

Analisis Sentimen Tweet Tentang UU Cipta Kerja Menggunakan Algoritma SVM Berbasis PSO

Trifebi Shina Sabrila ^{(1)*}, Yufis Azhar ⁽²⁾, Christian Sri Kusuma Aditya ⁽³⁾
Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang, Malang
e-mail : trifebiss@webmail.umm.ac.id, {yufis,christiaskaditya}@umm.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 12 Juni 2021, direvisi 26 September 2021, diterima 21 Oktober 2021, dan dipublikasikan 25 Januari 2022.

Abstract

Support Vector Machine (SVM) is one of the most widely used classification algorithms for sentiment analysis and has been shown to provide satisfactory performance. However, despite its advantages, the SVM algorithm still has weaknesses in selecting the right SVM parameters to optimize the performance. In this study, sentiment analysis was done with the use of data called tweets about Undang-Undang Cipta Kerja which reap many pros and cons by the people in Indonesia, especially the laborers. The classification method used in this study is the Support Vector Machine algorithm which is optimized using the Particle Swarm Optimization method for the SVM parameters selection in the hope of optimizing the performance generated by the SVM algorithm in sentiment analysis. The results of the study using 10 k-fold cross-validations using the SVM algorithm resulted in an accuracy of 92,99%, a precision of 93,24%, and a recall of 93%. Meanwhile, the SVM and PSO algorithms produce an accuracy of 95%, precision of 95,08%, and recall of 94,97%. The results show that the Particle Swarm Optimization method can overcome the weaknesses of the Support Vector Machine algorithm in the problem of parameter selection and has succeeded in improving the resulting performance where the SVM-PSO is more superior to SVM without optimization in sentiment analysis.

Keywords: Sentiment Analysis, Twitter, UU Cipta Kerja, SVM, PSO

Abstrak

Support Vector Machine (SVM) tergolong sebagai satu di antara algoritma klasifikasi yang paling sering dimanfaatkan dalam analisis sentimen dan terbukti memberikan performa yang memuaskan. Akan tetapi, terlepas dari keunggulan yang diberikan, algoritma SVM masih mempunyai kelemahan dalam penentuan parameter SVM yang tepat untuk mengoptimalkan performa yang dihasilkan. Pada penelitian ini, dilakukan analisis sentimen pada tweet tentang Undang-Undang Cipta Kerja yang menuai banyak pro dan kontra oleh masyarakat di Indonesia terutama bagi para buruh. Metode klasifikasi yang diterapkan pada penelitian ialah algoritma Support Vector Machine yang dioptimasi menggunakan metode Particle Swarm Optimization sebagai seleksi pada parameter SVM dengan harapan dapat mengoptimalkan performa yang dihasilkan oleh algoritma Support Vector Machine dalam analisis sentimen. Hasil penelitian menggunakan 10 k-fold cross validation menggunakan algoritma SVM menghasilkan akurasi sebesar 92,99%, presisi sebesar 93,24%, dan recall sebesar 93%. Sementara pada algoritma SVM dan PSO menghasilkan akurasi sebesar 95%, presisi sebesar 95,08%, dan recall sebesar 94,97%. Hasil menunjukkan bahwa metode optimasi Particle Swarm Optimization dapat mengatasi kelemahan algoritma Support Vector Machine dalam masalah pemilihan parameter dan berhasil meningkatkan performa yang dihasilkan di mana SVM-PSO lebih unggul dibandingkan SVM biasa dalam analisis sentimen.

Kata Kunci: Analisis Sentimen, Twitter, UU Cipta Kerja, SVM, PSO

1. PENDAHULUAN

Pertumbuhan media sosial belakangan ini amat berpengaruh serta memberikan peran besar bagi beragam aspek kehidupan masyarakat di era digital modern ini. Twitter merupakan salah satu media sosial terpopuler karena kemudahan dalam penggunaan dan pengaksesannya. Hingga saat ini, diperkirakan Twitter telah mencapai 330 juta pengguna aktif setiap bulannya dan terus



meningkat setiap harinya (Rekha et al., 2019). Melalui Twitter, pengguna dapat melakukan berbagai macam aktivitas, komunikasi antar individu maupun kelompok, menuliskan kegiatan sehari-hari, mempromosikan dagangan, beradu argumen, hingga menyampaikan opini terkait suatu topik hanya dengan membuat unggahan berupa pesan status (biasa disebut *tweet*) yang dibatasi hingga 280 karakter. *Tweet* yang dituliskan oleh pengguna tersebut merupakan sumber yang bisa dimanfaatkan untuk menganalisis sentimen masyarakat umum terhadap isu yang dibahas karena tulisan tersebut berisikan sentimen yang dapat dimanfaatkan sebagai sumber evaluasi dan pertimbangan dalam menarik suatu keputusan ke depannya.

