

IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI OPINI PUBLIK PADA MEDIA SOSIAL TWITTER TERHADAP KEBOCORAN DATA DI INDONESIA

Iwan Kurniawan¹, Desi Arisandi², Tri Sutrisno³

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

E-mail: iwan.ku20@gmail.com, desia@fti.untar.ac.id², tris@fti.untar.ac.id³

ABSTRACT

Personal data is sensitive and private, it can only be seen by an individual. However, as time goes by, data leaks often occur on the internet, especially in Indonesia. This data leak was busy on Twitter social media and was used as a forum for opinions regarding data leaks that occurred in Indonesia. Therefore, an automatic method is needed to classify tweets based on their sentiment, whether positive, negative, or neutral. By using classification methods such as Support Vector Machine (SVM) with TF-IDF feature extraction, we can automate this process and gain a more accurate understanding of public opinion regarding the issue of personal data breaches in Indonesia. The classification method used for this research is Support Vector Machine (SVM) with TF-IDF feature extraction. The dataset was obtained through the results of scraping Twitter and getting 5000 tweets. The dataset is manually labeled as Positive, Negative, and Neutral before entering the SVM Classification stage. And based on the results of the SVM classification, SVM produces an accuracy of 88%.

Keywords: SVM, Sentiment Analysis, Opinion Classification, Twitter

ABSTRAK

Data pribadi merupakan hal yang sensitif dan privasi, hanya dapat dilihat oleh seorang diri. Namun seiring berkembangnya zaman, sering terjadi kebocoran data di internet terutama Indonesia. Kebocoran data ini sempat ramai di sosial media Twitter dan dijadikan sebagai wadah untuk berpendapat terkait kebocoran data yang terjadi di Indonesia. Oleh karena itu, diperlukan metode otomatis untuk mengklasifikasikan tweet berdasarkan sentimennya, yaitu positif, negatif, atau netral. Dengan menggunakan metode klasifikasi seperti Support Vector Machine (SVM) dengan ekstraksi fitur TF-IDF, kita dapat mengotomatisasi proses ini dan mendapatkan gambaran yang lebih akurat tentang opini publik terkait masalah kebocoran data pribadi di Indonesia. Metode klasifikasi yang digunakan untuk penelitian ini adalah Support Vector Machine (SVM) dengan ekstraksi fitur TF-IDF. Dataset didapatkan melalui hasil scraping Twitter dan mendapatkan 5000 tweet. Dataset diberi label manual yaitu Positif, Negatif, dan Netral sebelum masuk ke tahap Klasifikasi SVM. Dan berdasarkan hasil klasifikasi SVM, SVM menghasilkan akurasi 88%.

Kata kunci: Analisis Sentimen, Klasifikasi Opini, SVM, Twitter.

1. PENDAHULUAN

Data Pribadi adalah data perseorangan tertentu yang disimpan, dirawat, dan dijaga kebenaran serta dilindungi kerahasiaannya. Perlindungan data pribadi merupakan salah satu hak asasi manusia yang merupakan bagian dari perlindungan diri pribadi (Wijaya, 2020). Perlindungan data pribadi ditujukan untuk menjamin hak warga negara atas perlindungan diri pribadi dan menumbuhkan kesadaran masyarakat serta menjamin pengakuan dan penghormatan atas pentingnya perlindungan data pribadi. Artinya setiap orang tidak bisa sembarangan memberikan data pribadi kepada orang lain, agar tidak terjadinya manipulasi data untuk kepentingan pribadi pada pihak yang tidak bertanggung jawab. Data pribadi ini biasanya digunakan untuk kepentingan tertentu, seperti pembuatan Kartu SKCK, Kartu SIM untuk berkendara, Kartu SIM dan sebagainya.

Dan belakangan ini, ada kebocoran data yang terjadi pada Indonesia, kebocoran data ini diperjual belikan di suatu forum yang ada di internet oleh hacker yang bernama Bjorka. Kebocoran data ini meliputi kartu SIM, Data Polri, Dokumen rahasia negara, provider internet.

