

Optimasi naive Bayes classifier untuk klasifikasi teks pada e-government menggunakan particle swarm optimization

Naive Bayes classifier optimization for text classification on e-government using particle swarm optimization

Kuncahyo Setyo Nugroho^{*)}, Istiadi, Fitri Marisa

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Widyagama
Jl. Borobudur No.35, Malang, Indonesia 65142

Cara sitasi: K. S. Nugroho, I. Istiadi, and F. Marisa, "Optimasi naive Bayes classifier untuk klasifikasi teks pada e-government menggunakan particle swarm optimization," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 1, pp. 21-26, 2020. doi: [10.14710/jtsiskom.8.1.2020.21-26](https://doi.org/10.14710/jtsiskom.8.1.2020.21-26), [Online].

Abstract - One of the public e-government services is a web-based online complaints portal. Text of complaint needs to be classified so that it can be forwarded to the responsible office quickly and accurately. The standard classification approach commonly used is the Naive Bayes Classifier (NBC) and k-Nearest Neighbor (k-NN), which still classifies one label and needs to be optimized. This research aims to classify the complaint text of more than one label at the same time with NBC, which is optimized using Particle Swarm Optimization (PSO). The data source comes from the Sambat Online portal and is divided into 70 % as training data and 30 % as testing data to be classified into seven labels. NBC and k-NN algorithms are used as a comparison method to find out the performance of PSO optimization. The 10-fold cross-validation shows that NBC optimization using PSO achieves an accuracy of 87.44 % better than k-NN of 75 % and NBC of 64.38 %. The optimization model can be used to increase the effectiveness of services to e-government in society.

Keywords - online public services; web mining; complaint text classification optimization; naive Bayes

Abstrak - Salah satu bentuk e-government adalah portal pengaduan online berbasis web. Teks pengaduan perlu diklasifikasikan sehingga dapat diteruskan ke dinas yang bertanggung jawab dengan cepat dan akurat. Pendekatan klasifikasi standar yang umum digunakan adalah Naive Bayes Classifier (NBC) dan k-Nearest Neighbor (k-NN) yang masih mengklasifikasikan satu label dan perlu dioptimasi. Penelitian ini bertujuan mengklasifikasi teks pengaduan lebih dari satu label sekaligus dengan NBC yang dioptimalkan menggunakan Particle Swarm Optimization (PSO). Sumber data berasal dari portal Sambat Online dan dibagi menjadi 70 % sebagai data latih dan 30 % sebagai data uji untuk

diklasifikasikan ke dalam tujuh label. Algoritma NBC dan k-NN digunakan sebagai metode pembandingan untuk mengetahui kinerja optimasi PSO. Hasil pengujian menggunakan 10-fold cross-validation menunjukkan bahwa optimasi NBC menggunakan PSO mencapai akurasi sebesar 87,44 % yang lebih baik dari k-NN sebesar 75 % dan NBC 64,38 %. Model optimasi ini diharapkan dapat meningkatkan efektifitas layanan e-government pada masyarakat.

Kata kunci - layanan publik online; web mining; optimasi klasifikasi teks pengaduan; naive Bayes

I. PENDAHULUAN

Istilah *e-government* sering dipandang sebagai pemanfaatan sumber daya internet dalam urusan pemerintahan dan pelayanan publiknya kepada masyarakat. Salah satu bentuk nyata *e-Government* yang disediakan oleh pemerintah untuk mendukung tujuan demokratis adalah layanan *online* yang dapat menampung pesan dari publik [1]. Layanan tersebut berisi pesan pengaduan dalam bentuk teks berupa saran, kritik, keluhan, dan pertanyaan. Pengaduan tersebut akan dikategorikan sesuai Organisasi Perangkat Daerah (OPD) yang bertanggung jawab untuk ditanggapi dan ditindaklanjuti.

