



Optimasi Algoritma Support Vector Machine Berbasis PSO Dan Seleksi Fitur Information Gain Pada Analisis Sentimen

Sharazita Dyah Anggita¹, Ferian Fauzi Abdulloh²

^{1,2}Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta

¹sharazita@amikom.ac.id*, ²ferian@amikom.ac.id

Abstract

Sentiment analysis is a method for processing consumer reviews. This study examines the application of the Support Vector Machine (SVM) algorithm based on PSO and Information Gain as feature selection to filter attributes as a form of optimization. Algorithm implementation in sentiment analysis is carried out by applying a test scenario to measure the level of accuracy of the several parameters used. Selection of the Information Gain feature using the top-k parameter yields an accuracy value of 85.3%. Algorithm optimization applying information gain feature selection on the PSO-based SVM resulted in an optimal accuracy rate of 86.81%. The resulting increase in accuracy is 18.84% compared to the application of classic SVM without PSO-based information gain feature selection. Applying information gain feature selection on the PSO-based SVM algorithm can increase the accuracy value in the online sentiment review analysis.

Keywords: Information Gain, SVM, PSO

Abstrak

Analisis sentimen merupakan sebuah metode untuk mengolah review konsumen. Penelitian ini mengkaji penerapan algoritma Support Vector Machine (SVM) berbasis PSO dan Information Gain sebagai seleksi fitur untuk dapat memfilter atribut sebagai bentuk optimasi. Implementasi algoritma pada analisis sentimen dilakukan dengan menerapkan scenario pengujian untuk mengukur tingkat akurasi dari beberapa parameter yang digunakan. Seleksi fitur Information Gain menggunakan parameter top-k menghasilkan nilai akurasi 85.3%. Optimasi algoritma dengan penerapan seleksi fitur information gain pada SVM berbasis PSO menghasilkan tingkat akurasi optimal sebesar 86.81%. Peningkatan akurasi yang dihasilkan sebesar 18.84% dibandingkan penerapan SVM default tanpa seleksi fitur information gain berbasis PSO. Penerapan seleksi fitur information gain pada algoritma SVM berbasis PSO mampu meningkatkan nilai akurasi pada analisis sentimen review online.

Kata Kunci: Information Gain, SVM, PSO

1. Pendahuluan

Opini konsumen mengenai suatu produk merupakan salah satu parameter untuk mengembangkan sebuah bisnis. Ulasan yang diberikan oleh konsumen akan berbanding lurus pada sebuah produk dan akan berpengaruh dalam pemilihan produk tersebut kepada konsumen lain [1]. Minat konsumen dalam menulis sebuah *review* produk saat ini juga meningkat [2]. Penulisan opini banyak juga dilakukan melalui media social, Twitter menjadi salah satu media social yang menghasilkan lebih dari 4 miliar *tweet* tiap tahunnya. Berdasarkan data tersebut maka Twitter mampu dijadikan sebagai salah satu media pengambilan opini public yang layak [3]. *Review Online* yang diberikan pelanggan biasanya disertai dengan *rating* penilaian, namun kadang antara *review* dan skala penilaian yang diberikan berbanding terbalik. *Review* yang diberikan mengarah ke positif namun *rating* penilaian yang diberikan terbilang kurang baik [4].

Salah satu metode yang dapat digunakan untuk mengolah *review* konsumen yaitu analisis sentimen. Analisis sentimen akan mampu menentukan pola opini ke arah positif, negatif atau netral sehingga akan mampu mengimbangi peningkatan *review* konsumen yang ada saat ini [5]. Metode algoritma yang dapat digunakan untuk analisis sentimen salah satunya adalah *Support Vector Machine* (SVM).

Berdasarkan penelitian sebelumnya metode SVM tidak selalu menghasilkan nilai yang baik. Penelitian tentang komparasi Naive Bayes dan SVM menyatakan bahwa hasil yang didapatkan dari metode SVM yaitu masih terdapat banyak prediksi yang tidak tepat [6]. Penelitian lain melakukan pengujian metode SVM dengan beberapa parameter terbaik dan iterasi 200 kali menghasilkan nilai akurasi sebesar 90% [7].

