

Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes

Ratino*¹, Noor Hafidz², Sita Anggraeni³, Windu Gata⁴
^{1,2,4}Magister Ilmu Komputer STMIK Nusa mandiri Jakarta
³Teknik Komputer STMIK Nusa Mandiri Jakarta
e-mail: *¹14002319@nusamandiri.ac.id, ²14002298@nusamandiri.ac.id,
³sita.sia@nusamandiri.ac.id, ⁴windu@nusamandiri.ac.id

Abstrak

Informasi saat ini banyak disampaikan melalui media sosial. Salah satu media sosial yang saat ini banyak digunakan adalah instagram. Berbagai sentimen masyarakat disampaikan melalui komentar pada media sosial instagram terhadap informasi COVID-19. Maka dari itu perlu dilakukan sentimen analisis untuk mengetahui sentimen dari setiap komentar. Adapaun algoritma klasifikasi yang digunakan yaitu Naïve Bayes dengan hasil akurasi 78,02% dan AUC 0,714, sedangkan Support Vector Machine menghasilkan akurasi sebesar 80,23% dan AUC 0,904. Memiliki selisih akurasi 2,21%. Setelah di optimasi dengan operator Particle Swarm Optimization, algoritma Naïve Bayes (PSO) menghasilkan akurasi sebesar 79,07% dan AUC 0,729, sedangkan algoritma Support Vector Machine (PSO) menghasilkan akurasi sebesar 81,16% dan AUC 0,903. Memiliki selisih akurasi sebesar 2,09%. Hasil pengujian algoritma, Support Vector Machine berbasis PSO maupun tidak, selalu dapat menghasilkan akurasi yang lebih tinggi.

Kata kunci— Covid-19, Naïve Bayes, Support Vector Machine, Particle Swarm Optimization.

Abstract

Information is now widely conveyed through social media. One of the social media that is currently widely used is Instagram. Various public sentiments were conveyed through comments on social media Instagram about information COVID-19. Therefore it is necessary to do sentiment analysis to find out the sentiment of each comment. The classification algorithm used is Naïve Bayes with the results of accuracy of 78,02% and AUC 0,714, while the Support Vector Machine produces an accuracy of 80,23% and AUC 0,904. Have accuracy differences 2.21%. After optimization with Particle Swarm Optimization operator. Naïve Bayes (PSO) algorithm produces accuracy 79,07% and AUC 0,729, while the Support Vector Machine (PSO) algorithm produces an accuracy of 81,16% and AUC 0,903. Have accuracy differences 2,09%. Algorithm test results, support vector machine with PSO or without PSO, can always produce higher accuracy

Keywords— Covid-19, Naïve Bayes, Support Vector Machine, Particle Swarm Optimization.

1. PENDAHULUAN

Organisasi kesehatan Dunia menyatakan virus baru terus bermunculan yang merupakan masalah bagi kesehatan masyarakat. Pada Desember 2019 sebuah epidemi kasus pernafasan

rendah terdeteksi. *World Health Organization* (WHO) menyatakan penyakit tersebut disebabkan

oleh coronavirus (CoV) baru. CoV baru ini adalah coronavirus 2019 (COVID-19) [1]. Virus COVID-19 ini telah menyebar dengan cepat hampir diseluruh dunia. Meskipun penyebarannya cepat diseluruh dunia, namun karakteristik penyakit pernafasan akut 2019-nCoV sebagian masih belum jelas [2].

Informasi dan cara pencegahan virus ini pun sudah tersebar diberbagai media sosial. Media sosial yang saat ini banyak digunakan salah satunya adalah Instagram. Instagram merupakan media yang kerap digunakan sebagai sarana berbagi informasi tentang berbagai hal. Hal ini karena Instagram menjadi media yang populer saat ini [3]. Dengan berbagai pengguna dari berbagai dunia dari mulai anak-anak hingga orang dewasa turut serta dalam popularitas Instagram.

Analisis setimen adalah proses mengklasifikasi dokumen teks yang berupa opini berdasarkan sentimen untuk menentukan apakah suatu tanggapan tertentu bersifat positif, netral atau negatif [4]. Analisis sentimen juga merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat [5].

Algoritma yang digunakan dalam melakukan klasifikasi pada penelitian ini yaitu *Support Vector Machine* dan *Naïve Bayes*. Kelebihan dari Algoritma *Support Vector Machine* yaitu mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda [6]. Namun algoritma *Support Vector Machine* memiliki kekurangan yaitu pada masalah pemilihan fitur yang sesuai. *Particle Swarm Optimization* (PSO) ditambahkan untuk meningkatkan kinerjanya.

