



Implementasi Algoritma Support Vector Machine untuk Meingdentifikasi Komentar Negatif dalam Gambar di Media Sosial

Acep Razif Andriyan^{1,*}, Cecep Nurul Alam², Dian Sa'adillah³, Maylawati⁴, Mohamad Irfan⁵, Nur Lukman⁶

^{1,2,3,4,5,6}Universitas Islam Negeri Sunan Gunung Djati Bandung, Bandung, Indonesia

Informasi Artikel

Sejarah Artikel:

Dikirim : 28 Maret 2023

Revisi : 04 April 2023

Diterima : 17 Mei 2023

Diterbitkan: 30 Juni 2023

Kata Kunci

Social Media, OCR, SVM, RBF, Negative comments, Classification, memes.

Koresponden

E-mail: aceprazif33@gmail.com*

A B S T R A K

Sebanyak 191 juta orang sebagai pengguna aktif media sosial di Indonesia, dengan banyaknya pengguna sering kali menyampaikan pendapat atau berkomentar di media sosial yang bersifat positif maupun negatif seperti menghujat, membuly, mencaci dan lain sebagainya. Salah satu bentuk komentar tersebut disajikan melalui gambar (meme) yaitu gambar yang mengandung teks di dalamnya. Maka dari itu diperlukan sebuah sistem untuk mengklasifikasi dua jenis gambar yang bersifat positif dan negatif menggunakan metode algoritma SVM dengan kernel RBF dan teknologi OCR untuk pengambilan teks dalam gambar. Algoritma SVM berfungsi untuk melakukan klasifikasi dan teknologi OCR berfungsi untuk mengekstrak text yang berada pada sebuah gambar. Pengujian dilakukan dengan menggunakan split validation yang menghasilkan akurasi dari model terbaik dengan menggunakan perbandingan data 90:10 dan menghasilkan akurasi 85.7%.

Abstract

As many as 191 million people are active users of social media in Indonesia, with many users often expressing opinions or making comments on social media that are positive or negative, such as blaspheming, bullying, insulting and so on. One form of comment is presented through images (memes), namely images that contain text in them. Therefore, a system was created to classify two types of images, positive and negative, using the SVM algorithm method with RBF kernel and OCR technology for retrieving text in images. The SVM algorithm functions to carry out classification and OCR technology functions to extract text from an image. Testing was carried out using split validation which produced the accuracy of the best model using a data comparison of 90:10 and produced an accuracy of 85.7%.

This is an open access article under the CC-BY-SA license



1. Pendahuluan

Hasil penelitian dari UNESCO menyimpulkan bahwa 4 dari 10 orang Indonesia aktif di media sosial seperti Facebook yang memiliki 3,3 juta pengguna, WhatsApp dengan jumlah 2,9 juta pengguna, kemudian Twitter salah satu media sosial yang digunakan oleh masyarakat Indonesia. Country Industry Head Twitter Indonesia mengklaim bahwa Indonesia merupakan negara dengan pertumbuhan pengguna aktif harian Twitter [1]. Dengan banyaknya akses media sosial sering menimbulkan perbedaan cara berpendapat tidak sedikit juga ditemukan komentar atau pendapat negatif berupa hujatan, cacian, bullyan dan komentar negatif lainnya di media sosial dalam bentuk gambar (meme).

