

Terakreditasi SINTA Peringkat 4

Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 28/E/KPT/2019
masa berlaku mulai Vol.3 No. 1 tahun 2018 s.d Vol. 7 No. 1 tahun 2022

Terbit online pada laman web jurnal:
<http://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs>



Vol. 5 No. 2 (2020) 61 - 70

JOINTECS

(Journal of Information Technology and Computer Science)

e-ISSN:2541-6448

p-ISSN:2541-3619

Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis Particle Swarm Optimization

Anas Faisal¹, Yuris Alkhalifi², Achmad Rifai³, Windu Gata⁴

Program Studi Ilmu Komputer, STMIK Nusa Mandiri Jakarta

¹14002356@nusamandiri.ac.id, ²14002360@nusamandiri.ac.id, ³achmad.acf@nusamandiri.ac.id,

⁴windu@nusamandiri.ac.id

Abstract

The use of the internet, especially social media, has become a part of national life. This is partly because many members of the House of Representatives of the Republic of Indonesia (DPR RI) convey ideas, policies and provide comments on government policies through social media. This research was conducted to measure opinions or separate positive and negative sentiments towards the DPR RI. The data used in this study was obtained by crawling on social media twitter. The study was conducted using two algorithms namely Support Vector Machine (SVM) and Naive Bayes (NB). Two algorithms are used in this study, Support Vector Machine (SVM), Naive Bayes (NB), and each is optimized using Particle Swarm Optimization (PSO). The accuracy values of cross-validation test of SVM and NB without PSO are 71.04% and 70.69% with Area Under the Curve (AUC) values of 0.817 and 0.661. The accuracy values of cross-validation test using PSO, for SVM and NB are 75.03% and 73.49%, with AUC values of 0.808 and 0.719. The use of PSO can increase the accuracy of SVM 3.99% and 2.8% for NB. The highest value of accuracy recorded, come from SVM using PSO, measured at 75.03%.

Keywords: legislative member; analysis sentiment; support vector machine; naive bayes; particle swarm optimization

Abstrak

Penggunaan internet terutama media sosial telah menjadi bagian dari kehidupan bernegara. Hal ini salah satunya karena Anggota Dewan Perwakilan Rakyat Republik Indonesia (DPR RI) banyak yang menyampaikan ide, kebijakan maupun memberikan komentar atas kebijakan pemerintah melalui media sosial. Penelitian ini dilakukan untuk mengukur pendapat atau memisahkan antara sentimen positif dan sentimen negatif terhadap DPR RI. Data yang digunakan dalam penelitian ini didapatkan dengan melakukan *crawling* pada media sosial *twitter*. Penelitian dilakukan dengan menggunakan dua Algoritma yaitu Algoritma *Support Vector Machine (SVM)* dan *Naive Bayes (NB)*. Kedua algoritma tersebut masing-masing dioptimasi menggunakan *Particle Swarm Optimization (PSO)*. Hasil pengujian *k-fold cross validation SVM* dan *NB* mendapatkan nilai *accuracy* 71,04% dan 70,69% dengan nilai *Area Under the Curve (AUC)* 0,817 dan 0,661. Sedangkan hasil pengujian *k-fold cross validation* dengan menggunakan *PSO*, untuk *SVM* dan *NB* masing-masing mendapatkan nilai *accuracy* 75,03% dan 73,49% dengan nilai *AUC* 0,808 dan 0,719. Penggunaan *PSO* mampu meningkatkan nilai *accuracy* algoritma *SVM* sebesar 3,99% dan 2,8% pada algoritma *NB*. Hasil dari pengujian kedua algoritma tersebut nilai *accuracy* tertinggi adalah *SVM* dengan *PSO* sebesar 75,03%.

Kata kunci: anggota legislatif; sentimen analisis; support vector machine; naive bayes; particle swarm optimization.

© 2020 Jurnal JOINTECS

1. Pendahuluan

Anggota DPR RI sebagai perwakilan yang telah dipilih oleh warga negara melalui mekanisme Pemilihan

Umum harus kredibel agar menghasilkan kebijakan yang dapat diterima oleh masyarakat. Anggota DPR RI yang kredibel akan bertindak sesuai dengan peraturan

Diterima Redaksi : 12-05-2020 | Selesai Revisi : 18-05-2020 | Diterbitkan Online : 30-05-2020

perundang-undangan. Mereka juga menyampaikan ide dan gagasannya pada tempatnya, selalu mengutamakan kepentingan warga negara, tidak berpolemik, tidak kontroversial, dan selalu berpikir kritis dan komprehensif. Jika tidak demikian, maka akan selalu mendapatkan respon secara spontanitas dari warga negara. Respon tersebut dapat berbentuk dukungan maupun berbentuk kritik yang negatif. Saat ini, anggota DPR RI banyak yang menyampaikan idenya yang belum dituangkan dalam bentuk kebijakan, maupun menyikapi kebijakan yang sudah terbit melalui media sosial. Warga negara pun memberikan respon atas hal tersebut.

Hal ini menjadi menarik untuk mengklasifikasikan sentimen masyarakat terhadap DPR RI. Pada Sub bagian Analisis Media DPR RI terdapat kegiatan yang rutin dilakukan setiap hari yaitu melakukan analisis media. Analisis media dibagi menjadi dua yaitu analisis media online dan analisis media cetak. Sumber berita disesuaikan dengan jenis medianya[1]. Untuk mempermudah dan mempercepat melakukan analisis media atau media monitoring terutama media online perlu dilakukan dengan menggunakan metode dan alat analisis yang tepat.

Internet saat ini sudah menjadi bagian dari kehidupan dan menjadi kebutuhan masyarakat, tidak terkecuali masyarakat di Indonesia. Berdasarkan hasil riset platform yang direlease oleh *We Are Social* pada tahun 2020, disebutkan bahwa pengguna internet di Indonesia sekitar 175,4 juta atau mencapai 64% dari penduduk Indonesia[2]. Salah satu penggunaan internet adalah terkait dengan media sosial. Media sosial dijadikan sarana interaksi sosial di dunia maya. Para pengguna internet menggunakan media sosial sebagai sarana untuk menyampaikan ide-ide mereka, berbagi rutinitas kegiatan, berbagi pengetahuan, dan lain-lain. Disamping itu, media sosial juga digunakan oleh instansi-instansi pemerintahan untuk menyampaikan pesan atas kebijakan-kebijakan yang dikeluarkan sehingga dengan cepat terinformasikan kepada masyarakat.

