

Sentimen Analisis Vaksinasi Covid-19 pada Twitter Menggunakan Algoritma Klasifikasi Berbasis PSO

Sentiment Analysis of Covid-19 Vaccination on Twitter Using Classification Algorithms based on PSO

¹Fuji Astuti*, ²Resi Taufan

Program Studi Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Nusa Mandiri

^{1,2}Jl. Kramat Raya No. 18, Senen 10450, DKI Jakarta, Indonesia

*e-mail: 14002194@nusamandiri.ac.id

(*received*: 25 November 2021, *revised*: 2 Februari 2022, *accepted*: 19 Februari 2022)

Abstrak

Media sosial *Twitter* banyak digunakan pengguna internet untuk memberikan opini terhadap suatu kejadian. Pada tahun 2021 pendapat tentang vaksinasi Covid-19 banyak diperbincangkan masyarakat Indonesia pada media sosial tersebut. Keberagaman pendapat dapat dikategorikan sentimen positif dan sentimen negatif. Untuk mengkategorikan sentimen negatif maupun positif dapat digunakan proses *data mining*. Penelitian ini membahas sentimen vaksinasi Covid-19 menggunakan algoritma klasifikasi. Data dari *Twitter* dikumpulkan dalam proses *crawling*, lalu dilakukan pelabelan menjadi dua kelas yaitu sentimen positif dan negatif. Setelah dilakukan pelabelan (*polarity*) tahap berikutnya adalah *preprocessing data* yang terdiri dari *transform case*, *removal annotation*, *tokenizing* dan *stemming*. Algoritma klasifikasi yang digunakan yaitu *Naïve Bayes* (NB) dan *Support Vector Machine* (SVM) kemudian dibandingkan dengan algoritma klasifikasi yang menggunakan *Particle Swarm Optimization* (PSO). Pengujian dilakukan menggunakan *k-Fold Cross Validation* untuk mendapat nilai akurasi, tabel *Confusion Matrix* dan *Area Under Curve*. Hasil pengujian pada algoritma klasifikasi menggunakan PSO lebih baik dibandingkan tanpa menggunakan PSO. Hasil akurasi algoritma NB dan SVM adalah 64,04% dan 72,55%, sedangkan hasil akurasi setelah menggunakan PSO pada algoritma SVM dan NB adalah 70,43% dan 76,38%.

Kata kunci: Sentimen Analisis, *Support Vector Machine*, *Naïve Bayes*, *Particle Swarm Optimization*, Vaksinasi Covid-19

Abstract

Twitter social media is widely used by internet users to provide opinions on an event. In 2021, the opinion about the Covid-19 vaccination was widely discussed by the Indonesian people on social media. Togetherness of opinion can be categorized as positive sentiment and negative sentiment. To categorize negative and positive sentiments, data mining processes can be used. This study discusses the sentiment of Covid-19 vaccination using classification. Data from Twitter is collected in a crawling process, then labeled into two classes, namely positive and negative sentiments. After labeling (polarity) the next stage is data preprocessing which consists of transform case, annotation removal, tokenizing and stemming. The classification algorithm used is Naïve Bayes (NB) and Support Vector Machine (SVM) and then compared with the classification algorithm using Particle Swarm Optimization (PSO). Tests were carried out using k-Fold Cross Validation to obtain accuracy values, Confusion Matrix tables and Area Under Curve. The test results on the classification using PSO are better than those without using PSO. The results of the accuracy of the NB and SVM algorithms are 64.04% and 72.55%, while the accuracy results after PSO on the SVM and NB algorithms are 70.43% and 76.38%, respectively.

Keywords: *Sentiment Analysis, Support Vector Machine, Naïve Bayes, Particle Swarm Optimization, Covid-19 Vaccination*

1 Pendahuluan

Kemudahan mengakses internet sebagai tanda kemajuan teknologi dan komunikasi mengakibatkan banyak *platform* media sosial untuk menampung komentar dan bertukar informasi.

Salah satu media sosial yang banyak digunakan oleh masyarakat Indonesia yaitu *Twitter* [1]. Dengan *Twitter* memudahkan penggunanya dalam berbagi opini, acara maupun saling berinteraksi [2]. Pada tahun 2021 seiring dengan mewabahnya virus Covid-19 mengharuskan pemerintah Indonesia untuk memberlakukan wajib vaksin sebagai upaya pencegahan penularan virus Covid-19 agar tidak semakin meluas. Pengembangan vaksin Covid-19 banyak dilakukan sebagai salah satu cara dalam pengendalian pandemi Covid-19, bahkan berbagai vaksin telah menjalani uji klinis dengan hasil yang menjanjikan di berbagai negara [3]. Vaksinasi Covid-19 menjadi hal mengandung polemik, hal ini ditandai dengan banyaknya negara melakukan aksi penolakan karena dianggap tidak efektif. Salah satu pemicu penolakan dari masyarakat yaitu karena kurangnya informasi dan sikap siaga pemerintah untuk mengedukasikan terkait Covid-19 [4].

Beragam tanggapan tentang vaksinasi Covid-19 banyak bermunculan di media *Twitter* bahkan menjadi topik utama yang banyak dibagikan. Meskipun *Twitter* belum bisa dinyatakan sebagai percakapan yang komprehensif, namun *Twitter* dapat memberikan wawasan tentang topik populer termasuk momen yang penting seperti pemberian vaksinasi Covid-19 [5].

Tanggapan pada *Twitter* dapat dikatakan sebagai suatu sentimen, tanggapan yang mendukung suatu kejadian disebut sentimen positif, sebaliknya tanggapan yang menentang atau menolak suatu kejadian disebut sentimen negatif. Pada penelitian ini munculnya reaksi masyarakat Indonesia khususnya pada media sosial *Twitter* menarik untuk dilakukan analisis. Analisis sentimen pada data *Twitter* dapat diimplementasikan dengan algoritma klasifikasi. Algoritma klasifikasi yang banyak digunakan dalam penelitian analisis sentimen yaitu *Naïve Bayes* (NB) dan *Support Vector Machine* (SVM). Algoritma klasifikasi tersebut dilakukan optimasi menggunakan *Particle Swarm Optimization* (PSO) sehingga meningkatkan akurasi. Metode optimasi PSO yang dipadukan pada algoritma NB dan SVM terbukti meningkatkan akurasi dengan baik seperti pada penelitian sentimen pemindahan ibukota negara [6][7].