Kebocoran data ini sangat beresiko karena membuka peluang bagi pihak yang tidak bertanggung jawab untuk menggunakan data yang sudah bocor. Oleh karena itu, maka pemerintah perlu melakukan investigasi terhadap kebocoran data yang terjadi diharapkan kedepannya tidak ada kejadian seperti ini lagi. Komentar-komentar yang ada di sosial media dapat digunakan untuk analisa opini masyarakat. Klasifikasi opini merupakan teknik dari machine learning yang dapat memproses untuk memahami, mengekstrak dan mengolah data secara otomatis untuk mendapatkan informasi yang berasal dari

kalimat opini di sosial media. Klasifikasi opini merupakan metode yang bertujuan untuk menentukan klasifikasi dari berbagai aspek misalnya, ulasan produk, opini masyarakat pada sosial media dan sebagainya yang bisa digunakan untuk klasifikasi.

Klasifikasi terhadap aspek dapat menentukan opini ke dalam kategori positif, negatif, maupun netral (Zuriel & Fahrurrozi, 2021). Penelitian ini memerlukan dataset sebelum melakukan klasifikasi. Maka dari itu, Twitter menjadi salah satu penelitian untuk mendapatkan komentar-komentar dengan cara metode scraping untuk pengambilan komentar.

Berdasarkan permasalahan yang ada, diperlukannya klasifikasi opini terhadap Kebocoran data yang terjadi di Indonesia. Adapun permasalahan untuk klasifikasi sebuah kalimat untuk menentukan ke dalam kategori yang dapat di selesaikan dengan metode Support Vector Machine (SVM). Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti Support Vector Classification) dan regresi (Support Vector Regression).

Klasifikasi komentar terkait kebocoran data sangat penting untuk menilai kecemasan masyarakat terhadap masalah ini. Dengan mengklasifikasikan komentar menjadi kategori positif, negatif, atau netral, kita dapat memahami sentimen publik terhadap masalah kebocoran data dan menilai seberapa besar dampaknya bagi masyarakat. Hasil klasifikasi ini dapat digunakan oleh pihak yang berkepentingan, seperti pemerintah atau perusahaan, untuk merespon masalah ini dengan cara yang lebih efektif.

2. METODE

2.1. Text Mining

Penelitian ini menggunakan metode pengambilan data atau *text mining*. Text mining adalah proses mengekstrak informasi yang berguna dari koleksi dokumen teks dengan menemukan pola, tren, atau hubungan antar teks. Namun, karena hasil text mining biasanya berupa teks yang memiliki tingkat noise yang tinggi, maka perlu dilakukan text preprocessing untuk mengurangi noise tersebut (Pamungkas & Kharisudin, 2021). Data yang digunakan dalam penelitian ini diperoleh melalui scraping Twitter untuk mendapatkan komentar-komentar terkait kebocoran data di Indonesia.

2.2. Praproses

Dataset yang diperoleh dari text mining biasanya memiliki struktur yang tidak teratur dan tidak terstruktur. Oleh karena itu, sebelum data dimasukkan ke dalam model, data harus melalui tahap preprocessing untuk membersihkan data dari noise dan mengubahnya menjadi data yang terstruktur (Irfani, 2020). Tahap ini sangat penting untuk memastikan bahwa data yang digunakan dalam

model adalah data yang berkualitas dan dapat memberikan hasil yang akurat.

2.3. Pembuatan TF-IDF

Pembobotan TF-IDF (Term Frequency-Inverse Document Frequency) digunakan agar data yang digunakan dapat dikategorikan dan diolah oleh algoritma. Pembobotan TF-IDF adalah ekstraksi fitur di mana dataset tweet adalah kalimat. Oleh karena itu, agar mesin dapat memproses informasi yang digunakan, setiap kata harus diubah menjadi angka (Padhana & Sadikin, 2021). Pada titik ini, setiap kata diberi nilai berdasarkan seberapa sering mereka muncul dalam dataset.