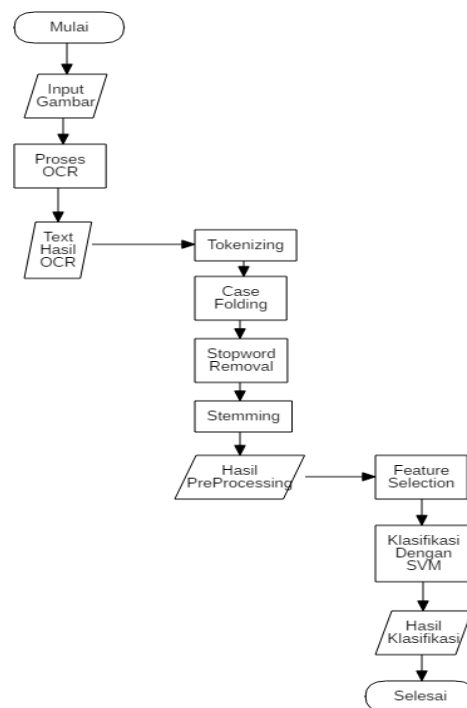
Tujuan dari penelitian kali ini adalah untuk mengkombinasikan algoritma SVM dan teknologi OCR dalam mengklasifikasi komentar negatif yang terkandung dalam sebuah gambar yang mengandung teks di dalamnya. Dengan perkembangan dari teknologi tentu saja dapat dimanfaatkan untuk mengklasifikasikan komentar negatif yang terkandung dalam sebuah gambar dengan melalui beberapa proses dalam OCR dan algoritma SVM. Optical Character Recognition (OCR) sebuah aplikasi

komputer yang digunakan untuk mengidentifikasi tulisan maupun angka dalam bentuk gambar yang dikonversi menjadi bentuk file teks. OCR merupakan cabang dari *Artificial Intelligence* (AI) dan *computer* [2].

Support Vector Machine (SVM) yaitu teknik dalam melakukan prediksi, baik untuk klasifikasi maupun regresi. SVM berada dalam satu kelas dengan *Artificial Neural Network* (ANN) dalam hal fungsi dan kondisi permasalahan yang mampu diselesaikan. Keduanya masuk dalam kelas supervised learning [3].

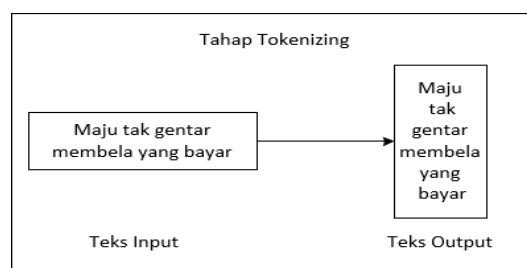
2. Metodologi Penelitian

2.1. Analisis Sistem



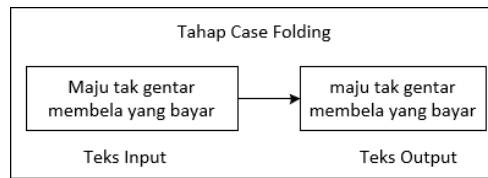
Gambar 1. Tahapan Sistem

Pada tahap awal, dokumen gambar di-*input* ke dalam sistem untuk di klasifikasi. Kemudian selanjutnya dari gambar tersebut akan melewati proses pengambilan kata kata yang ada di gambar tersebut dengan menggunakan *library* OCR (*Optical Character Recognition*). Setelah mendapatkan hasil dari OCR , selanjutnya akan melewati beberapa tahapan untuk *text preprocessing*, yaitu : *Tokenizing* adalah proses pemotongan urutan karakter menjadi potongan yang disebut token juga membuang tanda baca dan selain angka [4].Tahap *tokenizing* pada prinsipnya adalah pemisahan setiap kata yang menyusun suatu dokumen oleh karakter spasi.



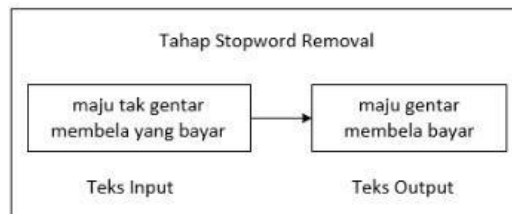
Gambar 2. Tahap Tekonizing

Case folding adalah proses perubahan semua huruf menjadi huruf kecil [4]. Fungsi *case folding* dibutuhkan dalam mengubah keseluruhan teks dalam dokumen yang digunakan menjadi huruf kecil (*lower case*).