Sekarang orang lebih sering menggunakan media sosial. Media sosial telah menjadi bagian dari masyarakat global, di mana orang-orang dari seluruh dunia membagikan pendapat mereka secara bebas, mereka berbagi pandangan di situs belanja *online*, blog pribadi, *Facebook*, *YouTube*, *Twitter*, *Weibo*, *WeChat*, dll. Banyak pengguna berbagi pemikiran mereka, dan mendiskusikan aspek politik dan ekonomi masyarakat mereka. Di antara media sosial, *Twitter* adalah platform media sosial paling populer untuk berbagi masalah politik dan masalah sosial[3]. *Twitter* merupakan layanan *micro blogging* yang populer di mana pengguna dapat membuat pesan status (disebut "*tweets*"). Kicauan ini mengandung banyak ekspresi suka, tidak suka, dan kontribusinya pada berbagai topik[4].

Penggunaan internet terutama media sosial telah memiliki dampak yang besar pada kehidupan bernegara. Dengan kebebasan setiap orang melakukan akses dan kegiatan dalam berpendapat di dunia maya, apakah pendapat tersebut bernilai negatif ataukah bernilai positif. Hal ini menjadi sangat penting untuk mengetahui nilai dari informasi yang disampaikan oleh para pengguna media sosial tersebut.

Indonesia adalah salah satu negara yang menganut sistem demokrasi. Salah satu bentuk jalannya sistem demokrasi ditandai dengan dilaksanakannya pemilihan umum secara periodik. Pada tahun 2019 Indonesia melaksanakan pemilihan umum serentak untuk memilih Presiden, DPR, DPRD, dan DPD. Sebelumnya telah dilakukan penelitian menggunakan metode *naive Bayes* untuk mengklasifikasikan sentimen dari hasil *crawling* data pada media sosial *twitter* dengan topik hasil *quick count* pemilu di Indonesia tahun 2019[5].

Setiap kebijakan yang dikeluarkan oleh DPR RI dan pemerintah akan selalu kelihatan menguntungkan sebagian masyarakat dan merugikan sebagian lainnya. Sehingga dengan cepat akan mendapatkan respon dari masyarakat. Sarana yang paling mudah bagi masyarakat untuk memberikan respon adalah dengan memberikan dukungan, komentar, kritik maupun saran untuk perbaikan. Oleh karena itu, untuk meningkatkan kualitas kebijakan yang dihasilkan, salah satunya dapat dilakukan dengan menganalisis sentimen masyarakat terhadap isu-isu kebijakan yang telah dikeluarkan.

Analisis sentimen dengan menggunakan algoritma *SVM* dan Algoritma *SVM* berbasis *PSO* pernah dilakukan terhadap data pengguna *fintech*. Penelitian tersebut memberikan kesimpulan bahwa pengujian dengan Algoritma *SVM* berbasis *PSO* menghasilkan nilai akurasi yang lebih baik yaitu 82,33% dengan toleransi kesalahan 7% dibandingkan menggunakan algoritma *SVM* tanpa menggunakan *PSO* yaitu 75,33% dengan toleransi kesalahan 7,77%[6].

Penelitian yang lain dengan menggunakan *feature selection* pada algoritma *SVM* pernah digunakan untuk melakukan analisis sentimen terhadap Komisi Pemilihan Umum. Pada penelitian tersebut disimpulkan bahwa penggunaan *feature selection weight by correlation* dapat meningkatkan nilai akurasi dan *AUC* [7]. Disamping itu, penelitian analisis sentimen dengan menggunakan *SVM linearly separable* data terhadap penggunaan kotak suara pada pemilu tahun 2019 yang sempat menuai kontroversi juga pernah dilakukan. Berdasarkan penelitian tersebut menyimpulkan bahwa sentimen positif terhadap penggunaan kotak suara kardus adalah aman, hemat, kedap, tahan, didukung dan disetujui[8]. Sehingga seharusnya pemilu tersebut menghasilkan anggota dewan yang kredibel.

Selain itu, Hernawati (2019) telah melakukan penelitian sentimen analisis terhadap operasi tangkap tangan oleh KPK menggunakan algoritma *SVM*, *Naive Bayes*

berbasis *Particle Swarm Optimization*. Pada penelitian tersebut disimpulkan bahwa hasil pengujian terhadap data *tweet* dengan menggunakan metode *SVM* baik sebelum dioptimasi maupun setelah dioptimasi dengan *PSO* selalu menghasilkan akurasi yang lebih baik dari metode *NB*[9]. Bobot atribut dan akurasi suatu algoritma dapat ditingkatkan dengan menggunakan Algoritma *PSO*. *PSO* memiliki konsep yang sederhana, konvergensi cepat, implementasinya mudah dan dapat diterapkan di berbagai bidang untuk memecahkan masalah terkait dengan optimasi [10].

Dengan melakukan *training* tentang klasifikasi *SVM*, maka akan menghasilkan nilai atau pola yang akan digunakan untuk proses pengujian *SVM*, yang bertujuan untuk pemberian label sentimen di *tweet*. Analisis sentimen merupakan proses memahami, mengekstraksi informasi subyektif dari sumber dan memproses kontekstual data secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam kalimat opini[11]. Menurut Wardani (2018), kelebihan dari *SVM* adalah memiliki kemampuan generalisasi yang tinggi[12]: 1). Mampu menghasilkan model klasifikasi yang sangat baik meskipun menggunakan set data *training* yang relatif sedikit dengan pengaturan parameter sederhana. Konsep dan formulasi *SVM* sudah jelas dengan melakukan pengaturan pada beberapa parameter. 2). *SVM* relatif mudah diimplementasikan.

Data pada *website* dan media sosial sangat banyak, sehingga sangat sulit untuk melakukan deteksi sentimen[13]. Pengguna *twitter* membuat kata-kata mereka sendiri dan dengan cara menggunakan ejaan dan tanda baca, melakukan kesalahan dalam mengeja, menggunakan kata-kata gaul, kata-kata baru, menambahkan *url*, dan istilah dan singkatan khusus sesuai dengan klasifikasi umurnya.

Data pada media sosial *twitter* menjadi bahan yang sangat menarik untuk dianalisis karena beberapa hal diantaranya, 1) sebagian besar anggota DPR RI memiliki akun dan aktif pada media sosial tersebut, 2) anggota DPR RI banyak yang secara personal menyampaikan ide dan gagasannya maupun memberikan komentar terhadap kebijakan yang dikeluarkan oleh pemerintah, 3) masyarakat dengan bebas menyampaikan ide, dukungan, tanggapan dan kritik kepada DPR RI.