2.3. Pembuatan Model

Setelah data melalui tahap preprocessing, langkah selanjutnya adalah mengubah data menjadi data latih dan data uji. Data latih akan digunakan untuk membuat model, sedangkan data uji akan digunakan untuk mengevaluasi atau menyesuaikan model (Darusman, Arifiyanti, & Wati, 2022). Ada beberapa metode yang dapat digunakan untuk menyimpan data, tetapi salah satu yang paling efektif adalah dengan menggunakan metode penyimpanan data secara tidak transparan.

2.4. Klasifikasi Opini

Klasifikasi opini adalah analisis yang bertujuan untuk menentukan sentimen terhadap aspek tertentu secara otomatis (Fitriyah, Warsito, & I Maruddani, 2020). Klasifikasi opini akan digunakan untuk penentuan untuk melihat pendapat seseorang terhadap suatu kebijakan, ulasan atau produk. Klasifikasi opini dapat dikategorikan menjadi 2 kelas, yaitu positif dan negatif atau lebih dari 2 kelas.

2.5. Support Vector Machine

Support Vector Machine adalah metode yang membandingkan parameter standar dengan seperangkat nilai diskrit yang disebut kandidat set untuk menemukan parameter yang memberikan akurasi klasifikasi tertinggi (Hakim, et al., 2020). SVM telah digunakan dalam klasifikasi teks, hiperteks dan gambar. SVM dapat bekerja dengan karakter tulisan tangan dan algoritma ini telah digunakan di laboratorium biologi untuk melakukan tugas seperti menyortir protein. Algoritma ini juga dikenal sebagai Support Vector Network (SVN).

2.6. Alur Metode Penelitian

Metode Penelitian ini akan menjelaskan alur dari penelitian yang dibuat. Pertama adalah untuk pengambilan data twitter menggunakan library python yaitu Snsrape sebagai alat untuk pengambilan data tweet sebanyak 5000 Tweet dari 7 September 2022 sampai 29 September 2022 dan diberi label secara manual. Berikut adalah contoh hasil Scraping Twitter dapat dilihat pada Tabel 1.

Tabel 1. Contoh hasil scraping Twitter

No	Tweet
1	@javminded Urusan kebocoran data pengguna itu nomor sekian, giliran ginian cepet. emang itu kementerian isinya orang2 gubluk bin tolol
2	Seperti yg gw bilang fokus disini tuh emang harusnya soal kebocoran data pribadi.... Sekarang semakin melebar kan banyak idol lain juga kena
3	@amourtir Kebocoran data pribadi siapa yg bertanggung jwb ,,,,

Selanjutnya adalah tahapan preprocessing. Tahapan preprocessing ini digunakan untuk pembersihan dataset seperti menghilangkan angka, link url, merubah kedalam huruf kecil, penyeleksian kata pada data tweets yang ada sehingga akan menghasilkan data yang siap untuk diklasifikasi (Darwis, Pratiwi, & Pasaribu, 2020). Berikut beberapa tahapan preprocessing yaitu: casefolding, cleansing, tokenizing, stopword, stemming. Untuk melihat contoh preprocessing ada pada Tabel 2.

Tabel 2. Contoh hasil Preprocessing

No	Sebelum	Sesudah
1	@javminded Urusan kebocoran data pengguna itu nomor sekian, giliran ginian cepet. emang itu kementerian isinya orang2 gubluk bin tolol	['urus', 'bocor', 'data', 'guna', 'nomor', 'sekian', 'gilir', 'ginian', 'cepat', 'emang', 'tri', 'isi', 'orang', 'gubluk', 'bin', 'tolol']
2	Perlu difahami nih biar gak salah lagi, mengenai kebocoran data itu merupakan kewenangan BSSN bukan Kemkominfo lagi. https://t.co/e3S1UNf6Dm	['difahami', 'salah', 'bocor', 'data', 'wenang', 'bssn', 'kemkominfo']
3	Semoga semakin minim kasus kebocoran data #poinpentingpdp Aturan UU PDP https://t.co/cC3HkkQXAo	['moga', 'minim', 'bocor', 'data', 'poinpentingpdp', 'atur', 'uu', 'pdp']

