

# Implementasi SVM Untuk Deteksi Komentar Negatif Berbahasa Indonesia di Twitter

Abdul Rahman Iqbal<sup>1</sup>, Yusup Miftahuddin<sup>2</sup>

<sup>1, 2</sup> Program Studi Informatika Institut Teknologi Nasional Bandung, Indonesia

Email : abdulrahmaniqbal9a@mhs.itenas.ac.id

Received DD MM YYYY | Revised DD MM YYYY | Accepted DD MM YYYY

## ABSTRAK

*Perkembangan media sosial sudah semakin pesat, mengingat peran teknologi sudah tidak dapat dilepaskan dari setiap kehidupan manusia. Perbedaan latar belakang pengguna media sosial dapat menimbulkan perbedaan dalam cara berkomunikasi, menyampaikan pendapat, dan cara memandang pendapat dari perspektif yang berbeda, apakah pendapat tersebut positif atau negatif. Tahap awal yang dilakukan dalam penelitian ini adalah proses OCR dimana mengidentifikasi tulisan maupun angka dalam bentuk gambar yang dikonversi menjadi bentuk file teks. Selanjutnya pengolahan text preprocessing meliputi tokenizing, case folding, stopword removal and stemming. Kemudian feature selection bertujuan untuk mendapatkan fitur pada setiap kata untuk dijadikan parameter klasifikasi. Untuk pengambilan keputusan apakah komentar mengandung makna positif atau negatif menggunakan kernel SVM. Data diambil dari media sosial twitter dengan jumlah data komentar 254 data. Berdasarkan hasil dari eksperimen yang dilakukan, kernel Radial Basis Function (RBF) dengan nilai  $\gamma = 0.05$ ,  $cost = 10$  menghasilkan accuracy sebesar 88%, precision 100%, recall 50% dan f1-Score 67%.*

**Kata kunci:** SVM, OCR, Sosial Media, Komentar Negatif, Klasifikasi

## ABSTRACT

*The development of social networks is increasing rapidly, because the role of technology cannot but update in each person's life. Different backgrounds of social media users can lead to differences in the way you communicate, express different opinions and points of view, whether those opinions are positive or negative. The step taken in this search is the OCR process that identifies the text and numbers as images that originally appeared as a text file. In addition, the word processor preprocessing includes encoding, folding, stopping word deletion, and rooting. Next, object selection is to get the objects in each word to use as the classification parameter. To decide if comments contain positive or negative connotations using the SVM kernel. The data is taken from the social network Twitter with a total of 254 comment data. Based on the results of the experiments performed, the kernel of the radius basis function (RBF) with  $\gamma = 0.05$ ,  $cost = 10$  gives 88% accuracy, 100% precision, 50% recall and f1-Score is 67%.*

**Keywords:** SVM, OCR, Social Media, Negative Comment, Classification

## **1. PENDAHULUAN**

### **1.1. Latar Belakang**

Kehadiran media sosial berdampak dalam cara kita berkomunikasi di segala bidang, seperti komunikasi pemasaran, komunikasi politik dan komunikasi dalam sistem pembelajaran (Setiadi, 2016). Komunikasi di media sosial tidak dibatasi oleh jarak, waktu, dan ruang (Watie, 2016). Contoh media sosial, termasuk situs web jejaring sosial Facebook, blogging, Tinder, Instagram, Yik Yak, dan lainnya (Carr & Hayes, 2015).

Country Industry Head Twitter Indonesia mengklaim bahwa Indonesia merupakan negara dengan pertumbuhan pengguna aktif harian Twitter (Abdulloh & Hidayatullah, 2019).

Penggunaan media sosial yang positif adalah memudahkan pengguna untuk menyebarkan informasi di seluruh dunia. Namun, media sosial dapat memberikan dampak negatif seperti memposting tulisan kejam yang berhubungan dengan individu lain dengan tujuan mengintimidasi dan merusak nama baik korban (Pardede, Miftahuddin, & Kahar, 2020).

Pada zaman sekarang banyak terjadi orang-orang menuliskan sebuah komentar seperti komentar negatif yang berisi kata-kata kasar dan menghina yang tentunya mengganggu pemilik akun (Wicaksono, 2020). Untuk mengetahui suatu teks komentar mengandung makna positif atau negatif, metode yang diusulkan adalah kernel SVM.