Pada awal Oktober 2020 lalu, pengesahan UU Cipta Kerja oleh DPR RI menjadi salah satu persoalan yang menuai banyak perbincangan di dunia nyata bahkan di media sosial terlebih Twitter. Pengesahan UU Cipta Kerja ini menuai berbagai reaksi pro dan kontra dalam masyarakat terutama para buruh. Masyarakat pun dengan bebas menyampaikan opininya baik berupa penolakan atau dukungan dalam menanggapi keputusan DPR RI dalam pengesahan UU ini melalui *tweet*. Hal ini menarik perhatian untuk dilakukannya analisis sentimen terhadap *tweet* opini masyarakat mengenai pengesahan UU Cipta Kerja untuk mengetahui kecenderungan respon masyarakat terkait isu tersebut.

Analisis sentimen adalah proses penggalian, pemahaman, dan pengekstraksian informasi dari data berupa teks yang digunakan untuk mengidentifikasi sentimen yang terdapat dalam sebuah kalimat opini secara otomatis (Hayatin et al., 2020; Iqbal et al., 2019) dengan menggunakan metode *Natural Language Processing* (NLP), statistik, atau *Machine Learning* (ML), yang selanjutnya akan diklasifikasikan ke dalam polaritas positif ataupun negatif (Sharma & Daniels, 2020).

Algoritma *Support Vector Machine* (SVM) ialah satu di antara algoritma klasifikasi yang paling banyak dimanfaatkan dalam klasifikasi teks termasuk analisis sentimen dan terbukti memberikan performa yang lebih unggul dibandingkan algoritma klasifikasi lainnya (Basari et al., 2013). Namun, terlepas dari baiknya performasi yang diberikan, algoritma *Support Vector Machine* (SVM) masih mempunyai kelemahan dalam penentuan parameter dan fitur yang tepat untuk mengoptimalkan performa yang dihasilkan (Kristiyanti & Wahyudi, 2017; Windha Mega & Haryoko, 2019). Oleh karena itu, diperlukan metode optimasi untuk melakukan pemilihan parameter SVM yang tepat sehingga dapat menangani kelemahan algoritma tersebut.

Pada beberapa penelitian sebelumnya, analisis sentimen terhadap data Twitter dilakukan oleh Godara & Kumar (2019) dengan membandingkan lima jenis algoritma klasifikasi, yaitu *K-Nearest Neighbor* (KNN), *Decision Tree*, *Artificial Neural Network* (ANN), *Naive Bayes*, dan *Support Vector Machine* (SVM) terhadap dataset Twitter. Hasil uji coba memperlihatkan algoritma SVM mengungguli metode lainnya dengan hasil akurasi 92,34%. Penelitian lain juga dilakukan oleh Tineges et al. (2020) tentang analisis sentimen pada *tweet* terkait layanan IndiHome dengan memanfaatkan algoritma SVM serta metode ekstraksi fitur TF-IDF dengan hasil akurasi sebesar 87%. Selain itu, Rustam et al. (2019) melakukan klasifikasi sentimen pada *tweet* dengan melakukan perbandingan terhadap tiga metode ekstraksi fitur, TF, TF-IDF, dan Word2Vec pada algoritma pembelajaran mesin. Hasil pengujian menunjukkan bahwa TF-IDF merupakan metode ekstraksi fitur yang terbaik dalam penelitian ini. Penelitian terkait analisis sentimen terhadap *review* jasa maskapai penerbangan dengan memanfaatkan algoritma SVM berbasis PSO dikerjakan oleh Risawati et al. (2020) mendapatkan akurasi sebesar 87,34% yang mana lebih unggul jika dibandingkan hanya dengan menggunakan algoritma SVM saja dengan akurasi sebesar 84,25%. Berikutnya, Arsi et al. (2021) juga melakukan penelitian dengan menggunakan algoritma SVM yang dioptimasi dengan metode PSO mengenai analisis sentimen wacana pindah Ibu Kota Indonesia dengan nilai akurasi yang diperoleh sebesar 81,15%, yakni lebih unggul 2,09% jika dibandingkan hanya menggunakan algoritma SVM saja yang hanya memperoleh akurasi sebesar 79,06%. Analisis sentimen terhadap opini masyarakat mengenai transportasi *online* di Twitter juga dilakukan oleh Que et al. (2020) dengan menerapkan algoritma SVM dan metode optimasi PSO yang menjadi rujukan utama dalam penelitian ini. PSO terbukti sanggup untuk mengatasi kelemahan SVM terhadap pemilihan parameter yang tepat dalam klasifikasi

