Analisis Sentimen Terhadap *Tweet* Pelecehan Seksual Dengan Perbandingan Metode *Term Weighting* Menggunakan Klasifikasi SVM Terhadap Tagar Permendikbud30

1st Meira Reynita Putri
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
meirarp@student.telkomuniversity.ac.id

2nd Kemas Muslim Lhaksmana
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
kemasmuslim@telkomuniversity.ac.id

Abstrak

Twitter merupakan salah satu media sosial yang dijadikan sebagai sarana dalam berpendapat dan mengeskpresikan diri, baik dalam menyalurkan pendapat ataupun aspirasi masyarakat sebagai salah satu bentuk kegiatan demokrasi. Salah satu contohnya adalah mengenai pengesahan Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) No 30 Tahun 2021 tentang Pencegahan dan Penanganan Kekerasan Seksual (PPKS) di Perguruan Tinggi. Munculnya *Tweet* dengan tagar #permendikbud30 menuai pro dan kontra di kalangan pengguna media sosial *Twitter*. Untuk mengolah informasi *Tweet* tersebut, dilakukan analisis sentimen yang berfungsi untuk menentukan pendapat atau opini mengenai suatu produk atau peristiwa. Pada prosesnya, *Tweet* diolah menggunakan *data mining* yaitu klasifikasi. Dalam menentukan klasifikasi ada beberapa tahapan yang harus dilakukan yaitu dataset, pelabelan, confusion matrix, pembobotan dan hasil akurasi. Berdasarkan sistem yang dibangun akan dilihat metode pembobotan mana yang memiliki nilai akurasi tertinggi dalam analisis sentiment terhadap #permendikbud30. Berdasarkan hasil pengujian didapatkan nilai F1-Score tertinggi untuk TF-RF dengan fungsi SVM *kernel rbf* sebesar 51%.

Kata kunci : Analisis Sentimen, *Twitter*, permendikbud30, *Confusion Matrix, Dataset*

Abstract

Twitter is one of the social media that is used as a means of expressing opinions and expressing themselves, both in channeling opinions and aspirations of the community as a form of democratic activity. One example is the ratification of the Minister of Education and Culture Regulation (Permendikbud) No. 30 of 2021 concerning the Prevention and Handling of Sexual Violence (PPKS) in Higher Education. The emergence of Tweets with the hashtag #permendikbud30 reaps the pros and cons among Twitter social media users. To process the Tweet information, sentiment analysis is carried out which serves to determine opinions or opinions about a product or event. In the process, Tweets are processed using data mining, namely classification. In determining the classification there are several stages that must be done, namely dataset, labeling, confusion matrix, weighting and accuracy results. Based on the system built, it will be seen from which weighting method has the highest accuracy value in sentiment analysis against #permendikbud30. Based on the test results, the highest F1-Score value for TF-RF with the SVM kernel rbf function is 51%.

Keywords: *Sentiment Analysis, Twitter*, permendikbud30, *Confusion Matrix, Dataset*

I. PENDAHULUAN

Latar Belakang

Salah satu tanda kemajuan teknologi adalah munculnya berbagai macam *platform* media sosial yang telah berkembang pesat. Media sosial menjadi aspek kebebasan dalam berpendapat maupun berekspresi. Saat ini banyak media sosial yang digunakan sebagai wadah penyalur pendapat maupun aspirasi masyarakat sebagai salah satu bentuk kegiatan demokrasi. Salah satu contohnya adalah mengenai pengesahan Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) No 30 Tahun 2021 tentang Pencegahan dan Penanganan Kekerasan Seksual (PPKS) di Perguruan Tinggi. Munculnya Permendikbud No 30 Tahun 2021 ini menuai pro dan kontra di kalangan pengguna media sosial seperti twitter.

Twitter adalah media untuk berdiskusi, bersosialisasi, atau saling memberi pendapat terhadap suatu peristiwa yang ada. Twitter juga

dapat digunakan sebagai media untuk menyampaikan opini yang transparan [1].Twitter memiliki fitur yang mampu membuat penggunanya mengirim dan membaca tweet tanpa adanya batasan. Tweet tersebut mampu memuat teks dengan maksimal 280 karakter. Hal inilah yang membuat twitter menjadi salah satu sarana digital dalam menyampaikan opini mengenai kondisi yang terjadi, baik berupa tulisan, foto, maupun video yang memuat deskripsi sehingga dapat mengidentifikasi profil penggunanya.