Kernel SVM salah satu metode yang dapat melakukan pengklasifikasi data dengan baik, karena proses yang akan dilakukan bersifat non-linear (Muis & Affandes, 2015). Pendekatan supervised machine learning (ML) banyak digunakan untuk klasifikasi teks (Pardede, Miftahuddin, & Kahar, 2020). Konsep klasifikasi dengan SVM mendefinisikan sebagai pencarian hyperplane terbaik yang berfungsi sebagai pembagi dua kelas data di ruang input. Fungsi kernel digunakan untuk mengkonversi data menjadi ruang dimensi yang lebih tinggi untuk memungkinkan pemisahan (Hasibuan & Prahutama, 2017). Parameter SVM seperti parameter kernel memiliki pengaruh besar pada kompleksitas dan kinerja model prediksi (Tharwat, 2019).

### **1.2. Rumusan Masalah**

Berdasarkan identifikasi yang telah ditetapkan, maka muncul masalah yang akan ditemui yaitu:

1. Bagaimana menggunakan OCR (Optical Character Recognition) dan menerapkan metode kernel Support Vector Machine (SVM) untuk klasifikasi.
2. Bagaimana menghasilkan akurasi yang maksimal untuk mengidentifikasi komentar negatif menggunakan kernel Support Vector Machine (SVM).

### **1.3. Tujuan**

Tujuan dari penelitian ini adalah menerapkan metode kernel Support Vector Machine (SVM) untuk mengidentifikasi komentar negatif berbahasa Indonesia di sosial media berdasarkan dokumen gambar yang nantinya diolah melalui proses OCR (Optical Character Recognition) dengan mengambil parameter teks. Dengan mengidentifikasi komentar negatif ini, diharapkan pengguna sosial media lebih bijak lagi dalam berkomentar dan dapat menjaga etika dalam berkomunikasi.

### **1.4. Ruang Lingkup**

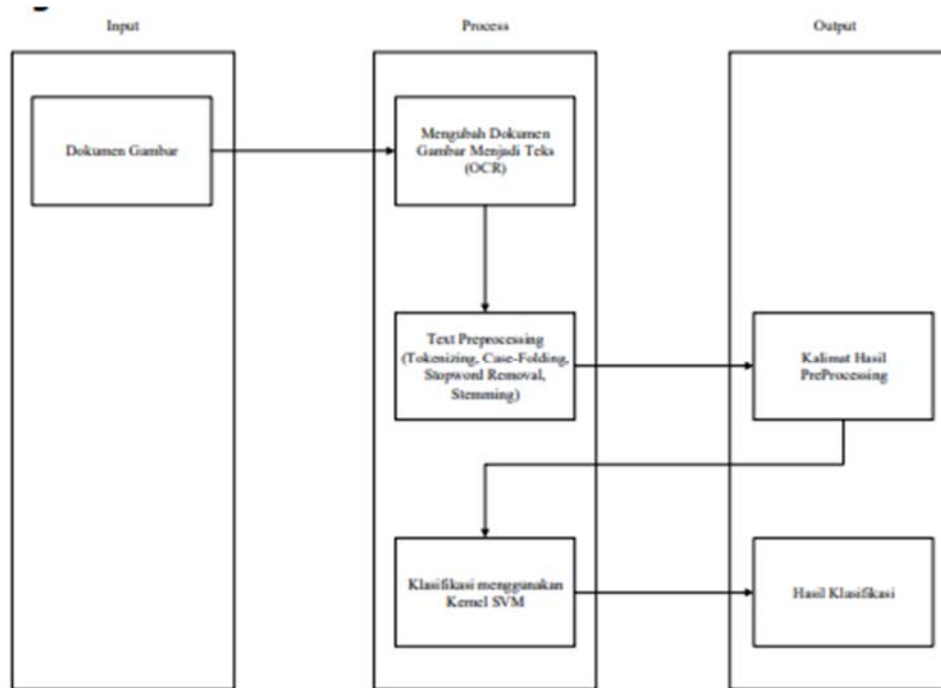
Dalam penelitian yang dilakukan, dibatasi ruang lingkup yang akan dibahas yaitu sebagai berikut:

1. Parameter teks dengan jumlah maksimal 30 kata.

2. Jenis font tebal atau tipisnya huruf dan background pada komentar gambar mempengaruhi terbacanya teks pada proses OCR.
3. Menggunakan kernel RBF (Radial Basis Function) untuk klasifikasi.
4. Jenis font tebal atau tipisnya huruf dan background gambar menjadi acuan untuk terbacanya teks pada komentar gambar di proses OCR.