Pada Berbagai penelitian tentang analisis dokumen tekstual paling banyak adalah menggunakan metode *Naïve Bayes*. *Naïve bayes* merupakan algoritma yang populer karena kemudahan penggunaannya [7]. Algoritma *Naïve Bayes* telah terbukti bekerja dengan memuaskan dibanyak domain [8].

Penelitian terkait yang membahas tentang analisis sentimen dengan menggunakan algoritma-algoritma klasifikasi telah banyak dilakukan. Penelitian yang dilakukan oleh Rofiqoh dan kawan-kawan dengan topik tingkat kepuasan telekomunikasi seluler Indonesia, penelitian tersebut menggunakan algoritma *Support Vector Machine* dengan *Luxicon Based Feature* sebagai pembaharuan fiturnya [9].

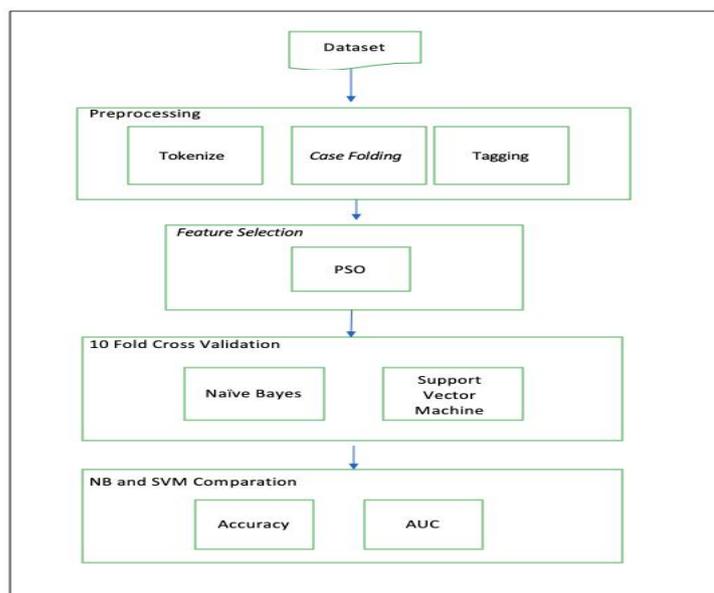
Penelitian lain yang dilakukan oleh Siswanto dan kawan-kawan dengan topik Analisis Klasifikasi komentar menggunakan *Support Vector Machine* dan *Naïve Bayes*. Penelitian tersebut menerapkan dua algoritma untuk menguji akurasi. Hasil akurasi *Support Vector Machine* sebesar 95,59% dan *Naïve Bayes* sebesar 93,00%. Penelitian tersebut menunjukkan bahwa *Support Vector Machine* memiliki akurasi yang lebih baik [10].

Penelitian lain yang dilakukan Buntoro dan kawan-kawan dengan judul Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter, penelitian tersebut menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* dengan menggunakan preprocessing tokenisasi, cleansing dan filtering [5].

Algoritma *Support Vector Machine* dengan menggunakan *Particle Swarm Optimization* pernah dilakukan oleh Wahyudi dan Kristiyanti dengan topik ulasan produk smartphone. Hasil akurasi *Support Vector Machine* adalah 82,00%. Sementara *Support Vector Machine* berbasis *Particle Swarm Optimization* (PSO) memperoleh tingkat akurasi 94,50% [11].

2. METODE PENELITIAN

Pada penelitian ini menggunakan metode *Support Vector Machine* dan *Naïve Bayes* dengan penambahan *Particle Swarm Optimization* (PSO) untuk menghasilkan akurasi terbaik dalam menganalisa komentar Instagram terhadap informasi berkaitan dengan COVID-19. Adapun kerangka berfikir pada penelitian ini seperti pada gambar berikut:



Gambar 1. Kerangka Berfikri

2.1. Pengumpulan Data

Data yang diperoleh merupakan sebuah data yang diambil dari komentar pengguna Instagram. Pengumpulan data komentar Instagram didapatkan melalui sebuah aplikasi Instagram Scraper.

