 8.

text_clean	token
malik ohoiwer penetapan tersangka irjen ferdy ...	['malik', 'ohoiwer', 'penetapan', 'tersangka', 'irjen', 'ferdy', '...']
apel pagi awal kinerja polri yang presisi divi...	['apel', 'pagi', 'awal', 'kinerja', 'polri', 'yang', 'presisi', 'divi', '...']
chimmykuu sampai detik ini saya masih kurang p...	['chimmykuu', 'sampai', 'detik', 'ini', 'saya', 'masih', 'kurang', 'p...', '...']
membuka langsung rakernis gabungan pusdokes p...	['membuka', 'langsung', 'rakernis', 'gabungan', 'pusdokes', 'p...', '...']

Gambar 8. Hasil Proses Tokenizing

c. Stopword Removal

Setelah proses *tokenizing* dilakukan, tahap selanjutnya adalah melakukan *stopword removal*, yaitu dengan menghapus kata-kata yang dianggap tidak berpengaruh terhadap kalimat seperti kata penghubung. Tahapan ini menggunakan *library* Sastrawi sebagai acuan untuk menentukan *token* yang dianggap kata penghubung. Contoh hasil *stopword removal* yang telah dilakukan dapat dilihat pada Gambar 9.

token	stop
['malik', 'ohoiwer', 'penetapan', 'tersangka', '...']	['malik', 'ohoiwer', 'penetapan', 'tersangka', '...']
['apel', 'pagi', 'awal', 'kinerja', 'polri', '...']	['apel', 'pagi', 'awal', 'kinerja', 'polri', '...']
['chimmykuu', 'sampai', 'detik', 'ini', 'saya', '...']	['chimmykuu', 'detik', 'kurang', 'puas', 'kine...']
['membuka', 'langsung', 'rakernis', 'gabungan', '...']	['membuka', 'langsung', 'rakernis', 'gabungan', '...']

Gambar 9. Hasil Proses Stopword Removal

d. Stemming

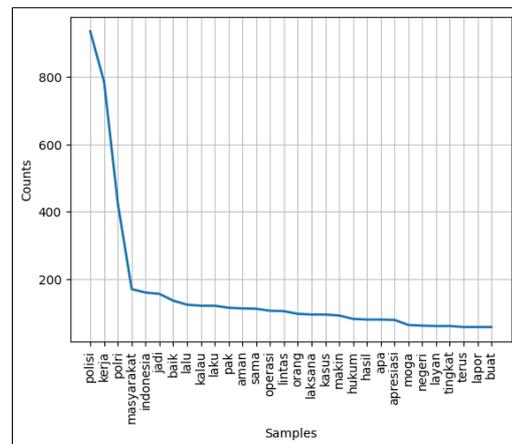
Tahap selanjutnya adalah melakukan *stemming* yaitu dengan mengubah kata pada setiap token menjadi kata dasar. Tahapan ini juga menggunakan *library* Sastrawi sebagai acuan untuk menentukan token yang akan dirubah menjadi kata dasar. Contoh hasil *stemming* yang telah dilakukan dapat dilihat pada Gambar 10.

stop	stemmed
['malik', 'ohoiwer', 'penetapan', 'tersangka', '...']	['malik', 'ohoiwer', 'tetap', 'sangka', 'irjen...']
['apel', 'pagi', 'awal', 'kinerja', 'polri', '...']	['apel', 'pagi', 'awal', 'kerja', 'polri', 'pr...']
['chimmykuu', 'detik', 'kurang', 'puas', 'kine...']	['chimmykuu', 'detik', 'kurang', 'puas', 'kerj...']
['membuka', 'langsung', 'rakernis', 'gabungan', '...']	['buka', 'langsung', 'rakernis', 'gabung', 'pu...']

Gambar 10. Hasil Proses Stemming

e. Hasil Text Preprocessing

Setelah serangkaian tahapan *text preprocessing* telah dilakukan, maka dapat dilihat grafik kemunculan *token* pada *dataset* dapat divisualisasikan pada grafik yang dapat dilihat pada Gambar 11.