Berdasarkan tanggapan pengguna *Twitter* tentang vaksinasi Covid-19 maka menarik dilakukan penelitian tentang sentimen analisis menggunakan algoritma klasifikasi. Tujuan dari penelitian ini agar sentimen mengenai vaksinasi Covid-19 dapat dikategorikan menjadi sentimen positif dan sentimen negatif menggunakan algoritma klasifikasi NB dan SVM. Selain itu dalam penelitian ini akan digunakan optimasi dengan *feature* optimasi PSO untuk membuktikan bahwa optimasi tersebut dapat meningkatkan akurasi.

2 Tinjauan Literatur

Sentimen analisis tentang *physical distancing* pada *Twitter* menggunakan algoritma NB dilakukan dalam penelitian [8]. Penelitian tersebut menghasilkan akurasi sebesar 50,26% menggunakan 547 data, dalam penelitian tersebut proses analisis menggunakan metode *text mining* yang diolah kedalam beberapa tahap, yaitu pengambilan data, pengolahan data, penerapan metode, validasi dan evaluasi. Selanjutnya penelitian sentimen opini publik mengenai Covid-19 dilakukan menggunakan algoritma NB dan KNN. Hasil akurasi algoritma NB sebesar 63,21% lebih baik dibandingkan dengan algoritma KNN sebesar 58,10% [9]. Penelitian lain mengkategorikan sentimen menjadi kelas positif, negatif dan netral. Hasil dari penelitian tersebut mendapat akurasi sebesar 74% [10].

Pada penelitian [11] algoritma NB yang dilakukan optimasi menggunakan PSO mendapat akurasi yang tinggi yaitu 93,50% dibandingkan tidak menggunakan PSO sebesar 51,50%. Dalam beberapa penelitian sentimen analisis yang dilakukan [12][13] membuktikan bahwa optimasi PSO dapat menghasilkan akurasi yang baik dibandingkan algoritma klasifikasi tanpa PSO.

Selain algoritma NB, algoritma *Support Vector Machine* (SVM) banyak digunakan dalam penelitian sentimen analisis. Seperti dalam penelitian [14] tentang sentimen pada kasus benih lobster 2020 yang mendapat akurasi 84,21% dari 156 data yang diolah. Penelitian lain [15] menggunakan algoritma SVM menghasilkan 79,6% dari 500 data yang diolah, penelitian yang dilakukan mengenai sentimen pelayanan TELKOM dan BIZNET.

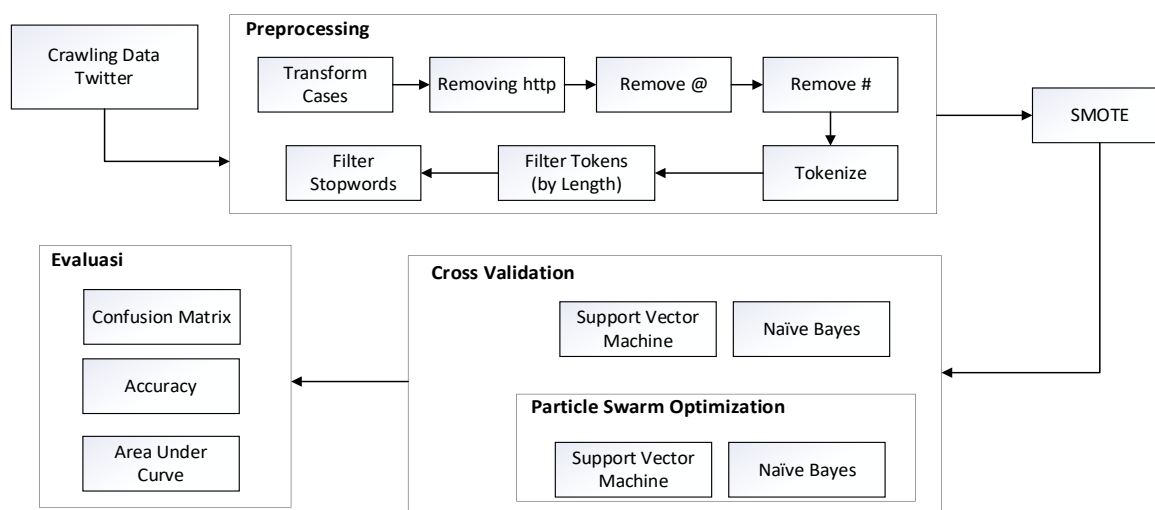
Dengan menggunakan optimasi PSO pada algoritma SVM berhasil meningkatkan akurasi dalam penelitian [16] akurasi sebelum menggunakan PSO sebesar 75,33% meningkat sebesar 7% setelah menggunakan PSO. Selanjutnya sentimen pemindahan ibukota negara RI menggunakan algoritma

SVM berbasis PSO mendapat akurasi 78,77% [6]. Dengan sentimen pada topik yang sama penelitian [7] menggunakan algoritma SVM berbasis PSO menghasilkan akurasi 81,15%.

Dari beberapa penelitian tentang sentimen analisis tersebut, pada penelitian ini melakukan analisis sentimen pada topik vaksinasi Covid-19 menggunakan algoritma NB, SVM, NB berbasis PSO dan SVM berbasis PSO.

3 Metode Penelitian

Metode penelitian menggunakan konsep *data mining* dengan algoritma klasifikasi dan digabungkan dengan optimasi PSO yang ditunjukkan pada Gambar 1. Metode PSO merupakan teknik optimasi mengikuti perilaku sekumpulan burung dan sekumpulan ikan untuk bertahan hidup [17]. Metode PSO yang diterapkan pada klasifikasi teks secara substansial dapat mengurangi dimensi fitur ruang sehingga meningkatkan akurasi [18]. Pada penelitian ini metode diawali dengan *crawling data* atau proses pengumpulan data, kemudian data akan diberikan label positif atau negatif yang disebut *polarity*, selanjutnya adalah *preprocessing* yang dibagi kedalam beberapa tahap yaitu *transform case*, *removal annotation*, *tokenizing*, *filtering* dan *stemming*. Kemudian dilakukan proses optimasi PSO untuk mengoptimalkan akurasi pada algoritma klasifikasi NB dan SVM.