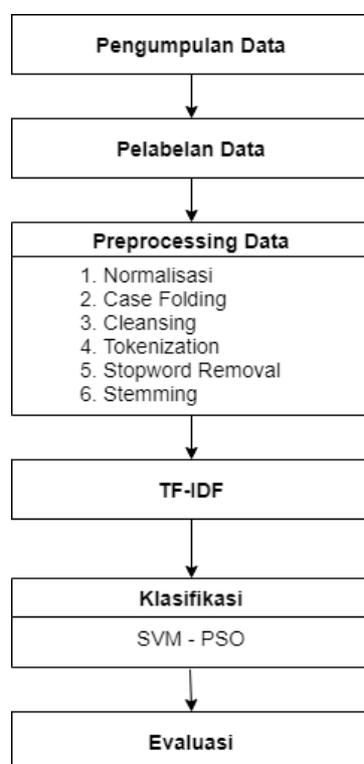


sentimen dengan meningkatkan hasil akurasi sebesar 96,04%, di mana akurasi tersebut unggul sebesar 0,58% jika dibandingkan dengan algoritma *Support Vector Machine* saja tanpa optimasi sebesar 95,46%.

Berdasarkan permasalahan yang telah dijabarkan sebelumnya, maka akan dilakukan penelitian terkait analisis sentimen terhadap *tweet* berbahasa Indonesia yang mengandung opini atau komentar masyarakat mengenai pengesahan UU Cipta Kerja oleh DPR RI dengan menerapkan algoritma *Support Vector Machine* (SVM) dan diintegrasikan dengan metode optimasi *Particle Swarm Optimization* (PSO) dengan harapan dapat mengoptimalkan performa yang dihasilkan oleh algoritma *Support Vector Machine* dalam proses analisis sentimen.

2. METODE PENELITIAN

Pada bagian ini, akan diuraikan langkah yang akan dilaksanakan pada penelitian. Secara garis besar alur penelitian ditampilkan pada Gambar 1.



Gambar 1 Alur Penelitian

2.1 Pengumpulan Data

Penelitian ini memanfaatkan data opini yang bersumber dari Twitter berupa *tweet* berbahasa Indonesia dengan topik UU Cipta Kerja. Proses pengumpulan data diproses dengan teknik *crawling* menggunakan 2 macam *tools*, yaitu Twitter API dan *snsrape* dengan total data sebanyak 1000 *tweet*.

2.2 Pelabelan Data

Data yang telah berhasil dikumpulkan kemudian akan diberikan label pada setiap datanya. Kategori yang digunakan dalam penelitian ini hanya terdiri atas label positif bagi data yang mengandung opini positif dan label negatif untuk data yang mengandung opini negatif. Proses pelabelan data ini dilakukan secara manual dengan bantuan 3 orang anotator yang terdiri atas 2 anotator primer dan 1 anotator sekunder.



2.3 Preprocessing Data

Preprocessing adalah salah satu tahap yang paling penting di dalam *text mining* (Alasadi & Bhaya, 2017) yang digunakan untuk melakukan perubahan data mentah menjadi data yang siap digunakan dan lebih terstruktur dengan pembersihan dan penyeragaman data karena data yang dikumpulkan biasanya masih berupa data kotor. Adapun tahap *preprocessing* yang diterapkan dalam penelitian ini adalah normalisasi, *case folding*, *cleansing*, *tokenization*, *stopword removal*, dan *stemming*.

2.3.1 Normalisasi

Normalisasi adalah proses yang bertujuan untuk memperbaiki kata yang memiliki kesalahan penulisan ataupun pengejaan serta kata yang dituliskan dengan disingkat.

2.3.2 Case Folding

Case folding adalah teknik guna mengganti huruf kapital yang ditemukan pada data sebagai huruf kecil seluruhnya.

2.3.3 Cleansing

Cleansing adalah proses yang bertujuan untuk menghilangkan berbagai informasi yang tidak dibutuhkan dalam proses analisis sentimen baik berupa *link* (http, https, pic.twitter), *hashtag*, *username* (dituliskan @username), maupun karakter spesial lainnya untuk memperoleh hasil analisis yang lebih baik.

2.3.4 Tokenization

Tokenization adalah proses pemecahan setiap kalimat yang terdapat pada data menjadi potongan-potongan kata. Caranya adalah dengan menjadikan spasi sebagai acuan untuk memisah setiap katanya.

2.3.5 Stopword Removal

Stopword ialah kata yang tidak bermakna penting, seperti kata “di”, “dan”, “dengan”, “oleh”, dan sebagainya. Sehingga, *stopword removal* ialah proses guna meniadakan kata yang tidak memiliki makna berharga terhadap data yang akan digunakan.

2.3.6 Stemming

Stemming adalah proses untuk meniadakan imbuhan yang didapati pada suatu kata menjadi sebuah kata dasar (Prasetyaningrum et al., 2020). Proses ini memiliki tujuan untuk mengurangi banyaknya varians dari suatu kata agar tidak menambah beban memori.

