

# **IMPLEMENTASI ANALISIS SENTIMEN TWITTER MENGENAI OPINI MASYARAKAT TERHADAP RKUHP TAHUN 2019**

Mungki Astiningrum<sup>1</sup>, Kadek Suarjuna Batubulan<sup>2</sup>, Lucky Alya Sias<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang  
<sup>1</sup>mungki.polinema@gmail.com, <sup>2</sup>kadeksuarjuna87@gmail.com, <sup>3</sup>luckyalyasias@gmail.com

**Abstrak**—Media sosial Twitter merupakan salah satu media yang digunakan untuk memberikan opini melalui tweet. Tweet tersebut dapat memuat opini dan komentar yang berkaitan dengan bidang politik, ekonomi, sosial, hiburan, pendidikan, olahraga, dan lain-lain. Salah satunya terjadi pada bulan September tahun 2019 banyak terjadi demo yang dilakukan oleh mahasiswa maupun masyarakat demi menolak RKUHP. Tujuan dibuat penelitian yaitu memberikan kemudahan bagi pengguna khususnya masyarakat umum mendapatkan informasi respon masyarakat di Twitter tentang RKHUP. Hasil penelitian bisa dijadikan tolak ukur bagaimana tingkat kepuasan dan penilaian masyarakat terhadap RKUHP. Dengan adanya tolak ukur tersebut, media sosial bisa dimanfaatkan sebagai obyek untuk dijadikan alat ukur tingkat kepuasan dan penelitian masyarakat dengan data yang valid tanpa adanya manipulasi data. Dalam penelitian ini, dilakukan analisis sentimen terhadap tweet dengan mengklasifikasikan kedalam kategori mendukung, menyanjung, netral, mengkritik atau menghujat menggunakan algoritma Support Vector Machine. Dataset berjumlah 1500 tweet yang didapatkan dari tweet mengenai RKUHP dalam periode tertentu. Tahapan penelitian yaitu setelah mendapatkan dataset dari Twitter akan mengalami proses preprocessing cleaning, casefolding dan stopword lalu menghitung TF IDF dan masuk ke tahap proses training menghasilkan model klasifikasi untuk proses testing sehingga dapat melakukan klasifikasi. Pada pengujian penelitian ini tingkat akurasi dari algoritma Support Vector Machine. Hasil akurasi klasifikasi algoritma Support Vector Machine menggunakan 70, 80, dan 90 persen data training dan hasil terbaiknya adalah 72%, 74% dan 81%.

**Kata kunci**— Analisis Sentimen, Opini, RKHUP, Support Vector Machine

## I. PENDAHULUAN

Indonesia adalah negara hukum sebagaimana ketentuan pasal 1 ayat (3) UUD 1945 amandemen ketiga. Negara menetapkan aturan-aturan serta kebijakan- kebijakan untuk mengatur masyarakat. Kekuasaan yang dijalankan oleh pemerintah/penguasa dapat diartikan sebagai tindakan negara. Termasuk dalam pemberlakuan Kitab Undang-Undang Hukum Pidana yang biasa disebut KUHP yang ditegaskan keberlakuannya berdasarkan Undang-Undang Nomor 73 tahun 1958 Tentang Menyatakan Berlakunya Undang-Undang Nomor 1 tahun 1946 Republik Indonesia Tentang Peraturan Hukum Pidana Untuk Seluruh Wilayah

Republik Indonesia dan Mengubah Kitab Undang-Undang Hukum Pidana [2].

Media jejaring sosial memberikan peran yang sangat besar khususnya perkembangan teknologi dalam bidang komunikasi yang menjadi tak terbatas. Twitter merupakan sebuah media sosial yang sangat populer di Indonesia. Hal ini terlihat dari jumlah pengguna Twitter yang mencapai 19,5 juta pengguna dari total 330 juta pengguna di dunia [3]. Kata yang terkandung dalam Twitter adalah bahasa alami manusia yang merupakan bahasa dengan struktur kompleks. Kebiasaan masyarakat mengutarakan pendapatnya melalui media sosial terutama Twitter dalam menanggapi kejadian atau hal-hal yang terjadi di lingkungannya dapat menjadi salah satu acuan untuk mengetahui sentimen masyarakat terhadap lingkungan atau kota tempat tinggal masyarakat tersebut berupa kritik atau saran [1]. Salah satunya terjadi pada bulan September tahun 2019 banyak terjadi demo yang dilakukan oleh mahasiswa maupun masyarakat demi menolak RKUHP. Namun juga banyak masyarakat yang mendukung RKHUP untuk disahkan. Selain turun ke jalan masyarakat juga menyampaikan pendapat berupa kritik saran maupun keresahan mereka melalui Twitter. Untuk memperoleh data respon masyarakat biasanya menggunakan survey, dimana beberapa hasil survey ini masih diragukan karena memungkinkan adanya manipulasi data.