Namun, dalam penerapannya layanan tersebut masih menggunakan sumber daya manusia untuk mengkategorikan seluruh pengaduan yang masuk ke layanan *online* tersebut. Pengaduan harus dipilah dan diteruskan ke OPD secara manual. Hal ini kurang efisien apabila terdapat pengaduan yang bersifat penting dan mendesak sehingga tidak dapat ditanggapi dengan cepat karena banyaknya jumlah pengaduan yang masuk dan tercampur dengan perihal pengaduan yang lain.

Untuk mempermudah dalam mengelola dan meningkatkan efisiensi waktu dalam mengkategorikan pengaduan diperlukan sebuah sistem yang dapat memilah pengaduan secara otomatis sesuai OPD yang bertanggung jawab. Konsep *text mining* dapat digunakan untuk menyelesaikan permasalahan seperti ini. *Text mining* menerapkan konsep dari teknik *data*

^{*)} Penulis korespondensi (Kuncahyo Setyo Nugroho)
Email: ksnugroho26@gmail.com

mining untuk mencari pola menarik dalam teks yang dapat digunakan untuk klasifikasi teks [2]. *Text mining* mengolah data tidak terstruktur. Dalam prosesnya, diperlukan tahap awal (*text preprocessing*) untuk mempersiapkan teks menjadi lebih terstruktur sehingga siap diolah dengan algoritma tertentu.

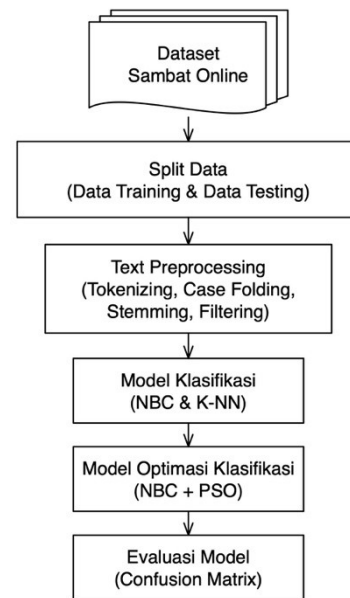
Proses klasifikasi dilakukan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas untuk kepentingan tertentu [3]. Metode klasifikasi yang populer dan banyak digunakan di antaranya adalah algoritma *Decision Tree*, *Naive Bayes Classifier* (NBC) dan *k-Nearest Neighbor* (k-NN) [4]. Beberapa kajian telah dilakukan untuk menyelesaikan kasus klasifikasi menggunakan algoritma ini. Nurajijah dan Riana [5] melakukan komparasi terhadap 3 algoritma sekaligus yaitu *decision tree*, NBC, dan SVM untuk menyelesaikan kasus klasifikasi persetujuan pembiayaan nasabah pada koperasi syariah. Handayani dan Pribadi [6] melakukan studi tentang klasifikasi teks menggunakan NBC pada layanan *call center* 110 dan memperoleh rata-rata akurasi sebesar 92,67 %. Algoritma k-NN telah digunakan dalam [7] untuk mengklasifikasikan teks pengaduan pada Sambat Online dengan memperoleh rata-rata akurasi sebesar 75,76 %.

Namun, algoritma yang digunakan dalam kajian tersebut di atas merupakan algoritma standar yang belum dioptimalkan untuk meningkatkan akurasi. Ada beberapa teknik yang dapat digunakan untuk meningkatkan hasil akurasi sebuah algoritma, salah satunya adalah *Particle Swarm Optimization* (PSO) dengan memodifikasikail beberapa parameternya [8]. Ketika dataset yang digunakan berskala kecil, maka algoritma sederhana yang dapat dipilih untuk dioptimasi adalah NBC.

Perpaduan algoritma NBC dengan PSO telah dilakukan dalam [9] untuk mengklasifikasi sel tunggal *pap smear* dalam deteksi awal keberadaan virus. Penelitian ini bertujuan untuk mengklasifikasikan teks pengaduan dengan memanfaatkan algoritma NBC yang dioptimalkan menggunakan PSO (NBC + PSO). Klasifikasi dilakukan pada 7 kategori sekaligus. Kedua metode standar (NBC dan k-NN) digunakan sebagai metode pembandingan untuk mengetahui kinerja algoritma NBC yang telah dioptimasi.