Lalu tahapan selanjutnya adalah pembobotan kata menggunakan TF-IDF. TF-IDF berfungsi untuk melakukan ekstraksi fitur dari dataset yang sudah di preprocessing, sehingga dari setiap kata pada dataset akan menghasilkan nilai-nilai yang sudah di ekstrasi menggunakan TF-IDF (Indraini & Ernawati, 2022). Untuk melihat contoh hasil TF-IDF dapat dilihat pada Gambar tabel 1.

TERM	TF			TF NORMALISASI			DF	D	IDF	TF-IDF		
	T1	T2	T3	T1	T2	T3				T1	T2	T3
urus	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
bocor	1	1	1	0.071	0.1	0.1	3	3	0	0	0	0
data	1	1	1	0.071	0.1	0.1	3	3	0	0	0	0
guna	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
nomor	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
sekian	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
gilir	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
ginian	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
tri	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
isi	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
orang	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
gubluk	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
bin	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
tolol	1	0	0	0.071	0	0	1	3	0.477	0.477	0	0
dipahami	0	1	0	0	0.1	0	1	3	0.477	0	0.477	0
nih	0	1	0	0	0.1	0	1	3	0.477	0	0.477	0
biar	0	1	0	0	0.1	0	1	3	0.477	0	0.477	0
gak	0	1	0	0	0.1	0	1	3	0.477	0	0.477	0
salah	0	1	0	0	0.1	0	1	3	0.477	0	0.477	0
wenang	0	1	0	0	0.1	0	1	3	0.477	0	0.477	0
bssn	0	1	0	0	0.1	0	1	3	0.477	0	0.477	0
kemkominfo	0	1	0	0	0.1	0	1	3	0.477	0	0.477	0
moga	0	0	1	0	0	0.1	1	3	0.477	0	0	0.477
minim	0	0	1	0	0	0.1	1	3	0.477	0	0	0.477
poin	0	0	1	0	0	0.1	1	3	0.477	0	0	0.477
penting	0	0	1	0	0	0.1	1	3	0.477	0	0	0.477
atur	0	0	1	0	0	0.1	1	3	0.477	0	0	0.477
uu	0	0	1	0	0	0.1	1	3	0.477	0	0	0.477
pdp	0	0	2	0	0	0.2	1	3	0.477	0	0	0.954
Panjang Dok	14	10	10									

Gambar 1. Contoh Hasil Tabel TF-IDF

Selanjutnya adalah tahapan Klasifikasi menggunakan Support Vector Machine. Setelah Dataset Twitter sudah di preprocessing dan melakukan pembobotan kata menggunakan TF-IDF. Artinya SVM sudah bisa diklasifikasikan. Setelah itu SVM menghasilkan akurasi berdasarkan dataset yang sudah diproses. Berikut adalah proses klasifikasi menggunakan SVM:

1. Melatih model SVM: Setelah dataset Twitter telah diproses dan kata-kata telah dibobot menggunakan TF-IDF, langkah selanjutnya adalah melatih model SVM. Untuk melakukan ini, data yang telah diproses dibagi menjadi dua set: set pelatihan dan set uji. Set pelatihan digunakan untuk melatih model SVM, sedangkan set uji digunakan untuk mengevaluasi akurasi model. Selama proses pelatihan, algoritma SVM mencoba menemukan batas keputusan terbaik yang memisahkan kelas-kelas yang berbeda dalam data pelatihan.
2. Mengevaluasi akurasi model: Setelah model SVM dilatih, langkah selanjutnya adalah mengevaluasi akurasinya. Untuk melakukan ini, model diterapkan pada data uji untuk membuat prediksi tentang label kelas dari instance uji. Akurasi model kemudian dihitung dengan membandingkan prediksinya dengan label kelas sebenarnya dari instance uji.
3. Menggunakan hasil klasifikasi: Setelah akurasi model SVM dievaluasi, hasil klasifikasinya dapat digunakan untuk berbagai tujuan. Misalnya, dalam kasus Anda, Anda dapat menggunakan hasil klasifikasi untuk memperoleh pemahaman yang lebih baik tentang opini publik mengenai kebocoran data di Indonesia. Dengan menganalisis distribusi tweet positif, negatif, dan netral, Anda dapat menilai seberapa besar

kekhawatiran masyarakat tentang masalah ini dan apa pendapat mereka.