2.2. Preprocessing

Preprocessing merupakan tahap awal yang diambil dalam proses pelatihan. Tujuan dari preprocessing yaitu menghindari dari data yang kurang sempurna, gangguan pada data dan data-data yang tidak konsisten [12].

2.3. Feature Selection

Setelah tahap preprocessing dan pelabelan, pada tahap selanjutnya adalah melakukan Feature Selection. Dalam tahap feature selection penelitian ini menggunakan algoritma Particle Swarm Optimization (PSO). Particle Swarm Optimization adalah metode optimisasi yang diperkenalkan oleh Kennedy dan Eberhart pada tahun 1995 yang dimotifasi oleh perilaku cerdas dari beberapa kawanan hewan [13][14].

2.4. Algoritma Support Vector Machine

Support Vector Machine adalah algoritma untuk melakukan prediksi dalam kasus klasifikasi maupun regresi [15]. Tingkat akurasi yang dihasilkan oleh SVM sangat bergantung terhadap pemilihan kernel dan parameter yang digunakan.

2.5. Algoritma Naïve Bayes

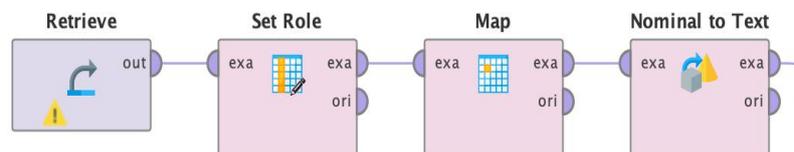
Naïve Bayes merupakan algoritma klasifikasi probabilistik berdasarkan teorema Bayes [8]. Naïve Bayes *Classification* memiliki akurasi yang tinggi dan kecepatan yang tinggi saat diaplikasikan ke dalam data yang besar.

3. HASIL DAN PEMBAHASAN

Pada bab ini akan menjelaskan proses yang dilakukan dalam penelitian ini. Adapun beberapa tahap yang dilakukan pada penelitian ini.

3.1. Pengumpulan Data

Dalam penelitian ini pengumpulan data mengambil dari Instagram pada akun *World Health Organization* (WHO) dengan postingan berkisar tentang covid-19. Hasil *Scrapping* data berkecenderungan file csv. Data tersebut kemudian dirubah menjadi file excel untuk mempermudah pengolahan datanya pada aplikasi RapidMiner.



Gambar 2. Proses pengolahan data pada RapidMiner

Proses *retrive* yang berfungsi untuk mengimpor data dengan beberapa kriteria penilaian terhadap sentimen, tahap selanjutnya yaitu *set role* untuk menentukan label dan atribut. Dan dilanjutkan ke tahap *map* untuk merubah label yang bernilai *true* menjadi positif dan *false* menjadi negatif. Lalu melakukan konversi nominal to string dengan menggunakan operator nominal to text. Setelah terkonversi tahap selanjutnya yaitu melakukan *preprocessing*.

3.2. Preprocessing

Pada tahap *preprocessing* adalah proses untuk menyiapkan sebuah data yang siap untuk dianalisa. Beberapa tahap *preprocessing* pada penelitian ini sebagai berikut.

a. Case Folding

Proses *case folding* adalah mengubah sebuah kalimat yang terdapat huruf besar menjadi huruf kecil. Hal ini dilakukan agar terdapat keseragaman huruf. Data pada penelitian ini berseragaman membentuk huruf kecil.

Tabel1. Proses Case Folding

Sebelum	Sesudah
STOP SMOKING NOW	stop smoking now

b. Tokenisasi

Tokenisasi adalah proses memecah sebuah kalimat menjadi sebuah kumpulan kata yang berdiri sendiri. Tokenisasi memecah teks yang semula kalimat menjadi beberapa kata. Tokenisasi menghilangkan delimeter seperti titik(.), Koma(,), spasi dan karakter angka pada kalimat tersebut.

Tabel 2. Proses Tokenisasi

Sebelum	Sesudah
Thank you W.H.O	thank you w h o

c. Tagging

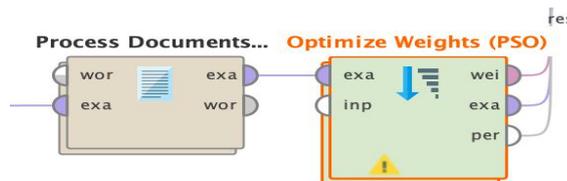
Menghilangkan nama pengguna yang disebutkan didalam komentar (@username), menghapus tanda hastag, menghapus karakter-karakter khusus seperti *emoticon*.