Penelitian tentang analisis sentiment pada dokumen Twitter telah banyak dilakukan salah satunya penelitian Analisis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur DKI Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine. Hasil akurasi yang didapatkan pada saat pengujian sebesar 87.80% [5]. Pada penelitian lain mengenai Analisis Sentimen *Twitter* dengan judul Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine didapatkan hasil akurasi tertinggi sebesar 83,3% [3].

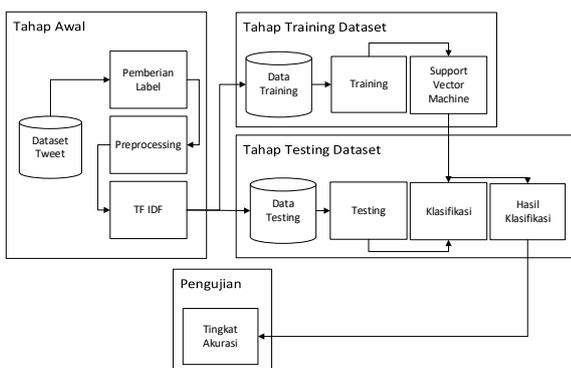
Berdasarkan penjelasan yang telah dipaparkan, akan dilakukan penelitian bertujuan untuk melakukan analisis sentimen tanpa adanya manipulasi data agar tidak disalah gunakan oleh oknum yang tidak bertanggung jawab tentang analisis sentimen twitter untuk mengklasifikasikan tweet mengenai opini masyarakat terhadap RKUHP tahun 2019. Data tersebut akan diolah menggunakan algoritma Support Vector Machine yang mempunyai tingkat akurasi lebih

tinggi dan akan digunakan untuk mengklasifikasi tweet menjadi mendukung, menyanjung, netral, mengkritik atau menghujat. Pengujian akan dilakukan berdasarkan hasil klasifikasi. Pengujian bertujuan untuk mengetahui tingkat akurasi dari klasifikasi. Hasil klasifikasi akan memberikan kemudahan bagi pengguna khususnya masyarakat umum mendapatkan informasi respon masyarakat di Twitter tentang RKHUP. Hasil penelitian bisa dijadikan tolak ukur bagaimana tingkat kepuasan dan penilaian masyarakat terhadap RKHUP. Dengan adanya tolak ukur tersebut, media sosial bisa dimanfaatkan sebagai obyek untuk dijadikan alat ukur tingkat kepuasan dan penelitian masyarakat dengan data yang valid tanpa adanya manipulasi data. Selain itu penelitian ini juga bertujuan untuk menghasilkan dataset baru berupa tweet opini masyarakat mengenai RKHUP tahun 2019 yang bisa digunakan menjadi bahan penelitian selanjutnya.

## II. METODE PENELITIAN

### A. Analisis Sistem

Sistem yang akan dibuat dapat mengklasifikasikan *tweet* ke dalam mendukung, menyanjung, netral, mengkritik atau menghujat. Sistem tersebut dapat menghasilkan hasil klasifikasi sesuai dengan data yang diperoleh. Data yang digunakan diperoleh dari hasil *crawling* pada *Twitter*. Sistem memiliki empat sub proses sehingga dapat menghasilkan satu proses klasifikasi yang mengimplementasikan algoritma *Support Vector Machine*. Berikut dijelaskan gambaran sistem yang akan dibuat dan tertera pada Gambar 1



Gambar 1. Gambaran Sistem.