Opinion Mining (OM) dan *Sentiment Analysis (SA)* adalah dua bidang yang muncul yang bertujuan untuk membantu pengguna menemukan informasi pendapat dan mendeteksi polaritas sentimen. *OM* dan *SA* umumnya digunakan secara bergantian untuk mengekspresikan makna yang sama. Namun, beberapa peneliti menyatakan bahwa mereka bertujuan untuk mengatasi dua masalah yang sedikit berbeda[14]. Analisis Sentimen membangun sistem yang mencoba mengidentifikasi dan mengekstraksi opini dalam teks [4]. Hal ini menyoroti klasifikasi teks dan berkaitan dengan penambahan teks. Polaritas klasifikasi

mengacu pada bentuk paling dasar, yaitu jika teks atau kalimatnya positif atau negatif[15].

Sentimen dapat ditemukan di komentar atau *tweet* untuk memberikan indikator yang berguna untuk berbagai tujuan. Analisis sentimen melalui *Twitter* baru-baru ini menjadi metode yang populer bagi organisasi dan individu untuk memantau pendapat publik terhadap merek dan bisnis mereka. Salah satu tantangan utama yang harus dihadapi oleh metode analisis sentimen *Twitter* adalah *noise* dari data yang dihasilkan *Twitter*. *Twitter* hanya memungkinkan untuk 140 karakter di setiap posting, yang memengaruhi penggunaan singkatan, ekspresi tidak teratur, dan kata-kata yang jarang.

Fenomena ini meningkatkan tingkat sparsitas data, yang memengaruhi kinerja pengklasifikasi sentimen *Twitter*[14]. Metode yang terkenal untuk mengurangi *noise* data tekstual adalah penghapusan *stopwords*. *Stopwords* merupakan kata-kata yang sangat umum yang muncul dalam teks namun memiliki sedikit makna dan hanya memenuhi fungsi sintaksis tetapi tidak menunjukkan substansi. *Stopwords* ini memiliki dua dampak berbeda pada proses pencarian informasi. Namun dapat memengaruhi efektivitas pengambilan karena memiliki frekuensi yang sangat tinggi dan cenderung mengurangi dampak perbedaan frekuensi di antara kata-kata yang kurang umum, yang memengaruhi proses pembobotan[16]. *Stopwords* dapat mengurangi akurasi dan menurunkan kinerja jika dimasukkan dalam pemrosesan teks karena mereka hanya sejumlah kecil dari kata-kata penghubung yang merupakan sebagian kecil dari total teks. Oleh karena itu diperlukan untuk menghilangkan *stopwords* selama fase *pre-processing* [17].

Analisis Sentimen *Twitter* dapat menjadi salah satu alat ampuh untuk menganalisis refleksi berharga dari persepsi publik. Analisis sentimen *Twitter* telah menarik banyak perhatian karena pertumbuhan cepat dalam popularitas *Twitter* sebagai sebuah platform bagi orang-orang untuk mengekspresikan pendapat dan sikap mereka terhadap beragam topik. Pendekatan analisis sentimen *Twitter* cenderung berfokus pada identifikasi sentimen *tweet* individu[18].

Analisis sentimen merupakan proses mengekstrak dan mengolah data secara otomatis dengan menggunakan algoritma tertentu untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat pendapat[19]. Analisis data media sosial dapat membantu bisnis, pemerintah, organisasi keamanan, dan lingkungan untuk mengetahui masalah, saran, dan kritik orang dan menemukan solusi yang tepat untuk masalah mereka. Menganalisis data media sosial perlu menyesuaikan dengan metode dan alat baru, juga perlu pemahaman yang lebih baik tentang pendapat dan kritik serta wawasan masyarakat. Analisis sentimen merupakan bidang penelitian yang muncul dari *Natural*

Language Processing (NLP) untuk mengekstraksi pendapat, pemikiran, dan pandangan orang[3].

Di antara *supervise machine*, SVM merupakan mesin pembelajaran yang populer karena memiliki keunggulan yaitu mampu mengidentifikasi yang terpisah *hyperplane* yang memaksimalkan *margin* antar kedua kelas yang berbeda. SVM adalah pilihan untuk mendapatkan hasil yang baik, karena SVM dapat memproses secara efisien dan efektif koleksi data. SVM memiliki dukungan mesin *vector* untuk mendeteksi sentimen *tweet*[15]. NB adalah algoritma pembelajaran yang didasarkan pada teori *Bayes* dengan menggunakan asumsi yang kuat. Teori *Bayes* merupakan teori tentang menemukan probabilitas tertinggi sesuatu berdasarkan data yang sudah ada[11].

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap lembaga DPR RI di media sosial *twitter*. Dengan adanya analisis sentimen terhadap lembaga negara dan kebijakan yang dikeluarkan, diharapkan dapat secara lebih cepat dan akurat menilai informasi dan komentar yang diberikan oleh masyarakat terhadap kebijakan-kebijakan yang akan diterbitkan berikutnya. Hal tersebut juga digunakan sebagai fungsi pengendalian dan pengawasan oleh masyarakat terhadap wakil-wakilnya yang telah dipilih menjadi anggota DPR RI. Manfaat yang didapat dengan melakukan sentimen analisis adalah mencari informasi berharga yang diperlukan orang lain dari data yang tidak terstruktur [7]. Di samping itu, penelitian ini juga untuk menguji kembali apakah penggunaan algoritma SVM berbasis PSO mendapatkan nilai akurasi dan AUC yang lebih baik dibandingkan dengan Algoritma SVM tanpa PSO, NB tanpa PSO dan NB berbasis PSO.