Tabel 3. Proses Tagging

Sebelum	Sesudah
Only WHO could mess this message up ♂	only who could mess this message up
@lchf_keto4lyfe @ichuzlife @_thereal_coke Follow	follow

3.3. Particle Swarm Optimization (PSO)

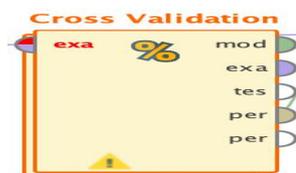
Setelah melalui preprocessing tahap selanjutnya adalah melakukan pengoptimasian, diantaranya dengan meningkatkan bobot attribut terhadap semua attribut yang digunakan, menyeleksi attribut dan *feature selection*. PSO merupakan algoritma yang baik dan efektif dalam permasalahan optimasi, hal ini dikarenakan memiliki performa yang konsisten.



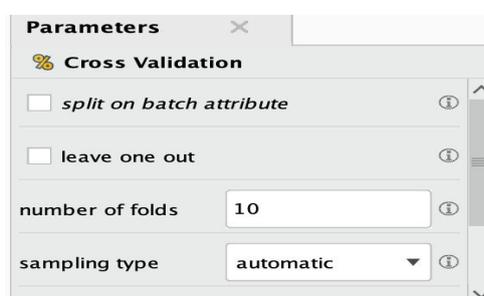
Gambar 4. Proses Optimasi

3.4. Proses Validasi

Pada proses validasi menggunakan operator *cross validation* dengan *k-10 fold cross validation*. Berikut operator yang digunakan pada aplikasi RapidMiner.

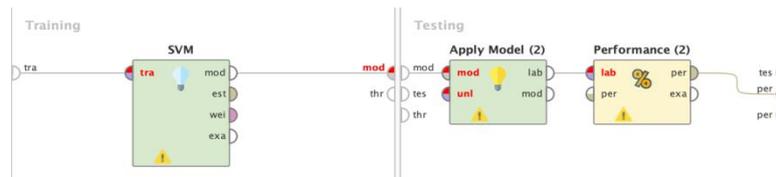


Gambar 5. Operator Cross Validation

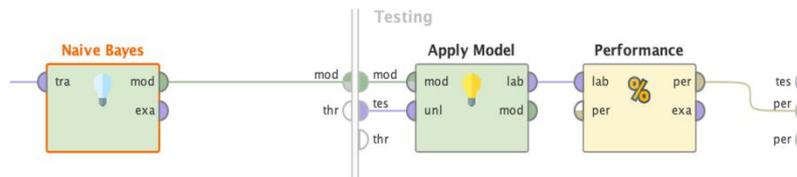


Gambar 6. Parameter Cross Validation

Tujuannya adalah untuk menentukan nilai fungsi dari model yang berhasil dilakukan ditahap sebelumnya. Didalam operator *cross validation* tersebut terdapat beberapa operasi diantaranya sebagai berikut:



Gambar 7. Operator Cross Validation dengan algoritma SVM



Gambar 8. Operator Cros Validation dengan algoritma NB

3.5. Evaluasi

Pada penelitian ini menggunakan algoritma *Support Vector Machine* dan *Naïve Bayes* dalam melakukan sentimen, serta melakukan evaluasinya menggunakan *accuracy* dan *AUC (Area Under Curve)*.