### B. Pengumpulan Data

Metode pengambilan data digunakan untuk mengumpulkan data-data pendukung yang dibutuhkan dalam proses pembuatan aplikasi. Beberapa cara yang dapat digunakan adalah melalui studi literatur, yaitu dengan mengumpulkan dan mempelajari beberapa referensi dari berbagai sumber yang berkaitan dengan judul penelitian yang dilakukan. Selain itu, untuk dataset yang digunakan diambil melalui Twitter dengan proses *crawling*. Proses *crawling* menggunakan aplikasi pihak ketiga yaitu *Scrapy TweetScraper*. Data tweet yang diambil berdasarkan 10 hastag yaitu *revisikuhp*, *revisirkuhp*, *rkhup*, *ruu*, *ruukuhp*, *ruupertanahan*, *tolakrukuhp*, *tolakruupks*, *tolakruupk*, *tolakruhpngawur*. Dataset yang diambil berupa file *sql*. Data yang akan digunakan dalam periode bulan Agustus 2019 sampai dengan November 2019 dan telah di validasi hasil pelabelannya oleh ahli hukum. Masing-masing hastag

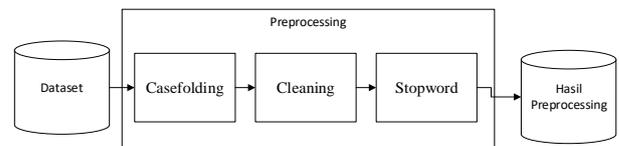
berisi data sebanyak 150, sehingga total data sebanyak 1500 tweet.

### C. Pengolahan Data

Data yang didapat dari Twitter merupakan data yang langsung diambil dari tweet pengguna di Twitter. Data tersebut perlu dilakukan pengolahan agar menjadi data yang mudah digunakan dalam proses *sentiment analysis*. Tweet akan mengalami penyeleksian kata-kata sehingga tweet menjadi lebih ringkas. Beberapa komponen dari tweet akan dihapus untuk menyeleksi tweet. Proses ini dapat disebut dengan *preprocessing*. Setelah melalui *preprocessing*, data yang berupa teks akan diubah ke dalam bentuk angka melalui perhitungan *TF IDF*. Nilai *TF IDF* ini yang akan menjadi masukan untuk algoritma *Support Vector Machine*. Algoritma *Support Vector Machine* akan menghasilkan klasifikasi untuk data yang dimasukkan.

#### 1) Preprocessing

Pada gambar 2 merupakan tahapan *preprocessing* dan selanjutnya akan dijelaskan cara kerja masing-masing *preprocessing*:



Gambar 2. Tahapan *Preprocessing*.

- *Casefolding*

*Casefolding* adalah mengubah semua huruf besar atau kapital dalam Twitter menjadi huruf kecil. Hanya huruf 'a' sampai 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter. Delimiter adalah urutan satu karakter atau lebih yang dipakai untuk membatasi atau memisahkan data yang disajikan dalam plain text.

- *Cleaning*

Tahap *cleaning* akan dilakukan pembersihan tweet dari spesial karakter, URL link, username, serta emoticon. Emoticon disini dihilangkan juga karena akan mempengaruhi hasil *preprocessing*.

- *Stopword*

Tahap *stopword* menghilangkan kata-kata yang tidak bermakna dalam *tweet* yang di *crawling*. Proses *stopword* kata pada *dataset* disesuaikan dengan koleksi *stopword list* yang dimiliki. Pada penelitian ini menggunakan *stopword list* dari Tala [9].

Tabel 1 merupakan contoh hasil *preprocessing* yang menjadi masukan untuk *term frequency*.

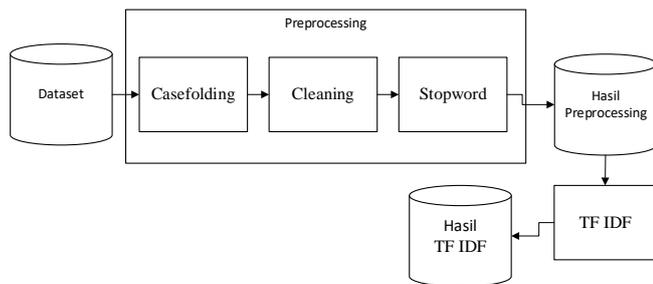
TABEL 1. HASIL *PREPROCESSING*.

No.	Preprocessing	Hasil
1.	<i>Tweet Awal</i>	Wah ternyata banyak sekali pasal ngawur #TundaRKUHP #RevisiKUHP #HapusPasalNgawur https://youtu.be/icS8KPaTErY

2.	Casefolding	wah ternyata banyak sekali pasal ngawur #tundarku #revisiku #hapusasalngawur https://youtu.be/ics8kpatery
3.	Cleaning	wah ternyata banyak sekali pasal ngawur
4.	Stopword	pasal ngawurr

2) TF IDF

TF IDF merupakan salah satu cara yang digunakan untuk menghitung bobot setiap kata dalam sentiment analysis. Perhitungan ini dilakukan dengan cara menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap kata dalam setiap tweet. Hasil dari perhitungan disimpan dalam basis data sistem untuk digunakan untuk proses selanjutnya.