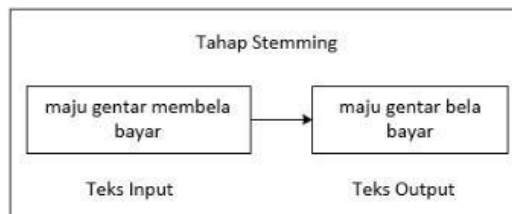
Gambar 3. Tahap Case Folding

Stopword Removal adalah proses penghapusan kata-kata yang tidak berguna yang tidak mempunyai arti [5]. *Stopword Removal* merupakan salah satu tahap untuk mendapatkan kata-kata penting dari proses sebelumnya.



Gambar 4. Tahap Stopword Removal

Stemming adalah teknik menemukan kata dasar dengan menghapus imbuhan [5].



Gambar 5. Tahap Stemming

Tahap selanjutnya adalah *feature selection* adalah ekstraksi data pada sebuah dokumen teks untuk mendapatkan fitur setiap kata yang muncul sebagai kriteria klasifikasi. Salah satu teknik ekstraksi fitur dalam sebuah teks dokumen adalah *Bag-of-words* dimana fitur ini terdiri dari individu atau kelompok kata yang ditemukan dalam teks[4].

Tabel 1. Feature Selection

Teks	Saya	pulang	ke	rumah	dia	bagus
Saya pulang ke rumah	1	1	1	1	0	0
Rumah dia bagus	0	0	0	1	1	1

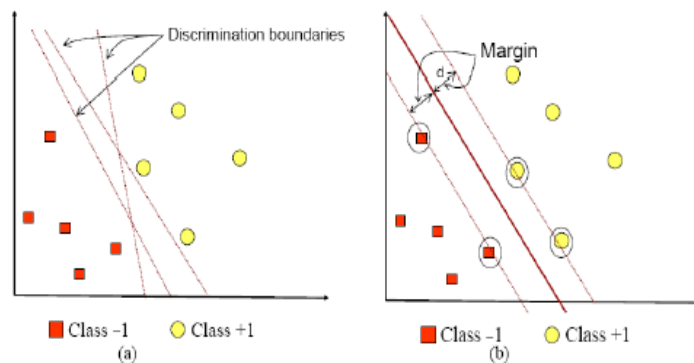
Terakhir, yaitu tahap klasifikasi. Hasil dari keluaran nantinya akan membedakan komentar positif dan komentar negatif. Pada penelitian ini, proses klasifikasi menggunakan algoritma SVM dengan fungsi *kernel Radial Basis Function*. *Support Vector Machine* (SVM) adalah teknik untuk melakukan prediksi, baik untuk klasifikasi maupun regresi. SVM berada dalam satu kelas dengan *Artificial Neural Network* (ANN) dalam hal fungsi dan kondisi permasalahan yang bisa diselesaikan. Keduanya masuk dalam kelas *supervised learning* [6] Untuk menyelesaikan masalah *non-linear*, SVM dimodifikasi dengan memasukkan fungsi *kernel*. Berbeda dengan metode klasifikasi pada umumnya yang justru mengurangi dimensi awal untuk menyederhanakan proses komputasi dan memberikan akurasi prediksi yang lebih baik.

2.2. Penggunaan Dataset

Penelitian ini menggunakan *datasets* yang dibuat secara manual yang diperoleh dari sosial media *twitter* dengan jumlah 280 komentar, dimana 168 komentar positif dan 112 komentar negatif. Data dibagi menjadi 90:10 untuk data latih dan data uji.

2.3. Algoritma SVM

Metode *Support Vector Machine* (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian konsep unggulan dalam bidang *pattern recognition*. SVM memiliki tujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah class pada *input space* [7].