Gambar 11. Grafik Frekuensi Kemunculan Token

3.4 POS Tagging

Setelah proses *text preprocessing* dilakukan, maka tahap selanjutnya adalah dengan pelabelan *tag* aspek kinerja pada *dataset*. Beberapa *library* tambahan digunakan pada tahapan ini, yang meliputi *sklearn*, *sklearn_crfsuite*, dan *pycrfsuite*. Aspek kinerja yang ditentukan yaitu 'penanganan kejahatan', 'kecepatan respon', 'interaksi terhadap masyarakat', dan 'tidak berespek'. Proses *tagging* aspek melalui *POS Tagging* pada penelitian ini menggunakan metode *Bio-Tagging*, yaitu dengan menentukan beberapa koleksi kata yang tergolong dalam setiap aspek yang ditentukan. Adapun, beberapa koleksi kata yang digunakan sebagai acuan pelabelan *tag* kinerja Kepolisian Negara Republik Indonesia melalui *Bio-Tagging* pada *dataset* ditunjukkan pada Tabel 1.

Tabel 1. Koleksi Kata pada Setiap Aspek Kinerja

Aspek Kinerja	Koleksi Kata
Penanganan Kejahatan	"jahat", "tangan", "penanganan", "kejahatan", "tangkap", "serah", "periksa", "kasus", "tindak", "investigasi", "menangkap", "kriminal", "pidana"
Kecepatan Respon	"cepat", "respon", "merespon", "langsung", "giat", "gerak", "tuntas", "responsif", "sigap", "tanggap", "segera", "respons"
Interaksi Terhadap Masyarakat	"interaksi", "ramah", "bantu", "membantu", "masyarakat", "rakyat", "orang", "himbauan", "selamat", "lapor", "bimbing", "bimbingan", "komunikatif"

Tabel 1 menunjukkan koleksi kata yang digunakan sebagai acuan dalam pembobotan kata untuk *tag* aspek kinerja. Kata yang muncul selain dari yang ditunjukkan Tabel 1 akan masuk ke dalam kategori 'tidak beraspek'. Proses dari *POS Tagging* dilakukan menggunakan metode Conditional Random Field (CRF). CRF merupakan salah satu metode probabilistik yang digunakan dalam pengembangan *POS Tagging*. Proses penentuan *tag* terhadap *dataset* pada tahap ini dilakukan dengan menggunakan *Python*, yang hasilnya dapat dilihat pada Gambar 12.

full_text	sentiment_aspects
Membuka langsung Rakernis Gabungan PUSDOKKES, ...	kecepatan respon
Kapolda Aceh Irjen Pol. Drs. Ahmad Haydar, S. ...	penanganan kejahatan
@CNNIndonesia Tenaga medis polri dan TNI, beke...	interaksi terhadap masyarakat
Gimana mau percaya sama kinerja polri, kalo in...	penanganan kejahatan
@DivHumas_Polri wow, semoga masyarakat papua t...	interaksi terhadap masyarakat
@btxtdrsunarto @portraperture @galih_husodo @co...	interaksi terhadap masyarakat
@_AnakKolong @ListyoSigitP @DivHumas_Polri SI...	interaksi terhadap masyarakat
@MurtadhaOne1 @PolresJaksel @DivHumas_Polri @L...	interaksi terhadap masyarakat
@GunRomli Percayakan pada kinerja polri, sudah...	penanganan kejahatan
@StefanAntonio_ @mohmahfudmd ini netizen Indo...	penanganan kejahatan

Gambar 12. Hasil Pelabelan Aspek dengan POS Tagging

3.5 Analisis Sentimen

Setelah pemodelan untuk klasifikasi aspek kinerja telah dilakukan, selanjutnya merancang model untuk klasifikasi sentimen pada *dataset*. *Tag* aspek kinerja yang disimpan sebelumnya akan menjadi fitur yang digunakan dalam proses analisis sentimen. Proses perancangan model klasifikasi sentimen berbasis aspek dilakukan melalui beberapa sub tahapan yang meliputi pembagian *dataset*, ekstraksi fitur, dan pembuatan model klasifikasi.

a. Pembagian Data

Tahapan pembagian data merupakan proses untuk membagi *dataset* menjadi *dataset training* serta *dataset testing*. Metode SVM menggunakan variabel *x* sebagai fitur dan

variabel *y* sebagai label yang akan diprediksi. Variabel *x* pada pemodelan klasifikasi aspek kinerja yaitu fitur data *tweet* serta variabel *y* yang merupakan label aspek kinerja yang sebelumnya diproses melalui POS Tagging, sedangkan variabel *x* pada pemodelan klasifikasi analisis sentimen terdiri dari dua fitur yang meliputi data *tweet* dan *tag* aspek kinerja serta variabel *y* yang merupakan label sentimen yang sebelumnya diproses melalui pelabelan manual. *Dataset* dibagi dengan persentase 80% data *training* dan 20% data *testing*. Adapun, komposisi dari data *training* yang berjumlah 80% dari keseluruhan data dapat dilihat pada Gambar 13.