Beberapa penelitian menjelaskan bahwa optimasi pada metode SVM dapat dilakukan untuk menghasilkan akurasi yang maksimal yaitu dengan metode *Particle*



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

Swarm Optimization (PSO). Optimasi menggunakan algoritma PSO mampu menghasilkan peningkatan akurasi dari 68.36% menjadi 75% [8]. Penelitian lain yang melakukan optimasi SVM menggunakan PSO menghasilkan tingkat akurasi sebesar 82.5% setelah sebelumnya menghasilkan akurasi senilai 80.14% [9].

Sebuah klasifikasi data mempunyai poin utama pada pemilihan atribut dan parameter yang tepat [10]. Performa sebuah algoritma klasifikasi dipengaruhi oleh banyak parameter, salah satunya adalah pemilihan fitur. Hambatan yang terdapat pada klasifikasi analisis sentimen berbasis *text* yaitu atribut yang diimplementasikan pada dataset mempunyai jumlah yang banyak, sehingga akan menghasilkan tingkat akurasi yang rendah [9]. Sebuah fitur yang dapat digunakan dalam sebuah klasifikasi adalah fitur yang memiliki relevansi yang maksimal dengan luaran yang didapatkan [11].

Pemilihan fitur dapat digunakan sebagai solusi dari permasalahan di atas dengan menggunakan *Information Gain*. Melalui *Information Gain* fitur yang dianggap kurang relevan akan dikurangi, dengan cara tersebut nantinya dimensi data juga akan berkurang dan dapat menghasilkan performa yang maksimal [9]. Beberapa penelitian sudah membuktikan bahwa penerapan *Information Gain* pada algoritma klasifikasi mampu meningkatkan nilai akurasi yang dihasilkan [12][13][14]. Optimasi algoritma SVM berbasis PSO tanpa menerapkan seleksi fitur menghasilkan nilai akurasi 80.3% [15].

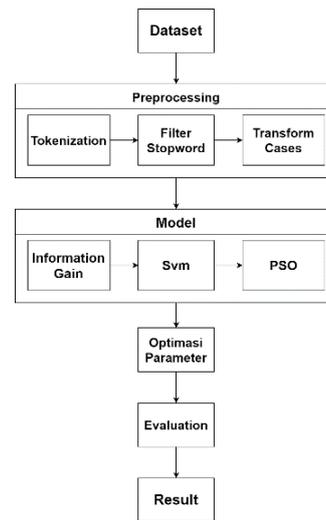
Berbeda dengan penelitian sebelumnya, penelitian ini akan melakukan seleksi fitur untuk dapat meminimalisir atribut pada dataset. Melalui penerapan seleksi fitur menggunakan algoritma *Information Gain* pada algoritma SVM berbasis PSO, tingkat akurasi akan dikaji untuk dapat melihat hasil yang paling maksimal.

2. Metode Penelitian

Metode penelitian yang diterapkan pada penelitian ini adalah metode eksperimen dengan tahapan, (1). Pengumpulan data dilakukan dengan *crawling* data opini *twitter* menggunakan *rapid miner studio*. Data yang digunakan dalam penelitian ini adalah data opini jasa ekspedisi yang bersifat teks dan berbahasa Inggris. (2). *Preprocessing* data dilakukan untuk dapat menghindari data tidak lengkap, *noise*, dan tidak konsisten [16]. Tahapan yang dilakukan pada *preprocessing* meliputi (a) *Tokenization* untuk memfilter tanda baca dan symbol yang bukan huruf (b) *Filter Stopword* merupakan eliminasi kata-kata yang sering ditampilkan dalam dataset (c) *Transform cases* untuk merubah data menjadi *lower case*. Data yang telah dikumpulkan akan dikelompokkan apakah opini tersebut mengarah ke positif atau negatif. (3). Metode *feature selection* menggunakan *Information Gain*. Melalui *Information Gain* akan dilakukan seleksi fitur yang dianggap kurang

relevan untuk dapat mengoptimasi algoritma klasifikasi yang digunakan, dalam hal ini adalah algoritma *Support Vector Machine*. (4). *Support Vector Machine* berbasis *Particle Swarm Optimization*, Klasifikasi sentimen dilakukan menggunakan algoritma *classifier Support Vector Machine*. Sebagai obyek optimasi diimplementasikan juga algoritma *Particle Swarm Optimization* yang akan melakukan optimasi global dengan generasi yang baru.

