

Analisis Sentimen Vaksinasi Covid-19 Pada Twitter Menggunakan *Naive Bayes Classifier* Dengan *Feature Selection Chi-Squared Statistic* Dan *Particle Swarm Optimization*

Ristasari Dwi Septiana¹, Agung Budi Susanto², Tukiyat³

^{1,2,3}Teknik Informatika, Pascasarjana Universitas Pamulang, Tangerang Selatan, Indonesia
¹ristasari.dwis@gmail.com, ²agungpsmk@gmail.com, ³dimastuky@gmail.com

Diterima : 31 Agustus 2021

Disetujui : 28 September 2021

Abstract— *Tingginya penyebaran Covid-19 semakin berdampak pada bidang kesehatan, ekonomi, bahkan bidang pendidikan di Indonesia, sehingga pemerintah Indonesia melakukan tindakan vaksinasi Covid-19 guna menekan tingkat penyebaran Covid-19 di Indonesia. Namun hal tersebut dinilai kontroversial sehingga menarik perhatian masyarakat untuk memberikan opini di berbagai media seperti media sosial twitter. Sehingga membutuhkan analisa sentimen masyarakat terhadap upaya pemerintah pada tindakan vaksinasi Covid-19 untuk mencapai hasil prediksi dengan nilai akurasi paling optimal. Proses crawling secara otomatis menggunakan tools Rapidminer akan mengambil data tweets yang mengandung 5 (lima) kata kunci, yaitu “Vaksin Sinovac”, “Vaksin Astrazeneca”, “Vaksin Moderna”, “Vaksin Merah Putih”, dan “Vaksinasi Covid-19”. Dataset tweets didapatkan dari tanggal 4 Agustus 2021 sampai 12 Agustus 2021. Dataset diperoleh sejumlah 2060 tweets dan diberi label secara manual didapatkan jumlah tweet sebanyak 1193 sentimen positif, 73 negatif, dan 794 netral. Data tersebut dianalisa dengan menggunakan Metode Feature Selection Chi-Squared Statistic dan Particle Swarm Optimization (PSO) untuk mengurangi atribut yang kurang relevan pada saat proses klasifikasi dengan algoritma Naive Bayes Classifier (NBC). Hasil pengujian menunjukkan bahwa Algoritma Naive Bayes Classifier (NBC) tanpa Feature Selection mendapatkan nilai akurasi 63,69%. Hasil penelitian menunjukkan bahwa Algoritma Naive Bayes Classifier (NBC) dengan Feature Selection Chi-Squared Statistic mempunyai tingkat akurasi 69,13%. Sedangkan hasil pengujian algoritma Naive Bayes Classifier (NBC) dengan Particle Swarm Optimization mempunyai tingkat akurasi 66,02%. Dengan demikian hasil seleksi fitur Chi-Squared Statistic mendapatkan nilai akurasi yang lebih baik jika dibandingkan dengan Particle Swarm Optimization untuk proses klasifikasi algoritma Naive Bayes Classifier (NBC) dengan selisih akurasi 3,11%.*

Keywords— *Analisis Sentimen, Chi-Squared Statistic, NBC, PSO, Twitter*

I. PENDAHULUAN

Wabah covid-19 di Cina pada bulan Desember tahun 2019 merebak ke berbagai negara di dunia sehingga *World Health Organization* (WHO) menjadikan wabah tersebut sebagai pandemi global. *Corona virus* menjadi virus yang sangat berbahaya dengan tingkat penyebaran yang sangat cepat hingga meluas ke

seluruh dunia termasuk Indonesia. Dampak wabah covid-19 di Indonesia selain bidang kesehatan juga berdampak pada bidang ekonomi dan pendidikan. Data WHO pada tanggal 5 Juni 2021 tercatat 173.357.945 kasus positif covid-19, 156.121.170 kasus sembuh dan 3.728.668 kasus meninggal dunia. Sedangkan kasus covid-19 di Indonesia tercatat 1.850.206 kasus positif, 1.701.784 kasus sembuh dan 51.449 kasus