Gambar 1. Metode Penelitian Sentimen Analisis

Tahapan metode penelitian pada gambar 1 dijelaskan kedalam beberapa proses sebagai berikut:

3.1. Crawling Data

Tahap pertama adalah proses pengumpulan data pada *Twitter* atau disebut *crawling data*. *Crawling data* pada *Twitter* yaitu suatu proses mengunduh data berupa user atau *tweet* dari server *Twitter* dengan bantuan *Application Programmin Integration (API)* [19]. Pada proses ini kata kunci yang digunakan yaitu #Vaksin Covid-19. Data yang diolah dari proses *crawling data* kemudian diberikan label sentimen positif atau sentimen negatif pada setiap *tweet* secara manual, proses ini dinamakan *polarity*.

3.2. Preprocessing

Tujuan dari tahap *preprocessing* untuk menghilangkan *noise* serta mengubah data yang tidak terstruktur menjadi data terstruktur atau mengubah teks menjadi term indeks yang diproses menjadi beberapa tahap, yaitu *transform case*, *removal annotation*, *tokenizing* dan *filtering* [12]. Pada penelitian ini ditambahkan proses *stemming* yang bertujuan menghilangkan imbuhan pada suatu kata. Tahapan *preprocessing* dijelaskan sebagai berikut:

- a) *Transform Cases*
Transform case menyeragamkan semua teks menjadi huruf kecil yang bertujuan untuk menghindari masalah saat proses *tokenizing*.
- b) *Removal Annotation*
Removal annotation untuk menghapus *annotation* yang banyak dijumpai pada *tweet*, penghapusan dilakukan karena *annotation* termasuk ke dalam *noise* dan tidak memiliki makna.
- c) *Tokenizing*
Proses *tokenize* bertujuan untuk memecah kalimat menjadi satuan kata yang sering disebut dengan token.
- d) *Filtering*
Proses *filtering* bertujuan untuk mengambil kata yang memiliki makna atau mewakili dokumen dan menghilangkan kata yang tidak memiliki makna dari hasil token dengan pendekatan *bag-of-words*.
- e) *Stemming*
Stemming bertujuan menghilangkan imbuhan dari hasil *filtering* sehingga terbentuk menjadi kata dasar. *Stemming* adalah proses mencari kata dasar dari sebuah kata derivative. Inti dari proses stemming yaitu agar suatu kata berubah menjadi kata dasar dengan menghilangkan imbuhan [20].

3.3. Model Pengujian Data

Tahap selanjutnya yaitu proses pemodelan menggunakan algoritma NB dan SVM serta melakukan optimasi dengan PSO pada setiap algoritma. Kemudian data akan diuji dengan *k-Fold Cross Validation* dilanjutkan dengan mengevaluasi data menggunakan tabel *confusion matrix* dan kurva *Area Under Curve* (AUC).

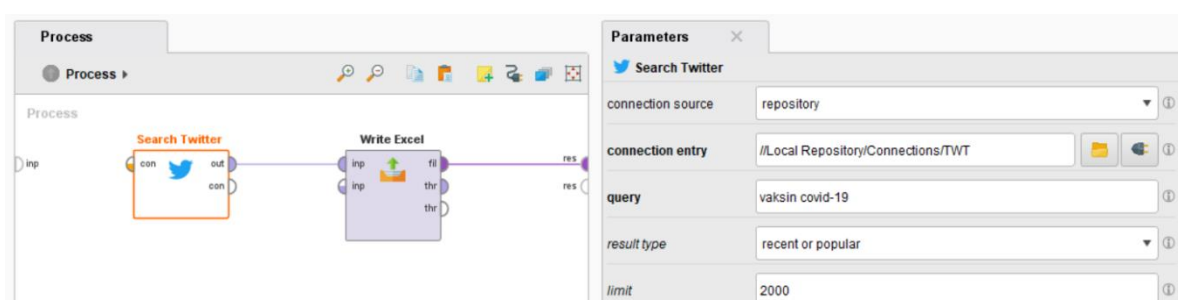
- a) SMOTE
Sebelum tahap pengujian data dilakukan, data hasil *preprocessing* dilakukan pengimbangan (*balancing*) dengan teknik SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE bertujuan untuk menghasilkan data yang baik dengan menangani ketidakseimbangan kelas akibat *overfitting* [21].
- b) SVM
Algoritma SVM merupakan metode yang banyak diterapkan pada klasifikasi *text mining* karena banyak menghasilkan performa yang lebih baik [22]. SVM merupakan algoritma klasifikasi linier berdasarkan prinsip memaksimalkan margin. SVM menggunakan *hyperplane* secara optimal untuk mengklasifikasikan data menjadi dua kelompok data dalam dimensi yang lebih tinggi [23].
- c) NB
NB yaitu salah satu algoritma yang menggunakan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, cara kerja NB yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [24].
- d) PSO
Particle swarm optimization atau disingkat menjadi PSO merupakan suatu algoritma yang sumber idenya bermula dari meniru perilaku sebuah kawanan serangga, seperti semut, rayap, lebah atau burung. Perilaku yang dimaksud adalah perilaku sosial organisme terdiri dari tindakan individu dan pengaruh dari individu-individu lain dalam suatu kelompok. Cara kerja PSO dengan menerapkan sifat-sifat individu dalam suatu kelompok besar, kemudian menyatukan sifat-sifat tersebut untuk menyelesaikan suatu permasalahan. Algoritma ini pertama kali dikenalkan oleh James Kennedy dan Russell Eberhart pada tahun 1995 [25]. PSO banyak diterapkan untuk berbagai masalah besar sebagai alternatif dalam mencari solusi yang optimal dengan menghitung bobot optimal untuk setiap *record* dan rata – rata seluruh sample pada klasifikasi teks [26].
- e) Evaluasi
Evaluasi dari hasil dari pengujian data menggunakan tabel *Confussion Matrix* yaitu *accuracy* dan AUC. *Confussion Matrix* merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma yang digunakan [27]. Akurasi digunakan untuk evaluasi karena untuk mengetahui rasio prediksi benar dengan data keseluruhan. Sedangkan, AUC digunakan

karena akan menunjukkan angka yang terkait langsung dengan data. Nilai AUC memberikan gambaran dari hasil keseluruhan pengukuran atas kesesuaian dari model yang digunakan [28].