2.4 Term Frequency-Inverse Document Frequency (TF-IDF)

Setelah dilakukan tahap *preprocessing*, data akan masuk ke dalam proses *term weighting* untuk dilakukan proses pembobotan atau pemberian nilai pada setiap *term* yang terdapat pada data. Salah satu metode *term weighting* yang paling sering diterapkan dalam *text mining* dan akan dimanfaatkan pada penelitian ini adalah *Term Frequency-Inverse Document Frequency* (TF-IDF).

TF-IDF adalah teknik untuk memberikan bobot pada kata guna mengekstraksi setiap kata pada *dataset* yang menggabungkan konsep *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF), di mana TF ialah total kemunculan *term* di suatu dokumen, semakin tinggi total suatu *term* yang ada di suatu dokumen, maka akan semakin dianggap penting pula dokumen tersebut (Jo, 2019), sedangkan IDF memperlihatkan kebersangkutan dari ketersediaan suatu *term* pada keseluruhan dokumen, semakin minim frekuensi dokumen yang memuat *term* yang



diimplikasikan, nilai IDF pun akan semakin tinggi. Rumus TF-IDF yang digunakan dapat dilihat pada persamaan berikut (Jo, 2019):

$$W_{ij} = tf_{ij} \times idf_j = tf_{ij} \times \log \frac{N}{df_j} \quad (1)$$

Di mana W_{ij} adalah bobot *term* ke- j terhadap dokumen- i yang akan kita cari dengan tf_{ij} sebagai frekuensi kemunculan *term*- j dalam dokumen- i , N berupa total dokumen secara keseluruhan, dan df_j berupa total dokumen yang mengandung *term*- j .

2.5 Klasifikasi

Setelah seluruh tahapan sebelumnya selesai dilakukan, selanjutnya akan dilakukan proses klasifikasi sentimen terhadap data yang sudah disiapkan. Klasifikasi yang dilakukan pada penelitian dilakukan dengan dua macam proses klasifikasi, yang pertama adalah klasifikasi sentimen dengan metode SVM saja, dan yang kedua adalah klasifikasi sentimen dengan metode SVM yang diintegrasikan dengan metode PSO.

2.5.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu teknik klasifikasi yang tergolong sebagai *supervised learning* dan merupakan bagian dari pembelajaran mesin yang prosesnya mengikuti aturan *Structural Risk Minimization* (SRM) guna mendapatkan *hyperplane* paling optimal dengan margin optimumnya yang membagi setiap kelas yang ada di dalam ruang *input* (Wardhani et al., 2018). SVM juga merupakan metode yang mampu mengatasi permasalahan secara *linier* maupun *non-linier*. Pada dasarnya, ketika memprediksi suatu kelas pada data, SVM akan memberikan label berdasarkan daerah kelas mana yang ditempati oleh data tersebut (Rana & Singh, 2016). Letak awal *hyperplane* pada SVM ditentukan oleh nilai dari parameter SVM yang diinisialisasikan di awal seperti nilai C , γ , jenis kernel, dan sebagainya, sehingga pemilihan parameter SVM yang tepat sangatlah berpengaruh dengan performa yang akan dihasilkan.

2.5.2 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) ialah suatu metode optimasi yang berdasar pada populasi atas sekumpulan partikel dengan kecepatan dan posisi yang selalu diperbarui dalam setiap iterasinya (Kiranyaz et al., 2014). PSO biasanya banyak digunakan dalam pemecahan masalah optimasi dan masalah seleksi fitur (Liu et al., 2011). Setiap partikel pada PSO akan melacak posisi di dalam ruang pencarian serta solusi terbaiknya yang disebut dengan *personal best* ($pbest$) serta *global best* ($gbest$) yang dicapai oleh populasi dengan indeks partikelnya (Kiranyaz et al., 2014). Persamaan yang digunakan untuk *update* kecepatan dan posisi dalam PSO ialah seperti berikut:

$$V_i(t) = \theta V_{i(t-1)} + c_1 r_1 [Pbest_i - X_{i(t-1)}] + c_2 r_2 [Gbest - X_{i(t-1)}] \quad (2)$$

Di mana $V_i(t)$ adalah kecepatan ke t , θ adalah nilai inersia, $c_1 c_2$ adalah *learning rate*, $r_1 r_2$ adalah bilangan *random* dengan *range* $0 - 1$, dan $X_{i(t-1)}$ adalah posisi partikel sebelumnya. Setelah mendapatkan $V_i(t)$ sebagai kecepatan yang baru, selanjutnya posisi partikel yang baru akan dihitung dengan menggunakan persamaan berikut:

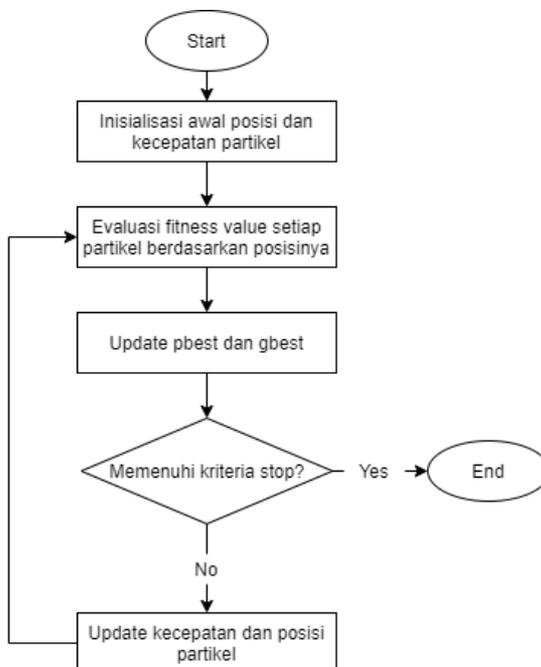
$$X_i(t) = V_i(t) + X_i(t - 1) \quad (3)$$

Di mana $X_i(t)$ adalah posisi partikel yang baru, $V_i(t)$ adalah kecepatan yang baru, dan $X_i(t - 1)$ adalah posisi partikel yang sebelumnya. Alur kerja PSO sendiri secara umum dapat dilihat pada diagram alur di Gambar 2.

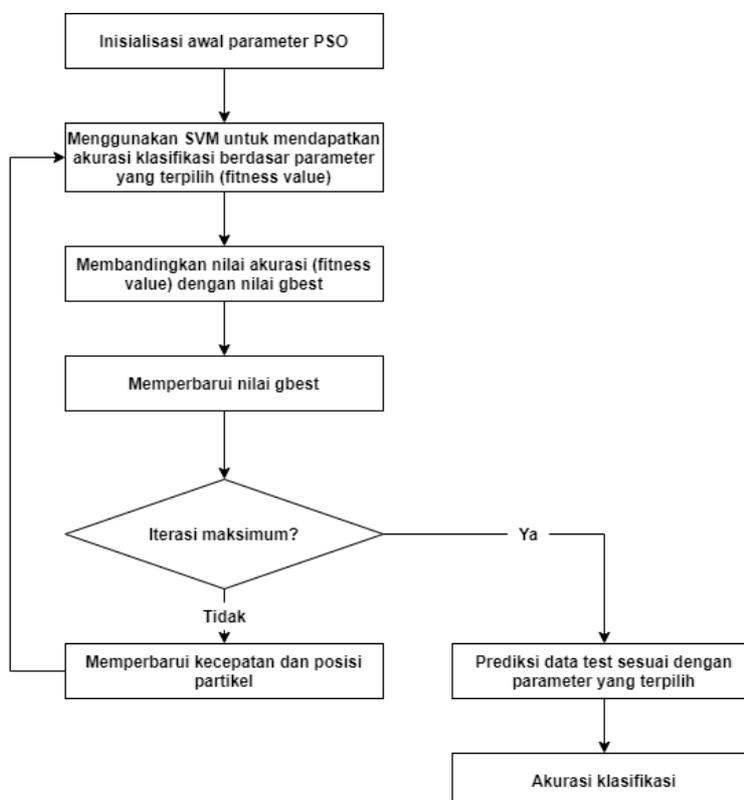
Sementara itu, untuk penerapan metode PSO sebagai seleksi parameter SVM pada klasifikasi menggunakan algoritma SVM pada penelitian ini, algoritma SVM akan mendapatkan hasil akurasi



klasifikasi berdasarkan nilai parameter SVM yang telah terpilih menggunakan metode PSO. Parameter SVM yang akan dioptimasi pada penelitian ini adalah nilai C dan nilai Gamma. Gambar 3 menunjukkan langkah-langkah untuk mengoptimalkan SVM menggunakan PSO dalam analisis sentimen (Hayatin et al., 2020).



Gambar 2 Alur Kerja Particle Swarm Optimization



Gambar 3 Alur Model Klasifikasi SVM-PSO



Pada tahap pertama, dilakukan inialisasi awal parameter PSO. Selanjutnya akan dilakukan evaluasi *fitness value* menggunakan algoritma SVM untuk mendapatkan akurasi klasifikasi berdasarkan parameter tiap partikel yang telah terpilih. Jika proses belum mencapai iterasi maksimumnya, maka kecepatan dan posisi tiap partikel akan terus diperbarui dengan menggunakan Persamaan (2) dan (3) hingga selesai diproses dan mencapai iterasi maksimum serta mencapai nilai akurasi yang terbaik untuk kemudian digunakan sebagai model untuk menguji data *test* sesuai dengan parameter yang terpilih.