Gambar 6. Metode Support Vector Machine

Support Vector Machine (SVM) yaitu teknik dalam melakukan prediksi, baik untuk klasifikasi maupun regresi. SVM berada dalam satu kelas dengan *Artificial Neural Network* (ANN) dalam hal fungsi dan kondisi permasalahan yang mampu diselesaikan. Keduanya masuk dalam kelas *supervised learning* [7].

SVM ini dimodifikasi dengan memasukkan fungsi *kernel* guna menyelesaikan masalah *non-linear*. *Kernel* dapat didefinisikan sebagai fungsi yang memetakan fitur data dari dimensi awal (rendah) ke fitur yang lebih tinggi. Pendekatan ini berbeda dengan metode klasifikasi pada umumnya yang justru mengurangi dimensi awal untuk menyederhanakan proses komputasi dan memberikan akurasi prediksi yang lebih baik.

Terdapat beberapa jenis *kernel* pada metode SVM, diantaranya *kernel Linear*, *Polynomial*, *Sigmoid*, *Radial Basis Function* (RBF), dan lainnya. Pada penelitian ini akan menggunakan *kernel* RBF untuk menyelesaikan permasalahan identifikasi komentar negatif.

Kernel RBF memiliki fungsi yang nilainya tergantung pada jarak asal atau dari beberapa titik. *Kernel* RBF memiliki rumus yang ditunjukkan pada persamaan 2.2 seperti berikut :

$$\begin{aligned} K(X_1, X_2) &= \text{exponent}(-\gamma (X_1 - X_2)^2), \text{ misalkan } \gamma = 1 \text{ maka:} \\ &= \text{exponent}(-1 (2.5 - 4)^2) \\ &= \text{exponent}(-1 (-1.5)^2) \\ &= \text{exponent}(-2.25) = 0.11 \end{aligned}$$

Kernel RBF mengacu pada dua pengukuran dosis yang berbeda. Selisih antara pengukuran kemudian dikuadratkan, memberi jarak antara dua pengamatan. Jadi, besarnya pengaruh satu pengamatan terhadap pengamatan lainnya adalah fungsi dari jarak kuadrat.

$$\begin{aligned} K(X_1, X_2) &= \text{exponent}(-\gamma (X_1 - X_2)^2), \text{ misalkan } \gamma \\ &= 2 \text{ maka:} \\ &= \text{exponent}(-2 (2.5 - 4)^2) \end{aligned}$$

$$\begin{aligned} &= \text{exponent}(-2 \cdot (-1.5)^2) \\ &= \text{exponent}(-2(2.25)) = 0.01 \end{aligned}$$

Hasil ini lebih kecil dari ketika $\gamma = 1$. Jadi, dengan menskalakan jarak gamma, besarnya pengaruh dua titik terhadap satu sama lain.

Kernel Gaussian RBF memiliki dua parameter, yaitu gamma dan sigma. Parameter gamma memiliki nilai default, yaitu $\gamma = 1 / (2\sigma)^2$. Ketika gamma tinggi, titik-titik di sekitar data kemungkinan akan dipertimbangkan dalam perhitungan. Parameter sigma digunakan untuk menemukan nilai optimal untuk setiap set data. Dimana γ (gamma) ditentukan oleh parameter gamma, harus lebih besar dari 0.

2.4. Optical Character Recognition (OCR)

Optical Character Recognition (OCR) sebuah aplikasi komputer yang digunakan untuk mengidentifikasi tulisan maupun angka dalam bentuk gambar yang dikonversi menjadi bentuk file teks. OCR merupakan cabang dari *Artificial Intelligence* (AI) dan computer [2]. Secara umum proses OCR sebagai berikut :

- File Input* : berupa gambar digital dengan format .jpg atau .png.
- Preprocessing* : proses menghilangkan bagian yang tidak diperlukan pada gambar.
- Segmentasi* : memisahkan area pengamatan (*region*) pada karakter.
- Ekstraksi Ciri* : mengambil ciri tertentu dari karakter yang diamati.
- Recognition* : proses mengenali karakter yang diamati.