II. METODE PENELITIAN

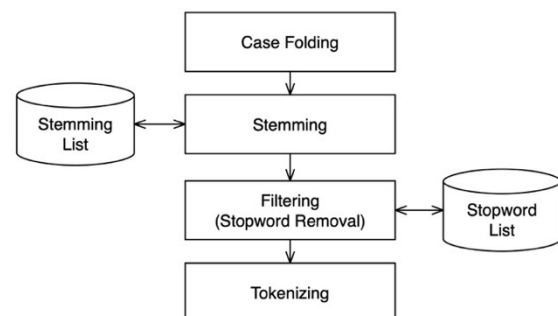
Kerangka penelitian pada Gambar 1 menjelaskan alur kerja dalam kajian ini. Secara garis besar, penelitian ini terdiri dari 3 tahap, yaitu *text preprocessing*, tahap klasifikasi, dan tahap optimasi untuk meningkatkan hasil klasifikasi. Dataset yang digunakan dalam penelitian ini berasal dari portal Sambat Online Kota Malang (www.sambat.malangkota.go.id) seperti halnya [7], yang dikumpulkan menggunakan metode *web scraping*. Dataset berupa teks yang dikonversi ke format *.xlsx*. Dataset terdiri dari 200 data dengan 7 kategori sebagai label yang mewakili OPD yang bertanggung jawab. Dataset dibagi menjadi 70 % sebagai data latih



Gambar 1. Kerangka penelitian

Tabel 1. Sebaran data pada dataset berdasarkan label kategori

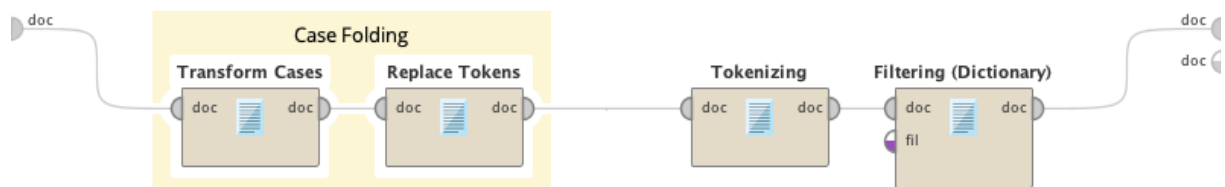
No	Kategori OPD	Jumlah Data
1	DISPENDUK	29
2	DLH	25
3	DPUPR	30
4	DISPENDIK	30
5	DISHUB	30
6	DPKP	28
7	SATPOL PP	28
Total Data		200



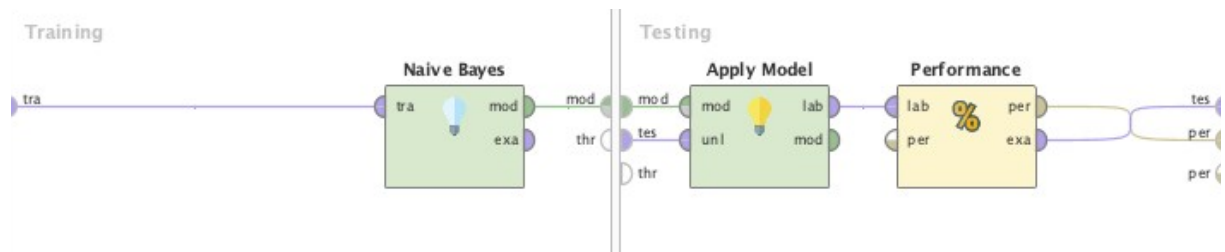
Gambar 2. Tahapan *text preprocessing*

dan 30 % sebagai data uji. Sebaran data pada dataset berdasarkan label kategori disajikan pada Tabel 1.