Pada penelitian ini akan dilakukan analisis sentimen terhadap *tweet* yang menggunakan tagar #permendikbud30. Analisis sentimen merupakan suatu proses yang dapat digunakan untuk menentukan pendapat atau opini mengenai suatu produk atau peristiwa [2]. Pada prosesnya *tweet* tersebut dapat di olah dengan teknik *data mining* yaitu klasifikasi. Dalam menentukan klasifikasi ada beberapa tahapan yang harus dilakukan yaitu pengumpulan *dataset*, *labelling,* pembobotan, hasil akurasi. Hasil akhir pada penelitian ini merupakan akurasi klasifikasi sentimen terhadap #permendikbud30.

Perumusan Masalah

Berdasarkan permasalahan yang dijelaskan pada latar belakang maka didapat rumusan masalah sebagaiberikut :

a. Bagaimana mendapatkan model klasifikasi sentiment menggunakan metode *term weighting* pada *tweet* dengan tagar #permendikbud30 ?

b. Bagaimana pengaruh metode *term weighting* yang digunakan terhadap nilai akurasi pada hasil prosesklasifikasi ?

Tujuan

Sesuai dengan latar belakang dan rumusan masalah yang dipaparkan, tujuan yang ingin dicapai padapenelitian ini adalah sebagai berikut :

a. Mendapatkan nilai akurasi dengan menggunakan algoritma *Support Vector Machine* terhadap hasil klasifikasi berdasarkan perbanding *term weighting* TF-IDF dan TF-RF

Batasan Masalah

Adapun batasan masalah yang dipakai pada penelitian ini adalah sebagai berikut :

a. Data yang diambil dari tanggal 16 November 2021 samapai dengan 20 Desember 2021.

b. Seluruh data bersumber dari twitter yang menggunakan tagar #permendikbud30.

II. KAJIAN TEORI

Terdapat penelitian terkait analisis sentimen terhadap beberapa hal diantaranya penelitian [13] menggunakan algoritma *Decision Tree* C4.5 dengan metode pembobotan TF-IDF dan TF-RF yang menghasilkan nilai *F-Measure* sebesar 0.2626 dengan data latih dan data uji 80:20 untuk pendekatan TF-RF. Sedangkan untuk pendekatan TF-IDF menghasilkan nilai *F-Measure* sebesar 0.4824 dengan data latih dan data uji 90:10. Dengan metode klasifikasi yang sama, penelitian [11] mendapat nilai akurasi sebesar 65.72% dengan pendekatan TF-RF.

Penelitian [9] menggunakan algoritma *Support Vector Machine* dengan metode pembobotan TF-IDF dan mendapatkan nilai terbesar sebesar 88.97%. Penelitian [8] menggunakan metode klasifikasi yang sama dan didapatkan tingkat akurasi terbaik sebesar 90% dengan kombinasi data latih dan data uji 50:50. Penelitian lain yang menggunakan metode klasifikasi yang sama adalah penilitian [14] dengan pendekatan *Multiclass One Vs Rest Support Vector Machine* dengan pemilihan fitur *Univariate Chi Square* yang menghasilkan nilai akurasi sebesar 91.8%.

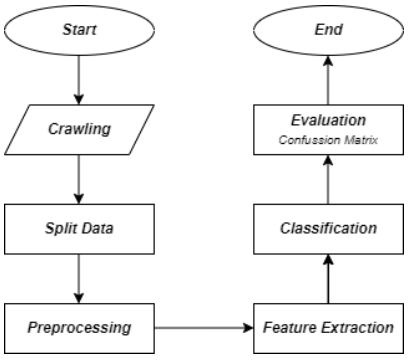
Penelitian lain yang menggunakan metode klasifikasi *Support Vector Machine* adalah penelitian [10] dengan membandingkan pendekatan *Lexion Based* dan *Confusion Matrix* yang mendapatkan nilai akurasi sebesar 84%. Penelitian [12] membandingkan 2 metode klasifikasi *Naïve Bayes* dan *Support Vector Machine* berbasis *Smote Technique.* Didapatkan metode SVM memiliki nilai akurasi terbesar 81.09%.