3. Hasil dan Pembahasan

3.1. Implementasi

Halaman utama merupakan halaman beranda dari sistem. Pada menu utama ini terdapat judul sistem dan navigasi untuk memilih ke halaman lain yaitu halaman OCR, *preprocessing*, klasifikasi, dan menu pengujian sistem. Gambar 4.1 merupakan tampilan halaman beranda dari sistem identifikasi komentar negatif berbahasa Indonesia di sosial media.



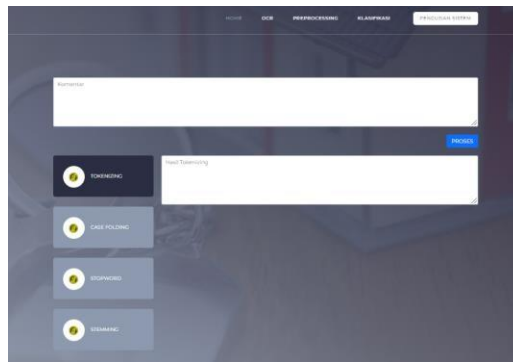
Gambar 7. Tampilan Halaman Beranda

Halaman OCR merupakan halaman untuk memasukkan komentar yang berupa file gambar berformat .jpg atau .png, kemudian diproses dengan menekan tombol "OCR" untuk mengambil parameter teks dari gambar tersebut. Gambar 4.2. merupakan tampilan halaman menu OCR.



Gambar 8. Tampilan Halaman Menu OCR

Halaman *preprocessing* terbagi menjadi 4 sub halaman, diantaranya *tokenizing*, *case folding*, *stopword removal*, dan *stemming*.



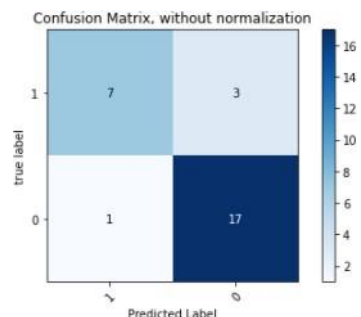
Gambar 9. Tampilan Halaman *Preprocessing*

Halaman Klasifikasi,, Hasil dari text *preprocessing* dimasukkan ke teks area yang disiapkan kemudian tekan tombol “KLASIFIKASI”, maka sistem akan menampilkan hasil klasifikasi berupa positif atau negatif.

3.2. Pengujian Sistem (Model)

Pengujian kinerja sistem dilakukan untuk mengukur akurasi, *precision*, *recall* dan *f1 score* dalam identifikasi komentar negatif berbahasa Indonesia di sosial media dengan menggunakan metode *kernel SVM*. Dalam studi kasus penelitian ini, pengukuran kinerja pada dataset yang digunakan dibagi menjadi rasio 90:10, 80:20, 70:30, 60:40, 50:50.

Pengujian pertama adalah dengan menggunakan perbandingan rasio pembagian 90:10 data *testing* dan *training*. Berdasarkan pengujian didapatkan hasil sebagai berikut :



Gambar 10. Pengujian Pertama

Kelas negatif :

$$Precision(0) = \frac{TP}{FP + TP} \times 100\% = \frac{7}{1 + 7} \times 100\% = \frac{7}{8} \times 100\% = 87,5\%$$

$$Recall(0) = \frac{TP}{FN + TP} \times 100\% = \frac{7}{3 + 7} \times 100\% = \frac{7}{10} \times 100\% = 70\%$$

$$F1 - Score (0) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{70 \times 87,5}{70 + 87,5} = 2 \times \frac{6125}{157,5} = 77,7\%$$

Kelas Positif :

$$Precision(1) = \frac{TN}{FP + TN} \times 100\% = \frac{17}{3 + 17} \times 100\% = \frac{17}{20} \times 100\% = 85\%$$