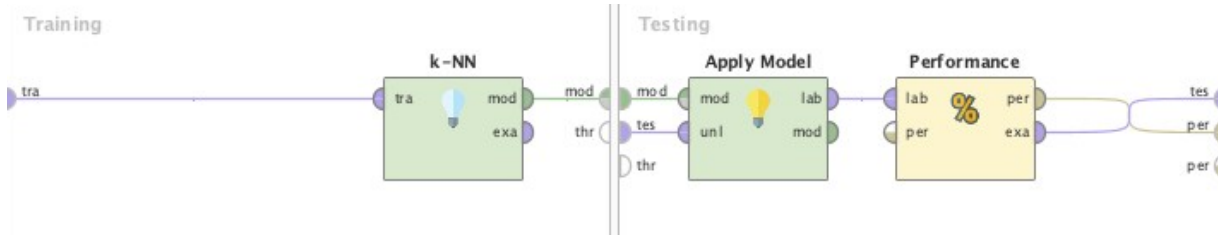
Dataset hasil *scraping* merupakan data mentah sehingga dibutuhkan tahap *text preprocessing* untuk menyiapkan dataset tersebut sebelum dilakukan proses klasifikasi. Tahapan *text preprocessing* yang dilakukan dalam penelitian ini meliputi *case folding*, *tokenizing*, *stemming*, dan *filtering* yang disajikan pada Gambar 2. *Case folding* mengubah bentuk teks pada dataset



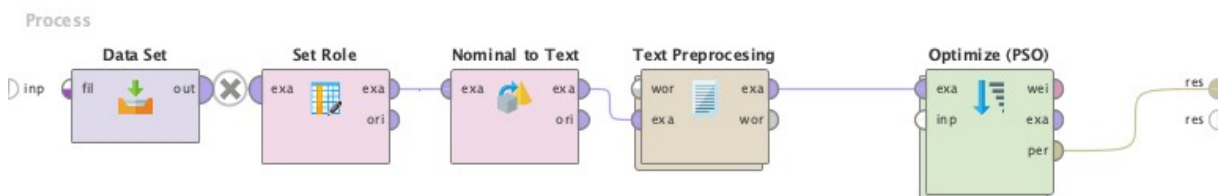
Gambar 3. Pemodelan text-preprocessing



Gambar 4. Pemodelan klasifikasi NBC



Gambar 5. Pemodelan klasifikasi k-NN



Gambar 6. Pemodelan optimasi klasifikasi NBC dan PSO

menjadi huruf kecil (*lowercase*) dan membuang karakter selain huruf a-z, termasuk angka kecuali spasi.

Setelah melewati *case folding*, dokumen teks pada dataset dipotong berdasarkan tiap kata penyusunnya dengan spasi sebagai delimiter (*tokenizing*). Dokumen yang telah terpotong menjadi kumpulan token kata dicari kata dasarnya untuk memperkecil indeks yang berbeda dari suatu dokumen pada dataset dengan teknik *stemming*. Tahap terakhir adalah *filtering*, yaitu dengan membuang token kata yang tidak sesuai untuk mengurangi indeks dan mempercepat waktu pemrosesan. Pemodelan *text-preprocessing* disajikan pada Gambar 3.

Data yang telah melewati tahap *text preprocessing* merupakan data yang telah bersih dan siap dilakukan proses klasifikasi. Model klasifikasi dibangun berdasarkan algoritma NBC dan k-NN. Dalam penelitian ini, algoritma NBC dipilih sebagai metode standar karena merupakan metode klasifikasi yang sederhana dan efisien. Selain itu, algoritma NBC memiliki performa yang sangat baik dalam beberapa

kasus klasifikasi teks [10]. Pemodelan klasifikasi algoritma NBC disajikan pada Gambar 4. Algoritma k-NN digunakan sebagai metode pembanding untuk metode standar NBC Nilai *k* pada algoritma k-NN bernilai 5. Pemodelan klasifikasi algoritma k-NN disajikan pada Gambar 5.

Hasil klasifikasi metode standar NBC dioptimasi menggunakan metode PSO. Parameter PSO yang digunakan untuk optimasi adalah ukuran populasi dan jumlah generasi maksimal. Hasil optimasi algoritma NBC dengan PSO dibandingkan kembali dengan metode pembanding k-NN. Pemodelan optimasi klasifikasi algoritma NBC dan PSO disajikan pada Gambar 6.