meninggal dunia. Banyaknya kasus covid-19 sangat berpengaruh di berbagai sektor kehidupan manusia, seperti ekonomi, pendidikan, politik, dan kehidupan sosial [1]. Upaya pemerintah Indonesia di antaranya dengan memberlakukan kebijakan protokol kesehatan, Pembatasan Sosial Berskala Besar (PSBB) hingga vaksinasi masyarakat untuk menekan penyebaran Covid-19. Tindakan vaksinasi masyarakat dinilai kontroversial dengan munculnya berbagai merek vaksin sehingga banyak kalangan mengutarakan pendapatnya di berbagai media. *Twitter* menjadi salah satu media bagi masyarakat untuk menyampaikan pendapat mereka tentang vaksinasi covid-19 di Indonesia. Pada kuartal ketiga tahun 2019 jumlah pengguna aktif *twitter* tercatat meningkat 17% menuju angka 145 juta pengguna [1].

Analisis sentimen berada di antara berbagai bidang penelitian lainnya seperti *Data Mining*, *Natural Language Processing* (NLP) dan *Machine Learning* untuk melakukan ekstraksi sentimen terhadap sebuah isi kalimat [2]. Analisis sentimen dilakukan untuk mengelompokkan dokumen, kalimat, atau pendapat berdasarkan polaritas teks di dalamnya. Polaritas tersebut merupakan pendapat yang memiliki aspek positif, negatif atau netral [3]. Klasifikasi dengan analisis sentimen membantu memberikan masukan dan tanggapan dari pelanggan dengan cepat [4].

Naive Bayes merupakan teknik klasifikasi yang paling sering digunakan untuk klasifikasi data yang disebut dengan *Naive Bayes Classifier* (NBC). Pemilihan algoritma *Naive Bayes* ini karena sangat cocok untuk *short data text* dan sederhana namun memiliki nilai akurasi yang tinggi dalam pengklasifikasian data teks [5]. Metode *Naive Bayes* banyak digunakan di berbagai penelitian tentang analisis dokumen tekstual dan merupakan algoritma populer karena kemudahan penggunaannya serta telah terbukti memuaskan di banyak domain penelitian [6].

Salah satu faktor penting untuk meningkatkan nilai akurasi klasifikasi adalah seleksi fitur. Apabila *dataset* berisi fitur yang banyak, maka dimensi ruang menjadi besar dan tingkat akurasi

klasifikasi menjadi rendah. Seleksi fitur digunakan untuk mengurangi jumlah *set* fitur yang besar menjadi *subset* fitur yang relatif lebih kecil agar nilai akurasi klasifikasi meningkat [4]. Seleksi fitur *chi-squared statistic* dan *particle swarm optimization* digunakan untuk mengurangi gangguan (*noise*) pada saat klasifikasi menggunakan teknik pembelajaran mesin *Naive Bayes Classifier*. Penelitian terdahulu menunjukkan bahwa frekuensi fitur dalam kategori benar dan dalam kategori palsu berperan penting dalam pemilihan fitur *chi-squared* [3].

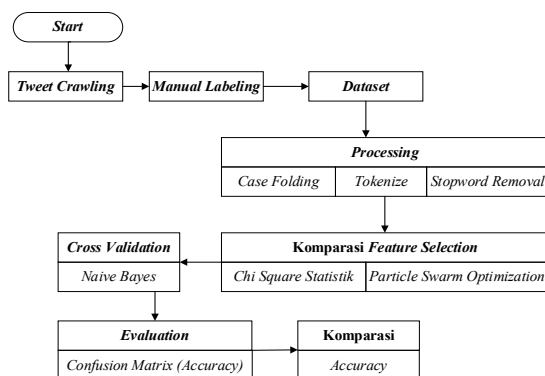
Seleksi fitur diperlukan untuk mendapatkan fitur-fitur yang relevan yang selanjutnya akan digunakan dalam proses konstruksi model probabilistik NBC. Penyeleksian fitur yang digunakan adalah *chi-squared* dan *particle swarm optimization*. Penelitian ini bertujuan untuk memprediksi sentimen masyarakat terhadap tindakan vaksinasi covid-19 dengan cara mengkomparasi metode *Feature Selection Chi Squared Statistic* dan *Particle Swarm Optimization* pada algoritma *Naive Bayes Classifier* ke dalam sentimen positif, negatif dan netral dengan menggunakan data *tweet* masyarakat dari media sosial *twitter*. Data yang diperoleh selanjutnya dibagi menjadi data pelatihan untuk memperoleh model klasifikasi dan data pengujian yang digunakan untuk menguji hasil klasifikasi dari algoritma NBC.