4 Hasil dan Pembahasan

4.1. Pengumpulan Data

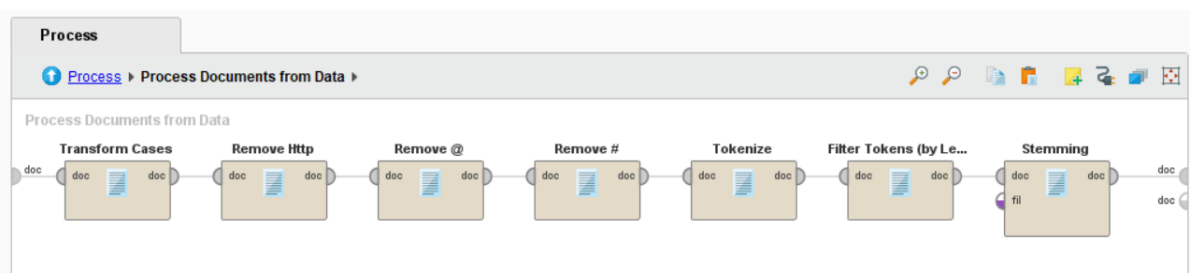
Tahap pertama adalah proses pengumpulan data pada *Twitter* atau disebut *crawling data*. Pada proses ini kata kunci yang digunakan yaitu #Vaksin Covid-19 yang terkumpul sebanyak 1.978 *tweet*. Data yang didapat dari proses ini masih terdapat duplikasi data sehingga perlu dilakukan penghapusan data agar data yang diolah menjadi unik tidak terjadi redudansi. Hasil dari penghapusan data duplikat sebanyak 470 *tweet*. Data yang diolah dari proses *crawling data* kemudian diberikan label sentimen positif atau sentimen negatif pada setiap *tweet* secara manual. Dari proses pelabelan jumlah sentimen positif sebanyak 192 dan sentimen negatif sebanyak 278. Proses *crawling data* pada *Twitter* menggunakan *software Rapid Miner* ditunjukkan pada Gambar 2 di bawah ini:



Gambar 2. Crawling Data Pada Twitter

4.2. Preprocessing Data

Preprocessing data meliputi tahap *transform case*, *removal annotation*, *tokenizing*, *filtering* dan *stemming* yang diterapkan pada *software Rapid Miner* seperti pada Gambar 3 berikut:



Gambar 3. Tahapan Pada Preprocessing

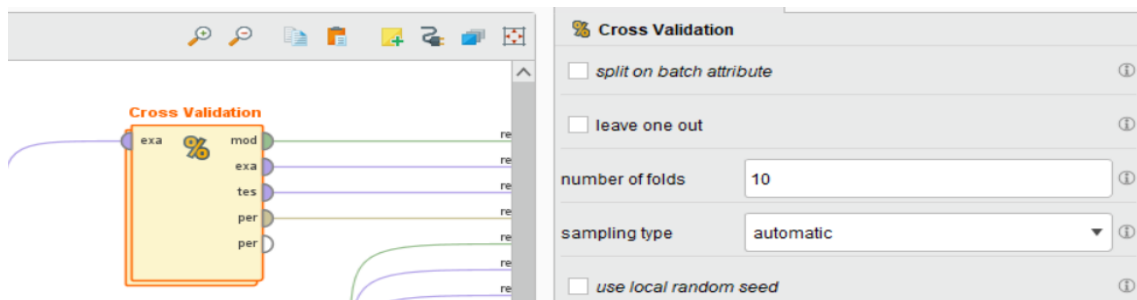
Penjelasan dari tahap *transform case* adalah menyeragamkan bentuk huruf menjadi huruf kecil agar menghindari *error* saat proses *tokenize*. Kemudian pada tahap selanjutnya yang termasuk pada tahap *removal annotation* pada Gambar 3 adalah removal http/https, removal @, removal #. Tahap *preprocessing* dilanjutkan dengan proses *tokenize* pada tahap ini memisahkan kata per kata dari teks kalimat kemudian proses *tokenize filter token by length* yang bertujuan mengambil data yang panjangnya 4 sampai 25 karakter dari hasil token. Pada tahap *preprocessing* yang terakhir dilakukan eliminasi pada kata yang tidak memiliki makna dan proses menghilangkan imbuhan pada kata dasar atau disebut *stemming*. Pada Tabel 1 menunjukkan contoh dari *tweet* sebelum dan sesudah *preprocessing*.

Tabel 1. Contoh Tweet Sebelum Dan Sesudah Preprocessing

Tahap Preprocessing	Sebelum	Sesudah
<i>Transform Case</i>	Hanya sedikit kasus yang menunjukkan kontraindikasi terhadap vaksin jenis tertentu pada anak. Namun, detail penggunaan vaksin akan disusun bagi anak yang memiliki kondisi khusus. #kumparanNEWS https://t.co/XfT9oDm4Pp	hanya sedikit kasus yang menunjukkan kontraindikasi terhadap vaksin jenis tertentu pada anak. namun, detail penggunaan vaksin akan disusun bagi anak yang memiliki kondisi khusus. #kumparannews https://t.co/xft9odm4pp
<i>Transform Case - Removal Annoatation</i>	Hanya sedikit kasus yang menunjukkan kontraindikasi terhadap vaksin jenis tertentu pada anak. Namun, detail penggunaan vaksin akan disusun bagi anak yang memiliki kondisi khusus. #kumparanNEWS https://t.co/XfT9oDm4Pp	hanya sedikit kasus yang menunjukkan kontraindikasi terhadap vaksin jenis tertentu pada anak namun detail penggunaan vaksin akan disusun bagi anak yang memiliki kondisi khusus
<i>Transform Case - Removal Annoatation - Filter Token By Length</i>	Ada 12-13 persen kasus Covid-19 pada anak, dokter Aman Pulungan menyampaikan setidaknya 3 alasan kenapa anak harus divaksin Covid-19. https://t.co/U3lr16p57N	2-13 persen kasus covid-19 pada anak dokter aman pulungan menyampaikan setidaknya alasan kenapa anak harus divaksin covid-19
<i>Transform Case - Removal Annoatation - Filter Token By Length - Stemming</i>	Hanya sedikit kasus yang menunjukkan kontraindikasi terhadap vaksin jenis tertentu pada anak. Namun, detail penggunaan vaksin akan disusun bagi anak yang memiliki kondisi khusus. #kumparanNEWS https://t.co/XfT9oDm4Pp	sedikit kasus tunjuk kontraindikasi vaksin jenis tertentu anak namun detail guna vaksin susun anak milik kondisi khusus