## 2. METODOLOGI PENELITIAN

### 2.1. Blok Diagram



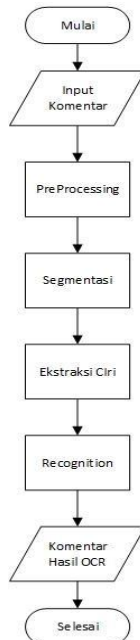
Gambar 1. Blok Diagram

Pada Gambar 1 diperlihatkan blok diagram dari sistem yang dibangun. Masukan pada sistem ini berupa komentar berupa dokumen gambar yang diambil dari komputer. Kemudian pada bagian proses akan melalui beberapa tahap yaitu OCR, kemudian *text preprocessing* yang meliputi *tokenizing*, *case folding*, *stopword removal*, dan *stemming*. Setelah dari tahap *preprocessing* tersebut menghasilkan *output* berupa kalimat hasil *preprocessing*. Proses selanjutnya yaitu klasifikasi menggunakan kernel SVM. Dari klasifikasi tersebut didapatkan hasil berupa kalimat yang bermakna positif atau negatif.

### 2.2. Optical Character Recognition (OCR)

*Optical Character Recognition* (OCR) merupakan 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 vision* (Galih, Erik, & Hakim 2014). Secara umum proses OCR sebagai berikut:

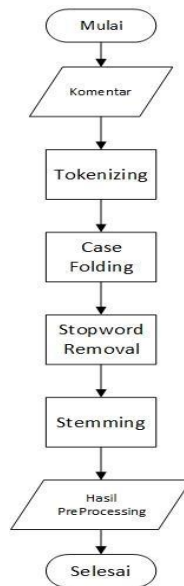
- *File Input*: berupa gambar digital dengan format .jpg atau .png.
- *Preprocessing*: proses menghilangkan bagian yang tidak diperlukan pada gambar.
- Segmentasi: memisahkan area pengamatan (*region*) pada karakter.
- Ekstraksi Ciri: mengambil ciri tertentu dari karakter yang diamati.
- *Recognition*: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



**Gambar 2. Flowchart OCR**

## 2.4 Text Preprocessing

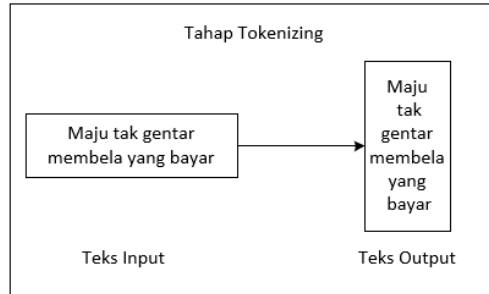
*Text preprocessing* merupakan tahapan untuk merepresentasikan dokumen dalam bentuk fitur vektor, yang artinya harus memisahkan teks menjadi kata terpisah. Proses ini penting dalam menentukan kualitas tahap klasifikasi (Ramyas & Pinakas, 2014). Tujuannya untuk menghilangkan noise yang terdapat pada dokumen teks dan mengambil fitur atau parameter penting pada dokumen teks ( Pardede, Miftahuddin, & Kahar, 2020). *Text Preprocessing* pada penelitian ini meliputi *tokenizing*, *case folding*, *stopword removal*, and *stemming*. Pada Gambar 4 dipresentasikan bagaimana *flowchart text preprocessing*.



**Gambar 3. Flowchart Text Preprocessing**

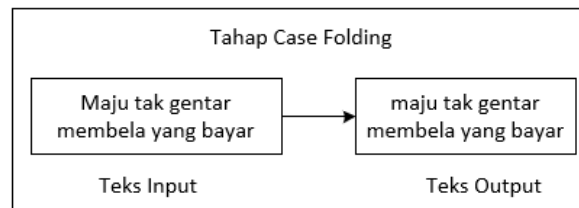
*Text preprocessing* dapat membantu dalam pengambilan keputusan secara baik berdasarkan pengolahan yang sudah dilakukan sebelumnya. Tahapannya sebagai berikut:

- *Tokenizing* adalah memotong urutan karakter menjadi potongan yang disebut token (Pardede, Miftahuddin, & Kahar, 2020). Contoh dari *tokenizing* ditunjukkan pada Gambar 4.