Gambar 3. Tahapan Term Frequency

3) Algoritma Support Vector Machine

Klasifikasi menggunakan Support Vector Machine dibagi menjadi 2 proses, yaitu proses training dan testing. Nilai TF IDF yang telah didapatkan melalui proses perhitungan akan menjadi masukan untuk proses training dan testing. Proses training digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai acuan dalam klasifikasi dengan data testing. Proses testing yaitu membandingkan kata-kata pada dokumen uji dengan nilai probabilitas kata pada setiap kelas dokumen yang telah tersimpan dalam data training. Berikut adalah algoritma klasifikasi sentimen menggunakan Support Vector Machine:

• Proses Training

- Pada persamaan (3.1) menghitung matriks kernel K dengan kernel Gaussian RBF Kernel yang digunakan adalah Gaussian RBF dengan rumus sebagai berikut.

$$K(x, x) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (3.1)$$

- Pada persamaan (3.2) menghitung Fungsi matriks kernel K menggunakan rumus sebagai berikut.

$$K(x_i, x_j) = \varphi(x) \cdot \varphi(i) \quad (3.2)$$

- Pada persamaan (3.3) menghitung nilai melalui persamaan Lagrange menggunakan persamaan berikut.

$$Ld = \sum_{i=1}^N \alpha_i - \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3.3)$$

- Pada persamaan (3.4) menghitung nilai hyperplane Menghitung nilai w dengan rumus berikut.

$$w = \sum_{i=1}^N \alpha_i y_i \varphi(x_i) \quad (3.4)$$

- Pada persamaan (3.5) menghitung nilai b dengan rumus sebagai berikut.

$$b = -1/2 (w \cdot x^+ + w \cdot x^- + w \cdot x^0) \quad (3.5)$$

• Proses Testing

Pada persamaan (3.6) Klasifikasi dilakukan dengan menggunakan pedoman persamaan hyperplane yang telah didapatkan. Klasifikasi dilakukan dengan cara substitusi nilai data testing ke dalam persamaan hyperplane. Apabila hasil lebih dari atau kurang dari persamaan, data akan masuk ke dalam kategori tertentu.

$$f(\varphi(x)) = \text{sign}(w \cdot \varphi(x) + b) \quad (3.6)$$

III. PENGUJIAN DAN ANALISA

Pada pengujian akurasi sistem dilakukan pengujian terhadap hasil klasifikasi pada proses testing. Untuk mengetahui keakuratan algoritma dalam proses klasifikasi maka akan dibandingkan hasil klasifikasi dengan hasil klasifikasi pelabelan manual oleh responden. Perbandingan tersebut akan dihitung tingkat keakurasiannya menggunakan accuracy, precision, recall dan F-Measure. Semakin tinggi nilai accuracy, precision, recall dan F-Measure pada sebuah algoritma yang digunakan maka menunjukkan bahwa algoritma tersebut berjalan dengan baik dan cocok untuk proses klasifikasi. Pengujian dilakukan dengan memilih data training dan data testing secara acak dari 1500 data yang akan digunakan.

Tabel 2 merupakan pengujian data training sebanyak 70% dan 30% untuk data testing. Pemilihan data training dan data testing dipilih secara acak dari 1500 dataset.

TABEL 2. PENGUJIAN DENGAN KOMPOSISI 70% DAN 30%.

Data Training 70% dan Data Testing 30%				
<b>Precision Mendukung</b>	<b>Precision Menyanjung</b>	<b>Precision Netral</b>	<b>Precision mengkritik</b>	<b>Precision menghujat</b>
62%	49%	78%	70%	27%
<b>Recall Mendukung</b>	<b>Recall Menyanjung</b>	<b>Recall Netral</b>	<b>Recall mengkritik</b>	<b>Recall menghujat</b>
57%	45%	77%	87%	28%
<b>F-Measure Mendukung</b>	<b>F-Measure Menyanjung</b>	<b>F-Measure Netral</b>	<b>F-Measure mengkritik</b>	<b>F-Measure menghujat</b>
57%	47%	78%	68%	22%
<b>Accuracy</b>				<b>74%</b>

Tabel 3 merupakan pengujian data training sebanyak 80% dan 20% untuk data testing. Pemilihan data training dan data testing dipilih secara acak dari 1500 dataset.