2. Metode Penelitian

2.1. Model Penelitian

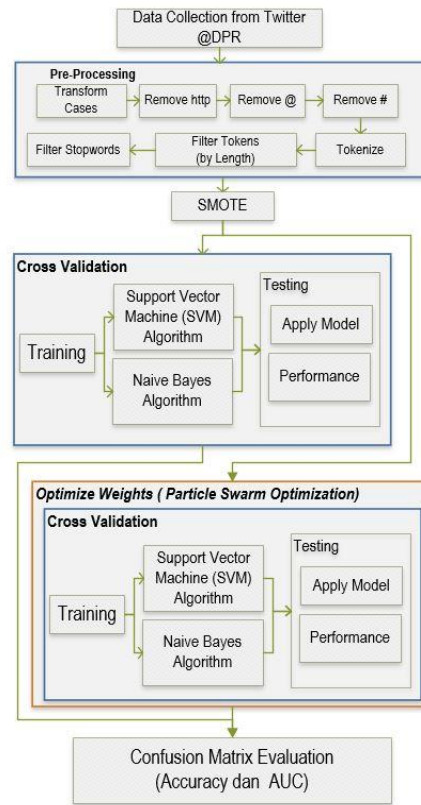
Langkah-langkah yang akan dilakukan pada penelitian ini dituangkan dalam sebuah model kerangka kerja penelitian. Kerangka kerja pada penelitian ini disajikan pada Gambar 1. Berdasarkan kerangka kerja penelitian tersebut, dilakukan pembuatan model analisis data. Tool yang digunakan untuk pemodelan ini menggunakan RapidMiner studio versi 9.2. Model analisis data yang dibuat sebagaimana disajikan pada Gambar 2.

2.2. Teknik Pengumpulan Data

Data yang digunakan untuk penelitian dapat berupa data primer maupun data sekunder. Data primer merupakan data yang diperoleh langsung dari sumber datanya, sedangkan data sekunder merupakan data yang didapatkan dari sumber yang sudah ada. Pada penelitian ini menggunakan data sekunder.

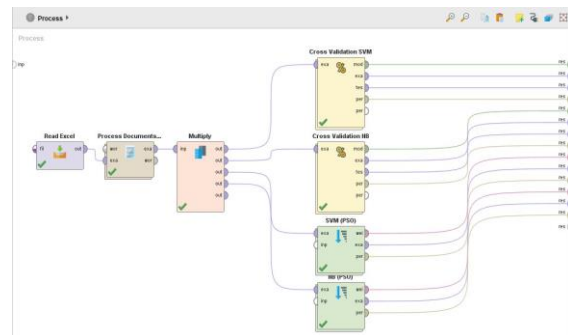
Pengumpulan data *twitter* dilakukan dengan *crawling tool* rapidminer menggunakan operator *search twitter*

dengan *setup parameter query* dpr, jumlah *record* yang *disetting* sebanyak 2.000 dan dengan tipe terkini atau



Gambar 1. Kerangka Kerja Analisis Sentimen

paling populer. Proses *crawling* data dari twitter sebagaimana disajikan pada Gambar 3. Data yang didapatkan pada proses tersebut sebanyak 2.000 baris. Data tersebut disimpan ke dalam bentuk file *excell*. Data pada file *excell* dibersihkan dari data duplikasi dan dilakukan penghapusan ‘RT’ (*Retweet*) pada kolom *text* secara manual sebelum dilakukan *pre-processing*. Penghapusan data duplikasi sebanyak 1.396 baris sehingga tersisa 604 baris. Sehingga dataset yang siap diproses untuk *pre-processing* sebanyak 604 baris.

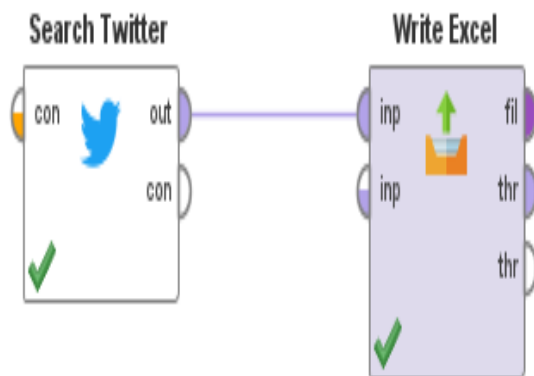


Gambar 2. Desain Model Komparasi SVM dan NB berbasis PSO

2.3. Pre-Processing

Data yang diperoleh dalam penelitian biasanya tidak langsung siap digunakan untuk proses pengujian.

Tahapan *pre-processing* merupakan tahapan mengubah data yang diperoleh dari sumber data menjadi dataset agar siap diproses lebih lanjut ke proses pengujian data.



Gambar 3. Proses *Crawling data dari Twitter*

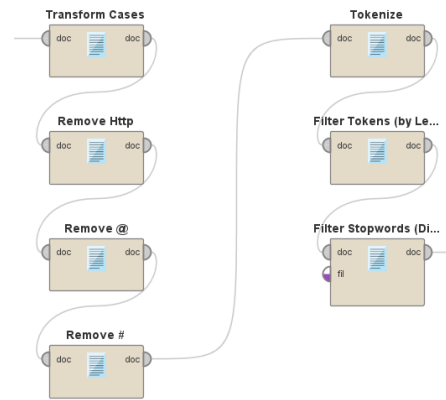
Pre-processing bertujuan untuk menghilangkan *noise* pada data, dan melakukan beberapa tahapan untuk mengubah data yang belum terstruktur menjadi terstruktur atau mengubah teks menjadi term indeks yang mewakili sebuah dokumen sehingga siap diproses lebih lanjut. Tahapan - tahapan *pre-processing* sebagaimana disajikan pada Gambar 4. Beberapa tahapan yang dilakukan dalam *pre-processing* sesuai dengan gambar tersebut yaitu:

Transform cases bertujuan untuk mengubah semua teks menjadi huruf kecil. Hal ini bertujuan agar tidak terjadi masalah ketika dilakukan proses *tokenize*. *Remove http* bertujuan untuk menghilangkan URL yang biasanya terdapat pada text *tweet*. Text sebelum dan sesudah diproses *Remove http* disajikan pada Tabel 1.

Remove Annotation bertujuan untuk menghapus tanda *annotation* yang seringkali terdapat pada text *tweet*. *Annotation* merupakan salah satu *noise* yang tidak memiliki arti. Contoh teks yang telah dihapus tanda *annotasinya* pada proses *Remove Annotation* disajikan pada Tabel 1. *Remove Hashtag* bertujuan untuk menghapus kata kunci *hashtag*. Contoh teks sebelum dan sesudah dilakukan proses penghapusan *hashtag* disajikan pada Tabel 1.

Proses *tokenize* merupakan proses pemotongan string input berdasarkan tiap kata penyusunnya. Tahapan ini bertujuan untuk memisahkan kata per kata dari teks kalimat. Tahapan ini bertujuan untuk melakukan filter atau hanya mengambil kata-kata yang panjangnya antara 4 sampai dengan 25 karakter berdasarkan dari hasil *token*.