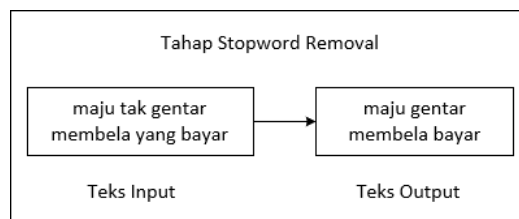
**Gambar 4. Contoh *Tokenizing***

- *Case folding* adalah mengubah semua huruf menjadi huruf kecil (Hasanah et al., 2018). Fungsi *case folding* dibutuhkan dalam mengubah keseluruhan teks dalam dokumen yang digunakan menjadi huruf kecil (*lower case*). Contoh dari *case folding* ditunjukkan pada Gambar 5.



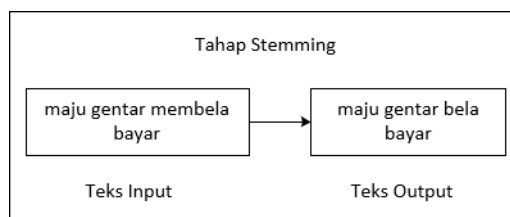
**Gambar 5. Contoh *Case Folding***

- *Stopword removal* adalah menghapus kata-kata yang tidak berguna yang tidak mempunyai arti (Al-khurayji & Sameh, 2017). *Stopword removal* merupakan salah satu tahap untuk mendapatkan kata-kata penting dari proses sebelumnya. Contoh dari *stopword removal* ditunjukkan pada Gambar 6.



**Gambar 6. Contoh *Stopword Removal***

- *Stemming* adalah teknik menemukan kata dasar dengan menghapus imbuhan (Hasanah et al., 2018). Contoh dari *stemming* ditunjukkan pada Gambar 7.



**Gambar 7. Contoh *Stemming***

## 2.5 Classification

Pada penelitian ini, proses klasifikasi menggunakan algoritma SVM dengan fungsi kernel *Radial Basis Function* dan *Polynomial*. *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* (Astuti et al., 2015). 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 (Prasetyo, 2012). Kernel RBF memiliki fungsi sebagai berikut:

$K(X_1, X_2) = \text{exponent}(-\gamma (X_1 - X_2)^2)$  (1) Kernel 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  $\gamma$  (gamma) ditentukan oleh parameter sigma, harus lebih besar dari 0.

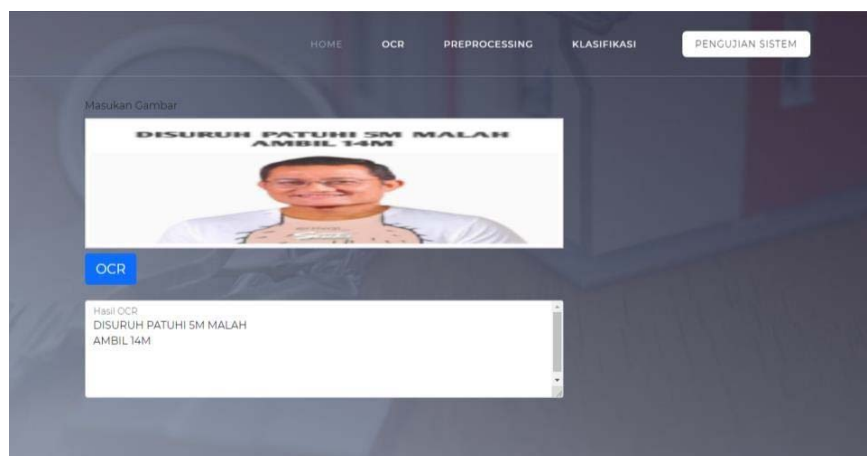


**Gambar 91. Flowchart Klasifikasi**

## 3. HASIL DAN PEMBAHASAN

### 3.1 OCR

Pada proses OCR merupakan pengkonversian komentar dari dokumen gambar menjadi dokumen teks. Pengguna (*user*) akan memasukkan komentar berupa gambar berformat .jpg atau .png kemudian klik tombol OCR, maka sistem akan menampilkan dari OCR tersebut. Proses OCR dapat dilihat pada Gambar 10.