Teknik pengujian menggunakan metode *k-fold cross validation* yang cocok digunakan untuk mengolah data dengan jumlah tidak terlalu banyak. Prinsip kerjanya membagi data sebanyak *k* sub-himpunan, dimana *k* adalah nilai dari *fold*. Tiap sub-himpunan tersebut dijadikan data uji dari hasil klasifikasi yang dihasilkan dari *k-1* sub-himpunan lainnya. Kajian ini menggunakan

Tabel 2. Matriks konfusi klasifikasi teks dengan algoritma NBC

		Nilai Benar						Class Precision	
		DIS PENDUK	DPUPR	DLH	DIS PENDIK	DPKP	SAT POL		DIS HUB
Nilai Prediksi	DISPENDUK	20	2	3	0	2	0	2	68,97 %
	DPUPR	0	14	0	1	3	1	0	73,68 %
	DLH	1	0	8	1	2	0	1	61,54 %
	DISPENDIK	0	1	1	17	0	1	0	85 %
	DPKP	1	4	2	0	11	3	1	50 %
	SATPOL	2	4	3	1	2	20	4	55,56 %
	DISHUB	0	2	1	2	2	1	13	61,9 %
	Class Recall	83,33 %	51,85 %	44,44 %	77,27 %	50 %	76,92 %	61,9 %	

Tabel 3. Matriks konfusi klasifikasi teks dengan algoritma k-NN

		Nilai Benar						Class Precision	
		DIS PENDUK	DPUPR	DLH	DIS PENDIK	DPKP	SAT POL		DIS HUB
Nilai Prediksi	DISPENDUK	22	0	0	0	0	0	0	100 %
	DPUPR	0	21	0	1	2	1	2	77,78 %
	DLH	0	1	11	0	3	2	1	61,11 %
	DISPENDIK	0	0	1	19	2	0	2	79,17 %
	DPKP	1	2	3	0	14	1	2	60,87 %
	SATPOL	1	1	2	1	0	19	0	79,17 %
	DISHUB	0	2	1	1	1	3	14	63,64 %
	Class Recall	91,67 %	77,78 %	61,11 %	86,36 %	63,64 %	73,08 %	66,67 %	

nilai *fold* 10 sehingga dari 200 data dibagi menjadi 10 blok dengan jumlah pelatihan yang sama yaitu 20 *instance*. Setiap data menjadi data uji sekali dan menjadi data latih sebanyak 19 kali (*k*-1).

Evaluasi model pada penelitian ini digunakan untuk mengukur keakuratan algoritma dalam klasifikasi. Seluruh kinerja algoritma klasifikasi dievaluasi menggunakan matriks konfusi untuk mendapatkan nilai *akurasi*, *precision*, dan *recall*. Implementasi kerangka penelitian ini menggunakan perangkat lunak Rapidminer Studio 9.3.

III. HASIL DAN PEMBAHASAN

Metode standar dengan algoritma NBC menghasilkan rata-rata *class precision* sebesar 65,23 % dan rata-rata *class recall* sebesar 63,67 %, seperti ditunjukkan dalam Tabel 2. Nilai akurasi yang didapatkan algoritma NBC diperoleh sebesar 64,38 % ± 14,15 % dengan waktu komputasi 3 detik.

Tabel 3 menunjukkan matriks konfusi yang dihasilkan dengan algoritma k-NN sebagai metode pembandingan. Metode ini menghasilkan rata-rata *class precision* sebesar 74,54 % dan rata-rata *class recall* sebesar 74,33 %. Nilai akurasi yang didapatkan algoritma k-NN diperoleh sebesar 75 % ± 13,18 % dengan waktu komputasi 2 detik.