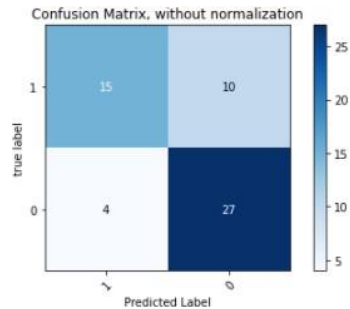
$$Recall(1) = \frac{TN}{FN + TN} \times 100\% = \frac{17}{1 + 7} \times 100\% = \frac{17}{8} \times 100\% = 94,4\%$$

$$F1 - Score (1) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{94,4 \times 85}{94,4 + 85} = 2 \times \frac{8024}{179,4} = 89,4\%$$

Sementara nilai akurasi yang dihasilkan model ini adalah

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{7 + 17}{7 + 17 + 1 + 3} \times 100\% = \frac{24}{28} \times 100\% = 85,7\%$$

Pengujian Kedua adalah dengan menggunakan perbandingan rasio pembagian 80:20 data *testing* dan *training*. Berdasarkan pengujian didapatkan hasil sebagai berikut :



Gambar 11. Pengujian Kedua

Kelas negatif :

$$Precision(0) = \frac{TP}{FP + TP} \times 100\% = \frac{15}{4 + 15} \times 100\% = \frac{15}{19} \times 100\% = 78,9\%$$

$$Recall(0) = \frac{TP}{FN + TP} \times 100\% = \frac{15}{10 + 15} \times 100\% = \frac{15}{25} \times 100\% = 60\%$$

$$F1 - Score (0) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{60 \times 78,9}{60 + 78,9} = 2 \times \frac{4734}{138,9} = 67,3\%$$

Kelas Positif :

$$Precision(1) = \frac{TN}{FP + TN} \times 100\% = \frac{27}{10 + 27} \times 100\% = \frac{27}{37} \times 100\% = 72,9\%$$

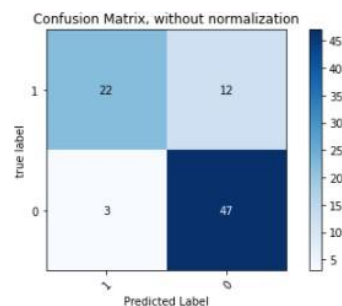
$$Recall(1) = \frac{TN}{FN + TN} \times 100\% = \frac{27}{4 + 27} \times 100\% = \frac{17}{31} \times 100\% = 87\%$$

$$F1 - Score (1) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{87 \times 72,9}{87 + 72,9} = 2 \times \frac{6264}{159} = 78,7\%$$

Sementara nilai akurasi yang dihasilkan model ini adalah

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{15 + 27}{15 + 27 + 4 + 10} \times 100\% = \frac{42}{56} \times 100\% = 75\%$$

Pengujian Ketiga adalah dengan menggunakan perbandingan rasio pembagian 70:30 data *testing* dan *training*. Berdasarkan pengujian didapatkan hasil sebagai berikut :



Gambar 12. Pengujian Ketiga

Kelas negatif :

$$Precision(0) = \frac{TP}{FP + TP} \times 100\% = \frac{22}{3 + 22} \times 100\% = \frac{22}{25} \times 100\% = 88\%$$

$$Recall(0) = \frac{TP}{FN + TP} \times 100\% = \frac{22}{12 + 22} \times 100\% = \frac{22}{34} \times 100\% = 64,7\%$$

$$F1 - Score (0) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{64,7 \times 88}{64,7 + 88} = 2 \times \frac{5693}{152} = 74,9\%$$

Kelas Positif :

$$Precision(1) = \frac{TN}{FP + TN} \times 100\% = \frac{47}{12 + 47} \times 100\% = \frac{47}{59} \times 100\% = 79,6\%$$