2.1 Rumus Persamaan SVM dan TF-IDF

Untuk mencari persamaan rumus SVM dengan kernel linear dan RBF dan TF-IDF, dapat dilihat pada rumus dibawah ini.

Rumus Kernel Linear:

$$f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b \quad (1)$$

Keterangan:

a_i = alfa ke-i

y_i = kelas data katih ke-i

M = jumlah data

$K(x_i, x)$ = fungsi kernel yang dipakai

Dengan x = data uji

x_i = data latih ke-i

b = bias

Rumus Kernel RBF:

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2) \quad (2)$$

Rumus Persamaan TF-IDF:

$$W_{j,i} = \frac{n_{j,i}}{\sum_k n_{k,i}} \cdot \log_2 \frac{D}{d_j} \quad (3)$$

Keterangan:

$W_{j,i} W_{j,i}$ = Pembobotan TF-IDF untuk term ke j pada dokumen ke i .

$n_{j,i} n_{j,i}$ = Jumlah kemunculan term ke j pada dokumen ke i .

$\sum_k n_{k,i} \sum_k n_{k,i}$ = Jumlah kemunculan seluruh term pada dokumen ke i .

DD = Banyaknya dokumen yang dibandingkan.

$d_j d_j$ = Banyaknya dokumen yang mengandung term ke j .

3. PEMBAHASAN

3.1. Pengujian Aplikasi

Pengujian dan Pembahasan ini akan dibahas lebih lanjut lewat aplikasi yang dibuat pada penelitian ini.

1. Scraping Twitter

Proses scraping twitter menghasilkan 5000 data tweet dengan kata kunci “Kebocoran Data” yang diambil pada waktu 7 September 2022 sampai 29 September 2022. Setelah itu diberi label secara manual dengan 3 kategori, yaitu “Positif”, “Negatif”,

“Netral”. Untuk contoh hasil scraping twitter dapat dilihat pada **Tabel 1**.

2. Preprocessing Data

Preprocessing merupakan tahapan untuk pembersihan data tweet yang sudah didapat sebelumnya. Tahapan Preprocessing meliputi:

1) Casefolding

Casefolding adalah tahapan proses untuk mengubah huruf pada data tweet menjadi huruf kecil. Berikut hasil Casefolding dapat dilihat pada Tabel 4.

Tabel 4. Hasil Casefolding

Text	Hasil
@javminded Urusan kebocoran data	@javminded urusan kebocoran data
pengguna itu nomor sekian, giliran ginian	pengguna itu nomor sekian, giliran ginian
cepat. emang itu kementerian isinya	cepat. emang itu kementerian isinya
orang2 gubluk bin tolol	orang2 gubluk bin tolol

2) Cleansing

Tahapan cleansing adalah proses untuk menghilangkan angka, HTML, emoticon, hashtag (#), mention username (@username) untuk mengurangi noise pada teks sehingga teks menjadi bersih dari noise. Berikut hasil cleansing dapat dilihat pada Tabel 5.

Tabel 5. Hasil Cleansing

Text	Hasil
@javminded Urusan kebocoran data	urusan kebocoran data
pengguna itu nomor sekian, giliran ginian	pengguna itu nomor sekian giliran ginian
cepat. emang itu kementerian isinya	cepat emang itu kementerian isinya
orang2 gubluk bin tolol	orang gubluk bin tolol

3) Tokenizing

Tokenizing adalah proses untuk memecahkan dan menyeleksi setiap kata pada dataset twitter. Berikut hasil *Tokenizing* dapat dilihat pada Tabel 6.