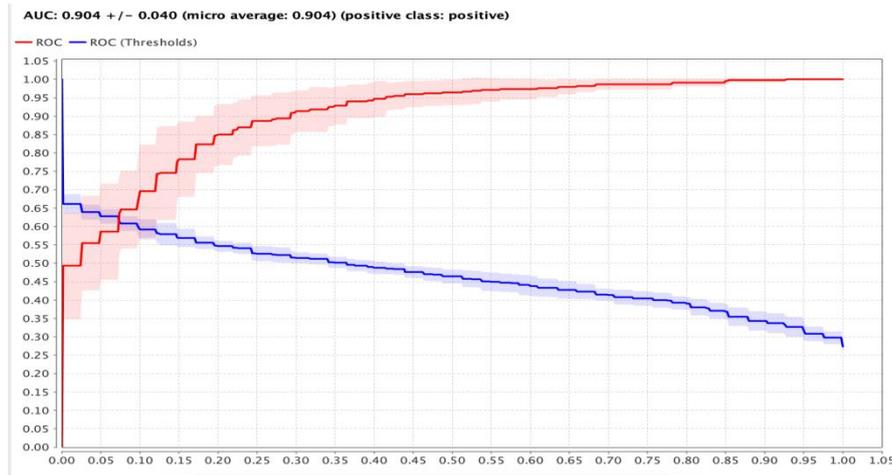
a. Support Vector Machine

Confusion matrik yang didapatkan dari hasil pemodelan dengan algoritma *Supprt Vector Machine*. Hasil akurasi yang didapat adalah 80,23%. Data negatif yang sesuai dengan prediksi negatif yaitu 268 data. Data Positif yang masuk prediksi negatif yaitu 30 data. Sedangkan data negatif yang masuk prediksi positif yaitu 140 data dan data positif yang sesuai prediksi positif yaitu 422 data.

Tabel 4. Hasil Akurasi Support Vector Machine

Accuracy : 80,23% +/- 5,05% (micro average : 80,23%			
	true negatif	true positif	class precision
prediksi negatif	268	30	89,93%
prediksi positif	140	422	75,09%
class recall	65,69%	93,36%	

Kurva ROC hasil pemodelan dengan *Support Vector Machine*. Mendapatkan hasil dengan grafik *Area Under Curve (AUC)* sebesar 0,904.



Gambar 9. Grafik AUC Support Vector Machine

b. *Naïve bayes*

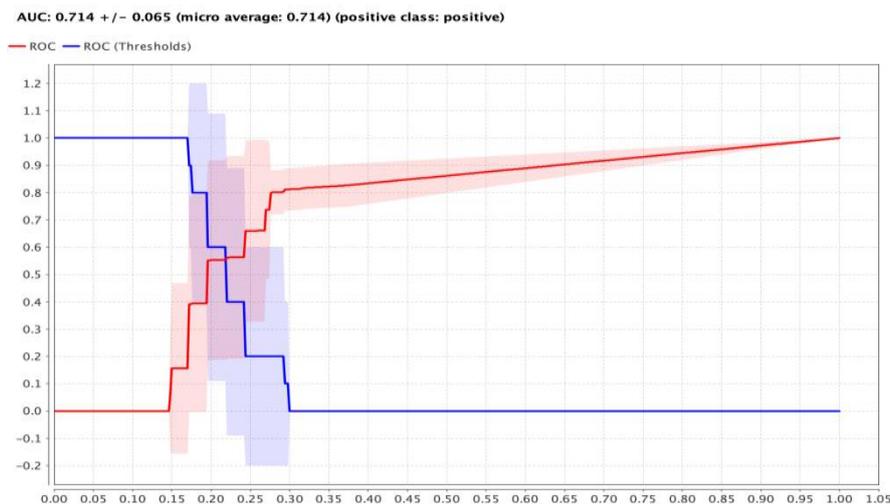
Confusion matrik yang didapatkan dari hasil pemodelan dengan algoritma *Naïve Bayes*. Hasil akurasi yang didapat adalah 78,02%. Data negatif yang sesuai dengan prediksi negatif yaitu 324 data. Data positif yang masuk prediksi negatif yaitu 105 data. Sedangkan data negatif yang masuk prediksi positif yaitu 84 data dan positif yang sesuai prediksi positif yaitu 347 data.

Tabel 4. Hasil akurasi *Naïve Bayes*

Accuracy : 78,02% +/- 5,66% (micro average : 78,02%)			
	true negatif	true positif	class precision
prediksi negatif	324	105	75,52%
prediksi positif	84	347	80,51%
class recall	79,41%	76,77%	

Kurva ROC hasil pemodelan dengan Support Vector Machine. Mendapatkan hasil dengan grafik Area Under Curve (AUC) sebesar 0,714.

Gambar 10. Grafik AUV *Naïve Bayes*



Sedangkan untuk akurasi dan AUC untuk algoritma *Support Vector Machine* dan *Naïve bayes* dengan tambahan *Particle Swarm Optimization* adalah sebagai berikut:

a. *Support Vector Machine* dengan *Feature Selection*

Confusion matrik yang didapatkan dari hasil pemodelan dengan algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization*. Hasil akurasi yang didapat adalah 81,16%. Data negatif yang sesuai dengan prediksi negatif yaitu 277 data. Data Positif yang masuk prediksi negatif yaitu 31 data. Sedangkan data negatif yang masuk prediksi positif yaitu 131 data dan data positif yang sesuai prediksi positif yaitu 421 data.