Demi mendapatkan hasil penelitian yang tepat guna, peneliti merujuk beberapa penelitian terkait yang pernah dilakukan sebelumnya sebagai komparasi dan data-data penunjang dalam penelitian ini. Penelitian yang dilakukan oleh Laurensz dan Sedyono dengan Metode *Naive Bayes* dan *Support Vector Machine* (SVM) untuk menganalisis sentimen masyarakat terhadap tindakan vaksinasi dalam upaya mengatasi pandemi covid-19 [7]. Klasifikasi metode *Naive Bayes* mendapatkan rata-rata akurasi 85,59%, sedangkan SVM sebesar 84,41%. Penelitian yang dilakukan Naraswati dengan Metode *Naive Bayes Classification* tentang Analisis Sentimen Publik dari *Twitter* tentang Kebijakan Penanganan Covid-19 di Indonesia [8]. Nilai akurasi klasifikasi yang didapatkan sebesar 87,34%,

dengan sensitivitas sebesar 93,43%, dan spesifisitas 71,76%. Penelitian yang dilakukan Yono Cahyono menggunakan *NBC* dengan *Feature Selection PSO* dan *Term Frequency* untuk menganalisis sentiment masyarakat pada sosial media *twitter* terhadap layanan Telkom Indihome [4]. Hasil pengujian seleksi fitur *Particle Swarm Optimization* (PSO) terbukti dapat meningkatkan akurasi algoritma *Naive Bayes Classifier* sebesar 97,48%.

II. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini meliputi tiga proses utama yaitu *preprocessing*, pelatihan dan pengujian.



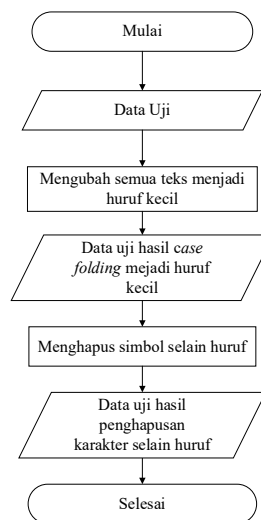
Gambar 1. Metodologi Penelitian

3.1 Preprocessing

Tweet crawling adalah tahapan yang dilakukan untuk mendapatkan data *tweet* masyarakat tentang vaksinasi covid-19 dari media sosial *twitter*. Sebelum menjadi *dataset* terlebih dahulu dilakukan tahapan *manual labelling*, yaitu memberikan label terhadap kalimat *tweet* tersebut kedalam kategori positif, netral atau negatif secara manual. *Preprocessing* adalah tahap yang dilakukan sebelum proses pelatihan pada algoritma NBC dan tahapan yang penting bagi *dataset* pada proses *mining*. *Preprocessing* dilakukan untuk menghindari *dataset* yang kurang sempurna, terdapat *noise* pada *dataset*, data-data yang tidak konsisten dan mempercepat pemrosesan terhadap dokumen [6]. Proses-proses yang dilakukan *preprocessing* yaitu *case folding*, *tokenize*, dan *stopwords removal*.

a. Case Folding

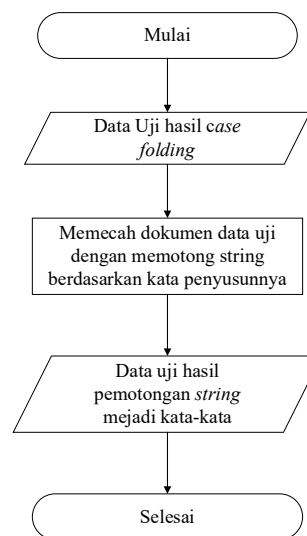
Tahap ini merupakan proses untuk mengubah semua huruf pada dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' saja yang diterima. Karakter yang dianggap sebagai delimiter seperti titik (.), koma (,), spasi dan karakter angka pada dokumen akan dihilangkan. Berikut adalah *flowchart* tahap *case folding*.