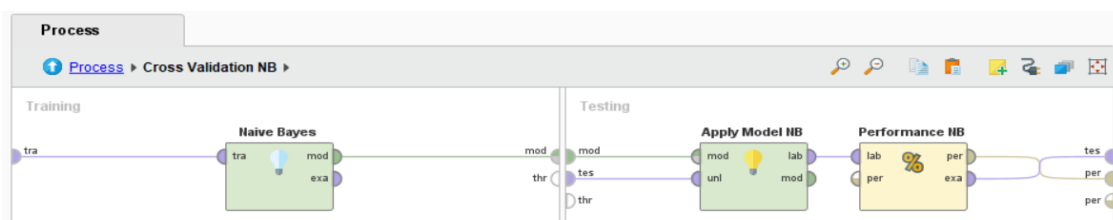
4.3. Proses Pengujian Algoritma

Dalam proses pengujian algoritma, 4 algoritma yang digunakan yaitu NB, SVM, NB berbasis PSO dan SVM berbasis PSO akan dilakukan pengujian menggunakan *k-Fold Cross Validation* dengan nilai $k=10$ yang ditunjukkan pada Gambar 4. Pada model *cross-validation* yang ditunjukkan pada Gambar 4 terdapat beberapa operator yaitu algoritma untuk *training* yang menghasilkan model dan *testing performance* model. Operator pada model NB ditunjukkan pada Gambar 5, sedangkan Operator pada model SVM ditunjukkan pada Gambar 6.



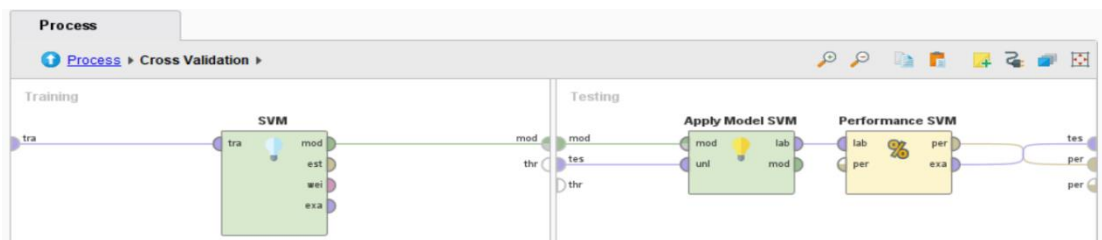
Gambar 4. Proses Pengujian *k-Fold Cross Validation*

Pada Gambar 5 menunjukkan rangkain operator pada model *cross validation* untuk algoritma NB yang direpresentasikan *software Rapid Miner*. Dataset dari hasil preprocessing akan divalidasi oleh model dari algoritma Naïve Bayes pada metode *cross validation*. Dimana didalamnya terdapat *performance* dengan menggunakan *confusion matrix* dan kurva *Receiver Operating Characteristic* (ROC) sebagai model evaluasi dari kinerja algoritma.



Gambar 5. Operator Pada Algoritma NB

Sedangkan pada Gambar 6 menunjukkan rangkain operator pada model *cross validation* untuk algoritma SVM yang direpresentasikan *software Rapid Miner*. Dataset dari hasil preprocessing akan divalidasi oleh model dari algoritma SVM menggunakan metode *cross validation*. Dimana didalamnya terdapat *performance* dengan menggunakan *confusion matrix* dan kurva ROC sebagai model evaluasi dari kinerja algoritma.



Gambar 6. Operator Pada Algoritma SVM

4.4. Hasil Pengujian dan Evaluasi

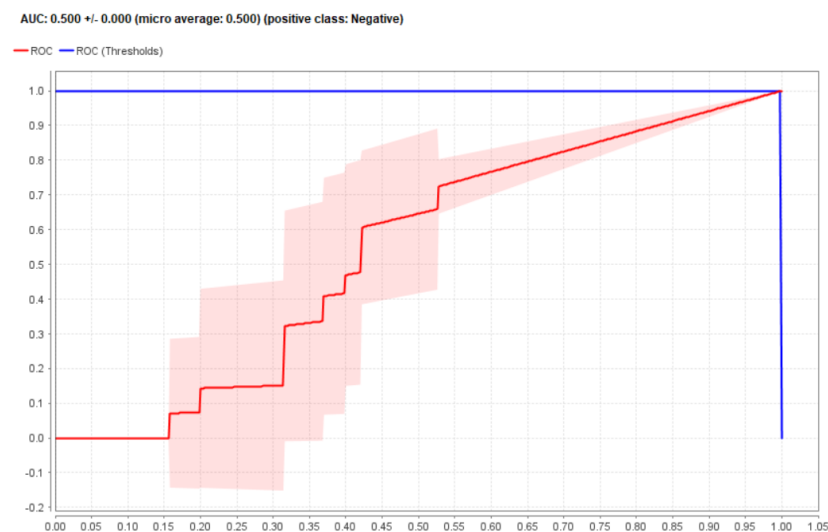
a) Algoritma Naïve Bayes

Hasil pengujian dengan algoritma NB menghasilkan akurasi yang ditampilkan pada Tabel 2, sedangkan hasil pengujian *performance* algoritma NB menghasilkan kurva ROC dengan nilai *Area Under Cover* (AUC) sebesar 0,500 pada Gambar 7.

Tabel 2. *Confusion Matrix* Algoritma NB

<i>Accuracy: 64,04% +/- 7,67% (micro average: 72,55%)</i>			
	<i>True Negative</i>	<i>True Positive</i>	<i>Class Precision</i>
<i>Prediction Negative</i>	175	66	72,61%
<i>Prediction Positive</i>	103	126	55,02%
<i>Class Recall</i>	62,95%	65,62%	

Pada Tabel 2 menunjukkan nilai akurasi yang dihasilkan sebesar 64,04% dengan tingkat *error* sebesar 7,67%. *True positive* sebesar 126 dan *true negative* sebesar 175 mengartikan bahwa sebanyak 126 *record* diklasifikasikan sesuai dengan sentimen positif dan sebanyak 175 *record* diklasifikasikan sesuai sentimen negatif. Sedangkan *false negative* sebesar 103 dan *false positive* sebesar 66 mengartikan bahwa sentimen positif diklasifikasikan negatif sebanyak 103 *record* dan sentimen negatif diklasifikasikan positif sebanyak 66 *record*.