Tahapan ini merupakan pemilihan kata-kata penting dari hasil *token*, yaitu kata-kata apa saja yang akan digunakan untuk mewakili dokumen. Sehingga pada proses ini akan menghapus kata-kata yang tidak deskriptif (tidak penting) yang dapat dibuang dengan pendekatan *bag-of-words*.



Gambar 4. Tahapan-tahapan pada *Pre-Processing*

2.4. Model Pengujian Data

Data yang telah dilakukan *pre-processing* sangat mungkin terjadi *imbalance* data. *Imbalance* data merupakan kondisi dimana di antara kedua kelas memiliki selisih yang sangat signifikan. Pemodelan dengan algoritma yang tidak memperhatikan ketidakseimbangan data akan didominasi oleh kelas mayor dan tidak memperhatikan kelas minornya. Hal ini akan menyebabkan hasil penelitian pada batas-batas keputusan yang diragukan dalam klasifikasi[20]. Untuk mengatasi hal ini, maka diperlukan *SMOTE* (*Synthetic Minority Oversampling Technique*), dengan melakukan *up-sampling* atau membangkitkan data buatan kelas minor agar setara dengan data kelas mayor.

Data kemudian akan dilakukan pemodelan dengan algoritma *SVM*, *NB* dan masing-masing menggunakan *PSO*. Tipe *kernel* yang digunakan dalam pemodelan ini adalah *Linear splines in one-dimension (dot)*. Tipe *kernel* ini sangat berguna ketika berhadapan dengan data vektor dan sering digunakan dalam kategorisasi teks. Untuk selanjutnya dilakukan pengujian data dengan *k-Fold Cross Validation*. *K-Fold cross validation* merupakan metode statistik yang digunakan untuk mengevaluasi kinerja model atau algoritma yang akan memisahkan data menjadi dua *subset* yaitu sebagai data proses pembelajaran (data *training*) dan data validasi/ evaluasi (data *testing*). Model akan dilatih oleh *subset training* dan divalidasi oleh *subset testing*.

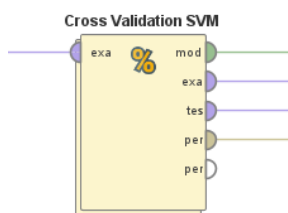
Hasil dari pengujian data akan dievaluasi menggunakan tabel *Confussion Matrix* yaitu *accuracy* dan *Area Under the Curve (AUC)*. *Confussion Matrix* merepresentasikan prediksi dan kondisi sebenarnya dari data yang dihasilkan oleh algoritma yang digunakan. *Accuracy* digunakan untuk evaluasi karena untuk mengetahui rasio prediksi benar (*true positif* dan *true negatif*) dengan data keseluruhan. Sedangkan, *AUC* digunakan karena akan menunjukkan angka yang terkait langsung dengan data. Nilai *AUC* menggambarkan hasil keseluruhan pengukuran atas kesesuaian dari model yang digunakan. Nilai *AUC* yang semakin besar menunjukkan bahwa variabel yang diteliti semakin baik dalam memprediksi kejadian [21].

Tabel 1. Regular (URL, Annotation, Hashtag) Expressions Removal

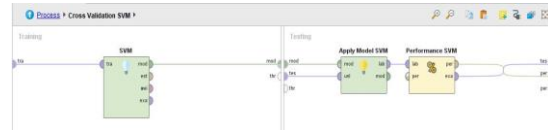
No.	Sebelum	Sesudah
1	@KemensetnegRI Buat saya gak masalah lockdown karena semua kebutuhan hidup saya sudah terpenuhi. Beras ambil di lumbung. Sayur dan buah petik di kebun. Ikan mancing di danau. Tapi gimana orang miskin yang ekonominya harus keluar rumah? @jokowi @setkabgoid @Kiyai_MarufAmin @KSPgoid @DPR_RI https://t.co/YzWQCmnaZP	Buat saya gak masalah lockdown karena semua kebutuhan hidup saya sudah terpenuhi Beras ambil di lumbung Sayur dan buah petik di kebun Ikan mancing di danau Tapi gimana orang miskin yang ekonominya harus keluar rumah
2	RT @AskaraOfficial: DPR: Keadaan Makin Parah, Pemerintah Tergopoh-gopoh Buat Aturan Karantina Wilayah #NEWS https://t.co/7mmN4uhjyX	DPR Keadaan Makin Parah Pemerintah Tergopoh-gopoh Buat Aturan Karantina Wilayah
3	Ketua DPR Puan Maharani mengatakan rapat paripurna harus dilakukan agar DPR bisa mulai bekerja melakukan fungsi pengawasan, budgeting dan legislasi terutama di saat darurat wabah Covid-19 sekarang ini. #TolakOmnibusLaw #AtasiVirusCabutOmnibusLaw #gagalkanomnibuslaw https://t.co/UftIXnUICD @marzukialie_MA @DPR_RI Bukannya dah lama pak, anggota DPR punya habit physical distancing... angka kehadiran minus dll... hebat anggota DPR.. punya indera ke 9... bahwa phys.dis/ soc.dis... akan di berlakukan di Ind... ????	Ketua DPR Puan Maharani mengatakan rapat paripurna harus dilakukan agar DPR bisa mulai bekerja melakukan fungsi pengawasan budgeting dan legislasi terutama di saat darurat wabah Covid sekarang ini Bukannya dah lama pak anggota DPR punya habit physical distancing angka kehadiran minus dll hebat anggota DPR punya indera ke bahwa phys dis soc dis akan di berlakukan di Ind

2.5. Pengujian Data

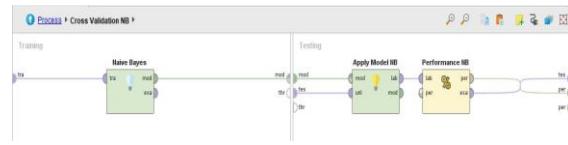
Proses pengujian data menggunakan algoritma SVM, SVM dengan PSO, NB tanpa PSO dan NB dengan PSO. *K-fold cross validation* merupakan metode statistik untuk mengevaluasi kinerja model dengan algoritma yang telah ditentukan [22]. Pada penelitian ini nilai *k-fold cross validation* sebagaimana disajikan pada Gambar 5, diatur pada angka 10. Sehingga dataset dibagi menjadi 10 area, dengan masing-masing aspek memberikan informasi yang sama persentase setiap jenis data. Sembilan area data digunakan dalam metode pelatihan sehingga membentuk model, sedangkan yang tinggal digunakan dalam prosedur pengujian.