Gambar 2. Flowchart Tahapan Case Folding

b. Tokenize

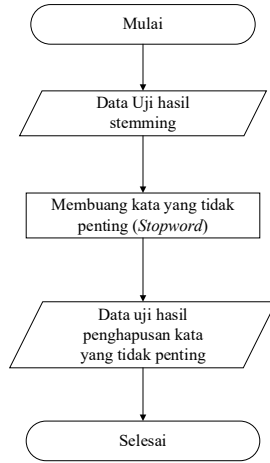
Tahap memotong *string input* berdasarkan kata yang menyusunnya serta membedakan karakter-karakter tertentu yang diperlukan sebagai pemisah kata atau bukan. Tahap ini dilakukan setelah melewati tahap *case folding*. Berikut adalah *flowchart* tahap *tokenizing*.



Gambar 3. Flowchart Tahapan Tokenize

c. *Stopwords Removal*

Tahap menghilangkan kata-kata yang tidak memiliki arti, berdasarkan daftar *stopwords* bahasa Indonesia sebanyak 758 kata. Berikut adalah *flowchart* tahap *stopwords*.



Gambar 4. *Flowchart* Tahapan *Stopwords*

3.2 Pelatihan

a. Seleksi Fitur *Chi-Squared Statistic*

Nilai seleksi fitur *Chi-Squared Statistic* diperoleh dengan menggunakan persamaan (1).

$$X^2(t, c) = \frac{Nx(AD - CB)^2}{(A + C) \times (B + D) \times (A + B) \times (C + D)} \quad (1)$$

Keterangan:

- A: Banyaknya dokumen dalam kategori *c* yang mengandung *term t*
- B: Banyaknya dokumen yang bukan kategori *c* tetapi mengandung *term t*
- C: Banyaknya dokumen dalam kategori *c* tetapi tidak mengandung *term t*
- D: Banyaknya dokumen yang bukan kategori *c* dan tidak mengandung *term t*
- N: Total keseluruhan dokumen

Melalui *Chi-Squared Statistic* kesesuaian (*goodnes of fit*) dari kategori terhadap *terms* dapat diamati [4]. Dalam statistika uji *Chi-Squared Statistic* diterapkan untuk melihat independensi terhadap dua peristiwa. Sedangkan dalam seleksi fitur melihat kemunculan dari fitur dan kemunculan dari kategori.

b. *Particle Swarm Optimization (PSO)*

Particle Swarm Optimization (PSO) adalah teknik optimasi stokastik berbasis populasi yang

dikembangkan oleh Eberhart dan Kennedy pada tahun 1995, yang terinspirasi oleh perilaku sosial kawanan burung atau ikan [9]. Teknik ini diawali dari populasi sejumlah *term* yang dibangkitkan secara acak dan selanjutnya dicari solusi optimum dengan memperbaiki *term* untuk sejumlah kategori tertentu. Posisi (*xi, d*) dan kecepatan (*ci, m*) dibangkitkan dari kumpulan partikel secara acak dengan menggunakan batas bawah (X_{min}) dan batas atas (X_{max}). Selanjutnya *update velocity* (kecepatan) semua *term* berdasarkan nilai *fitness* untuk menentukan *term* yang memiliki nilai *global* terbaik (*global best*) dan posisi terbaik (*local best*) dari setiap *term* pada semua waktu sekarang dan sebelumnya. Lakukan pengulangan sampai kriteria terpenuhi.

$$V_{i,m} = w.V_{i,m} + C_1 * R * (pbest_{i,m} - X_{i,m}) + C_2 * R * (gbest_m - X_{i,m}) \quad (2)$$

$$xid = xi, m + vi, m \quad (3)$$

Dimana:

n : jumlah partikel dalam kelompok

d : dimensi

vi, m : kecepatan partikel ke-*i* pada iterasi ke-*i*

w : faktor bobot inersia

c1, c2 : konstanta akselerasi (*learning rate*)