Tabel 6. Hasil Tokenizing

Text	Hasil
@javminded Urusan kebocoran data	['urusan', 'kebocoran', 'data', 'pengguna', 'itu',
pengguna itu nomor sekian, giliran ginian	'nomor', 'sekian', 'giliran', 'ginian',
cepat. emang itu kementerian isinya	'cepat', 'emang', 'itu', 'kementerian', 'isinya',
orang2 gubluk bin tolol	'orang', 'gubluk', 'bin', 'tolol']

4) Stopword Removal

Stopword Removal adalah proses untuk menghilangkan kata-kata yang tidak ada

hubungannya dengan analisis sentimen. Berikut hasil dapat dilihat pada Tabel 7.

Tabel 7. Hasil Stopword

Text	Hasil
@javminded	['urusan', 'kebocoran', 'kebocoran data', 'pengguna itu nomor sekian, giliran ginian cepet. emang itu kementerian isinya orang2 gubluk bin tolol']

5) Stemming

Stemming adalah proses untuk merubah setiap kata pada dataset menjadi kata dasar, seperti kebocoran menjadi bocor, urusan menjadi urus, dll. Hasil Stemming dapat dilihat pada Tabel 8.

Tabel 8. Hasil Stemming

Text	Hasil
@javminded	['urus', 'bocor', 'data', 'kebocoran data', 'guna', 'nomor', 'sekian', 'pengguna itu nomor', 'gilir', 'ginian', 'cepet', 'emang', 'tri', 'isi', 'kementerian isinya', 'orang', 'gubluk', 'bin', 'orang2 gubluk bin tolol']

Setelah dataset dipreprocessing, dataset diperkecil menjadi 420 data tweet, dikarenakan ketidakseimbangan data pada setiap kelas, maka jumlah kelas "Positif" 140, jumlah kelas "Negatif" 140, jumlah "Netral" 140.

3. TF-IDF Extraction

Pada tahap TF-IDF, aplikasi akan melakukan ekstraksi fitur TF-IDF untuk memberi nilai dari setiap kata dari Dataset Twitter yang sudah di preprocessing. Berikut untuk hasil TF-IDF dilakukan pada Visual Studio Code pada Gambar 1.

Gambar 1. Hasil proses TF-IDF pada Aplikasi

4. Klasifikasi SVM

Setelah Dataset Twitter sudah di preprocessing dan melakukan pembobotan kata menggunakan TF-IDF, langkah selanjutnya adalah tahapan Klasifikasi menggunakan Support Vector Machine. Untuk melakukan pengujian dan pelatihan model SVM, pada saat scraping mendapatkan data 5000 tweet yang

sebelum di preprocessing, dan 5000 tweet tersebut di preprocessing menjadi 420 di dikarenakan ketidakseimbangan data yang ada. Dataset tersebut dibagi menjadi dua set: set pelatihan dan set uji dengan perbandingan 90:10. Set pelatihan digunakan untuk melatih model SVM, sedangkan set uji digunakan untuk mengevaluasi akurasi model. Akurasi yang disebutkan di bagian kesimpulan merupakan hasil dari pengujian model SVM pada data uji. Untuk hasil akurasi dapat dilihat pada Gambar 2.

	precision	recall	f1-score	support
negative	0.76	0.81	0.79	16
neutral	0.91	0.91	0.91	11
positive	0.86	0.80	0.83	15
accuracy			0.83	42
macro avg	0.84	0.84	0.84	42
weighted avg	0.84	0.83	0.83	42

	precision	recall	f1-score	support
negative	0.79	0.94	0.86	16
neutral	1.00	0.91	0.95	11
positive	0.92	0.80	0.86	15
accuracy			0.88	42
macro avg	0.90	0.88	0.89	42
weighted avg	0.89	0.88	0.88	42

Gambar 2. Hasil Akurasi Kernel Linear dan RBF

Pada gambar di atas, dapat dilihat akurasi dengan kernel RBF lebih baik dibandingkan dengan hasil akurasi dengan kernel Linear. Dan hasil evaluasi dengan confusion matrix kernel linear dan RBF dapat dilihat pada Tabel 10.