Tabel 5. Akurasi SVM dengan *Feature Selection*

Accuracy : 81,16% +/- 3,46% (micro average : 81,16%)			
	true false	true true	class precision
prediksi false	277	31	89,94%
prediksi true	131	421	76,27%
class recall	67,89%	93,14%	

Kurva ROC hasil pemodelan dengan *Support Vector Machine*. Mendapatkan hasil dengan grafik *Area Under Curve (AUC)* sebesar 0,903.



Gambar 11. Grafik AUC *Support Vector Machine* dengan *Feature Selection*

b. *Naïve Bayes* dengan *Feature Selection*

Confusion matrik yang didapatkan dari hasil pemodelan dengan algoritma *Naïve Bayes* berbasis *Particle Swarm Optimization*. Hasil akurasi yang didapat adalah 79,07%. Data negatif yang sesuai dengan prediksi negatif yaitu 324 data. Data positif yang masuk

prediksi negatif yaitu 97 data. Sedangkan data negatif yang masuk prediksi positif yaitu 83 data dan data positif yang sesuai prediksi positif yaitu 355 data.

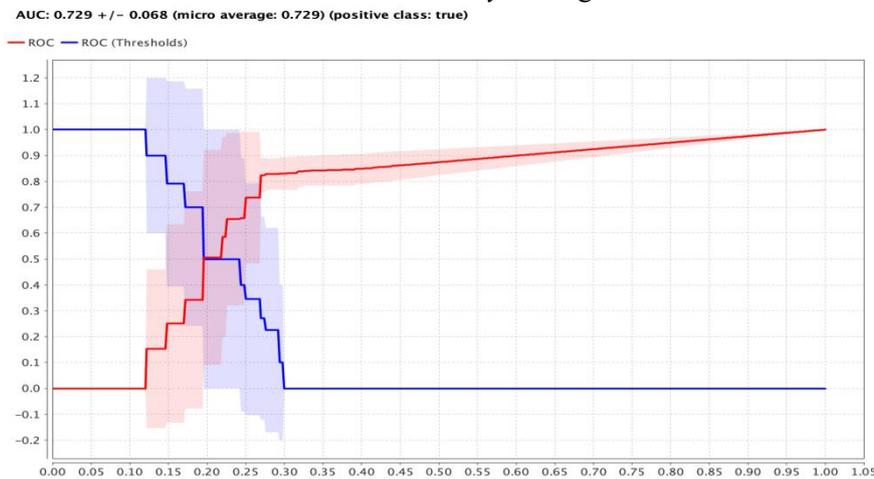
Ratino, Sentimen Analisis Informasi Covid-19 menggunakan Support Vector 9

Tabel 6. Akurasi *Naïve Bayes* dengan *Feature Selection*

Accuracy : 79,07% +/- 3,99% (micro average : 79,07%)			
	true false	true true	class precision
prediksi false	324	97	77,01%
prediksi true	83	355	81,05%
class recall	79,66%	78,54%	

Kurva ROC hasil pemodelan dengan *Naïve Bayes*. Mendapatkan hasil dengan grafik Area Under Curve (AUC) sebesar 0.729.

Gambar 12. Grafik AUC *Naïve Bayes* dengan *Feature Selection*



Berdasarkan hasil penelitian, ada perbedaan nilai akurasi maupun AUC ketika membandingkan algoritma yang menggunakan *particle Swarm Optimization* (PSO) dengan algoritma tanpa PSO. Dimana algoritma yang ditambahkan PSO ada peningkatan akurasi. Hasil pengujian dapat dilihat pada tabel berikut.

Tabel 7. Perbandingan akurasi dan AUC algoritma

	NB	SVM	NB + PSO	SVM + PSO
accuracy	78,02%	80,23%	79,07%	81,16%
AUC	0,714	0,904	0,729	0,903

Pada tabel 7 menunjukkan bahwa algoritma *Naïve Bayes* untuk data sejumlah 860 komentar Instagram, Menghasilkan akurasi sebesar 78,02% dan AUC 0,714. Sedangkan algoritma *Support Vector Machine* menghasilkan akurasi sebesar 80,23% dan AUC 0,904. Memiliki selisih akurasi sebesar 2,21%. Dan setelah dioptimasi dengan penambahan *Particle Swarm Optimization* untuk algoritma *Naïve Bayes* menghasilkan akurasi sebesar 79,07% dan AUC 0,729. Sedangkan untuk *Support Vector Machine* menghasilkan akurasi sebesar 81,16% dan AUC 0,903.