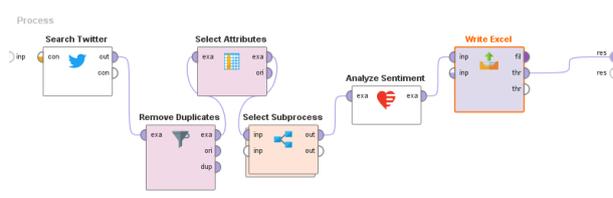
Diagram alur penelitian terdapat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

3. Hasil dan Pembahasan

Dataset yang akan digunakan pada implementasi model yang diusulkan didapatkan melalui proses *crawling data twitter* menggunakan *Rapid Miner*. Proses *crawling* data terdapat pada Gambar 2.

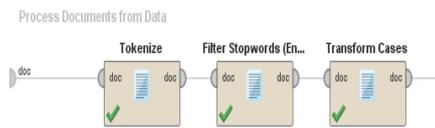


Gambar 2. Model Pengumpulan dan Pelabelan Data

Operator *search twitter* digunakan untuk menentukan banyak dataset yang akan diambil beserta *query* yang akan digunakan sebagai *keyword*. Dataset yang diambil sebanyak 500 data, dengan menggunakan *query* Fedex sebagai *keyword* pencarian data. Proses *filtering* data terdapat pada operator *remove duplicates* dan *select attributes* yang akan menghapus data duplikat dan membatasi akses data pada tipe data *text*. Proses selanjutnya adalah pelabelan data, dimana terdapat 2 label yaitu positif dan negatif pada operator *analyze sentiment*. Melalui keseluruhan proses, didapatkan hasil

akhir sebanyak 457 dataset yang terdiri dari 153 data berlabel negative dan 304 data berlabel positif.

Tahap awal penelitian diawali dengan proses *preprocessing data* melalui 3 tahapan pada Gambar 3 :



Gambar 3. Preprocessing Data

Tahapan *preprocessing* diawali dengan memotong tiap kata pada kalimat yang menyusunnya menggunakan operator *tokenize*. Tahap *filtering* data untuk dapat mengambil kata-kata yang diperlukan dari hasil *tokenize* dan yang terakhir adalah perubahan data menjadi *lower case* melalui operator *transform case*.

Penerapan model yang diusulkan dengan mencari tingkat akurasi performa algoritma *default Support Vector Machine*. Metode SVM diketahui menghasilkan prediksi *class* yang tidak tepat [4]. Salah satu factor penyebabnya adalah SVM memiliki kesulitan untuk melihat pemilihan parameter pada proses *training* [17]. Percobaan dilakukan dengan melakukan perbandingan nilai akurasi model *default* algoritma yang digunakan dengan model yang sudah dioptimasi [18] . Pada penelitian [19] evaluasi penerapan algoritma dilakukan menggunakan 10 *fold*. Evaluasi pengujian dilakukan dengan 2 tahap yaitu pada nilai *k-fold* 5 dan *k-fold* 10 sebagai perbandingan dengan hasil pada Tabel 1 dan Tabel 2.

Tabel 1. Hasil *Confusion Matrix SVM K-Fold 5*

Accuracy = 65.77% +/- 9,25% (mikro average : 65,79%)		
	True Positive	True Negative
pred Positive	152	77
pred Negative	79	148

Hasil pengujian algoritma SVM pada Tabel 1 dengan nilai *k-fold* = 5 menghasilkan nilai akurasi sebesar 65.79%. 152 data diklasifikasikan benar positif dan 148 data diklasifikasikan benar negative.

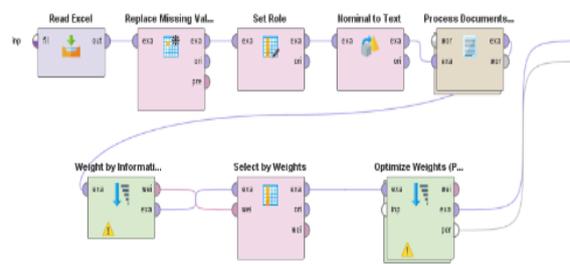
Tabel 2. Hasil *Confusion Matrix SVM K-Fold 10*

Accuracy = 67.97% +/- 6,99% (mikro average : 67,98%)		
	True Positive	True Negative
pred Positive	148	63
pred Negative	83	162

Tabel 2 menunjukkan hasil pengujian algoritma SVM pada nilai *k-fold* = 10 yang menghasilkan nilai akurasi sebesar 67,97%. 148 data diklasifikasikan benar positif dan 162 data diklasifikasikan benar negative. Jika dibandingkan dengan pengujian algoritma SVM pada nilai *k-fold*=5,

nilai akurasi yang dihasilkan hanya mengalami sedikit nilai kenaikan.