Gambar 7. Grafik ROC Pada Model NB

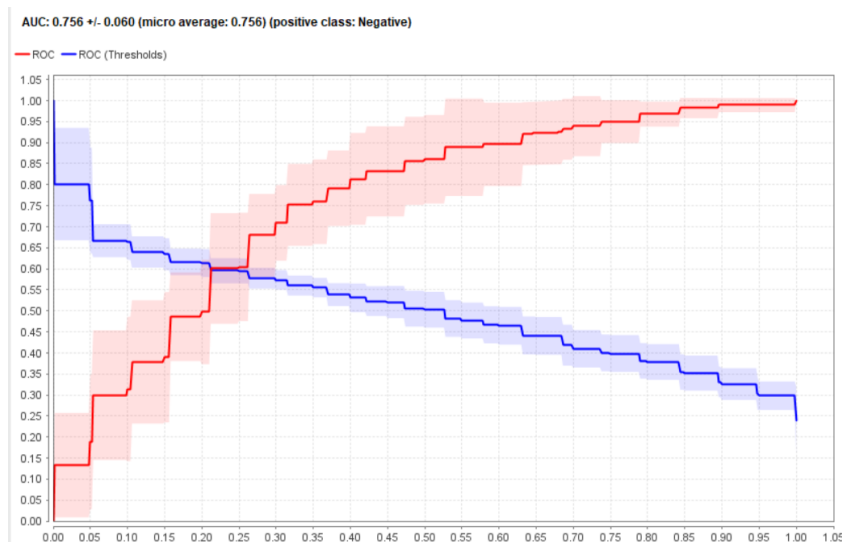
b) Algoritma *Support Vector Machine* (SVM)

Hasil pengujian dengan algoritma SVM menghasilkan akurasi yang ditampilkan pada Tabel 3, sedangkan hasil pengujian *performance* algoritma SVM menghasilkan kurva ROC dengan nilai AUC sebesar 0,756 pada Gambar 8.

Tabel 3. Confusion Matrix Algoritma SVM

Accuracy: 72,55% +/- 6,76% (micro average: 72,55%)			
	<i>True Negative</i>	<i>True Positive</i>	<i>Class Precision</i>
<i>Prediction Negative</i>	242	93	72,44%
<i>Prediction Positive</i>	36	99	73,33%
<i>Class Recall</i>	87,05	51,56	

Pada Tabel 3 menunjukkan nilai akurasi yang dihasilkan sebesar 72,55% dengan tingkat *error* sebesar 6,76%. *True positive* sebesar 99 dan *true negative* sebesar 242 mengartikan bahwa sebanyak 242 *record* diklasifikasikan sesuai dengan sentimen positif dan sebanyak 36 *record* diklasifikasikan negative. Sedangkan *false negative* sebesar 36 dan *false positive* sebesar 93 mengartikan bahwa sentimen positif diklasifikasikan negatif sebanyak 36 *record* dan sentimen negatif diklasifikasikan positif sebanyak 93.



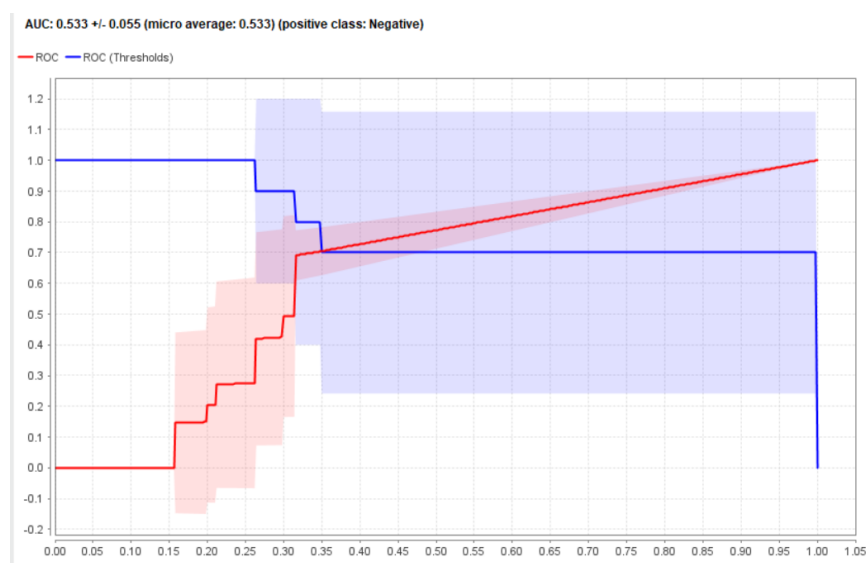
Gambar 8. Grafik ROC Pada Model SVM

c) Algoritma NB Berbasis PSO

Pada Tabel 4 menunjukkan hasil pengujian dengan algoritma NB berbasis PSO, sedangkan hasil pengujian *performance* algoritma NB berbasis PSO menghasilkan kurva ROC dengan nilai AUC sebesar 0,533 pada Gambar 9.

Tabel 4. Confusion Matrix NB Berbasis PSO

Accuracy: 70,43% +/- 5,89% (micro average: 72,55%)			
	<i>True Negative</i>	<i>True Positive</i>	<i>Class Precision</i>
<i>Prediction Negative</i>	187	48	79,57%
<i>Prediction Positive</i>	91	144	61,28%
<i>Class Recall</i>	67,27%	75,00%	



Gambar 9. Grafik ROC Pada Model NB Berbasis PSO

Pada Tabel 4 menunjukkan nilai akurasi yang dihasilkan sebesar 70,43% dengan tingkat *error* sebesar 5,89%. *True positive* sebesar 144 dan *true negative* sebesar 187 mengartikan bahwa sebanyak 144 *record* diklasifikasikan sesuai dengan sentimen positif dan sebanyak 187 *record* diklasifikasikan

negative. Sedangkan *false negative* sebesar 91 dan *false positive* sebesar 48 mengartikan bahwa sentimen positif diklasifikasikan negatif sebanyak 91 *record* dan sentimen negatif diklasifikasikan positif sebanyak 48.

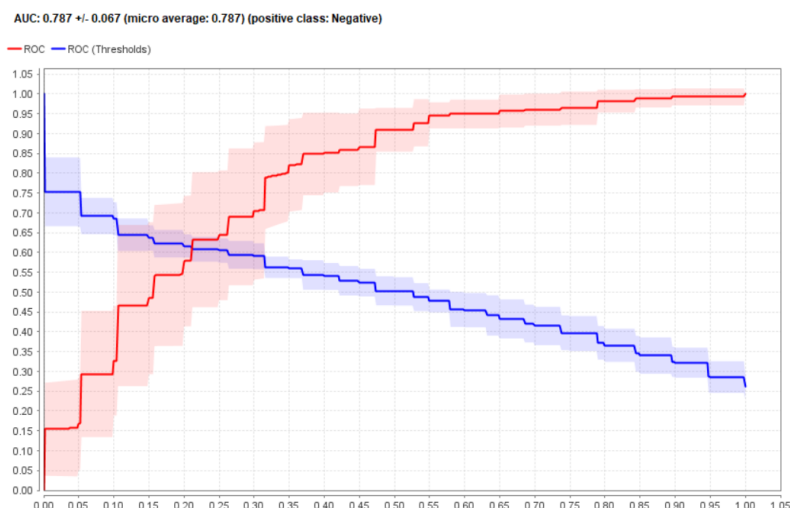
d) Algoritma SVM Berbasis PSO

Pada Tabel 5 menunjukkan hasil pengujian dengan algoritma SVM berbasis PSO, sedangkan hasil pengujian *performance* algoritma SVM berbasis PSO menghasilkan kurva ROC dengan nilai AUC sebesar 0,787 pada Gambar 10.