Perbandingan nilai akurasi dan perhitungan luas daerah di bawah kurva (AUC) dari algoritma NBC dan k-NN disajikan pada Tabel 4. Algoritma k-NN memberikan nilai yang lebih baik daripada NBC, yaitu dengan nilai akurasi sebesar 75 % dan AUC sebesar

Tabel 4. Hasil perbandingan algoritma NBC dan k-NN

Parameter	NBC	k-NN
Akurasi	64,38 % ± 14,15 %	75 % ± 13,18 %
AUC	0,582	0,707

0,707. Hasil ini tidak sejalan dengan [11] yang menyatakan bahwa NBC lebih unggul daripada k-NN dengan nilai akurasi NBC sebesar 70 % dibandingkan k-NN sebesar 40 %. Kajian tersebut membandingkan kinerja algoritma NBC dan k-NN untuk klasifikasi artikel jurnal berbahasa Indonesia dengan menggunakan dataset sebanyak 40 artikel dengan 4 kategori jurnal yang masing-masing kategori berisi 10 artikel. Metode NBC mendapatkan nilai akurasi yang baik karena keunggulan NBC sendiri, yaitu mampu melakukan klasifikasi meskipun memiliki data latih sedikit untuk estimasi parameternya, sedangkan k-NN tidak efektif melakukan klasifikasi apabila data latih yang digunakan sedikit.

Hasil klasifikasi NBC yang lebih rendah daripada k-NN (Tabel 4) sesuai dengan [12] yang menyatakan bahwa NBC memiliki kelemahan pada seleksi atribut sehingga dapat mempengaruhi nilai akurasi. Oleh karena itu, NBC perlu dioptimasi dengan cara memberikan bobot pada atribut agar dapat bekerja lebih efektif. Untuk menyelesaikan permasalahan tersebut, metode PSO dapat digunakan untuk melakukan pembobotan atribut untuk meningkatkan akurasi NBC.

Model yang diusulkan pada kajian ini menerapkan optimasi metode PSO ke metode NBC standar. Penerapan optimasi mengacu pada penentuan nilai

parameter besar populasi dan jumlah maksimum generasi yang tepat dengan memperhatikan waktu komputasi. Nilai akurasi yang optimal dari parameter tersebut membentuk struktur optimasi yang ideal untuk pemecahan masalah. Untuk mendapatkan nilai parameter terbaik, pengujian optimasi pertama dilakukan dengan memberi nilai pada parameter besar populasi dari 10-80 dan jumlah maksimum generasi bernilai konstan 100. Besar populasi menunjukkan jumlah individu pada setiap generasi, sedangkan jumlah maksimum generasi menunjukkan jumlah generasi maksimal untuk menghentikan jalannya algoritma.

Hasil dari pengujian pertama pada Rapidminer Studio dengan memberikan nilai parameter besar populasi disajikan pada Tabel 5. Nilai parameter besar populasi terbaik yang didapatkan adalah 70 dengan akurasi $86,73\% \pm 7,3\%$ dengan waktu komputasi 8 menit 45 detik. Waktu komputasi semakin meningkat saat nilai pada parameter besar populasi bertambah, walaupun nilai akurasi yang dihasilkan tidak selalu meningkat.

Pengujian kedua dilakukan dengan memberi nilai parameter besar populasi konstan 70 yang didapatkan dari nilai terbaik pada pengujian pertama dan memberi nilai parameter jumlah maksimum generasi dari 100-500. Hasil dari pengujian kedua disajikan pada Tabel 6. Pada pengujian optimasi kedua, akurasi tertinggi dan waktu komputasi terendah terjadi pada saat parameter jumlah maksimum generasi bernilai 300 dengan akurasi $87,44\% \pm 7,72\%$ dengan waktu komputasi 19 menit 29 detik. Akurasi cenderung menurun setelah diberikan nilai optimal pada parameter jumlah maksimum generasi sebesar 300, sedangkan waktu komputasi meningkat jika nilai pada parameter jumlah maksimum generasi ditambah. Matriks konfusi dari optimasi PSO dengan kombinasi kedua parameter terbaik disajikan pada Tabel 7.