R : bilangan random (0-1)

Xi, d : posisi saat ini dari partikel ke-*i* pada iterasi ke-*i*

pbesti : posisi terbaik sebelumnya dari partikel ke-*i*

gbest : partikel terbaik diantara semua partikel dalam satu kelompok atau populasi

c. *Naïve Bayes Classifier (NBC)*

Algoritma NBC digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling tepat [10]. Klasifikasi *naïve bayes* juga digolongkan ke dalam pembelajaran terawasi (*supervised learning*) di mana untuk penentuan kelas objek uji didasarkan dari data latihnya dan setiap objek di data latih telah diketahui kelasnya masing-masing [11]. Rumus teorema *bayes* dapat dilihat seperti berikut ini.

$$P(H|X) = \frac{P(X|H) * P(H)}{P(X)} \quad (4)$$

Dengan:

H: hipotesisnya

X: kejadian yang merealisasikan ke hipotesis

P(H): peluang hipotesis

P(X): peluang kejadian

P(X | H): peluang banyak X di dalam H

P(H | X): peluang banyaknya H di dalam X

3.3 Pengujian

a. Pengujian *K-Fold Cross Validation*

Cross Validation merupakan metode untuk menemukan *parameter* terbaik dengan menguji besarnya galat pada data pengujian. Metode ini membagi *dataset* secara acak dengan ukuran yang sama ke dalam *k* bagian dan tiap-tiap bagian akan dilakukan proses klasifikasi [12]. Untuk pengujian yang biasa di gunakan adalah dengan $k=10$, atau *10-fold cross validation*. Sebuah partisi *k-Fold* dari *dataset* dibuat. Setiap percobaan *k*, *k-1* lipatan digunakan untuk pelatihan dan sisanya satu untuk pengujian. Maka rata-rata *error E* di semua cobaan *k* dihitung sebagai berikut.

$$E = \frac{1}{k} \sum_{i=1}^k E_i \quad (5)$$

b. *Confusion Matrix*

Proses prediksi yang dilakukan diharapkan mampu melakukan klasifikasi semua *dataset* dengan benar, tetapi tidak dapat dipungkiri jika kinerja suatu sistem tidak 100% benar sehingga perlu dilakukan pengukuran kinerja klasifikasi dengan menggunakan matriks konfusi (*confusion matrix*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi [13]. Evaluasi performasi dilakukan untuk menguji hasil dari klasifikasi dengan mengukur nilai performasi dari model yang telah dibuat. *Parameter* pengujian yang digunakan untuk evaluasi yaitu akurasi yang perhitungannya diperoleh dari tabel *Confussion Matrix*.

Tabel 1. *Confussion Matrix*

	<i>True Negatif</i>	<i>True Positif</i>	<i>True Netral</i>
<i>Pred.Positif</i>	FN	TP	Fnet
<i>Pred.Netral</i>	FN	FP	Tnet
<i>Pred.Negatif</i>	TN	FP	Fnet

Nilai akurasi menunjukkan ketepatan hasil proses klasifikasi data secara benar. Dapat dikatakan nilai akurasi merupakan perbandingan data yang terklasifikasi benar terhadap keseluruhan data. Nilai akurasi dihitung dengan persamaan berikut.

$$Accuracy = \frac{TP + TN + Tnet}{TP + TN + FP + FN + FNet + Tnet} \quad (6)$$

III. HASIL DAN PEMBAHASAN

Penelitian dilakukan dengan kombinasi metode atau melakukan penentuan jumlah fitur pada proses seleksi fitur. Tujuan dari proses seleksi fitur ini agar semakin sedikit jumlah fitur yang digunakan, semakin rendah waktu komputasi dan semakin tinggi akurasi yang dicapai. Tahap seleksi fitur akan membandingkan seleksi fitur menggunakan *Chi-Squared Statistic* dan *Particle Swarm Optimization* (PSO).