Pada Tabel 5 menunjukkan nilai akurasi yang dihasilkan sebesar 76,38% dengan tingkat *error* sebesar 4,65%. *True positive* sebesar 101 dan *true negative* sebesar 258, mengartikan bahwa sebanyak 101 *record* diklasifikasikan sesuai dengan sentimen positif dan sebanyak 258 *record* diklasifikasikan negatif. Sedangkan *false negative* sebesar 20 dan *false positive* sebesar 91, mengartikan bahwa sentimen positif diklasifikasikan negatif sebanyak 20 *record* dan sentimen negatif diklasifikasikan positif sebanyak 91.

Tabel 5. Confusion Matrix SVM Berbasis PSO

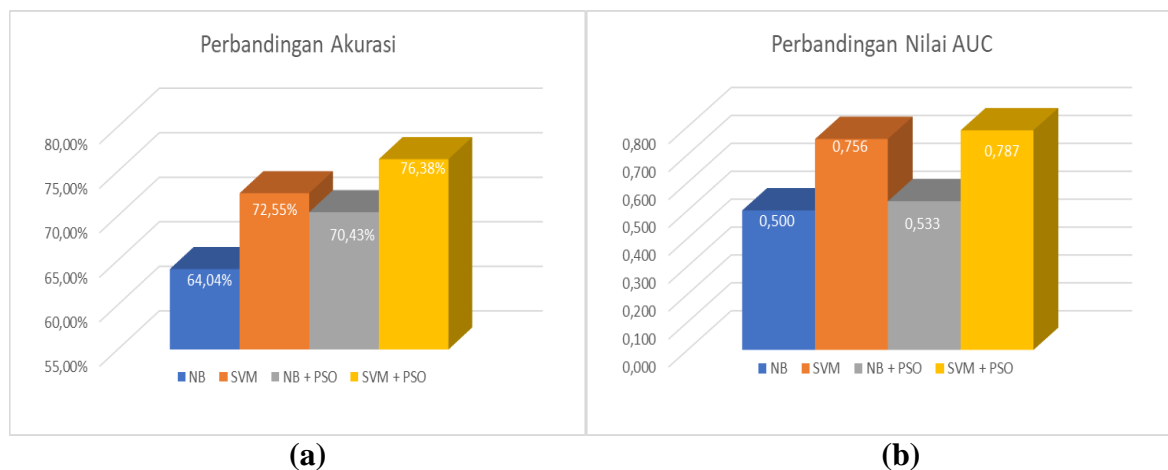
<i>Accuracy: 76,38% +/-4,65% (micro average: 76,38%)</i>			
	<i>True Negative</i>	<i>True Positive</i>	<i>Class Precision</i>
<i>Prediction Negative</i>	258	91	73,91%
<i>Prediction Positive</i>	20	101	83,47%
<i>Class Recall</i>	92,81%	52,60%	



Gambar 10. Grafik ROC Pada Model SVM Berbasis PSO

4.5. Perbandingan Hasil Pengujian

Hasil dari evaluasi pengujian algoritma klasifikasi yang telah dilakukan, maka diperoleh nilai akurasi dan kurva AUC. Perbedaan nilai akurasi pada setiap algoritma dapat dilihat pada grafik Gambar 11(a), sedangkan hasil dari kurva AUC ditampilkan dalam grafik pada gambar 11(b)



Gambar 11. (a) Grafik Perbandingan Akurasi (b) Grafik Perbandingan Nilai AUC

Pada Gambar 11(a) hasil akurasi algoritma NB sebesar 64,04% sedangkan akurasi algoritma NB berbasis PSO sebesar 70,43% dan hasil akurasi algoritma SVM sebesar 72,55% sedangkan algoritma SVM berbasis PSO sebesar 76,38%. Hasil akurasi algoritma NB berbasis PSO mengalami kenaikan sebesar 6,38% dibandingkan dengan hasil akurasi algoritma NB tanpa PSO, begitu juga dengan algoritma SVM berbasis PSO berhasil mengalami kenaikan sebesar 3,83% dibanding dengan algoritma SVM tanpa PSO.

Kenaikan akurasi pada algoritma berbasis PSO berbanding lurus dengan hasil kurva AUC yang ditunjukkan pada Gambar 11(b), nilai AUC pada algoritma NB sebesar 0,500 dan pada algoritma NB berbasis PSO sebesar 0,533. Sedangkan nilai AUC pada algoritma SVM sebesar 0,756 dan nilai AUC pada algoritma SVM berbasis PSO sebesar 0,787. Ditinjau dari perbandingan hasil akurasi dan nilai AUC maka optimasi menggunakan *feature* optimasi PSO pada setiap algoritma mendapatkan hasil lebih baik jika dibandingkan dengan algoritma tanpa PSO.

5 Kesimpulan

Penelitian tentang sentimen analisis dengan sumber data *Twitter* menggunakan konsep *data mining* yang diimplementasikan menggunakan *software Rapid Miner* dengan algoritma klasifikasi mendapat hasil yang baik. Hasil akurasi tertinggi diperoleh nilai akurasi sebesar 76,38% dan nilai AUC sebesar 0,787 pada algoritma SVM berbasis PSO. Penelitian ini juga membuktikan bahwa *feature* optimasi PSO dapat meningkatkan akurasi. Pada algoritma NB berbasis PSO mengalami kenaikan akurasi sebesar 6,38% dibandingkan dengan algoritma NB tanpa PSO, sedangkan pada algoritma SVM kenaikan akurasi sebesar 3,83% dibandingkan dengan algoritma SVM tanpa PSO.