negative	322
positive	291
neutral	245
Name: count, dtype: int64	

Gambar 13. Komposisi Pembagian Data *Training*

Gambar 13 merupakan komposisi dari 80% data *tweet* yang digunakan sebagai data *training*. Terlihat pada Gambar 13 bahwa pada data *training*, komposisi data yang digunakan adalah data *tweet* dengan label sentimen positif sejumlah 291, label sentimen negatif sejumlah 322, dan label sentimen netral sejumlah 245, dengan total keseluruhan data *training* yaitu sebanyak 858 data.

b. Ekstraksi Fitur

Proses ekstraksi fitur yang dilakukan pada tahapan ini yaitu menggunakan TF-IDF, yang merupakan metode untuk melakukan pembobotan vektor pada setiap kata pada *dataset tweet*. TF-IDF dapat dilakukan pada *Python* dengan menggunakan *method TfidfVectorizer* yang diperoleh dari *library Sklearn*. Hasil dari pembobotan vektor ini akan

digunakan untuk proses pemodelan selanjutnya.

c. Pemodelan

Setelah tahapan ekstraksi fitur dilakukan, maka tahap terakhir adalah membangun perancangan model klasifikasi aspek kinerja dan model klasifikasi sentimen berbasis aspek. Model dibangun dengan menggunakan algoritma SVM. Tujuan dari penggunaan SVM adalah yaitu variabel x yang digunakan sebagai komponen prediktor pada pemodelan klasifikasi sentimen berbasis aspek, menggunakan dua fitur yang meliputi teks *tweet* dan label aspek kinerja.

Perhitungan SVM dengan dua fitur dilakukan pada berdasarkan pada persamaan (2). Diberikan sampel fitur yang telah dilakukan proses ekstraksi fitur sehingga menghasilkan nilai numerik, dengan nilai fitur sebagai berikut:

Tabel 2. Sampel Data Fitur

X_{tweet}	X_{aspek}
0,89	1

Maka diperoleh perhitungan untuk setiap data ke- i dengan mengimplementasikan notasi persamaan (2) yaitu sebagai berikut:

$$f(x) = w_{tweet} \cdot 0,89 + w_{aspek} \cdot 1 + b$$

Hasil dari persamaan ini akan digunakan dalam langkah-langkah optimisasi selanjutnya untuk mencari vektor bobot dan bias yang optimal. Implementasi perhitungan ini akan dijalankan dengan bahasa pemrograman Python untuk memperoleh proses perhitungan yang kompleks dengan memanfaatkan library Scikit-Learn.

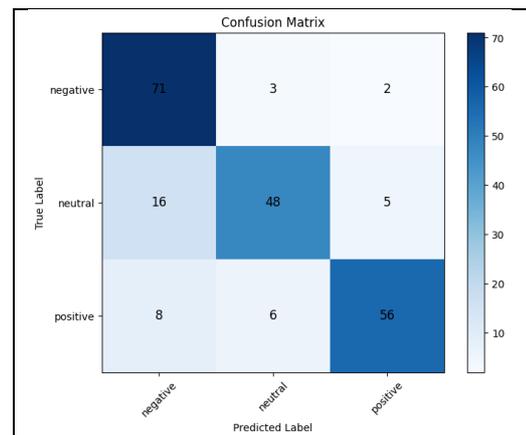
Pemodelan klasifikasi melalui Support Vector Machine pada Python dilakukan dengan pengujian dua kernel yaitu linear dan RBF, dan memilih hasil pemodelan terbaik dari kedua kernel tersebut. Metode pemilihan kernel dengan pengujian dua kernel ini disebut dengan *hyperparameter optimization*. Selain itu, digunakan *library* GridSearchCV dari modul *model_selection* di *library* Scikit-learn. GridSearchCV adalah method yang digunakan untuk melakukan pencarian parameter terbaik untuk model, dengan mencoba semua kombinasi parameter yang mungkin dari suatu kumpulan nilai yang ditentukan.