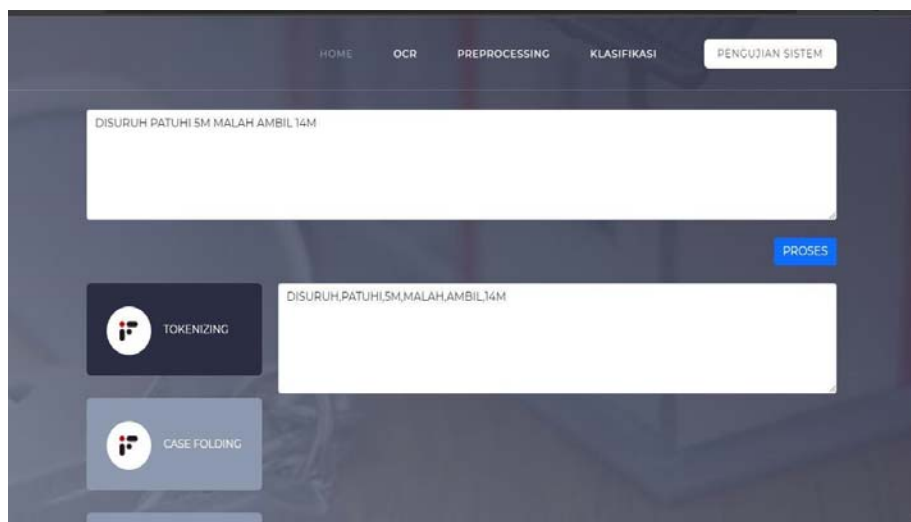
Gambar 20. Hasil OCR

### 3.2 Text Preprocessing

Pada proses *text preprocessing* ini meliputi 4 tahapan, yaitu *tokenizing*, *case folding*, *stopword removal*, dan *stemming*.

#### a. Tokenizing

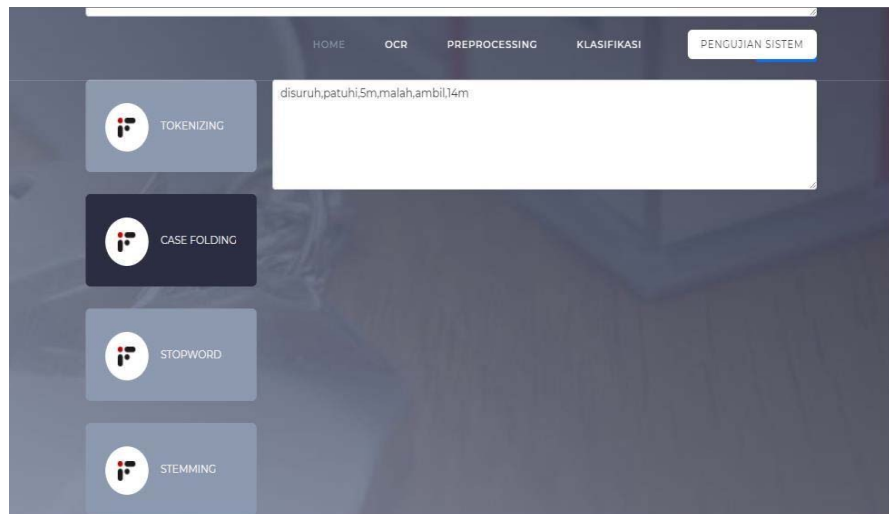
Proses pertama adalah *tokenizing*, yaitu setelah mendapatkan hasil OCR kalimat dipecah menjadi kata (token). Adapun hasilnya dapat dilihat pada Gambar 11.