Pada penelitian tugas akhir ini, metode Support Vector Machine digunakan sebagai metode klasifikasi dengan bantuan beberapa metode preprocessing dan ekstraksi fitur yang dapat digunakan untuk memperbaiki struktur data dan mengubah data menjadi suatu vektor data sebelum dilakukan klasifikasi. Hal lain yang membedakan dengan penelitian ini adalah setiap kata menggunakan pembobotan TF-IDF dan TF-RF.

III. METODE

Dalam penelitian ini sistem yang akan dibangun merupakan sistem yang mampu mengklasifikasi teks sentimen terhadap pengesahan #permendikbud30. Berikut

gambaran sistem yang dibangun pada penelitian ini :



Gambar 3.1 Flowchart Gambaran Sistem Klasifikasi

Crawling

Crawling data adalah salah satu teknik pengambilan data secara otomatis dari suatu tempat penyimpanan data. Pada penelitian ini data yang dikumpulkan diambil dari platform media sosial *Twitter*. Data yang didapat berupa tweet dari masyarakat yang memuat topik mengenai #permendikbud30. Data yang dikumpulkan merupakan data yang belum diberi label sehingga pada tahap selanjutnya data akan masuk ke proses pelabelan data. Pelabelan data menggunakan angka (-1) untuk sentimen negatif, (0) untuk sentiment netral, dan (1) untuk sentiment positif.

Split Data Train Test

Proses split data dilakukan dengan membagi dataset menjadi dua, yaitu data latih dan data uji. Data latih digunakan untuk dipelajari oleh sistem yang nantinya sistem akan membuat *learning model* berdasarkan metode klasifikasi yang digunakan. Kemudian, data uji akan menggunakan *learning model* tersebut untuk klasifikasi teks pada data yang diambil. Pada penelitian ini terdapat 2230 data dengan data positif = 1.278, data negatif = 519, dan data netral = 433, pembagian data yang dilakukan akan memakai perbandingan 70:30.

Preprocessing

c. *Cleaning Data,* merupakan proses untuk menghilangkan *noisy* seperti tanda baca, emotikon, simbol, dsb. *Case Folding* juga masuk ke dalam proses *cleaning* yaitu untuk merubah huruf kapital menjadi huruf kecil (*lower case*) dari sebuah dokumen. Selain *case folding*, remove number juga dilakukan pada proses cleaning, sehingga data yang dimiliki hanya berisikan huruf alfabet dari a-z saja. Hal ini dilakukan agar data yang diproses menjadi bersih dari komponen-komponen yang tidak berhubungan dan tidak diperlukan.

d. *Filtering Stopword,* merupakan proses untuk menghilangkan kata-kata yang kurang penting. Contoh *stopword* dalam Bahasa Indonesia adalah “pun”, “ke”, “atau”, “yang”, dsb. Dengan dihilangkannya kata-kata yang kurang penting tersebut, proses klasifikasi akan lebih mudah karena dapat lebih fokus pada kata-kata yang diperlukan saja [3].

e. *Stemming,* merupakan proses penghilangan atau pemotongan dari suatu kata menjadi bentuk kata dasar. Contoh *stemming* dalam Bahasa Indonesia adalah “berkata → kata”, “membela → bela”, “dibandingkan → dibanding”, dsb.

f. *Word Normalization,* dilakukan untuk membuang kata yang tidak memiliki arti, salah ketik, dan mengubah kata baku menjadi kata yang lebih baku.

g. *Tokenizing,* merupakan proses untuk membagi teks yang berupa kalimat agar menjadi term (kata). Tujuan dari *tokenizing* adalah untuk memisahkan kata-kata pada sebuah paragraph, kalimat ataupun halaman untuk menjadi kata tunggal agar mempermudah proses

pengertian kata. *Tokenizing* juga memuat normalisasi kata yang dilakukan untuk membuang kata yang tidak memiliki arti, salah ketik, dan mengubah kata tidak baku menjadi kata baku. Kata baku yang digunakan mengacu pada Kamus Besar Bahasa Indonesia (KBBI).