3.6 Evaluasi

Setelah model klasifikasi berhasil dirancang, tahap terakhir sebelum dilakukan pengembangan sistem adalah dengan melakukan evaluasi terhadap model analisis sentimen berbasis aspek. Evaluasi dilakukan dengan menghitung nilai *confussion matrix* yang terdiri dari akurasi, presisi, *recall*, dan *F1-score*. Proses evaluasi terhadap model klasifikasi sentimen berbasis aspek dilakukan melalui *Python* dengan hasil evaluasi yang berturut-turut dapat dilihat pada Gambar 14 dan Gambar 15.

Akurasi Model Sentimen (setelah tuning): 0.81				
	precision	recall	f1-score	support
negative	0.75	0.93	0.83	76
neutral	0.84	0.70	0.76	69
positive	0.89	0.80	0.84	70
accuracy			0.81	215
macro avg	0.83	0.81	0.81	215
weighted avg	0.82	0.81	0.81	215

Gambar 14. Hasil Evaluasi Model Klasifikasi Sentimen Berbasis Aspek



Gambar 15. Grafik Hasil Evaluasi Model Klasifikasi Sentimen Berbasis Aspek

Gambar 14 menampilkan hasil dari evaluasi model klasifikasi sentimen berbasis aspek, yang dapat dilihat bahwa model yang dibangun memiliki nilai akurasi sebesar 81%, nilai rata-rata presisi sebesar 83%, nilai rata-rata *recall* sebesar 81%, dan *F1-score* sebesar 81%. Selain itu, Gambar 15 menampilkan visualisasi hasil evaluasi *confussion matrix* pada model klasifikasi sentimen berbasis aspek, yang terlihat bahwa terdapat sebanyak 71 data diprediksi sebagai kelas 'negatif' dapat terprediksi dengan benar, sebanyak 48 data diprediksi sebagai kelas 'netral' dapat terprediksi dengan benar, dan

anyak 56 data diprediksi sebagai kelas 'positif' dapat terprediksi dengan benar.

Perhitungan manual dari nilai *confusion matrix* yang meliputi akurasi, presisi, *recall*, dan F1-score dilakukan untuk memastikan bahwa hasil yang diperoleh pada sistem merupakan nilai yang benar. Perhitungan nilai akurasi dilakukan dengan menggunakan persamaan (5) sehingga diperoleh perhitungan sebagai berikut.

$$\text{Akurasi} = \frac{71+48+56}{71+48+56+(3+2)+(16+5)+(8+6)}$$

$$\text{Akurasi} = 0,81395$$

Hasil perhitungan secara manual menunjukkan nilai akurasi yaitu sebesar 0,81395, yang apabila diubah ke dalam bentuk persen maka nilai akurasi menjadi 81%.

Nilai ini sama seperti hasil perhitungan yang dilakukan pada proses evaluasi. Selanjutnya, perhitungan nilai presisi dilakukan dengan menggunakan persamaan (6) sehingga diperoleh perhitungan sebagai berikut.

$$\text{Presisi (positif)} = \frac{56}{56+2+5} = 0,8889$$

$$\text{Presisi (negatif)} = \frac{71}{71+16+8} = 0,7473$$

$$\text{Presisi (netral)} = \frac{48}{48+3+6} = 0,8421$$

Hasil perhitungan secara manual menunjukkan nilai presisi pada masing-masing sentimen yaitu sentimen positif sebesar 0,8889, yang apabila dibulatkan ke dalam desimal dua angka maka menjadi 0,89, sentimen negatif sebesar 0,7473, yang apabila dibulatkan ke dalam desimal dua angka maka menjadi 0,75, dan sentimen netral sebesar 0,8421, yang apabila dibulatkan ke dalam desimal dua angka maka menjadi 0,84, sehingga diperoleh rata-rata nilai presisi yaitu:

$$\text{Presisi} = \frac{0,89+0,75+0,84}{3} = 0,8267$$

Berdasarkan perhitungan tersebut maka hasil perhitungan secara manual menunjukkan nilai rata-rata presisi sebesar 0,8267, yang apabila diubah ke dalam bentuk persen maka menjadi 83%.