Gambar 11. Hasil *Tokenizing*

#### b. Case Folding

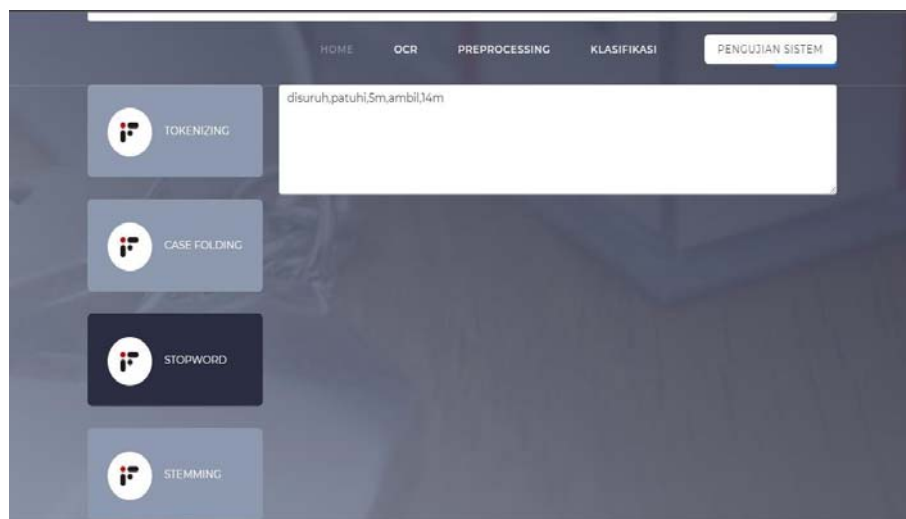
Pada tahap ini merupakan lanjutan dari tahap *tokenizing* dengan tujuan mengubah teks dari huruf kapital menjadi huruf kecil seperti pada Gambar 12.



**Gambar 32. Hasil *Case Folding***

### **c. Stopword Removal**

Stopword removal merupakan lanjutan dari tahap *case folding* dengan tujuan menghilangkan kata-kata yang tidak mempunyai arti yang ditunjukkan pada Gambar 13.

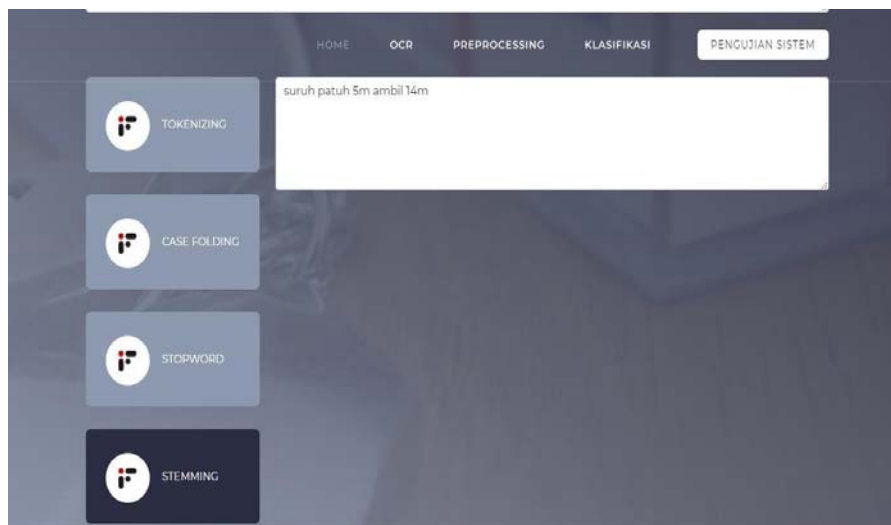


**Gambar 4. Hasil *Stopword Removal***

### **d. Stemming**

*Stemming* merupakan tahap terakhir dari *text preprocessing* yang bertujuan menemukan kata dasar dari komentar seperti yang ditunjukkan pada Gambar 14.