TABEL 3. PENGUJIAN DENGAN KOMPOSISI 80% DAN 20%.

Data Training 80% dan Data Testing 20%				
<b>Precision Mendukung</b>	<b>Precision Menyanjung</b>	<b>Precision Netral</b>	<b>Precision mengkritik</b>	<b>Precision menghujat</b>
68%	50%	80%	72%	29%
<b>Recall Mendukung</b>	<b>Recall Menyanjung</b>	<b>Recall Netral</b>	<b>Recall mengkritik</b>	<b>Recall menghujat</b>
60%	47%	80%	90%	30%
<b>F-Measure Mendukung</b>	<b>F-Measure Menyanjung</b>	<b>F-Measure Netral</b>	<b>F-Measure mengkritik</b>	<b>F-Measure menghujat</b>
58%	49%	80%	70%	20%
<b>Accuracy</b>				<b>74%</b>

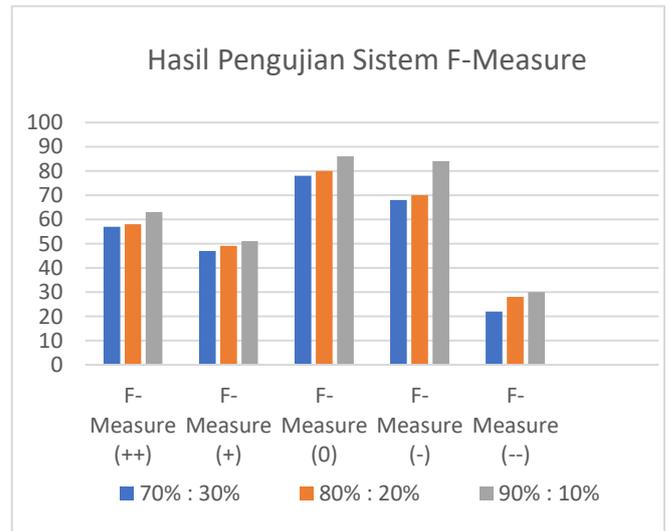
Tabel 4 merupakan pengujian data training sebanyak 90% dan 10% untuk data testing. Pemilihan data training dan data testing dipilih secara acak dari 1500 dataset.

TABEL 4. PENGUJIAN DENGAN KOMPOSISI 90% DAN 10%.

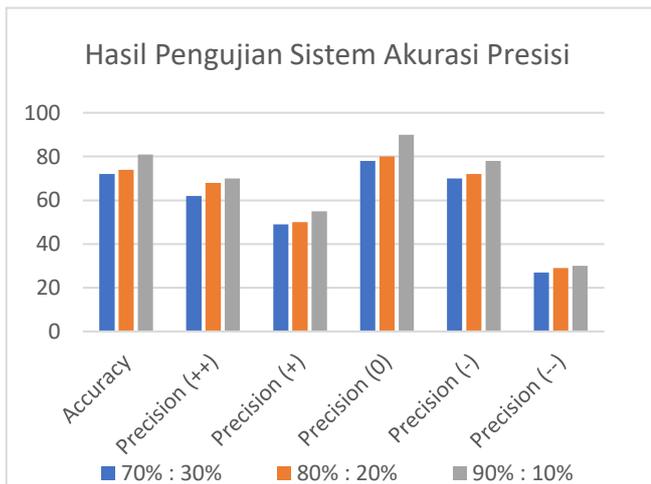
Data Training 90% dan Data Testing 10%				
<b>Precision Mendukung</b>	<b>Precision Menyanjung</b>	<b>Precision Netral</b>	<b>Precision mengkritik</b>	<b>Precision menghujat</b>
70%	55%	90%	78%	30%

Recall Mendukung	Recall Menyanjung	Recall Netral	Recall mengkritik	Recall menghujat
61%	51%	83%	97%	30%
F-Measure Mendukung	F-Measure Menyanjung	F-Measure Netral	F-Measure mengkritik	F-Measure menghujat
63%	51%	86%	84%	30%
Accuracy			81%	