Gambar 5. K-Folds Cross Validation



Gambar 6. Training dan Testing Support Vector Machine



Gambar 7. Training dan Testing Naive Bayes

Di dalam model *cross-validation* ini terdapat beberapa operator yaitu algoritma untuk training sehingga menghasilkan model dan *testing performance* model. Operator *training* untuk menghasilkan model dan *testing performance* sebagaimana disajikan pada Gambar 6 dan Gambar 7.

3. Hasil dan Pembahasan

Proses selanjutnya yaitu melakukan evaluasi atas hasil pengujian data. Evaluasi dilakukan dengan menggunakan nilai akurasi pada tabel *confusion matrix* dan nilai *AUC* pada kurva *ROC*.

3.1. Support Vector Machine Algorithm

Confusion matrix yang dihasilkan dari pengujian algoritma SVM tanpa PSO sebagaimana disajikan pada Tabel 2.

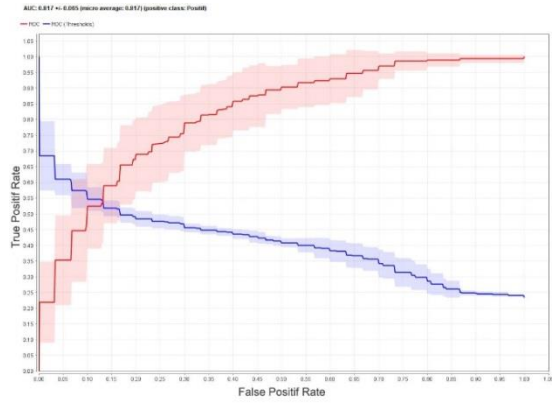
Tabel 2. Confusion Matrix Support Vector Machine Tanpa PSO

accuracy: 71,04% +/- 6,68% (micro average: 71,03%)			
	true Negatif	true Positif	class precision
pred. Negatif	249	122	67,12%
pred. Positif	53	180	77,25%
class recall	82,45%	59,60%	

Pada Tabel 2 tersebut menunjukkan bahwa nilai akurasi yang didapatkan sebesar 71,04% dengan toleransi kesalahan sebesar 6,68%, dengan nilai *true positif* 180 records dan *true negatif* 249 records. Hal ini menunjukkan bahwa data yang diklasifikasikan sesuai dengan sentimen positif sebanyak 180 records dan data yang diklasifikasikan sesuai dengan sentimen negatif sebanyak 249 records. Sedangkan data yang seharusnya sentimen positif namun diklasifikasikan negatif (*false negatif*) sebanyak 53 records dan data yang seharusnya diklasifikasikan sentimen negatif namun diklasifikasikan positif (*false positif*) sebanyak 122 records. Berdasarkan hasil pengujian *performance* menghasilkan Kurva *Receiver Operating Characteristic (ROC)* sebagaimana pada Gambar 8 dan nilai *Area Under Curve (AUC)* yang didapatkan adalah sebesar 0,817 (*good classification*).

3.2. Naive Bayes Algorithm

Confusion matrix yang dihasilkan dari pengujian algoritma NB tanpa PSO sebagaimana disajikan pada Tabel 3.



Gambar 8. Grafik ROC Support Vector Machine

Tabel 3. Confusion Matrix Naive Bayes Tanpa PSO

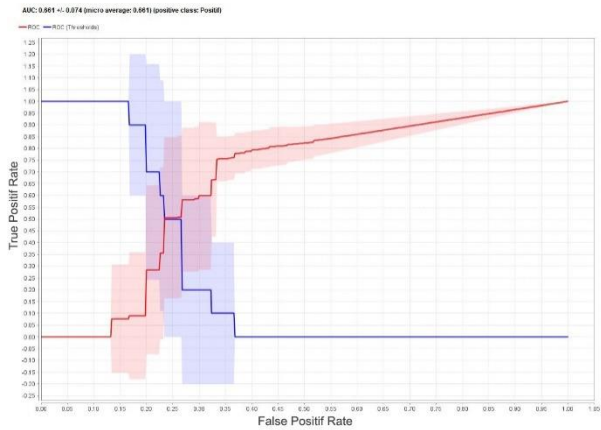
accuracy: 70,69% +/- 5,56% (micro average: 70,70%)			
	true Negatif	true Positif	class precision
pred. Negatif	231	106	68,55%
pred. Positif	71	196	73,41%
class recall	76,49%	64,90%	

Pada Tabel 3 tersebut menunjukkan bahwa nilai akurasi yang didapatkan sebesar 70,69% dengan toleransi kesalahan sebesar 5,56%, dengan nilai true positif 196 records dan true negatif 231 records. Hal ini menunjukkan bahwa data yang diklasifikasikan sesuai dengan sentimen positif sebanyak 196 records dan data yang diklasifikasikan sesuai dengan sentimen negatif sebanyak 231 records. Sedangkan data yang seharusnya sentimen positif namun diklasifikasikan negatif (*false negatif*) sebanyak 71 records dan data yang seharusnya diklasifikasikan sentimen negatif namun diklasifikasikan positif (*false positif*) sebanyak 106 records. Berdasarkan hasil pengujian performance menghasilkan Kurva Receiver Operating Characteristic (ROC) sebagaimana pada Gambar 9 dan nilai Area Under Curve (AUC) yang didapatkan adalah sebesar 0,661 (*poor classification*).

3.3. Support Vector Machine Algorithm (PSO)

Confusion matrix yang dihasilkan dari pengujian menggunakan algoritma SVM dengan dioptimasi PSO sebagaimana disajikan pada Tabel 4.

Pada Tabel 4 tersebut menunjukkan bahwa nilai akurasi yang didapatkan sebesar 75,03% dengan toleransi kesalahan sebesar 6,80%, dengan nilai true positif 216 records dan true negatif 237 records. Hal ini menunjukkan bahwa data yang diklasifikasikan sesuai dengan sentimen positif sebanyak 216 records dan data yang diklasifikasikan sesuai dengan sentimen negatif sebanyak 237 records. Sedangkan data yang seharusnya sentimen positif namun diklasifikasikan negatif (*false negatif*) sebanyak 65 records dan data yang seharusnya diklasifikasikan sentimen negatif namun diklasifikasikan positif (*false positif*) sebanyak 86 records. Berdasarkan hasil pengujian performance menghasilkan Kurva Receiver Operating Characteristic

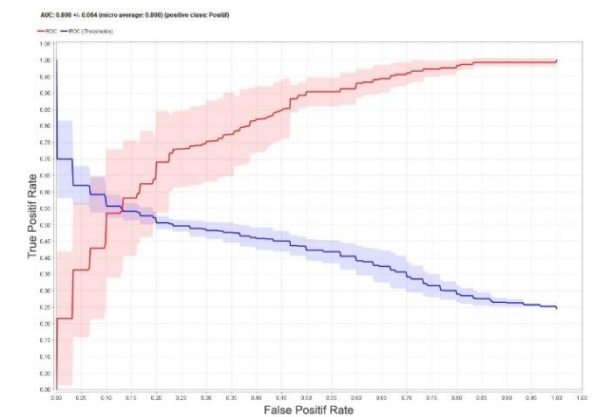


Gambar 9. Grafik ROC Naive Bayes

(ROC) sebagaimana pada Gambar 10 dan nilai Area Under Curve (AUC) yang didapatkan adalah sebesar 0,817 (*good classification*).