Feature Extraction

a. *Term Frequency - Inverse Document Frequency* (TF-IDF)

TF-IDF diterapkan untuk menentukan bobot dari setiap kata sehingga bobot yang paling tinggi dapat diasumsikan sebagai topik utama dalam kalimat [4]. TF-IDF merupakan kombinasi dari *Term Frequency* (TF) *dan Invers Document Frequency* (IDF). TF diguanakan untuk mengukur berapa kali sebuah kata muncul dalam dokumen, sedangkan IDF merupakan jumlah dokumen yang mengandung kata *t* [5]. Perhitungan TF-IF dapat dirumuskan sebagai berikut :

$$TF - IDF(t,d) = tf_{t,d} \times log\left(\frac{N}{d} \quad (1)\right)$$

$tf_{t,d}$ merupakan frekuensi sebuah kata *t* didalam dokumen *d* dan *N* merupakan jumlah dokumen didalam *dataset*. df_t merupakan jumlah dokumen yang ada didalam *dataset* yang mengandung kata *t*.

b. *Term Frequency – Relevance Frequency* (TF-RF)

Relevance frequency merupakan pemberian bobot kata dengan mempertimbangkan relevansi dari sebuah dokumen untuk setiap kata yang muncul dalam kelas tertentu. Seringga RF hanya akan meningkatkan bobot dari kata positif dan bobot kata dengan kategori negatif akan semakin berkurang. Perhitungan TF-RF dapat dirumuskan sebagai berikut :

$$TF \times RF(t,c) = TF(d,t) \times log_2\left(2 \quad (2)\right)$$

c merupakan kelas kategori, A dan C merupakan dokumen dalam kelas c yang mengandungsebuah kata, dan *t* adalah kata.

Klasifikasi

Support Vector Machine (SVM) merupakan salah teknik klasifikasi yang bertujuan untuk menemukan *hyperplane* dengan *margin* yang paling besar. *Margin* merupakan jarak antara titik *support vector* ke *hyperplane*. Perhitungan *hyperplane* dapat dirumuskan sebagai berikut :

$$\vec{w} \cdot \vec{x} + b = 0 \quad (3)$$

Untuk memperoleh nilai *margin* dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dengan *support vector*, yaitu $1/\|\vec{w}\|^2$ yang berarti mencari titik minimal dari persamaan (4) dengan memperhatikan *constraint* persamaan (5) [6].

$$min\frac{1}{2}\|\vec{w}\|^2 \quad (4)$$

$$y_i(\vec{w} \cdot \vec{x_i} + b) - 1 \geq 0, \forall i \quad (5)$$

Dengan data masukan SVM adalah sebuah *vector* angka (0,1). Oleh karena itu , dataset yang digunakan perlu diubah menjadi nilai numerik yang dibantu oleh ektraksi fitur. Pada SVM terdapat beberapa fungsi yang dinamakan fungsi *kernel.* Fungsi *kernel* adalah suatu fungsi yang umum digunakan untuk *polynomial, gaussian RBF,* dan *linear* [7].

Matrix Uji

Evaluasi dilakukan untuk mendapatkan nilai akurasi terbaik dari hasil klasifikasi. Nilai akurasi tersebuat akan digunakan untuk melihat efektifitas performa metode yang digunakan. Perhitungan nilai

performansi digunakan *confusion matrix.*

Tabel 3.1 Contoh *Confusion Matrix 3x3*

Data Aktual	Data Prediksi		
	A	B	C
A	AA	AB	AC
B	BA	BB	BC
C	CA	CB	CC

Dalam *confusion matrix* terdapat 4 istilah nilai hasil klasifikasi yang nantinya akan digunakan untuk menghitung nilai akurasi, presisi. *Recall,* dan *F-Measure* yaitu, *True Positife* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) [15]. Dimana :

a. *True Positive* (TP) adalah nilai hasil klasifikasi yang benar.
b. *True Negative* (TN) adalah nilai hasil klasifikasi yang benar.
c. *False Negative* (FN) adalah nilai dimana hasil yang di prediksi *negative* namun sebenarnya *positive*.
d. *False Positive* (FP) adalah nilai nilai dimana hasil yang di prediksi *positive* namun sebenarnya *negative*.