Pengujian algoritma SVM berbasis PSO dengan mengimplementasikan seleksi fitur *Information Gain* dirancang dengan desain pada Gambar 4.



Gambar 4. Pemodelan Penerapan *Information Gain* pada Algoritma SVM Berbasis PSO

Bentuk optimasi algoritma dilakukan dengan memanfaatkan parameter yang terdapat pada beberapa operator pada Gambar 4. Operator *weight by information* dan *select by weights* berfungsi sebagai operator seleksi fitur yang diterapkan sebelum implementasi algoritma SVM berbasis PSO berjalan. Pengujian seleksi fitur dilakukan dengan menggunakan parameter *weight relation top K* dengan 7 kali percobaan nilai K pada Gambar 5.



Gambar 5. Hasil Pengujian Seleksi Fitur

Melalui 15 kali pengujian dengan nilai K yang berbeda, dihasilkan tingkat akurasi tertinggi pada nilai top-13. Tingkat akurasi pada nilai top-13 sebesar 85.3%. Penggunaan nilai k terbukti mempengaruhi klasifikasi yang dilihat dari perbedaan tingkat akurasi yang dihasilkan. Hasil pengujian ini selanjutnya dijadikan dasar untuk mengimplementasikan algoritma SVM berbasis PSO guna melihat apakah penerapan seleksi fitur berpengaruh pada nilai akurasi yang dihasilkan dibandingkan dengan pemodelan klasifikasi menggunakan SVM *default*.

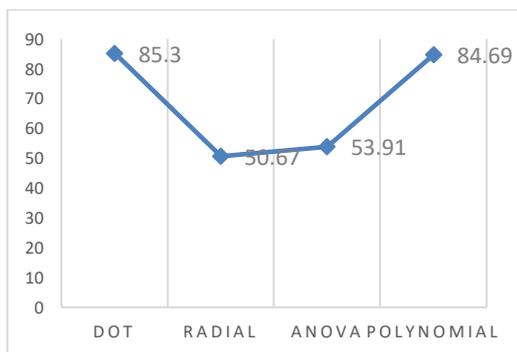
Pengujian algoritma SVM *Information Gain* berbasis PSO dilakukan dengan membandingkan nilai akurasi beberapa kernel yang terdapat didalamnya berbeda dengan penelitian lainnya [5][7][6] . Adapun penelitian [20] yang melakukan optimasi SVM menggunakan PSO dengan mencari akurasi terbaik dari kernel *linier*, *radial*, *polynomial* dan *sigmoid*. Kernel yang digunakan sebagai

perbandingan pada penelitian ini adalah kernel *dot*, *radial*, *anova* dan *polynomial*. Penelitian ini akan mencoba membuktikan apakah penggunaan kernel pada algoritma SVM *Information Gain* akan mempengaruhi tingkat akurasi. Hasil dari pengujian terdapat pada Table 3.

Tabel 3. Hasil Pengujian Kernel SVM

No	Kernel	Folds	Accuracy (%)
1	Dot	10	85.3
2	Dot	5	84.44
3	Radial	10	59.67
4	Radial	5	57.1
5	Anova	10	53.28
6	Anova	5	53.91
7	Polynomial	10	84.69
8	Polynomial	5	82.45

Melalui pengujian pada table 3, kernel *dot* menghasilkan nilai akurasi terbaik pada nilai *folds* = 10 dengan tingkat akurasi 85.3%. Kernel *radial* menghasilkan nilai akurasi terbaik pada nilai *folds*= 10 dengan tingkat akurasi 59.67%. Kernel *anova* menghasilkan nilai terbaik pada nilai *folds*=5 dengan tingkat akurasi 53.91%. Sedangkan kernel *polynomial* menghasilkan tingkat akurasi terbaik pada nilai *folds* = 10 dengan tingkat akurasi 84.69%. Berdasarkan hasil pengujian tiap kernel dihasilkan nilai akurasi terbaik dari setiap kernel pada Gambar 6.