Berdasarkan evaluasi dari proses pengujian optimasi pertama dan kedua menggunakan 10-fold cross validation dapat dinyatakan bahwa metode optimasi menggunakan PSO terbukti dapat meningkatkan hasil akurasi algoritma NBC dengan memberikan kombinasi nilai yang ideal pada parameter besar populasi dan jumlah maksimum generasi walaupun waktu komputasi berjalan lebih lama. Waktu komputasi berjalan sesuai

Tabel 5. Hasil pengujian optimasi NBC menggunakan PSO dengan parameter *population size*

Ukuran Populasi	Akurasi	AUC	Waktu (Menit)
10	76,6% ± 11,96%	0,718	1:33
20	78,33% ± 15,67%	0,74	2:43
30	82,18% ± 10,89%	0,787	3:58
40	83,78% ± 14,21%	0,805	4:57
50	81,09% ± 11,44%	0,714	6:47
60	85,13% ± 4,45%	0,82	7:29
70	86,73% ± 7,3%	0,84	8:45
80	81,47 ± 9,65%	0,777	9:47

Tabel 6. Hasil pengujian optimasi NBC menggunakan PSO dengan parameter *maximum number of generation*

Jumlah Generasi	Akurasi	AUC	Waktu (Menit)
100	86,73% ± 7,3%	0,84	8:45
200	86,67% ± 7,56%	0,84	16:53
300	87,44% ± 7,72%	0,849	19:29
400	86,73% ± 10,29%	0,841	26:37
500	86,6% ± 7,66%	0,838	25:22

dengan nilai yang diberikan pada parameter PSO. Hal ini sejalan dengan [10] yang menyatakan bahwa optimasi menggunakan PSO dapat meningkatkan hasil akurasi karena mampu mengurangi dimensi data yang sangat besar.

Perbandingan hasil matriks konfusi dengan metode standar (NBC & k-NN) dan metode optimasi terbaik (NBC + PSO) disajikan pada Tabel 8 berdasarkan Tabel 2, Tabel 3, dan Tabel 7. Algoritma NBC menggunakan PSO dapat meningkatkan akurasi klasifikasi teks sebesar 23,06 % dari akurasi 64,38 % yang menggunakan metode standar dengan algoritma NBC. Hasil optimasi NBC menggunakan PSO juga lebih baik jika dibandingkan metode pembandingan dengan algoritma k-NN.

Akurasi optimasi pada penelitian ini lebih baik jika dibandingkan dengan [7] yang menggunakan algoritma k-NN untuk mengklasifikasi teks pengaduan. Selaras dengan [5], algoritma NBC memberikan nilai akurasi paling rendah dibandingkan algoritma lain. Di sisi lain,

Tabel 7. Matriks konfusi klasifikasi teks menggunakan metode optimasi NBC dengan PSO dengan parameter terbaik

	Nilai Benar							
	DIS PENDUK	DPUPR	DLH	DIS PENDIK	DPKP	SAT POL	DIS HUB	Class Precision
Nilai Prediksi	DISPENDUK	24	0	0	1	0	0	96 %
	DPUPR	0	14	0	0	0	1	93,33 %
	DLH	0	0	10	0	0	0	100 %
	DISPENDIK	0	1	0	10	1	1	71,43 %
	DPKP	0	3	2	0	21	1	75 %
	SATPOL	0	0	0	0	0	8	100 %
	DISHUB	0	2	0	0	0	1	25
Class Recall	100 %	70 %	83,33 %	90,91 %	95,45 %	66,67 %	92,59 %	

Tabel 8. Hasil perbandingan algoritma kinerja NBC, k-NN dan NBC+PSO

Parameter	NBC	k-NN	NBC + PSO
Akurasi	64,38 %	75 %	87,44 %
Presisi	65,23 %	74,54 %	89,29 %
Recall	63,67 %	74,33 %	85,56 %

algoritma NBC pada [6] memperoleh akurasi lebih tinggi meskipun tidak dilakukan optimasi menggunakan PSO. Hal ini disebabkan karena jumlah data jumlah data uji mempengaruhi nilai akurasi yang diperoleh. Semakin banyak jumlah data uji, semakin besar nilai akurasi yang didapatkan. Selain itu, jumlah kategori juga mempengaruhi kinerja klasifikasi teks menggunakan algoritma NBC.