Untuk *confusion matrix 3x3,* persamaan yang dapat digunakan untuk menghasilkan nilai akurasi, *recall, F-Measeure,* dan presisi adalah sebagai berikut [16] :

$$Akurasi = \frac{AA + BB}{AA + AB + AC + BA + BB} \quad (6)$$

$$Presisi = \left(\frac{AA}{AA + BA + CA} + \frac{BB}{AB + BB} \right) \quad (7)$$

$$Recall = \left(\frac{AA}{AA + AB + AC} + \frac{BB}{BA + BB} \right) \quad (8)$$

$$F - Measure = \frac{2 \times Presisi \times Recall}{Presisi \times Recall} \quad (9)$$

IV. HASIL DAN PEMBAHASAN

Pengujian ini dilakukan berdasarkan *term weighting* TF-IDF dan TF-RF menggunakan algoritma *Support Vector Machine* (SVM) dengan fungsi yang umum digunakan seperti, *kernel linear, gaussian radial basis function (rbf)*, dan *polynomial*. Kemudian menggunakan parameter C, *gamma*, dan *degree*. Perbandingan data latih dan data uji yang digunakan adalah 70:30. Dibawah ini merupakan hasil skenario pengujian menggunakan fungsi *Support Vector Machine* :

1. *Kernel Linear*

Tabel 4.1 F1-Score Kernel Linear

Nilai C	*F1-Score*	
	TF-IDF	TF-RF
0,01	39%	39%
0,1	43%	43%
1	**43%**	**47%**
10	44%	44%
100	44%	44%

Berdasarkan hasil pengujian yang dilakukan, terdapat nilai *F1-Score* pada *kernel linear* SVM dengan menggunakan nilai C sebagai sebagai parameter. Dapat dilihat pada table 4.1 bahwa hasil akurasi terbaik dari SVM *linear* yaitu saat menggunakan nilai konstanta C = 1, didapatkan nilai *F1-Score* pada TF-IDF = 43% dan TF-RF = 47%.

2. *Kernal Polynomial*

Tabel 4.2 F1-Score Kernel Polynomial

d	**γ**									
	0,01		0,1		1		10		100	
	TF-IDF	**TF-RF**	**TF-IDF**	**TF-RF**	**TF-IDF**	**TF-RF**	**TF-IDF**	**TF-RF**	**TF-IDF**	**TF-RF**

1	39%	39%	43%	43%	**47%**	**47%**	44%	44%	44%	44%
2	39%	39%	43%	43%	**47%**	**47%**	46%	46%	46%	46%
3	39%	39%	43%	43%	46%	46%	46%	46%	46%	46%

Berdasarkan pengujian yang dilakukan dapat dilihat pada tabel 4.2 nilai *F1-Score* dari *kernel polynomial* dengan menggunakan nilai *degree* (d) dan *gamma*, terlihat bahwa ada kenaikan dan penurunan terhadap nilai *F1-Score* yang dihasilkan. Nilai *F1-Score* tertinggi dihasilkan ketika nilai *gamma* = 1 dan untuk nilai *degree* = 1 dan 2. Berdasarkan hasil yang didapatkan dapat disimpulkan bahwa semakin kecil nilai *degree* maka semakin besar nilai *F1-Score* yang didapatkan. Sehingga dibutuhkan nilai *gamma* yang optimal untuk mendapatkan nilai *F1-Score* yang lebih besar.