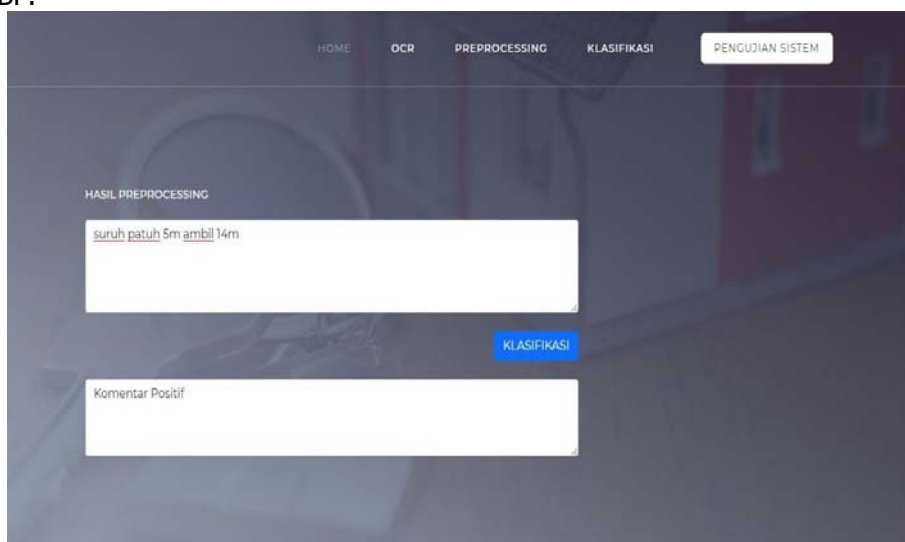




Gambar 54. Hasil *Stemming*

### 3.3 Classification

Proses klasifikasi adalah menentukan apakah komentar yang diidentifikasi mengandung makna positif atau negatif seperti yang terlihat pada Gambar 16 dengan menggunakan kernel RBF.



Gambar 6. Hasil Klasifikasi

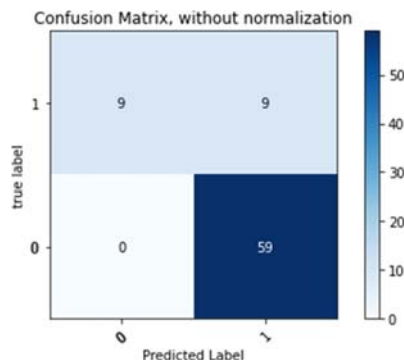
Hasil dari pengujian pada 10 komentar dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pengujian 10 Komentar

No	Komentar	Kategori	Hasil Klasifikasi
1	Andai saja aku tidak lupa untuk mematikan mic ku	Positif	Positif
2	MATI LO MATI LO ANJENG	Negatif	Positif
3	bubar anjing	Negatif	Negatif
4	orang bacot sekali halu banget	Negatif	Positif
5	Gagitu juga tolo!	Negatif	Negatif

6	GINI AMAT HIDUP	Positif	Positif
7	indonesia kalo ga ada depok	Positif	Positif
8	Dah sesat Kite dah sesat	Negatif	Positif
9	PARGOY DULU SAYY	Positif	Positif
10	ANJING KAGA NYAMBUNG	Negatif	Negatif

Berdasarkan hasil pengujian terhadap 10 komentar terdapat beberapa kesalahan dalam pengujian, misalnya pada komentar ke-2, ke-4 dan ke-8 dengan kategori negatif, tetapi sistem mengklasifikasikan komentar tersebut positif. Dengan begitu dapat dihitung nilai *accuracy*, *precision*, *recall*, dan *f1-score* menggunakan *confusion matrix* sesuai data komentar yang diuji. Adapun hasil pengujian kinerja sistem dengan *confusion matrix* dapat dilihat pada Gambar 18.



**Gambar 7. Confusion Matrix**

Untuk mengukur kinerja sistem secara keseluruhan meliputi nilai *accuracy*, *precision*, *recall* dan *f1-score* dapat dihitung dengan menggunakan persamaan berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

Keterangan:

TP = *True Positive*

FP = *False Positive*

FN = *False Negative*

TN = *True Negative*

Kernel RBF

$$Accuracy = \frac{9 + 59}{9 + 59 + 0 + 9} = \frac{68}{77} = 0.88 = 88\%$$

$$Precision = \frac{9}{9 + 0} = \frac{9}{9} = 1 = 100\%$$

$$Recall = \frac{9}{9 + 9} = \frac{9}{18} = 0.50 = 50\%$$

$$F1 - Score = 2 \times \frac{(1 \times 0.50)}{(1 + 0.50)} = 2 \times \frac{1}{1.50} = 0.67 = 67\%$$

#### **4. KESIMPULAN**

Penelitian ini telah berhasil membangun model untuk mendeteksi deteksi komentar negatif berdasarkan teks menggunakan kernel SVM. Metode yang diusulkan mampu melakukan identifikasi terhadap komentar negatif dan positif berdasarkan teks 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 tahap selanjutnya. Dari hasil pengujian, penggunaan kernel RBF dengan nilai  $\gamma=0.05$ ,  $cost=10$  menghasilkan  $accuracy$  sebesar 88%,  $precision$  100%,  $recall$  508% dan  $f1-Score$  67%.