Referensi

- [1] T. Krisdiyanto, E. Maricha, and O. Nurharyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naive Bayes Clasifiers," vol. 7, no. 1, pp. 32–37, 2021.
- [2] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, "International Journal of Information Management Social media analytics – Challenges in topic discovery , data collection , and data preparation," *Int. J. Inf. Manage.*, vol. 39, no. December 2017, pp. 156–168, 2018, doi: 10.1016/j.ijinfomgt.2017.12.002.
- [3] S. Nugroho and I. Hidayat, "Efektivitas Dan Keamanan Vaksin Covid-19 : Studi Refrensi," *J. Keperawatan Prof.*, vol. 9, 2021.
- [4] N. P. Astuti, E. G. Z. Nugroho, J. C. Lattu, I. R. Potempu, and D. A. Swandana, "Persepsi Masyarakat terhadap Penerimaan Vaksinasi Covid-19: Literature Review," *J. Keperawatan*, vol. 13, no. 3, pp. 569–580, 2021, doi: 10.32583/keperawatan.v13i3.1363.
- [5] R. Yanuarti, "Jurnal Sistem dan Teknologi Informasi Analisis Media Sosial Twitter Terhadap Topik Vaksinasi Covid-19," vol. 6, no. 2, pp. 121–130, 2021.

- [6] F. Zamachsari, G. V. Saragih, Susafa'ati, and W. Gata, "Analisis Sentimen Pemindahan Ibu Kota Negara dengan Feature Selection," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 3, pp. 504–512, 2020.
- [7] P. Arsi, R. Wahyudi, and R. Waluyo, "Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 231–237, 2021, doi: 10.29207/resti.v5i2.2698.
- [8] N. Hardi, Y. Alkahfi, P. Handayani, W. Gata, and M. R. Firdaus, "Analisis Sentimen Physical Distancing pada Twitter Menggunakan Text Mining dengan Algoritma Naive Bayes Classifier," *Sistemasi*, vol. 10, no. 1, p. 131, 2021, doi: 10.32520/stmsi.v10i1.1118.
- [9] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn," *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.
- [10] Yuyun, N. Hidayah, and S. Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 10, pp. 820–826, 2021.
- [11] R. Aulianita and A. Rifai, "Optimasi Particle Swarm Optimization Pada Naive Bayes Untuk Sentiment Analysis Furniture," *Inf. Manag. Educ. Prof.*, vol. 3, no. 1, pp. 31–40, 2018.
- [12] A. Faisal, Y. Alkhalifi, A. Rifai, and W. Gata, "Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis Particle Swarm Optimization," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 5, no. 2, p. 61, 2020, doi: 10.31328/jointecs.v5i2.1362.
- [13] A. E. Augustia, R. Taufan, Y. Alkhalifi, and W. Gata, "Analisis Sentimen Omnibus Law Pada Twitter Dengan Algoritma Klasifikasi Berbasis Particle Swarm Optimization," *Paradigma*, vol. 23, no. 2, pp. 158–166, 2021.
- [14] M. N. Ardiansyah, R. Umar, and Sunardi, "Analisis Sentimen pada Twitter Menggunakan Metode Support Vector Machine," *Semin. Nas. Teknol. Fak. Tek. Univ. Krisnadwipayana*, vol. 1, no. 1, pp. 739–742, 2019, [Online]. Available: <https://jurnal.teknikunkris.ac.id/index.php/semnastek2019/article/view/343/342>.
- [15] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.
- [16] Warjiyono, S. Aji, and Fandhilah, "The Sentiment Analysis of Fintech Users Using Support Vector Machine and Particle Swarm Optimization Method," *7th Int. Conf. Cyber IT Serv. Manag. (CITSM 2019) Fig.*, 2019.
- [17] Q. BAI, "Analysis of Particle Swarm Optimization Algorithm Qinghai," *Comput. Inf. Sci.*, vol. 3, no. 1, pp. 180–184, 2010, doi: 10.1007/s13369-019-03991-8.
- [18] X. Bai, X. Gao, and B. Xue, "Particle Swarm Optimization Based Two-Stage Feature Selection in Text Mining," *2018 IEEE Congr. Evol. Comput. CEC 2018 - Proc.*, pp. 1–8, 2018, doi: 10.1109/CEC.2018.8477773.
- [19] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, "Data Crawling Otomatis pada Twitter," no. September, pp. 11–16, 2016, doi: 10.21108/indosc.2016.111.
- [20] A. Guterres, Gunawan, and J. Santoso, "Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based," *Teknika*, vol. 8, no. 2, pp. 142–147, 2019, doi: 10.34148/teknika.v8i2.224.
- [21] S. A. Putri, "Prediksi Cacat Software Dengan Teknik Sampel Dan Seleksi Fitur Pada Bayesian Network," *J. Kaji. Ilm.*, vol. 19, no. 1, p. 17, 2019, doi: 10.31599/jki.v19i1.314.
- [22] Styawati and K. Mustofa, "A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification," vol. 13, no. 3, pp. 219–230, 2019.
- [23] R. Wijayanti and A. Arisal, "Ensemble approach for sentiment polarity analysis in user-generated Indonesian text," *Proc. - 2017 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Comput. Sci. Eng. IC3INA 2017*, vol. 2018-Janua, pp. 158–163, 2017, doi: 10.1109/IC3INA.2017.8251759.
- [24] A. Amolik, N. Jivane, M. Bhandari, and M. Venkatesan, "Twitter sentiment analysis of movie reviews using machine learning technique," *Int. J. Eng. Technol.*, vol. 7, no. 6, pp. 2038–2044, 2016.
- [25] I. Assayyis, I. Cholissodin, and Tibyani, "Optimasi Travelling Salesman Problem Pada Angkutan Sekolah Menggunakan Algoritme Genetika (Studi Kasus : Sekolah MI Salafiyah <http://sistemasi.ftik.unisi.ac.id>

- Kasim Blitar),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 1, pp. 454–461, 2020.
- [26] M. H. Rifqo and N. D. M. Veronica, “Implementasi Algoritme Naïve Bayes Berbasis Particle Swarm Optimization Dalam Penentuan Pemberian Kredit,” *Pseudocode*, vol. 6, no. 1, pp. 1–12, 2019, doi: 10.33369/pseudocode.6.1.1-12.
- [27] J. Winahyu and I. Suharjo, “Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes,” *Kumpul. Artik. Mhs. Pendidik. Tek. Inform.*, vol. 10, no. 2, p. 206, 2021, doi: 10.23887/karmapati.v10i2.36609.
- [28] 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, doi: 10.15395/mkb.v50n4.1342.