Tabel 4. Confusion Matrix Support Vector Machine (PSO)

accuracy: 75,03% +/- 6,80% (micro average: 75,00%)			
	true Negatif	true Positif	class precision
pred. Negatif	237	86	73,37%
pred. Positif	65	216	76,87%
class recall	78,48%	71,52%	



Gambar 10. Grafik ROC Support Vector Machine (PSO)

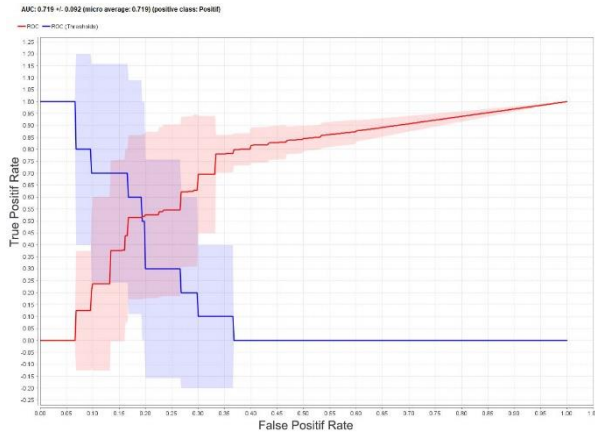
3.4. Naive Bayes Algorithm (PSO)

Confusion matrix yang dihasilkan dari pengujian algoritma NB dengan menggunakan PSO sebagaimana disajikan pada Tabel 5.

Tabel 5. Confusion Matrix Naive Bayes (PSO)

accuracy: 73,49% +/- 6,52% (micro average: 73,51%)			
	true Negatif	true Positif	class precision
pred. Negatif	249	107	69,94%
pred. Positif	53	195	78,63%
class recall	82,45%	64,57%	

Pada Tabel 5 tersebut menunjukkan bahwa nilai akurasi yang didapatkan sebesar 73,49% dengan toleransi kesalahan sebesar 6,52%, dengan nilai true positif 195 records dan true negatif 249 records. Hal ini menunjukkan bahwa data yang diklasifikasikan sesuai



Gambar 11. Grafik ROC Naive Bayes (PSO)

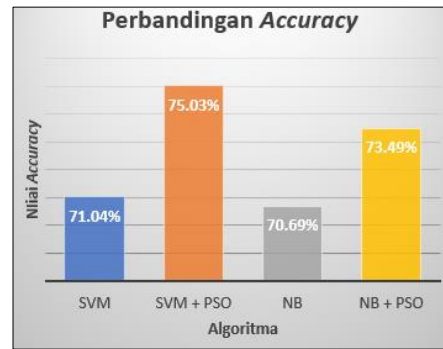
dengan sentimen positif sebanyak 195 records dan data yang diklasifikasikan sesuai dengan sentimen negatif sebanyak 249 records. Sedangkan data yang seharusnya sentimen positif namun diklasifikasikan menjadi negatif (*false negatif*) sebanyak 107 records dan data yang seharusnya diklasifikasikan sentimen negatif namun diklasifikasikan ke dalam positif (*false positif*) sebanyak 249 records. Berdasarkan hasil pengujian *performance* menghasilkan Kurva Receiver Operating Characteristic (ROC) sebagaimana pada Gambar 11 dan nilai Area Under Curve (AUC) yang didapatkan adalah sebesar 0,719 (*fair classification*).

Tabel 6. Komparasi Accuracy and AUC

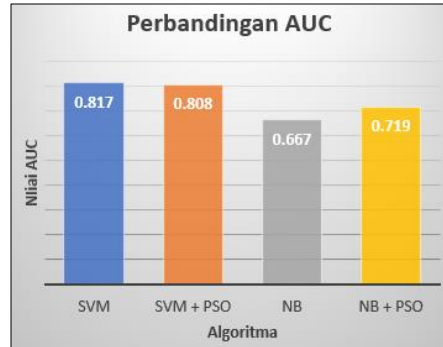
Algoritma	Accuracy	AUC
SVM	71,04%	0,817
NB	70,69%	0,661
SVM + PSO	75,03%	0,808
NB + PSO	73,49%	0,719

Rekapitulasi hasil *accuracy* dan *AUC* yang didapatkan diantara keempat algoritma tersebut diatas disajikan pada Tabel 6. Pada Tabel 6 menunjukkan bahwa berdasarkan pengujian yang dilakukan dengan menggunakan algoritma SVM mendapatkan nilai *accuracy* 71,04% dan nilai *AUC* sebesar 0,817, sedangkan pengujian yang dilakukan dengan menggunakan algoritma Naive Bayes mendapatkan nilai *accuracy* 70,69% dan nilai *AUC* sebesar 0,661. Hasil *accuracy* dari kedua algoritma tersebut terdapat selisih sebesar 0,35%. Sedangkan pengujian yang dilakukan dengan menggunakan algoritma SVM yang dioptimalisasi dengan operator PSO mendapatkan nilai *accuracy* sebesar 75,03% dan nilai *AUC* sebesar 0,808, sedangkan pengujian yang dilakukan dengan menggunakan algoritma Naive Bayes yang telah dioptimalisasi dengan PSO mendapatkan nilai *accuracy* sebesar 73,49% dan nilai *AUC* sebesar 0,719. Nilai *AUC* setelah dioptimalisasi terdapat selisih sebesar 1,54%.

Grafik perbandingan nilai akurasi dan nilai *AUC* diantara keempat algoritma tersebut sebagaimana disajikan pada Gambar 12 dan Gambar 13.