2.6 Evaluasi

Hasil dari klasifikasi yang berhasil dilakukan berikutnya masuk ke dalam tahap evaluasi dengan tujuan untuk menguji dan mengetahui performa dari model yang telah diusulkan. Evaluasi yang dilaksanakan pada penelitian ini adalah melalui perhitungan nilai akurasi, presisi, serta *recall* sesuai dengan nilai yang terdapat pada *confusion matrix* berdasarkan pengujian model yang diproses dengan menerapkan *k-fold cross validation* dengan nilai k sebesar 10.

K-fold cross validation adalah metode guna melaksanakan evaluasi performa suatu model di mana data dipecah ke dalam dua bagian yaitu data *train* dan data *test*. *K-fold cross validation* dengan nilai k sebesar 10, data akan dipecah ke dalam 10 bagian data di mana 9 bagian akan berperan sebagai data latih dan satu bagian akan berperan sebagai data uji, proses ini dilakukan secara bergantian sebanyak 10 kali menggunakan model yang telah dibentuk.

Sedangkan *confusion matrix* sendiri adalah konsep dalam pembelajaran mesin yang mengandung informasi mengenai nilai yang sesungguhnya serta prediksi hasil klasifikasi yang sudah diproses oleh suatu sistem (Deng et al., 2016). *Confusion matrix* digunakan untuk mengevaluasi kinerja dari sistem yang telah dibangun. Contoh dari *confusion matrix* ditampilkan pada Gambar 4.

| | | Predicted Class | |
|--------------|----------|-----------------|----------|
| | | Positive | Negative |
| Actual Class | Positive | TP | FN |
| | Negative | FP | TN |

Gambar 4 Contoh Confusion Matrix

Terdapat istilah nilai hasil klasifikasi yang terdapat dalam *confusion matrix*, yaitu *True Positive* (TP) yang memberikan banyak data positif yang berhasil diprediksi secara benar, *True Negative* (TN) yang memberikan banyak data negatif yang diprediksi secara benar, *False Positive* (FP) yang memberikan banyak data positif namun diprediksikan salah, dan *False Negative* (FN) yang memberikan banyak data negatif namun diprediksikan salah (Al Rivan et al., 2020). Berdasarkan data dari *confusion matrix* tersebut, kinerja suatu sistem dapat diketahui dengan melakukan perhitungan terhadap nilai akurasi, presisi, dan *recall* dari sistem tersebut dengan menggunakan persamaan berikut (Al Rivan et al., 2020):

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%. \quad (4)$$

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (6)$$



3. HASIL DAN PEMBAHASAN

Pada bab ini akan dipaparkan hasil dari penelitian yang telah dilaksanakan. Tahap pengujian pada penelitian menerapkan bahasa pemrograman *python* untuk proses implementasinya.

3.1 Dataset

Dataset yang dimanfaatkan pada penelitian ini adalah data berupa *tweet* berbahasa Indonesia yang memuat topik mengenai UU Cipta Kerja. Pengumpulan data dikerjakan dengan proses *crawling* menggunakan dua macam *tools*, yaitu Twitter API dan *snsrape*. *Dataset* yang dikumpulkan berjumlah 1000 data. Proses pelabelan data dikerjakan secara manual oleh tiga orang anotor yang merupakan ahli bahasa Indonesia sesuai dengan bahasa pada data yang dikumpulkan, yang mana bertujuan agar interpretasi maksud atas setiap teks dapat masuk ke dalam kelas sentimen yang sesuai. Setelah dilaksanakan pelabelan data, sebanyak 500 data masuk ke dalam label positif dan 500 data masuk ke dalam label negatif.

3.2 Hasil Klasifikasi

Pengujian klasifikasi dalam penelitian ini dilakukan dalam empat skenario pengujian dengan menggunakan 10 *fold cross validation*. Di mana dalam skenario pengujian kesatu, data yang telah siap digunakan diproses untuk klasifikasi sentimen menggunakan algoritma *Logistic Regression* yang mana ialah satu di antara metode klasifikasi yang sering dipergunakan dalam klasifikasi teks sebagai pembanding di antara algoritma SVM dan SVM-PSO. Pada skenario pengujian kedua, dilakukan proses pengujian dengan menggunakan algoritma *Support Vector Machine* tanpa optimasi dengan menggunakan parameter *default* SVM. Pada skenario pengujian ketiga, dilakukan proses pengujian dengan menggunakan algoritma SVM dan metode optimasi PSO sebagai pemilihan parameter SVM dengan *setting* parameter untuk SVM dan PSO secara *default* seperti yang dilakukan pada penelitian sebelumnya. Kemudian pada skenario pengujian keempat, dilakukan proses pengujian dengan menggunakan algoritma SVM dan metode optimasi PSO sebagai pemilihan parameter SVM dengan iterasi sebanyak 50 kali.