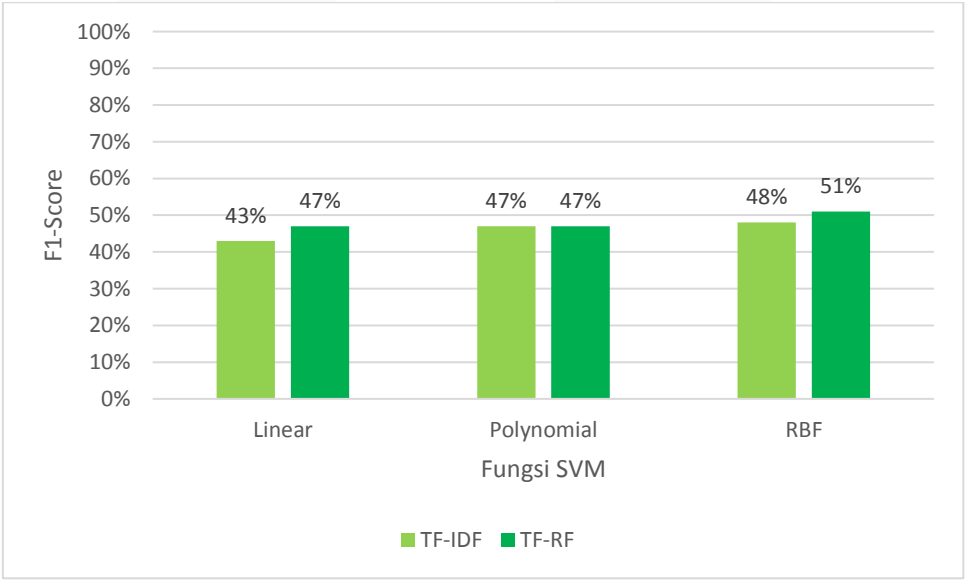
3. Kernel RBF

Tabel 4.3 F1-Score Kernel RBF

C	**γ**									
	0,01		0,1		1		10		100	
	TF-IDF	**TF-RF**	**TF-IDF**	**TF-RF**	**TF-IDF**	**TF-RF**	**TF-IDF**	**TF-RF**	**TF-IDF**	**TF-RF**
0.01	39%	39%	39%	39%	39%	39%	39%	39%	39%	39%
0.1	39%	39%	39%	39%	39%	39%	39%	39%	39%	39%
1	39%	39%	39%	44%	43%	40%	40%	40%	40%	40%
10	39%	45%	48%	**51%**	48%	40%	40%	40%	40%	40%
100	48%	49%	44%	50%	47%	40%	40%	40%	40%	40%

Berdasarkan pengujian yang dilakukan, dapat dilihat nilai *F1-Score* pada *kernel rbf* menggunakan nilai konstanta C dan *gamma* sebagai parameter. Nilai *gamma* menentukan bentuk *hyperplane*, dimana semakin kecil nilai gamma maka bentuk *hyperplane* akan semakin *linear*. Berdasarkan table 4.3, nilai *F1-Score* terbaik yang dihasilkan adalah ketika konstanta *gamma* = 0.1 dan C = 10.

4.1 Analisis Pengujian



Gambar 4.1.1 Hasil F1-Score Term Weighting

Berdasarkan hasil pengujian dengan menggunakan fungsi *kernel* serta parameter *gamma*, C, dan *degree* diperoleh nilai *F1-Score* tertinggi pada *kernel linear* untuk TF-IDF = 43% dan TF-RF = 47% dengan nilai C = 1. Nilai *F1-Score*

ditampilkan karena jumlah data yang tidak seimbang, kemudian *kernel polynomial* untuk setiap *term weighting* TF-IDF dan TF-RF mendapat nilai 47% dengan kombinasi *gamma* = 1 dan *degree* = 1 dan 2. Sedangkan *kernel rbf* untuk TF-IDF = 48% dan TF-RF = 51% dengan nilai *gamma* = 1.0 dan C = 10. Maka dari itu dengan menggunakan *kernel rbf* mendapat kenaikan nilai *F1-Score* sebesar 4%. Hal tersebuat disebabkan oleh semakin besar nilai konstanta C maka semakin besar nilai *F1-Score* yang dihasilkan, sedangkan untuk *gamma* semakin besar nilai konstanta nya tidak terlalu berpengaruh terhadap nilai *F1-Score* yang dihasilkan sehingga diperlukan nilai *gamma* yang optimum untuk dapat menghasilkan nilai *F1-Score* yang baik.

V. KESIMPULAN

5.1 Kesimpulan

1. Berdasarkan hasil pengujian nilai *F1-Score* untuk TF-IDF dengan menggunakan fungsi *kernel* SVM mengalami kenaikan yang signifikan.
2. Nilai *F1-Score* untuk TF-RF lebih baik dibandingkan dengan TF-IDF, mencapai nilai 57% ketika menggunakan fungsi *kernel rbf* dengan kombinasi nilai *gamma* = 1.0 dan C = 10.