Selanjutnya, perhitungan nilai *recall* dilakukan dengan menggunakan persamaan (7) sehingga diperoleh perhitungan sebagai berikut.

$$\text{Recall (positif)} = \frac{56}{56+8+6} = 0,8$$

$$\text{Recall (negatif)} = \frac{71}{71+3+2} = 0,9342$$

$$\text{Recall (netral)} = \frac{48}{48+16+5} = 0,6956$$

Hasil perhitungan secara manual menunjukkan nilai *recall* pada masing-masing sentimen yaitu sentimen positif sebesar 0,8, yang apabila dibulatkan ke dalam desimal dua angka maka menjadi 0,80, sentimen negatif sebesar 0,9342, yang apabila dibulatkan ke dalam desimal dua angka maka menjadi 0,93, dan sentimen netral sebesar 0,6956, yang apabila dibulatkan ke dalam desimal dua angka maka menjadi 0,70, sehingga diperoleh rata-rata nilai *recall* yaitu:

$$\text{Recall} = \frac{0,80+0,93+0,70}{3} = 0,81$$

Berdasarkan perhitungan tersebut maka hasil perhitungan secara manual yang apabila diubah ke dalam bentuk persen maka menunjukkan nilai rata-rata *recall* sebesar 81%.

Selanjutnya, perhitungan nilai F1-score dilakukan dengan menggunakan persamaan (8) sehingga diperoleh perhitungan sebagai berikut.

$$\text{F1-score (positif)} = 2 * \frac{0,89 * 0,8}{0,89 + 0,8} = 0,8426$$

$$\text{F1-score (negatif)} = 2 * \frac{0,75 * 0,93}{0,75 + 0,93} = 0,8303$$

$$\text{F1-score (netral)} = 2 * \frac{0,84 * 0,7}{0,84 + 0,7} = 0,7636$$

Hasil perhitungan secara manual menunjukkan nilai F1-score pada masing-masing sentimen yaitu sentimen positif sebesar 0,8426, yang apabila dibulatkan ke dalam desimal dua angka maka menjadi 0,84, sentimen negatif sebesar 0,8303, yang apabila dibulatkan ke dalam desimal dua angka maka menjadi 0,83, dan sentimen netral sebesar 0,7636, yang apabila dibulatkan ke dalam desimal dua angka maka menjadi 0,76, sehingga diperoleh rata-rata nilai F1-score yaitu:

$$F1\text{-score} = \frac{0,84+0,83+0,76}{3} = 0,81$$

Berdasarkan perhitungan tersebut maka hasil perhitungan secara manual yang apabila diubah ke dalam bentuk persen maka menunjukkan nilai rata-rata F1-score sebesar 81%.

Hasil perhitungan manual *confusion matrix* secara keseluruhan menunjukkan nilai yang sama dengan hasil evaluasi yang dilakukan pada pemodelan, sehingga dapat dikatakan bahwa nilai hasil evaluasi yang muncul merupakan nilai yang benar berdasarkan rumus perhitungan *confusion matrix*.

4. KESIMPULAN

Penelitian ini menjelaskan mengenai pemodelan klasifikasi sentimen berbasis aspek kinerja Polri, dengan mengimplementasi POS Tagging sebagai metode pelabelan aspek kinerja Polri pada data *tweet* dan metode SVM sebagai proses klasifikasi sentimen dengan menggunakan dua fitur meliputi data teks dan aspek kinerja. Berdasarkan hasil penelitian yang telah dilakukan, maka dapat disimpulkan bahwa analisis sentimen berbasis aspek kinerja Polri memberikan hasil klasifikasi yang baik. Evaluasi dari model yang dibangun POS memperoleh hasil *confusion matrix* dengan perolehan nilai akurasi sebesar 81%, nilai rata-rata presisi sebesar 83%, nilai rata-rata recall sebesar 81%, dan F1-score sebesar 81%. Saran yang dapat diberikan pada penelitian ini yaitu beberapa metode *machine learning* lainnya dapat digunakan sebagai perbandingan antara SVM penambahan jumlah dataset awal yang berkelanjutan, serta pengembangan aspek yang dianalisis selain aspek kinerja dapat dilakukan sebagai pengembangan penelitian kedepannya.