Berdasarkan pengujian yang telah dilakukan pada keempat skenario, didapatkan perbandingan hasil perhitungan akurasi, rata-rata presisi, dan rata-rata *recall* yang ditampilkan pada Tabel 1.

Tabel 1 Perbandingan Hasil Evaluasi

| Skenario | Akurasi | Presisi | Recall |
|---|---------|---------|--------|
| Logistic Regression | 91,70% | 91,90% | 91,70% |
| SVM tanpa optimasi | 92,99% | 93,24% | 93% |
| SVM-PSO parameter <i>default</i> (Que et al., 2020) | 94,89% | 95% | 94,88% |
| SVM-PSO 50 iterasi | 95% | 95,08% | 94,97% |

Pada tabel hasil evaluasi klasifikasi sentimen empat skenario di atas diperoleh nilai akurasi, pada skenario pengujian pertama dengan menggunakan algoritma *Logistic Regression* sebesar 91,70%, presisi sebesar 91,90%, dan *recall* sebesar 91,70. Selanjutnya pada skenario pengujian kedua diperoleh nilai akurasi sebesar 92,99%, presisi sebesar 93,24%, dan *recall* sebesar 93% di mana pada skenario ini proses klasifikasi hanya dilakukan menggunakan algoritma SVM tanpa optimasi menggunakan metode PSO. Kemudian, diperoleh nilai akurasi pada skenario pengujian ketiga yakni menggunakan SVM-PSO dengan parameter *default* sebesar 94,89%, presisi sebesar 95%, dan *recall* sebesar 94,88%. Terakhir pada skenario keempat yakni menggunakan SVM-PSO dengan iterasi PSO 50 kali didapatkan akurasi sebesar 95%, presisi sebesar 95,08%, dan *recall* sebesar 94,97%.

Berdasarkan hasil yang diperoleh tersebut, dapat terlihat bahwa pengujian dengan menggunakan algoritma *Support Vector Machine* baik yang hanya menggunakan *Support Vector Machine* saja ataupun yang dioptimasi menggunakan metode *Particle Swarm Optimization* mendapatkan hasil yang lebih tinggi jika dibandingkan dengan pengujian menggunakan algoritma *Logistic*



Regression. Hal berikut membuktikan bahwa algoritma SVM memang lebih unggul ketika digunakan dalam klasifikasi teks. Selanjutnya, algoritma SVM yang telah dioptimasi dengan PSO dapat meningkatkan performa yang dihasilkan dalam seluruh parameter pengujian jika dibandingkan dengan algoritma SVM saja tanpa optimasi dengan kenaikan akurasi sebesar 1,9%, presisi sebesar 1,76%, dan *recall* sebesar 1,88% ketika parameter PSO di *setting default*, dan kenaikan akurasi sebesar 2,01%, presisi sebesar 1,84%, dan *recall* sebesar 1,97% ketika iterasi PSO di *setting* sebanyak 50 kali. Hal tersebut menunjukkan bahwa PSO benar dapat meningkatkan performa yang dihasilkan oleh algoritma SVM dengan melakukan optimasi serta mengatasi kelemahan dalam pemilihan parameter SVM-nya sesuai dengan yang dinyatakan pada sejumlah penelitian terdahulu.

4. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan terkait analisis sentimen pada *tweet* tentang UU Cipta Kerja, didapatkan hasil evaluasi pada skenario pengujian pertama klasifikasi menggunakan algoritma SVM saja tanpa optimasi dengan nilai akurasi 92,99%, nilai rata-rata presisi 93,24%, dan nilai rata-rata *recall* 93%. Sedangkan untuk klasifikasi menggunakan algoritma SVM yang diintegrasikan dengan PSO untuk pemilihan parameter SVM, hasil evaluasi pada skenario pengujian kedua mendapat hasil akurasi sebesar 95%, nilai rata-rata presisi sebesar 95,08%, dan nilai rata-rata *recall* sebesar 94,97%.

Dari proses klasifikasi tersebut dapat disimpulkan bahwa metode optimasi *Particle Swarm Optimization* dapat mengatasi kelemahan algoritma *Support Vector Machine* dalam masalah pemilihan parameter dan meningkatkan seluruh parameter pengujian baik nilai akurasi, presisi, maupun *recall* terhadap model yang dibangun.