#### **UCAPAN TERIMA KASIH**

Penulis mengucapkan terima kasih kepada Prodi Informatika Institut Teknologi Nasional Bandung. Penulis juga berterima kasih kepada semua pihak yang telah mendukung sehingga penelitian ini terlaksana dengan baik.

#### **DAFTAR RUJUKAN**

- Abdulloh, N., & Hidayatullah, A. F. (2019). Deteksi Cyberbullying pada Cuitan Media Sosial Twitter. *Automata, Vol 1(1)*, 1–5.
- Al-khurayji, R., & Sameh, A. (2017). An Effective Arabic Text Classification Approach Based on Kernel Naive Bayes Classifier. *International Journal of Artificial Intelligence & Applications, 8(6)*, 01–10. <https://doi.org/10.5121/ijaia.2017.8601>
- Astuti, S. I., Arso, S. P., & Wigati, P. A. (2015). KLASIFIKASI DOKUMEN TEKS MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DENGAN PEMILIHAN FITUR CIRI CHI-SQUARE. *Analisis Standar Pelayanan Minimal Pada Instalasi Rawat Jalan Di RSUD Kota Semarang, 3(1210651073)*, 103–111.
- Hasanah, U., Astuti, T., Wahyudi, R., Rifai, Z., & Pambudi, R. A. (2018). An experimental study of text preprocessing techniques for automatic short answer grading in Indonesian. *Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018*, 230–234. <https://doi.org/10.1109/ICITISEE.2018.8720957>
- Hasibuan, C. A., & Prahutama, A. (2017). Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue (Dbd) Menggunakan Support Vector Machine (Svm) Berbasis Gui Matlab. *Jurnal Gaussian, 6(2)*, 171–180.
- Hayes, C. T. C. & R. A. (2015). Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication, 1-43, 8(5)*, 46–65.
- Imelda A.Muis & Muhammad Affandes, M. . (2015). Penerapan Metode Support Vector Machine ( SVM ) Menggunakan Kernel Radial Basis Function ( RBF ) Pada Klasifikasi Tweet. *Sains, Teknologi Dan Industri.UIN Sultan Syarif Kasim Riau, 12(2)*, 189–197.
- Pardede, Jasman; Miftahudin, Yusup; Kahar, W. (2020). Deteksi Komentar Cyberbullying Pada Media Sosial Berbahasa Inggris Menggunakan Naïve Bayes Classification. *Jurnal Informatika, 7(1)*, 46–54. <https://doi.org/10.31311/ji.v7i1.6920>
- R. Sandhika Galih A., Erik, M. L. H. (2014). Penerapan Teknik Ocr ( Optical Character Recognition ) Pada Aplikasi Terjemahan Kitab Fiqih Safinah an-Naja menggunakan

- Readiris. *Seminar Nasional Informatika, 2014*(semnasIF), 60–69.
- Rahmansyah, A., Dewi, O., Andini, P., Hastuti, T., Ningrum, P., & Suryana, M. E. (2018). Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 1–7.
- Ramya, M., & Pinakas, J. A. (2014). Different Type of Feature Selection for Text Classification. *International Journal of Computer Trends and Technology*, 10(2), 102–107. <https://doi.org/10.14445/22312803/ijctt-v10p118>
- Setiadi, A. (2016). Pemanfaatan media sosial untuk efektifitas komunikasi. *Jurnal Humaniora*, 16(2), 1–7. <https://ejournal.bsi.ac.id/ejurnal/index.php/cakrawala/article/download/1283/1055>
- Tharwat, A. (2019). Parameter investigation of support vector machine classifier with kernel functions. *Knowledge and Information Systems*, 61(3), 1269–1302. <https://doi.org/10.1007/s10115-019-01335-4>
- Watie, E. D. S. (2016). Komunikasi dan Media Sosial (Communications and Social Media). *Jurnal The Messenger*, 3(2), 69. <https://doi.org/10.26623/themessenger.v3i2.270>
- Wicaksono, I. (2020). *Sistem identifikasi komentar negatif pada instagram*. [http://eprints.ums.ac.id/82841/7/Naskah Publikasi - Upload Perpus.pdf](http://eprints.ums.ac.id/82841/7/Naskah%20Publikasi%20-%20Upload%20Perpus.pdf)