Gambar 12. Grafik Perbandingan Accuracy



Gambar 13. Grafik Perbandingan AUC

Berdasarkan analisis rekapitulasi tersebut menunjukkan bahwa evaluasi menggunakan algoritma SVM dengan PSO menjadi solusi paling baik untuk mendapatkan nilai *accuracy* tertinggi, sedangkan untuk hasil *AUC* tertinggi menggunakan algoritma SVM tanpa PSO.

4. Kesimpulan

Pengujian terhadap data hasil *crawling* dari media sosial twitter dengan *query* DPR telah dilakukan. Berdasarkan hasil analisis sentimen terhadap Dewan Perwakilan Rakyat Republik Indonesia di Media Sosial Twitter dengan menggunakan algoritma Support Vector Machine, Naive Bayes, Support Vector Machine dengan Particle Swarm Optimization, dan Naive Bayes dengan Particle Swarm Optimization, dapat disimpulkan bahwa penggunaan algoritma Support Vector Machine yang dioptimasi dengan Particle Swarm Optimization menghasilkan tingkat akurasi lebih baik dibandingkan dengan algoritma Support Vector Machine, Naive Bayes dan Naive Bayes yang dioptimasi dengan Particle Swarm Optimization. Sehingga algoritma Support Vector Machine yang dioptimasi dengan Particle Swarm Optimization menjadi solusi untuk melakukan klasifikasi sentimen masyarakat terhadap Dewan Perwakilan Rakyat Republik Indonesia. Penelitian selanjutnya agar dilakukan untuk kasus yang sama dengan menggunakan operator *feature selection by corellation* atau dengan *feature selection by gain information* sehingga *performancenya* dapat dibandingkan dan mendapatkan nilai akurasi yang paling optimal.

Daftar Pustaka

- [1] S. D. Cahyani, “*Media Monitoring (Analisis Media) Di Bagian Media Cetak Dan Media Sosial Dewan Perwakilan Rakyat Republik Indonesia (DPR RI)*,” Jakarta, 2019.
- [2] A. Tri, F. M. Laboratorium, O. Reno, and P. Resmi, “*Riset: Ada 175 , 2 Juta Pengguna Internet di Indonesia*,” Updated 20 Feb 2020, 2020.[Online].Available:https://inet.detik.com/cyberlife/d4907674/riset-ada-1752-juta-pengguna-internet-di-indonesia. [Accessed: 20-Apr-2020].
- [3] M. Kamyab, R. Tao, M. H. Mohammadi, and A. Rasool, “*Sentiment analysis on Twitter: A text mining approach to the Afghanistan status reviews*,” *ACM Int. Conf. Proceeding Ser.*, pp. 14–19, 2018.
- [4] A. Shelar and C. Y. Huang, “*Sentiment analysis of twitter data*,” *Proc. - 2018 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2018*, pp. 1301–1302, 2018.
- [5] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, “*Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier*,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019.
- [6] Warjiyono, S. Aji, Fandhilah, N. Hidayatun, H. Faqih, and Liesnaningsih, “*The Sentiment Analysis of Fintech Users Using Support Vector Machine and Particle Swarm Optimization Method*,” *2019 7th Int. Conf. Cyber IT Serv. Manag. CITSM 2019*, 2019.
- [7] I. Santoso, W. Gata, and A. B. Paryanti, “*JURNAL RESTI Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum*,” vol. 1, no. 10, pp. 5–11, 2019.
- [8] A. Dhika, P. Wardhani, and I. Slamet, “*Analisis Sentimen Penggunaan Kotak Suara Kardus*,” pp. 167–174, 2019.
- [9] M. I. Komputer and K. J. Pusat, “*Sentimen Analisis Operasi Tangkap Tangan KPK Menurut Masyarakat Menggunakan Algoritma Support Vector Machine , Naive Bayes Berbasis Particle Swarm Optimizition*,” vol. 12, no. 3, pp. 230–243, 2019.
- [10] Ridwansyah and E. Purwaningsih, “*Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Pemasaran Bank*,” *J. PILAR Nusa Mandiri*, vol. 14, no. 1, pp. 83–88, 2018.
- [11] B. Pratama et al., “*Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM and NB Methods*,” *J. Phys. Conf. Ser.*, vol. 1201, no. 1, 2019.
- [12] N. K. Wardhani et al., “*Sentiment analysis article news coordinator minister of maritime affairs using algorithm naive bayes and support vector machine with particle swarm optimization*,” *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 24, pp. 8365–8378, 2018.
- [13] M. Fernández-Gavilanes, T. Álvarez-López, J. Juncal-Martínez, E. Costa-Montenegro, and F. Javier González-Castaño, “*Unsupervised method for sentiment analysis in online texts*,” *Expert Syst. Appl.*, 2016.
- [14] A. Giachanou and F. Crestani, “*Like it or not: A survey of Twitter sentiment analysis methods*,” *ACM Comput. Surv.*, vol. 49, no. 2, 2016.
- [15] D. A. Kristiyanti, A. H. Umam, M. Wahyudi, R. Amin, and L. Marlinda, “*Comparison of SVM Naive Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter*,” *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1–6, 2019.
- [16] I. A. El-Khair, “*Effects of Stop Words Elimination for Arabic Information Retrieval: A Comparative Study*,” pp. 1–15, 2017.
- [17] J. Kaur and P. Kaur Buttar, “*A Systematic Review on Stopword Removal Algorithms*,” *Int. J. Futur. Revolut. Comput. Sci. Commun. Eng.*, no. April, pp. 207–210, 2018.
- [18] H. Saif, Y. He, M. Fernandez, and H. Alani, “*Contextual semantics for sentiment analysis of Twitter*,” *Inf. Process. Manag.*, vol. 52, no. 1, pp. 5–19, 2016.
- [19] G. A. Buntoro, “*Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter*,” *Integer J. Maret*, vol. 1, no. 1, pp. 32–41, 2017.
- [20] O. Heranova, “*Synthetic Minority Oversampling Technique pada Averaged One Dependence Estimators untuk Klasifikasi Credit Scoring*,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 443–450, 2019.
- [21] T. T. Maskoen and D. Purnama, “*Area Under the Curve dan Akurasi Cystatin C untuk Diagnosis Acute Kidney Injury pada Pasien Politrauma*,” *Maj. Kedokt. Bandung*, vol. 50, no. 4, pp. 259–264, 2018.
- [22] A. Rohani, M. Taki, and M. Abdollahpour, “*A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I)*,” *Renew. Energy*, 2018.

Halaman ini sengaja dikosongkan