DAFTAR PUSTAKA

- Al Rivian, M. E., Rachmat, N., & Ayustin, M. R. (2020). Klasifikasi Jenis Kacang-Kacangan Berdasarkan Tekstur Menggunakan Jaringan Syaraf Tiruan. *Jurnal Komputer Terapan*, 6(1), 89–98. <https://doi.org/10.35143/jkt.v6i1.3546>
- Alasadi, S. A., & Bhaya, W. S. (2017). Review of Data Preprocessing Techniques in Data Mining. In *Journal of Engineering and Applied Sciences* (Vol. 12, Issue 16, pp. 4102–4107). <https://doi.org/10.3923/jeasci.2017.4102.4107>
- Arsi, P., Wahyudi, R., & Waluyo, R. (2021). Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 231–237. <https://doi.org/10.29207/resti.v5i2.2698>
- Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, 53, 453–462. <https://doi.org/10.1016/j.proeng.2013.02.059>
- Deng, X., Liu, Q., Deng, Y., & Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*, 340–341, 250–261. <https://doi.org/10.1016/j.ins.2016.01.033>
- Godara, N., & Kumar, S. (2019). Opinion Mining using Machine Learning Techniques. *International Journal of Engineering and Advanced Technology*, 9(2), 4287–4292. <https://doi.org/10.35940/ijeat.B4108.129219>
- Hayatin, N., Marthasari, G. I., & Nuarini, L. (2020). Optimization of Sentiment Analysis for Indonesian Presidential Election using Naive Bayes and Particle Swarm Optimization. *Jurnal Online Informatika*, 5(1), 81–88. <https://doi.org/10.15575/join.v5i1.558>
- Iqbal, F., Hashmi, J. M., Fung, B. C. M., Batool, R., Khattak, A. M., Aleem, S., & Hung, P. C. K. (2019). A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction. *IEEE Access*, 7(c), 14637–14652. <https://doi.org/10.1109/ACCESS.2019.2892852>
- Jo, V. (2019). Introduction. *Seminars in Diagnostic Pathology*, 36(2), 83–84. <https://doi.org/10.1053/j.semmp.2019.02.002>
- Kiranyaz, S., Ince, T., & Gabbouj, M. (2014). *Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition*. Springer.



- Kristiyanti, D. A., & Wahyudi, M. (2017). Feature selection based on Genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review. *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, 1–6. <https://doi.org/10.1109/CITSM.2017.8089278>
- Liu, Y., Wang, G., Chen, H., Dong, H., Zhu, X., & Wang, S. (2011). An improved particle swarm optimization for feature selection. *Journal of Bionic Engineering*, 8(2), 191–200. [https://doi.org/10.1016/S1672-6529\(11\)60020-6](https://doi.org/10.1016/S1672-6529(11)60020-6)
- Prasetyaningrum, I., Fathoni, K., & Priyantoro, T. T. J. (2020). Application of recommendation system with AHP method and sentiment analysis. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 18(3), 1343. <https://doi.org/10.12928/telkomnika.v18i3.14778>
- Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 9(2), 162–170. <https://doi.org/10.22146/jnteti.v9i2.102>
- Rana, S., & Singh, A. (2016). Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques. *2016 2nd International Conference on Next Generation Computing Technologies (NGCT), October*, 106–111. <https://doi.org/10.1109/NGCT.2016.7877399>
- Rekha, V., Raksha, R., Patil, P., Swaras, N., & Rajat, G. L. (2019). Sentiment Analysis on Indian Government Schemes Using Twitter data. *2019 International Conference on Data Science and Communication (IconDSC)*, 1–5. <https://doi.org/10.1109/IconDSC.2019.8817036>
- Risawati, R., Ernawati, S., & Maryani, I. (2020). Optimasi Parameter PSO Berbasis SVM untuk Analisis Sentimen Review Jasa Maskapai Penerbangan Berbahasa Inggris. *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(2), 64–71. <https://doi.org/10.31294/evolusi.v8i2.9248>
- Rustam, F., Ashraf, I., Mehmood, A., Ullah, S., & Choi, G. S. (2019). Tweets Classification on the Base of Sentiments for US Airline Companies. *Entropy*, 21(11), 1078. <https://doi.org/10.3390/e21111078>
- Sharma, A., & Daniels, A. (2020). *Tweets Sentiment Analysis via Word Embeddings and Machine Learning Techniques*.
- Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(3), 650. <https://doi.org/10.30865/mib.v4i3.2181>
- Wardhani, N. K., Rezkiani, Kurniawan, S., Setiawan, H., Gata, G., Tohari, S., Gata, W., & Wahyudi, M. (2018). Sentiment analysis article news coordinator minister of maritime affairs using algorithm naive bayes and support vector machine with particle swarm optimization. *Journal of Theoretical and Applied Information Technology*, 96(24), 8365–8378.
- Windha Mega, P. D., & Haryoko. (2019). Optimization Of Parameter Support Vector Machine (SVM) using Genetic Algorithm to Review Go-Jek's Services. *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 6, 301–304. <https://doi.org/10.1109/ICITISEE48480.2019.9003894>