DAFTAR PUSTAKA

- [1] K. Tasaripa. "Tugas Dan Fungsi Kepolisian Sebagai Penegak Hukum," *J. Ilmu Huk. Leg. Opin.*, vol. 2 No. 1, pp. 2–9.2013.
- [2] Y. Sadyatma. "Analisis Sentimen Kinerja Kepolisian Republik Indonesia Di Twitter Menggunakan Naive Bayes Classifier", Universitas AMIKOM Yogyakarta, 2022.
- [3] *Undang-Undang Republik Indonesia Nomor 2 Tahun 2002 Tentang Kepolisian Negara Republik Indonesia*.
- [4] X. Liu. "A big data approach to examining social bots on Twitter," *J. Serv. Mark.*, vol. 33 No. 4, pp. 369–379.2019.
- [5] R. Hasya. "7 Media Sosial Paling Banyak Digunakan Warganet Indonesia Sepanjang 2022," *GoodStats*, 2023. [Daring]. Internet: <https://goodstats.id/article/whatsapp-teratas-ini-7-media-sosial-paling-banyak-digunakan-warganet-indonesia-sepanjang-2022-iklw>. [Apr 08, 2024].
- [6] J. Winahyu dan I. Suharjo. "Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes," *Kumpul. Artik. Mhs. Pendidik. Tek. Inform.*, vol. 10 No. 2, pp. 206.2021.
- [7] G. A. Buntoro, T. B. Adji, dan A. E. Purnamasari. "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation," *Citee*, pp. 39–43.2014.
- [8] X. Wu dkk. "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14 No. 1, pp. 1–37.2008.
- [9] N. H. Ovirianti, M. Zarlis, dan H. Mawengkang. "Support Vector Machine Using A Classification Algorithm," *Sinkron.*, vol. 7 No. 3, pp. 2103–2107.2022.
- [10] H. A. Putranto, O. Setyawati, dan W. Wijono. "Pengaruh Phrase Detection dengan POS-Tagger terhadap Akurasi Klasifikasi Sentimen menggunakan SVM," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 5 No. 4, pp. 252–259.2016.
- [11] A. Mulyanto, Y. A. Nurhuda, dan N. Wiyanto. "Penyelesaian Kata Ambigu Pada Proses POS Tagging Menggunakan Algoritma Hidden Markov Model (HMM)" Presented at *Seminar Nasional Metode Kuantitatif*, Lampung, Indonesia, 2017.
- [12] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, dan F. Syafria. "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5 No. 1, pp. 49–54.2022.
- [13] F. Saefulloh. "Part of Speech Tagger untuk Bahasa Indonesia Menggunakan Conditional Random Field (CRF)," 2017.
- [14] R. Banga dan P. Mehndiratta. "Tagging Efficiency Analysis on Part of Speech

- Taggers” Presented at *International Conference on Information Technology (ICIT)*, Bhubaneswar, India, 2017.
- [15] W. Ningsih, B. Alfianda, R. Rahmadden, dan D. Wulandari. “Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4 No. 2, pp. 556–562.2024.
- [16] D. D. Nur Cahyo dkk. “Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method,” *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 8 No. 2, pp. 90–95.2023.
- [17] T. Widyanto, I. Ristiana, dan A. Wibowo. “Komparasi Naïve Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter,” *SINTECH (Science Inf. Technol. J.)*, vol. 6 No. 3, pp. 147–161.2023.
- [18] A. Zilziana, A. A. Suryani, dan I. Asror. “Part of Speech Tagging Menggunakan Bahasa Jawa Dengan Metode Condition Random Fields,” *e-Proceeding Eng.*, vol. 7 No. 2, pp. 8103–8111.2020.
- [19] M. Kurniawan, K. Kusriani, dan M. R. Arief. “Part of Speech Tagging Pada Teks Bahasa Indonesia dengan BiLSTM + CNN + CRF dan ELMo,” *J. Eksplor Inform.*, vol. 11 No. 1, pp. 29–37.2022.
- [20] R. Melita, V. Amrizal, H. B. Suseno, dan T. Dirjam. “Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim),” *J. Tek. Inform.*, vol. 11 No. 2, pp. 149–164.2018.
- [21] N. W. S. Saraswati, I. P. K. S. Putra, I. D. M. K. Muku, dan G. D. Pramitha. “Support Vector Machine For Hoax Detection,” *SINTECH (Science Inf. Technol. J.)*, vol. 6 No. 2, pp. 107–117.2023.
- [22] D. K. Pathak, S. K. Kalita, dan D. K. Bhattacharya. “Hyperspectral image classification using support vector machine: a spectral spatial feature based approach,” *Evol. Intell.*, vol. 15 No. 3, pp. 1809–1823.2022.
- [23] D. Siregar, F. Ladayya, N. Z. Albaqi, dan B. M. Wardana. “Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter,” *J. Stat. dan Apl.*, vol. 7 No. 1, pp. 93–104.2023.
- [24] P. Bhatia. *Data Mining Data Warehouseing Principales and Practical Techniques*. Cambridge: Cambridge University Press, 2019.