Gambar 6. Hasil Pengujian Kernel Terbaik SVM

Melihat perbandingan nilai terbaik tiap kernel pada Gambar 6, ditarik kesimpulan bahwa penggunaan kernel pada algoritma SVM ternyata berpengaruh terhadap tingkat akurasi. Nilai akurasi tertinggi terdapat pada kernel *dot* menggunakan nilai *folds* = 10 yang menghasilkan nilai akurasi 85.3%. Jika dibandingkan dengan pengujian sebelumnya, tingkat akurasi tidak mengalami kenaikan. Selanjutnya model penerapan parameter dengan nilai yang terbaik ini akan digunakan sebagai dasar pengujian dengan membandingkan nilai akurasi dari 3 tipe *sampling* yaitu *linier*, *shuffled* dan *stratified*. Pengujian dilakukan dengan menggunakan kernel *dot* dengan nilai *folds* = 10. Hasil pengujian terdapat pada table 4.

Berdasarkan hasil pengujian *sampling type* algoritma SVM, kernel *linier* menghasilkan nilai akurasi 67.32%, *sampling type shuffled* menghasilkan nilai akurasi 85.72%

dan, *sampling type stratified* menghasilkan nilai akurasi 86.81%. Melalui 3 pengujian tersebut, nilai terbaik dihasilkan pada, *sampling type stratified* dengan tingkat akurasi 86.81% pada kernel *dot* dan nilai *folds* = 10. Tingkat akurasi mengalami kenaikan sebesar 1.51% dari pengujian sebelumnya.

Tabel 4. Hasil Pengujian *Sampling Type* SVM

Kernel	Folds	Sampling Type	Accuracy (%)
Dot	10	Linier	67.32
Dot	10	Shuffled	85.72
Dot	10	Stratified	86.81

Hasil dari pemodelan menggunakan kernel *dot*, nilai *folds* =10 dan *sampling type stratified* kemudian digunakan sebagai parameter dasar pengujian optimasi algoritma SVM berbasis PSO dengan implementasi seleksi fitur *information gain* yang berbeda dengan penelitian [6][7]. Optimasi dilakukan dengan melakukan pengujian pada parameter *validation*=10, *population size* dengan rentang 1-10, serta nilai *inertia weight* = 0.1. Tingkat akurasi yang berhasilkan dari 10 kali pengujian dalam pada Table 5.

Tabel 5. Hasil Pengujian Algoritma dengan Parameter *Population Size* dan *Inertia Weight*

Validation	Population Size (Q)	Inertia Weight	Accuracy (%)
10	1	0.1	85.75
10	2	0.1	86.81
10	3	0.1	85.29
10	4	0.1	85.31
10	5	0.1	85.53
10	6	0.1	86.4
10	7	0.1	85.09
10	8	0.1	85.31
10	9	0.1	86.19
10	10	0.1	85.96

Penerapan parameter *validation*, *population size* dan *inertia weight* pada pengujian ini berpengaruh pada tingkat akurasi yang dihasilkan, Tingkat akurasi terbaik pada table 7 terdapat pada parameter *validation* = 10, *population size* = 2 dan *inertia weight* = 0.1. Akurasi yang dihasilkan 86.81%.

Pengujian terakhir dilakukan untuk mengoptimasi model algoritma yang diusulkan adalah dengan menerapkan nilai *population size* = 2 yang merupakan nilai terbaik pada table 7. Selain itu parameter pengujian akan ditambahkan dengan parameter *max number* dengan rentang 10 sampai dengan 100 dengan hasil pengujian pada Table 6.

Tabel 6. Hasil Pengujian Algoritma dengan Parameter *Population Size*, *Inertia Weight* dan *Max Number*

Validat ion	Population Size (Q)	Inertia Weight	Max Number	Accuracy (%)
10	2	0.1	10	85.08
10	2	0.1	20	85.76
10	2	0.1	30	86.81
10	2	0.1	40	85.75
10	2	0.1	50	85.75
10	2	0.1	60	85.96

Validat ion	Population Size (Q)	Inertia Weight	Max Number	Accuracy (%)
10	2	0.1	70	86.18
10	2	0.1	80	85.75
10	2	0.1	90	85.96
10	2	0.1	100	86.18

Implementasi parameter *max number* pada algoritma SVM berbasis PSO dengan seleksi fitur *Information Gain* berpengaruh secara signifikan pada nilai akurasi yang dihasilkan. Hal ini terlihat pada hasil pengujian table 6, dimana dari 10 nilai parameter *max number*, tingkat akurasi yang dihasilkan berbeda-beda. Tingkat akurasi terbaik pada hasil pengujian di table 6 terdapat pada parameter *max number* = 30 dengan nilai *validation* = 10, *population size* = 2 serta *inertia weight* = 0.1. Akurasi yang dihasilkan dari penerapan model di atas sebesar 86.81%. Hasil akhir klasifikasi dari penerapan seleksi fitur *Information Gain* pada analisis sentiment menggunakan algoritma SVM berbasis PSO dijelaskan pada Tabel 7.