Tabel 10. Hasil Confusion Matrix Linear

Actual	Prediction		
	Positive	Negative	Neutral
Positive	13	1	3
Negative	1	10	0
Neutral	2	0	12

Untuk confusion matrix dengan kernel RBF dapat dilihat pada Tabel 11.

Tabel 11. Hasil Confusion Matrix RBF

Actual	Prediction		
	Positive	Negative	Neutral
Positive	15	1	3
Negative	0	10	0
Neutral	1	0	12

Berdasarkan dataset twitter yang ada, dapat disimpulkan bahwa untuk pengujian SVM dengan kernel linear pada pengujian memiliki hasil akurasi sebesar 83%. Untuk kernel RBF pada pengujian SVM memiliki hasil akurasi sebesar 88%.

4. KESIMPULAN

Dari penelitian yang sudah dijelaskan, maka dapat disimpulkan bahwa:

1. Hasil akurasi pada pengujian SVM dengan 2 kernel yaitu RBF dan Linear dengan perbandingan 90:10. Hasil kernel RBF memiliki akurasi yang lebih baik dibandingkan dengan kernel Linear. Kernel RBF memiliki akurasi 88% dan kernel Linear memiliki akurasi 83%.
2. Data yang didapat dari twitter merupakan data hasil dari scraping twitter sebanyak 5000 data tweet dan dilakukan preprocessing menjadi 420 data yang dapat diolah dikarenakan ketidakseimbangan data yang ada.

PUSTAKA

- Darusman, F., Arifiyanti, A., & Wati, S. (2022). Sentiment Analysis Pedulilindungi Tweet Using Support Vector Machine Method. *Applied Technology and Computing Science Journal*, 113-118.
- Darwis, D., Pratiwi, E., & Pasaribu, A. (2020). PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA. *Edutic - Scientific Journal of Informatics Education*.
- Fitriyah, N., Warsito, B., & I Maruddani, D. (2020). ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM). *Jurnal Gaussian*, 376-390.
- Hakim, S. R., Rizki, M. A., Zekha F, N. I., Fitri, N., A, Y. R., & Nooraeni, R. (2020). Analisis Sentimen Pengguna Instagram Terhadap Kebijakan Kemdikbud Mengenai. *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, 15.
- Indraini, A., & Ernawati, I. (2022). Analisis Sentimen Terhadap Pembelajaran Daring Di Indonesia Menggunakan Support Vector Machine (SVM). *Jurnal Ilmiah FIFO*, 68.
- Irfani, F. (2020). ANALISIS SENTIMEN REVIEW APLIKASI RUANGGURU MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *JBMI (Jurnal Bisnis, Manajemen, dan Informatika)*, 258-266.
- Padhana, K., & Sadikin, M. (2021). Analisis Sentimen Masyarakat Terhadap Kondisi Perekonomian di Indonesia Pada Masa Pandemi 2020. *Jurnal Ilmu Teknik dan Komputer*.
- Pamungkas, F., & Kharisudin, I. (2021). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan

Masyarakat Indonesia Terhadap . *PRISMA, Prosiding Seminar Nasional Matematika*, 628-634.

- Wijaya, G. (2020). PELINDUNGAN DATA PRIBADI DI INDONESIA: IUS CONSTITUTUM DAN IUS CONSTITUENDUM. *Law Review*, 326.
- Zuriel, H., & Fahrurrozi, A. (2021). IMPLEMENTASI ALGORITMA KLASIFIKASI SUPPORT VECTOR MACHINE UNTUK ANALISA SENTIMEN PENGGUNA TWITTER TERHADAP KEBIJAKAN PSBB. *Jurnal Ilmiah Informatika Komputer*, 149-162.