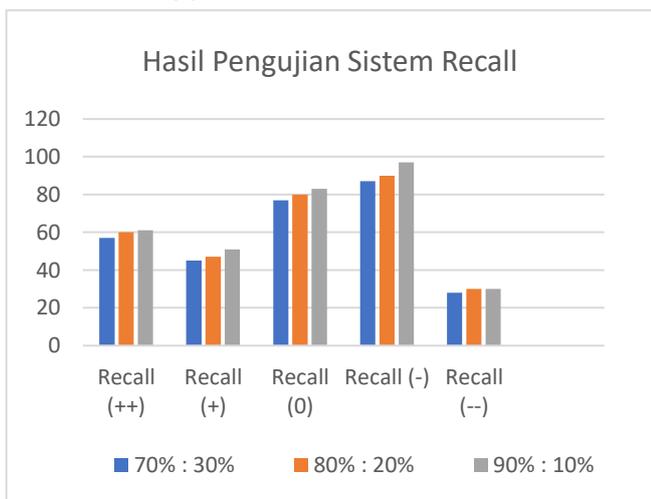
Dari hasil nilai algoritma Support Vector Machine berturut-turut didapatkan sebesar 72%, 74% dan 81%. Nilai akurasi terendah adalah 72% dan tertinggi adalah 81%. Setiap terjadi penambahan komposisi data training, maka nilai dari accuracy, precision, recall dan F-Measure juga mengalami peningkatan. Hal ini dikarenakan algoritma Support Vector Machine merupakan algoritma yang sangat bergantung pada data training, kemungkinan akurasi dapat ditingkatkan lagi dengan menambahkan data training yang lebih banyak lagi. Hasil pengujian dari penelitian ini setiap terjadi penambahan data training maka akurasinya cenderung mengalami peningkatan.



Gambar 4. Hasil Pengujian Sistem F-Measure.



Gambar 4. Hasil Pengujian Sistem Akurasi Presisi.



Gambar 4. Hasil Pengujian Sistem Recall.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

1. Algoritma Support Vector Machine dapat digunakan untuk mengklasifikasikan tweet kedalam mendukung, menyanjung, netral, mengkritik atau menghujat tentang opini masyarakat mengenai RKUHP.
2. Algoritma Support Vector Machine Dari hasil pengujian berturut-turut didapatkan sebesar 72%, 74% dan 81%. Nilai akurasi terendah adalah 72% dan tertinggi adalah 81%. Setiap terjadi penambahan komposisi data training, maka nilai dari accuracy, precision, recall dan F-Measure juga mengalami peningkatan. Hal ini dikarenakan algoritma Support Vector Machine merupakan algoritma yang sangat bergantung pada data training, kemungkinan akurasi dapat ditingkatkan lagi dengan menambahkan data training yang lebih banyak lagi. Hasil pengujian menunjukkan bahwa semakin besar data training yang digunakan, maka nilai accuracy, precision, recall dan F-Measure cenderung mengalami peningkatan.

#### DAFTAR PUSTAKA

- [1] A. A. Arifiyanti, "Klasifikasi Artikel Berita Berbahasa Indonesia Berbasis Naive Bayes Classifier Menggunakan Confix-Stripping Stemmer," Institut Teknologi Sepuluh Nopember Surabaya, 2014.
- [2] F. Fanny, "Tinjauan Yuridis Terhadap Kemerdekaan Menyampaikan Pendapat Dimuka Umum Sebagai Makar Menurut Kuhp" Universitas Andalas Padang, 2019.
- [3] F. Rahutomo, P. Y. Saputra and M. A. Fidyawan, "Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine," *Jurnal Informatika Polinema*, 2017.
- [4] Haviluddin, "Memahami Penggunaan UML (Unified Modelling Language)," *Jurnal Informatika*

*Mulawarman, 2011.*

- [5] E. P. Nuansa, "Analisis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur Dki Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine," Institut Teknologi Sepuluh Nopember, Surabaya, 2017.
- [6] A. O, "Effect of Training Size on SVM and Naive Bayes for Twitter Sentiment Analysis," *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2015.
- [7] P, Y. S., D, H. S., & F, Z. W, "Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes," *Jurnal Informatika Polinema*, 2019.
- [8] A. Sucipto & A. K. Zyen, "Pengembangan Model Support Vector Machines (Svm) Dengan Memperbanyak Dataset Untuk Prediksi Bisnis Forex Menggunakan Metode Kernel Trick," *Jurnal Informatika Upgris*, 2017.
- [9] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *Institute for Logic, Language and Computation Universeit Van Amsterdam*, 2003.