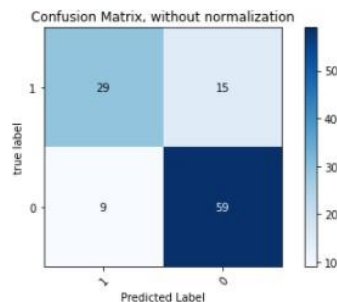
$$Recall(1) = \frac{TN}{FN + TN} \times 100\% = \frac{47}{3 + 47} \times 100\% = \frac{47}{50} \times 100\% = 94\%$$

$$F1 - Score (1) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{94 \times 79,6}{94 + 79,6} = 2 \times \frac{7428}{173} = 85,8\%$$

Sementara nilai akurasi yang dihasilkan model ini adalah

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{22 + 47}{22 + 47 + 3 + 12} \times 100\% = \frac{69}{84} \times 100\% = 82,1\%$$

Pengujian Keempat adalah dengan menggunakan perbandingan rasio pembagian 60:40 data *testing* dan *training*. Berdasarkan pengujian didapatkan hasil sebagai berikut :



Gambar 13. Pengujian Keempat

Kelas negatif :

$$Precision(0) = \frac{TP}{FP + TP} \times 100\% = \frac{29}{9 + 29} \times 100\% = \frac{29}{38} \times 100\% = 76,3\%$$

$$Recall(0) = \frac{TP}{FN + TP} \times 100\% = \frac{29}{15 + 29} \times 100\% = \frac{29}{44} \times 100\% = 65\%$$

$$F1 - Score (0) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{65 \times 76,3}{65 + 76,3} = 2 \times \frac{4959}{141,3} = 70,1\%$$

Kelas Positif :

$$Precision(1) = \frac{TN}{FP + TN} \times 100\% = \frac{59}{15 + 59} \times 100\% = \frac{59}{74} \times 100\% = 79,7\%$$

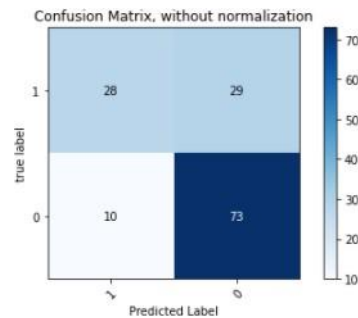
$$Recall(1) = \frac{TN}{FN + TN} \times 100\% = \frac{59}{9 + 59} \times 100\% = \frac{59}{68} \times 100\% = 86,7\%$$

$$F1 - Score (1) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{86,7 \times 79,7}{86,7 + 79,7} = 2 \times \frac{6909}{166,4} = 83,0\%$$

Sementara nilai akurasi yang dihasilkan model ini adalah

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{29 + 59}{29 + 59 + 9 + 15} \times 100\% = \frac{85}{112} \times 100\% = 75,9\%$$

Pengujian Kelima adalah dengan menggunakan perbandingan rasio pembagian 50:50 data *testing* dan *training*. Berdasarkan pengujian didapatkan hasil sebagai berikut :



Gambar 14. Pengujian Kelima

Kelas negatif :

$$Precision(0) = \frac{TP}{FP + TP} \times 100\% = \frac{28}{10 + 28} \times 100\% = \frac{28}{38} \times 100\% = 73,6\%$$

$$Recall(0) = \frac{TP}{FN + TP} \times 100\% = \frac{28}{29 + 28} \times 100\% = \frac{28}{57} \times 100\% = 49,1\%$$

$$F1 - Score (0) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{49,1 \times 73,6}{49,1 + 73,6} = 2 \times \frac{2140}{122,7} = 59,2\%$$

Kelas Positif :

$$Precision(1) = \frac{TN}{FP + TN} \times 100\% = \frac{73}{29 + 73} \times 100\% = \frac{73}{102} \times 100\% = 71,5\%$$

$$Recall(1) = \frac{TN}{FN + TN} \times 100\% = \frac{73}{10 + 73} \times 100\% = \frac{73}{83} \times 100\% = 87,9\%$$

$$F1 - Score (1) = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{87,9 \times 71,5}{87,9 + 71,5} = 2 \times \frac{6284}{159,4} = 78,8\%$$