Sehingga didapatkan bahwa akurasi *Support Vector Machine* dengan penambahan *Particle Swarm Optimization* memiliki akurasi yang lebih tinggi dari algoritma yang lain.

4. KESIMPULAN

Dari hasil pembahasan dapat diambil beberapa kesimpulan sebagai berikut :

1. *Support Vector Machine* dan *Naïve Bayes* dapat digunakan dalam menganalisa sentimen pada kolom komentar Instagram.
2. Analisis sentimen terhadap data komentar Instagram mengenai Covid-19 yang dilakukan dengan *Support Vector Machine* dengan nilai akurasi 80,23% dan *Naïve Bayes* dengan nilai akurasi 78,02%. Dan hasil *Support Vector Machine* dengan *Particle Swarm Optimization* dengan nilai akurasi 81,16% dan *Naïve Bayes* dengan *Particle Swarm Optimization* dengan nilai akurasi 79,07%.
3. Hasil algoritma yang menggunakan *Particle Swarm Optimization* terbukti dapat meningkatkan nilai akurasi.

5. SARAN

Adapun saran untuk dapat dilakukan penelitian selanjutnya adalah:

1. Menggunakan algoritma K-NN atau Decision Tree untuk lebih banyak menguji algoritma.
2. Menggunakan metode *Feature Selection* yang lain, seperti Chi Square, Information Gain, Luxicon Based Feature, dan lain-lain agar dapat membandingkan hasilnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada redaksi jurnal JUPITER yang telah memberi kesempatan kepada penulis sehingga artikel ini dapat diterbitkan.

DAFTAR PUSTAKA

- [1] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, "Features, Evaluation and Treatment Coronavirus (COVID-19).", 2020, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/32150360>.
- [2] W. J. Guan *et al.*, "Clinical Characteristics of Coronavirus Disease 2019 in China," *N. Engl. J. Med.*, 2020, doi: 10.1056/NEJMoa2002032.
- [3] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, 2018.
- [4] S. Kurniawan, W. Gata, D. A. Puspitawati, N. -, M. Tabrani, and K. Novel, "Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, 2019, doi: 10.29207/resti.v3i2.935.
- [5] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter,"

Integer J. Maret, 2017.

- [6] N. Yunita, "Analisis Sentimen Berita Artis Dengan Menggunakan Algoritma Support Vector Machine dan Particle Swarm Optimization," *J. Sist. Inf. STMIK Antar Bangsa*, 2016.

Ratino, *Sentimen Analisis Informasi Covid-19 menggunakan Support Vector*

11

- [7] M. S. Hadna, P. I. Santosa, and W. W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, 2016.
- [8] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier," *Int. J. Inf. Eng. Electron. Bus.*, 2016, doi: 10.5815/ijieeb.2016.04.07.
- [9] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, 2017.
- [10] Siswanto, Y. P. Wibawa, W. Gata, G. Gata, and N. Kusumawardhani, "Classification Analysis of MotoGP Comments on Media Social Twitter Using Algorithm Support Vector Machine and Naive Bayes," in *Proceedings of ICAITI 2018 - 1st International Conference on Applied Information Technology and Innovation: Toward A New Paradigm for the Design of Assistive Technology in Smart Home Care*, 2018, doi: 10.1109/ICAITI.2018.8686751.
- [11] M. Wahyudi and D. A. Kristiyanti, "Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization," *Journal of Theoretical and Applied Information Technology*. 2016.
- [12] S. Gusriani, K. D. K. Wardhani, and M. I. Zul, "Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka)," *4th Appl. Bus. Eng. Conf.*, 2016.
- [13] M. R. Bonyadi and Z. Michalewicz, "Particle swarm optimization for single objective continuous space problems: A review," *Evolutionary Computation*. 2017, doi: 10.1162/EVCO_r_00180.
- [14] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: an overview," *Soft Comput.*, 2018, doi: 10.1007/s00500-016-2474-6.
- [15] N. Hafidz, Sfenrianto, Y. Pribadi, E. Fitri, and Ratino, "ANN and SVM Algorithm in Divorce Predictor," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 3, pp. 2523–2527, Feb. 2020, doi: 10.35940/ijeat.C5902.029320.