Tabel 7. *Confusion Matrix* Penerapan *Information Gain* pada Algoritma SVM berbasis PSO

Accuracy = 86.80% +/- 5.08% (mikro average : 86.81%)		
	True Positive	True Negative
pred Positive	180	12
pred Negative	51	213
class recall	77.92%	94.67%

Hasil klasifikasi final pada table 7 diperoleh 180 opini positif dan 213 opini negative. Sedangkan kesalahan klasifikasi yang seharusnya bernilai positif terdapat 12 opini dan opini yang seharusnya bernilai negative terdapat 51 opini.

Perbandingan nilai akurasi algoritma SVM *default* dengan penerapan seleksi fitur *information gain* pada algoritma SVM berbasis PSO dijelaskan pada Table 8.

Tabel 8. Perbandingan Akurasi Algoritma SVM

Algoritma	Accuracy (%)
SVM default	67.97
SVM berbasis PSO + Information Gain	86.81

Implementasi algoritma SVM *default* pada analisis sentiment menghasilkan nilai akurasi sebesar 65.79% pada nilai *folds* = 10. Optimasi algoritma dengan penerapan seleksi fitur *information gain* pada SVM berbasis PSO menghasilkan tingkat akurasi optimal sebesar 86.81%. Nilai tersebut dihasilkan dengan optimasi parameter nilai *folds*=10, kernel *dot*, *sampling type* = *stratified*, *population size* = 2, *inertia weight* = 0.1 dan *max number* = 30. Peningkatan akurasi yang dihasilkan sebesar 18.84% dibandingkan penerapan SVM *default* tanpa seleksi fitur *information gain* berbasis PSO

4. Kesimpulan

Penerapan algoritma SVM pada analisis sentiment *review online* menghasilkan nilai akurasi sebesar 67.97% pada nilai *folds* = 10. Optimasi algoritma SVM

dilakukan dengan menerapkan seleksi fitur *information gain* dan algoritma PSO di dalamnya. Melalui beberapa kali pengujian, penerapan model tersebut mampu menghasilkan peningkatan akurasi. Tingkat akurasi terbaik penerapan seleksi fitur *information gain* pada algoritma SVM berbasis PSO terdapat pada parameter nilai *folds*=10, kernel *dot*, *sampling type* = *stratified*, *population size* = 2, *inertia weight* = 0.1 dan *max number* = 30. Nilai akurasi yang dihasilkan sebesar 86.81%, dimana peningkatan akurasi terjadi sebesar 18.84% dari pengujian SVM *default*. Berdasarkan pemaparan di atas dapat disimpulkan bahwa penerapan seleksi fitur *information gain* pada algoritma SVM berbasis PSO mampu meningkatkan nilai akurasi pada analisis *sentiment review online*.