Sementara nilai akurasi yang dihasilkan model ini adalah

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{28 + 73}{28 + 73 + 10 + 29} \times 100\% = \frac{101}{140} \times 100\% = 72,1\%$$

Hasil Keseluruhan sebagai berikut :

Tabel 2. Hasil Keseluruhan Pengujian Sistem

Rasio	Precision		Recall		F1 - Score		Akurasi
	1	0	1	0	1	0	
90:10	85	88	94	70	89	78	85.7
80:20	73	79	87	60	79	68	75
70:30	80	88	94	65	86	75	82.1
60:40	80	76	87	66	83	71	78.5
50:50	72	74	80	49	79	59	72
Rata-rata	78	81	88.4	62	83.2	70.2	78.6

4. Kesimpulan

Pada penelitian ini telah terimplementasikan algoritma SVM dengan salah satu karnel yaitu RBF untuk identifikasi komentar negatif dalam gambar di sosial media (*Twitter*). Metode yang diusulkan mampu melakukan identifikasi terhadap komentar negatif dan positif berdasarkan teks yang terkandung di dalam gambar dengan tahapan OCR, text *preprocessing* dan klasifikasi. Penggunaan OCR dapat membantu menghasilkan data berupa karakter (teks) dari sebuah komentar yang berbentuk file gambar sehingga memudahkan untuk diproses ke tahapan berikutnya. Setelah

dilakukan pengujian dengan beberapa perbandingan, penggunaan kernel RBF dengan perbandingan 90:10 yang di nilai adalah model yang paling baik diantara yang lain karena menghasilkan *accuracy* sebesar 85.7%. Performa algoritma SVM pada analisis sentiment ini di pengaruhi dengan berbagai macam faktor yaitu data hasil OCR yang tidak memungkinkan mendeteksi semua gambar yang mengandung *text* karena kualitas gambar itu sendiri.

Daftar Pustaka

- [1] N. Abdulloh and A. F. Hidayatullah, "Deteksi Cyberbullying pada Cuitan Media Sosial Twitter," *Automata*, vol. Vol 1, no. 1, pp. 1-5, 2019.
- [2] M. L. H. R. Sandhika Galih A., Erik, "Penerapan Teknik Ocr (Optical Character Recognition) Pada Aplikasi Terjemahan Kitab Fiqih Safinah an-Naja menggunakan Readiris," *Semin. Nas. Inform.*, vol. 2014, no. semnasIF, pp. 60-69, 2014.
- [3] D. Ariadi and K. Fithriasari, "Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer," *J. SAINS DAN SENI ITS Vol. 4, No.2*, vol. 4, no. 2, pp. 248-253, 2015.
- [4] J. Pardede, "Deteksi Komentar Cyberbullying Pada Media Sosial Berbahasa Inggris Menggunakan Naïve Bayes Classification," *J. Inform.*, vol. 7, no. 1, pp. 46-54, 2020, doi: 10.31311/ji.v7i1.6920.
- [5] U. Hasanah, T. Astuti, R. Wahyudi, Z. Rifai, and R. A. Pambudi, "An experimental study of text preprocessing techniques for automatic short answer grading in Indonesian," *Proc. - 2018 3rd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2018*, pp. 230-234, 2018, doi: 10.1109/ICITISEE.2018.8720957.
- [6] W. Pardede, Jasman; Miftahudin, Yusup; Kahar, "Deteksi Komentar Cyberbullying Pada Media Sosial Berbahasa Inggris Menggunakan Naïve Bayes Classification," *J. Inform.*, vol. 7, no. 1, pp. 46-54, 2020, doi: 10.31311/ji.v7i1.6920.
- [7] D. H. Anto Satriyo Nugroho, Arief Budi Witarto, "Support Vector Machine," *Proc. 2011 Chinese Control Decis. Conf. CCDC 2011*, pp. 842-847, 2003, doi:10.1109/CCDC.2011.5968300.