4.1 Persiapan *Dataset*

Twitter menyediakan data yang bisa diakses secara bebas dengan menggunakan *Twitter API* (*Application Programming Interface*) [14]. Tahap *crawling* mengambil data *tweets* yang mengandung 5 (lima) kata kunci, yaitu “Vaksin Sinovac”, “Vaksin Astrazeneca”, “Vaksin Moderna”, “Vaksin Merah Putih”, dan “Vaksinasi Covid-19”. Dengan proses *crawling* didapatkan *dataset tweets* dari tanggal 4 Agustus 2021 sampai 12 Agustus 2021 yang ditunjukkan pada tabel berikut.

Tabel 1. Total *Dataset Hasil Crawling Tweets*

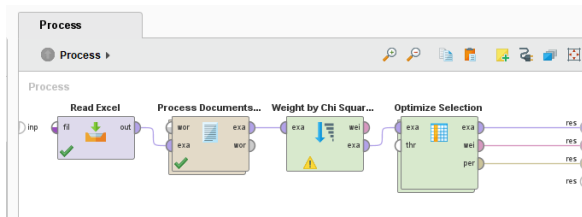
No.	Kata Kunci	Jumlah Records
1	Vaksin Sinovac	449
2	Vaksin Astrazeneca	465
3	Vaksin Moderna	351
4	Vaksin Merah Putih	152
5	Vaksinasi Covid-19	643
Total		2060

Data *tweets* selanjutnya diberi label secara manual ke dalam tiga kategori, yaitu kategori positif, netral, dan negatif. Data *tweets* disimpan dalam *file excel* yang terdiri dari 1193 positif, 73 negatif dan 794 netral.

4.2 Seleksi Fitur *Chi-Squared Statistic*

Tahap seleksi fitur menggunakan *Chi Squared Statistic* dilakukan dengan *optimize selection*. Pada *optimize selection* digunakan

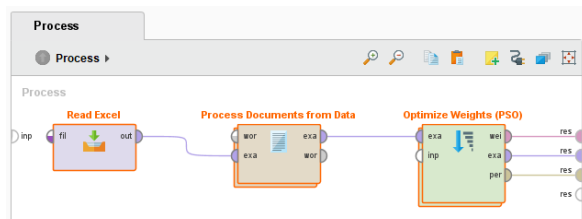
parameter *forward selection*. Hasil seleksi fitur ini mendapatkan 24 kata terseleksi.



Gambar 5. Proses Seleksi Fitur *Chi-Squared Statistic*

4.3 Seleksi Fitur *Particle Swarm Optimization*

Optimize weights PSO sebagai seleksi fitur dengan parameter *forward selection* dapat dilihat pada gambar berikut.

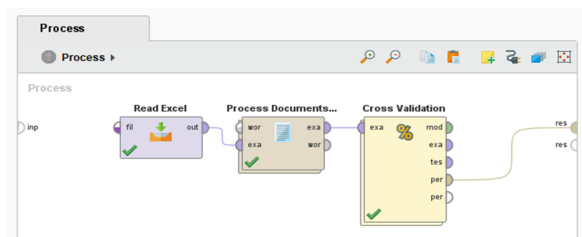


Gambar 6. Proses Seleksi Fitur *Optimize weights PSO*

Hasil seleksi fitur PSO dengan parameter *Term Frequency* didapatkan 28 kata yang terseleksi.

4.4 Hasil Pengujian dengan *Naive Bayes Classifier*

Pengujian pertama dilakukan untuk mengetahui tingkat akurasi dari penggunaan algoritma *Naive Bayes Classifier* tanpa seleksi fitur. Desain proses untuk pengujian klasifikasi dengan *Naive Bayes Classifier* dapat dilihat pada gambar berikut.



Gambar 7. Proses Pengujian Klasifikasi dengan NBC

Cross validation membagi data menjadi 10 bagian sama banyak, satu *subset* untuk data pelatihan dan *subset* lainnya untuk data pengujian. Hasil evaluasi pengujian dengan

Confussion Matrix pada algoritma *Naive Bayes Classifier* di atas ditunjukkan pada table berikut.