Daftar Rujukan

- [1] M. H. Azhar, P. P. Adikara, and Y. A. Sari, "Analisis Sentimen pada Ulasan Hotel dengan Fitur Score Representation dan Identifikasi Aspek pada Ulasan Menggunakan K-Modes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2777–2782, 2018.
- [2] D. A. Kristiyanti and M. Wahyudi, "Feature selection based on Genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review," *2017 5th Int. Conf. Cyber IT Serv. Manag. CITSM 2017*, 2017, doi: 10.1109/CITSM.2017.8089278.
- [3] I. Kurniawati and H. F. Pardede, "Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of SVM-Based Sentiment Analysis," *2018 Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2018 - Proc.*, pp. 1–5, 2018, doi: 10.1109/ICITSI.2018.8695953.
- [4] A. Supriyatna and W. P. Mustika, "Komparasi Algoritma Naive Bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kulit," no. 2, pp. 152–161, 2018.
- [5] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," vol. 2, no. 11, pp. 4704–4713, 2018.
- [6] A. Yudha, Y. Nuryaman, I. Nuddin, and A. Andhikawati, "Sentiment Analysis Pandangan Masyarakat Terhadap Tarif Tol Trans-Jawa Menggunakan Support Vector Machine dan Particle Swarm Optimization," pp. 13–22, 2019.
- [7] S. Antar, B. Vol, V. I. No, and E. Supriyadi, "Metode SVM Berbasis PSO untuk Meningkatkan Prediksi Ketepatan Waktu Kelulusan Mahasiswa," no. 2, pp. 113–120, 2017.
- [8] O. Somantri, D. Apriliani, J. T. Informatika, P. Harapan, and B. Tegal, "Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Support Vector Machine Based on Feature Selection for Sentiment Analysis Customer Satisfaction on Culinary," vol. 5, no. 5, pp. 537–548, 2018, doi: 10.25126/jtiik.20185867.
- [9] M. Hilman Aprilian Nurjaman and M. Syahrul Mubarak, "Analisis Sentimen Pada Ulasan Buku Berbahasa Inggris Menggunakan Information Gain Dan Support Vector Machine Sentiment Analysis on the English Book Reviews Using Information Gain and Support Vector Machine," vol. 4, no. 3, pp. 4900–4906, 2017.
- [10] M. Somvanshi, P. Chavan, S. Tambade, and S. V. Shinde, "A review of machine learning techniques using decision tree and support vector machine," *Proc. - 2nd Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2016*, 2017, doi: 10.1109/ICCUBEA.2016.7860040.
- [11] R. R. Setiawan et al., "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, 2017, doi: 10.1074/jbc.M209498200.
- [12] A. Novelty Octaviani Faomasi Daeli, "Sentiment Analysis on

- Movie Reviews Using Information Gain and K-Nearest Neighbor,” *J. Data Sci. Its Appl.*, vol. 3, no. 1, pp. 1–007, 2020, doi: 10.34818/JDSA.2020.3.22.
- [13] N. M. G. D. Purnamasari, M. Ali Fauzi, Indriati, and L. S. Dewi, “Cyberbullying identification in twitter using support vector machine and information gain based feature selection,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 3, pp. 1494–1500, 2020, doi: 10.11591/ijeecs.v18.i3.pp1494-1500.
- [14] S. Widya Sihwi, I. Prasetya Jati, and R. Anggrainingsih, “Twitter Sentiment Analysis of Movie Reviews Using Information Gain and Naïve Bayes Classifier,” *Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018*, pp. 190–195, 2018, doi: 10.1109/ISEMANTIC.2018.8549757.
- [15] Sharazita Dyah Anggita and Ikmah, “Algorithm Comparison of Naive Bayes and Support Vector Machine based on Particle Swarm Optimization in Sentiment Analysis of Freight Forwarding Services,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 2, pp. 362–369, 2020, doi: 10.29207/resti.v4i2.1840.
- [16] G. G. Gunadi, S. Rostianingsih, A. Gunawan, and J. S. Surabaya, “Prediksi Retensi Customer Berdasarkan Support Vector Machine dengan Preprocessing Menggunakan Hadoop,” pp. 1–4.
- [17] W. Zihao, L. Lan, X. Zongyi, and C. Guangtao, “The forecasting model of wheelset size based on PSO-SVM,” *Proc. 31st Chinese Control Decis. Conf. CCDC 2019*, pp. 2609–2613, 2019, doi: 10.1109/CCDC.2019.8832937.
- [18] F. R. Purwandari, “Sistem Informasi Penjualan Berbasis Web pada PT. Mustika Jati,” *Sist. Inf. Penjualan Berbas. Web Pada PT. Mustika Jati Farhan*, vol. 1, no. 4393, pp. 43–57, 2018, [Online]. Available: <http://research.kalbis.ac.id/Research/Files/Article/Full/ET722JFJEKQYRF2PKZC1UBQOU.pdf>.
- [19] Y. Setiowati and A. Helen, “Klasifikasi Analisis Sentimen Mengenai Hotel Di Yogyakarta,” *SCAN - J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, 2018, doi: 10.33005/scan.v13i1.1052.
- [20] Y. R. Nugraha, A. P. Wibawa, and I. A. E. Zaeni, “Particle Swarm Optimization-Support Vector Machine (PSO-SVM) Algorithm for Journal Rank Classification,” *Proc. - 2019 2nd Int. Conf. Comput. Informatics Eng. Artif. Intell. Roles Ind. Revolut. 4.0, IC2IE 2019*, pp. 69–73, 2019, doi: 10.1109/IC2IE47452.2019.8940822.