IV. KESIMPULAN

Optimasi PSO dapat diterapkan sebagai seleksi fitur yang dapat meningkatkan hasil akurasi algoritma NBC dalam melakukan klasifikasi teks dengan memberikan nilai akurasi terbaik dibandingkan NBC tanpa PSO dan k-NN. Model pada kajian ini dapat dipertimbangkan untuk diterapkan pada kasus klasifikasi teks di portal pengaduan online agar pengaduan dapat segera ditangani dan lebih terorganisir sehingga meningkatkan pelayanan *e-government* yang disediakan pemerintah untuk masyarakat.

DAFTAR PUSTAKA

- [1] E. A. Sosiawan, "Tantangan dan hambatan dalam implementasi e-government di Indonesia," *Seminar Nasional Informatika (SEMNASIF)*, vol. 1, no. 5, pp. 99-108, 2008.
- [2] R. Feldman and J. Sanger, *The text mining handbook*. Cambridge: Cambridge University Press, 2006.
- [3] S. Defiyanti, "Integrasi metode klasifikasi dan clustering dalam data mining," in *the 9th National Conference on Information Technology and Electrical Engineering*, Yogyakarta, Indonesia, Jul. 2017, pp. 39-44.
- [4] F. Gorunescu, *Data mining : concepts, models and techniques*. Berlin-Heidelberg: Springer-Verlag, 2011.
- [5] N. Nurajijah and D. Riana, "Algoritma naïve bayes, decision tree dan svm untuk klasifikasi persetujuan pembiayaan nasabah koperasi syariah," *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 2, pp. 77-82, 2019. doi: [10.14710/jtsiskom.7.2.2019.77-82](https://doi.org/10.14710/jtsiskom.7.2.2019.77-82)
- [6] F. Handayani and F. S. Pribadi, "Implementasi algoritma naïve bayes classifier dalam pengklasifikasian teks otomatis pengaduan dan pelaporan masyarakat melalui layanan call center 110," *Jurnal Teknik Elektro*, vol. 7, no. 1, pp. 19-24, 2015.
- [7] A. A. Prasanti, M. A. Fauzi, and M. T. Furqon, "Klasifikasi teks pengaduan pada sambat online menggunakan metode n-gram dan neighbor weighted k-nearest neighbor (NW-KNN)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, pp. 594-601, 2018.
- [8] A. Nurhadi, "Implementasi algoritma naïve bayes classifier berbasis particle swarm optimization (PSO) untuk klasifikasi konten berita digital bahasa Indonesia," *Speed - Sentra Penelitian Engineering dan Edukasi*, vol. 8, no. 3, pp. 48-56, 2016.
- [9] T. Hidayatulloh, A. Herliana, and T. Arifin, "Klasifikasi sel tunggal pap smear berdasarkan analisis fitur berbasis naïve bayes classifier dan particle swarm optimization," *Swabumi*, vol. 4, no. 2, pp. 186-193, 2016.
- [10] A. Taufik, "Optimasi particle swarm optimization sebagai seleksi fitur pada analisis sentimen review hotel berbahasa Indonesia menggunakan algoritma naïve bayes," *Jurnal Teknik Komputer*, vol. 3, no. 2, pp. 40-47, 2017.
- [11] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan kinerja metode naïve bayes dan k-nearest neighbor untuk klasifikasi artikel berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 4, pp. 427-434, 2018. doi: [10.25126/jtiik.201854773](https://doi.org/10.25126/jtiik.201854773)
- [12] H. Muhamad, C. A. Prasajo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, "Optimasi naïve bayes classifier menggunakan particle swarm optimization pada data iris," *Jurnal Teknologi Informasi Dan Ilmu Komputer*, vol. 4, no. 3, pp. 180-184, 2017. doi: [10.25126/jtiik.201743251](https://doi.org/10.25126/jtiik.201743251)