Tabel 2. *Confussion Matrix* pada Algoritma NBC

	<i>true Negatif</i>	<i>true Positif</i>	<i>true Netral</i>	<i>Class precision</i>
<i>Pred.Positif</i>	334	465	33	55,89%
<i>Pred.Netral</i>	31	100	19	12,67%
<i>Pred.Negatif</i>	828	229	21	76,81%
<i>class recall</i>	69,40%	58,56%	26,03%	

Tabel 2 menunjukkan bahwa penggunaan algoritma *Naive Bayes Classifier* tanpa seleksi fitur mendapat nilai akurasi 63,69%.

4.5 Hasil Pengujian *Chi-Squared Statistic* dengan *Naive Bayes Classifier*

Pengujian kedua dilakukan untuk mengetahui tingkat akurasi algoritma *Naive Bayes Classifier* dengan penggunaan seleksi fitur *Chi-Squared Statistic* berbasis *forward selection*.

Tabel 3. *Confussion Matrix Chi-Squared Statistic* Berbasis *Forward Selection* pada Algoritma NBC

	<i>true Negatif</i>	<i>true Positif</i>	<i>true Netral</i>	<i>Class precision</i>
<i>Pred.Positif</i>	372	603	39	59,47%
<i>Pred.Netral</i>	4	3	4	36,36%
<i>Pred.Negatif</i>	817	188	30	78,94%
<i>class recall</i>	68,48%	75,94%	5,48%	

Tabel 3 menunjukkan bahwa penggunaan seleksi fitur *Chi-Squared Statistic* berbasis *forward selection* mampu meningkatkan akurasi *Naive Bayes Classifier* hingga mencapai 69,13%.

4.6 Hasil Pengujian PSO dengan *Naive Bayes Classifier*

Pengujian ketiga dilakukan untuk mengetahui tingkat akurasi algoritma *Naive Bayes Classifier* dengan seleksi fitur PSO menggunakan parameter *Term Frequency* (TF).

Tabel 4. *Confussion Matrix PSO Menggunakan Parameter TF* pada NBC

	<i>true Negatif</i>	<i>true Positif</i>	<i>true Netral</i>	<i>Class precision</i>
<i>Pred.Positif</i>	223	417	31	62,15%

<i>Pred.Netral</i>	48	115	21	11,41%
<i>Pred.Negatif</i>	922	262	21	76,51%
<i>class recall</i>	77,28%	52,52%	28,71%	

Tabel 4 menunjukkan bahwa PSO dengan parameter *Term Frequency* (TF) meningkatkan akurasi algoritma *Naive Bayes Classifier* hingga mencapai 66,02%.

Confusion matrix pada tabel 5 menunjukkan bahwa pengujian yang dilakukan dapat meningkatkan akurasi algoritma *Naive Bayes Classifier*. Komparasi seleksi fitur *Chi Squared Statistic* dan *Particle Swarm Optimization* (PSO) dengan *Naive Bayes Classifier* dapat dilihat pada tabel 5.

Tabel 5. Komparasi Akurasi *Chi-Squared Statistic* dan *Particle Swarm Optimization* pada Algoritma NBC

Komparasi	Accuracy
Algoritma NBC	63,69%
<i>Chi-Squared Statistic</i> berbasis <i>Forward Selection</i> pada Algoritma NBC	69,13%
PSO dengan parameter <i>Term Frequency</i> (TF) pada Algoritma NBC	66,02%

Hasil pengujian seleksi fitur *Chi-Squared Statistic* dan *Particle Swarm Optimization* (PSO), keduanya terbukti mampu meningkatkan nilai akurasi algoritma *Naive Bayes Classifier*. Pengujian algoritma *Naive Bayes Classifier* menggunakan seleksi fitur *Chi-Squared Statistic* berbasis *forward selection* memiliki nilai akurasi yang lebih tinggi yaitu 69,13%, dibandingkan dengan seleksi fitur *Particle Swarm Optimization* (PSO) dengan parameter *term frequency* (TF) sebesar 66,02% dengan selisih akurasi sebesar 3,11%.

IV. KESIMPULAN

Simpulan yang dapat diambil dari penelitian ini adalah:

1. Proses seleksi fitur *chi squared statistic* dan *particle swarm optimization* (PSO) yang digunakan dalam penelitian ini terbukti mampu menyeleksi jumlah *term* menjadi lebih sedikit sehingga mempermudah proses seleksi klasifikasi *naive bayes* menjadi lebih cepat.