5.2 Saran

Pada penelitian selanjutnya dapat menambahkan dataset dan atau menggunakan metode klasifikasi yang lain untuk dapat meningkatkan nilai akurasi sehingga nilai akurasi yang diperoleh menjadi lebih tinggi.

REFERENSI

[1] N. Z. U. N. F. M. N. F. Kholiq Budiman, "Analysis of Sexsual Harassment Tweet Sentiment on Twitter in Indonesia using Naive Bayes Method through National Institute of Standard and Technology Digital Forensic Acquistition Approach," *Journal of Advances in Information Systems and Technology 2 (2),* vol. 21, pp. 21-30, 2020.

[2] X. L. Y. S. Emma Haddi, "The Role of Text Pre-processing in Sentiment Analysis," *Procedia Computer Science 17,* pp. 26-32, 2013.

[3] D. R. A. Vaishali Kalra, "Importance of Text Data Preprocessing & Implementation in RapidMiner," *Proceedings Of ICITKM,* vol. 14, pp. 71-75, 2017.

[4] A. H. D. P. Muhamad Fauzan Putra, "Analisis Pengaruh Normalisasi, TF-IDF, Pemilihan Feature-set Terhadap Klasifikasi Sentimen Menggunakan Maximum Entropy (Studi Kasus : Grab dan Gojek)," *e-Proceeding of Engineering,* vol. 6, p. 8520, 2019.

[5] R. A. Shahzad Qaiser, "Text Mining : Use of TF-IDF to Examine the Relevance of Word to Document," *International Journal of Computer Applications,* vol. 181, 2018.

[6] L. X.-G. G. C.-H. Chu Yan-Xu, "Multiscale models on time series of silicon content in blast furnace hot metal based on Hilbert-Huang transform," *2011 Chinese Control and Decision Conference (CCDC),* pp. 842-847, 2011.

[7] C. F. U. L. Y. Ariana Yunita, "Implementasi Metode Multiple Kernel Support Vector Machine Untuk Seleksi Fitur Dari Data Ekspresi Gen Dengan Studi Kasur Leukimia dan Tumor Usus Besar," *Matics,* vol. 4, 2017.

[8] I. C. R. S. P. Wanda Athira Luqyana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 2, pp. 4704-4713, 2018.

[9] A. S. A. F. Muhammad Yasin Fajari, "Negation Holding dalam Sentiment Analysis Menggunakan Algoritma Support Vector Machine pada Teks Ulasan Film Bahasa Indonesia".

[10] I. S. H. D. P. Fiktor Imanuel Tanesab, "Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine," *International Journal of Computer Science and Software Engineering (IJCSSE),* vol. 6, pp. 180-185, 2018.

[11] E. B. S. F. N. N. Willy, "Implementation of Decision Tree C4.5 for Big Five Personality Predictions with TF-RF and TF-CHI2 on Social Media Twitter," *2019*

International Conference on Computer, Control, Informatics and its Applications (IC3INA), pp. 114-119, 2019.

[12] A. Y. K. T. A. E. B. P. L. E. R. O. Hermanto, "Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique," *Journal of Physics: Conference Series,* vol. 1641, 2020.

[13] E. B. S. Maulina Gustiani Tambunan, "Prediksi Kepribadian DISC Pada Twitter Menggunakan Metode Decision Tree C4.5 dengan Pembobotan TF-IDF dan TF-RF," *eProceedings of Engineering,* vol. 7, p. 2725, 2020.

[14] L. K. B. A. A. S. T. H Syahputra, "Setiment Analysis of Public Opinion on The Go-Jek Indonesia Through Twitter Using Algorithm Support Vector Machine," *Journal of Physics: Conference Series,* vol. 1462, 202.

[15] T. S. S. Y. A. Mujaddid Izzul Fikri, "Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA JURNAL,* vol. 10 No 02, pp. 71-76, 2020.

[16] Y. K. S. Derick Iskandar, "Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C4.5 dan Naive Bayes," *Network Engineering Research Operation,* vol. 2 No. 1, 2016.