2. Hasil Pengujian algoritma *Naive Bayes Classifier* menggunakan seleksi fitur *Chi Squared Statistic* berbasis *forward selection* memiliki nilai akurasi yang lebih tinggi dibandingkan menggunakan seleksi fitur *Particle Swarm Optimization* (PSO) dengan parameter *term frequency* (TF) dengan selisih akurasi sebesar 3,11%.

DAFTAR PUSTAKA

- [1] H. Noviyani, . A. Y. Permana and A. Fauzi, "Komparasi Algoritma *Naive Bayes* Dan *K-Nearest Neighbor* Dalam Melihat Analisis Sentimen Terhadap Vaksinasi Covid-19," *Jurnal Ilmiah Informatika, Arsitektur dan Lingkungan*, vol. 14, no. 1, 2021.
- [2] M. Al-Ayyoub, A. A. Khamaiseh, Y. Jararweh and M. N. Al-Kabi, "A *Comprehensive Survey Of Arabic Sentiment Analysis*," *Information Processing and Management*, vol. 56, no. 2, pp. 320-342, 2019.
- [3] J. Ling, I. P. E. N. Kencana and T. B. Oka, "Analisis Sentimen Menggunakan Metode *Naive Bayes Classifier* dengan Seleksi Fitur *Chi Square*," *E-Jurnal Matematika*, vol. 3, no. 3, pp. 92-99, 2014.
- [4] Y. Cahyono, "Analisis Sentiment pada Sosial Media Twitter Menggunakan *Naive Bayes Classifier* dengan *Feature Selection Particle Swarm Optimization* dan *Term Frequency*," *Jurnal Informatika Universitas Pamulang*, vol. 2, no. 1, pp. 14-19, 2017.
- [5] P. Routray, C. K. Swain and S. P. Mishra, "A *Survey on Sentiment Analysis*," *International Journal of Computer Applications*, vol. 76, no. 10, pp. 1-8, 2013.
- [6] Ratino, N. Hafidz, S. Anggraeni and W. Gata, "Sentimen Analisis Informasi Covid-19 menggunakan *Support Vector Machine* dan *Naive Bayes*," *Jurnal JUPITER*, vol. 12, no. 2, pp. 1-11, 2020.
- [7] B. Laurensz and E. Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 2, pp. 118-123, 2021.
- [8] N. P. G. Naraswati, D. C. Rosmilda, D. Desinta, F. Khairi, R. Damaiyanti and R. Nooraeni, "Analisis Sentimen Publik dari *Twitter* Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan *Naive Bayes Classification*," *Jurnal Sistem Informasi*, vol. 10, no. 1, pp. 222-238, 2021.

- [9] X. Hu, R. C. Eberhart and Y. Shi, "*Particle Swarm with Extended Memory for Multiobjective Optimization*," in *IEEE Swarm Intelligence Symposium*, 2003.
- [10] R. Feldman and J. Sanger, "*The text mining handbook: advanced approaches in analyzing unstructured data*," Cambridge university press, 2007.
- [11] M. Arhami and M. Nasir, *Data Mining Algoritma dan Implementasi*, vol. 6, R. I. Utami, Ed., Lhokseumawe, Buketrata: Penerbit ANDI, 2020, pp. 104-111.
- [12] Rinawati, "Penerapan *Particle Swarm Optimization* Untuk Seleksi Atribut pada Metode *Support Vector Machine* untuk Penentuan Penilaian Kredit," in Seminar Nasional Ilmu Pengetahuan dan Teknologi Komputer (SNIPTEK), Jakarta, 2013.
- [13] E. Prasetyo, *Data Mining Konsep dan Aplikasi menggunakan MATLAB*, Yogyakarta: CV. Andi Offset, 2012.
- [14] E. S. Negara, R. Andryani and P. H. Saksono, "Analisis Data *Twitter*: Ekstraksi dan Analisis Data Geospasial," *INKOM*, vol. 10, no. 1, pp. 27-36, 2016.