

## PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA *TWITTER* KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA

Dedi Darwis<sup>1</sup>, Eka Shintya Pratiwi<sup>2</sup>, A. Ferico Octaviansyah Pasaribu<sup>3</sup>  
<sup>1,2,3</sup>Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia  
Bandar Lampung, Indonesia  
[darwisedi@teknokrat.ac.id](mailto:darwisedi@teknokrat.ac.id)

### Abstrak

KPK RI merupakan lembaga terdepan yang memiliki kuasa penuh dan diharuskan untuk memberikan kinerja yang baik dalam memberantas tindak pidana korupsi. Namun dengan berkembangnya zaman, menjadikan masyarakat semakin mudah berselancar di media sosial untuk mengetahui informasi, dan bertukar informasi atau opini ke publik tanpa dibatasi ruang dan waktu. Media sosial *twitter* merupakan salah satu media sosial yang dijadikan sebagai wadah menampung opini tersebut. Metode klasifikasi yang digunakan pada penelitian ini adalah *Support Vector Machine (SVM)* dan ekstraksi fitur menggunakan TF-IDF. Dari 2000 data hasil *twitter crawling*, penelitian ini menghasilkan 1890 data dan 3846 *term/kata* dari hasil *preprocessing* lalu dihitung nilai dari kemunculan kata untuk *labeling* yang menghasilkan sentimen positif, negatif dan netral. Berdasarkan hasil pengujian yang dihasilkan, penerapan metode *SVM* menghasilkan nilai Akurasi sebesar 82% dan menghasilkan sentimen dengan label negatif lebih besar dengan jumlah 77%, label positif 8% dan label netral 25%.

**Kata Kunci:** Analisis Sentimen, KPK RI, SVM, Twitter.

### Abstract

*The KPK RI is the foremost institution that has full power and is required to provide good performance in eradicating criminal acts of corruption. However, with the development of the times, it makes it easier for people to surf on social media to find out information, and exchange information or opinions with the public without being limited by space and time. Twitter social media is one social media that is used as a place to accommodate these opinions. The classification method used in this study is the Support Vector Machine (SVM) and feature extraction using TF-IDF. From 2000 data from twitter crawling, this research resulted in 1890 data and 3846 terms / words from the preprocessing results and then calculated the value of the word occurrences for labeling which resulted in positive, negative and neutral sentiments. Based on the resulting test results, the application of the SVM method resulted in an accuracy value of 82% and resulted in a sentiment with a negative label greater than 77%, 8% positive label and 25% neutral label.*

**Keywords:** Sentiment Analysis, KPK RI, SVM, Twitter.

## PENDAHULUAN

KPK (Komisi Pemberantasan Korupsi) merupakan garda terdepan dalam penanganan kasus tindak pidana korupsi, serta lembaga negara yang memiliki kuasa penuh dalam memberantas korupsi sesuai dengan Undang-Undang No.30 tahun 2002 tentang Komisi Pemberantasan Tindak pidana korupsi (Taufiq H.R, Wardoyo & Pratama, 2018). Kinerja KPK dalam memberantas korupsi mendapat sorotan utama dari masyarakat, serta menuai beragam opini dari masyarakat terkait kinerjanya.

*Twitter* merupakan media komunikasi yang dapat digunakan masyarakat untuk berekspresi dan lebih diminati oleh masyarakat dunia. Pada *twitter* terdapat istilah *tweet* yaitu pengguna *twitter* dapat memberikan kabar terbaru, berekspresi, berinspirasi, dan beropini yang ditulis oleh pengguna *twitter* lainnya terutama topik atau hal yang sedang menjadi perbincangan utama (Haranto & Sari, 2019). Hal tersebut menjadikan *twitter* sebagai salah satu wadah dalam menempatkan sumber data *text*. Oleh karena itu, peneliti dapat memanfaatkan *twitterscrapper* dalam mengambil data yang akan digunakan dalam penelitian ini. Penelitian ini, menggunakan *dataset* dari komentar *tweet* KPK RI yang menggunakan teks bahasa Indonesia sebagai topik penelitian. Opini akan diolah melalui metode-metode yang disesuaikan pada proses pengambilan teks atau *text mining*.

*Text mining* menggunakan proses *preprocessing* dokumen, proses ini dapat menghasilkan sentimen kelas positif, negatif, dan netral. Beberapa teknik dalam *text mining* antara lain untuk menangani masalah *classification*, *clustering*, *information extraction*, dan *information retrieval* (Rahutomo, Saputra & Fidyawan, 2018), (Surahman, Pasaribu, & Darwis, 2020). Pada penelitian ini akan dilakukan *feature selection* untuk menentukan pemilihan fitur dalam proses pengklasifikasian yang lebih efektif dengan mengurangi jumlah data analisa dan mengidentifikasi fitur yang lebih relevan (Alita, 2017), (Mulyanto & Nurhuda, 2017).

Pada penelitian ini metode yang digunakan untuk klasifikasi adalah *Support Vector Machine* (SVM). SVM merupakan model yang berasal dari teori pembelajaran statistika yang akan memberikan hasil yang lebih baik dibandingkan metode lainnya. Dalam SVM Linear, setiap data pelatihan dikenal sebagai  $(x_i, y_i)$ , di mana  $i = 1, 2, \dots, N$  dan  $x_i = \{x_{i1}, x_{i2}, \dots, x_{iq}\}$   $T$  adalah atribut untuk data pelatihan  $I$ ,  $y_i \{-1, +1\}$  adalah Kelas Label (Sulistiani, 2018), (Sulistiani & Muludi, 2018). Kemudian bagaimana *Support Vector Machine* dapat menghasilkan akurasi dari hasil klasifikasi pengujian tersebut.

Oleh karena itu, klasifikasi opini *tweet* pada *twitter* KPK diharapkan dapat menjadi bahan informasi dan acuan bagi KPK dalam

penilaian di masyarakat. Sehingga hal ini dapat dimanfaatkan sebagai tolak ukur dan bahan pertimbangan perbaikan kinerja KPK dalam memberantas tindak pidana korupsi di Indonesia.

Penelitian oleh (Rahutomo, Saputra & Fidyawan, 2018) meneliti tentang Implementasi *Twitter* Sentimen Analisis Untuk *Review* Film Menggunakan Algoritma *Support Vector Machine*. Proses klasifikasi memudahkan melihat opini positif, negatif, atau netral dengan proses klasifikasi menggunakan algoritma *Support Vectore Machine*. Evaluasi akurasi menghasilkan tingkat akurasi yang stabil hingga 76% dari seluruh proses *training* dan tingkat akurasi ini lebih baik dari algoritma *Naïve Bayes Classifier* yang hanya stabil sampai dengan 75%.

Penelitian oleh (Haranto & Sari, 2019), Mengkaji tentang Implementasi *Support Vector Machine* untuk Analisis Sentimen Pengguna *Twitter* Terhadap Pelayanan Telkom serta Biznet. Dengan tujuan menganalisis sentimen pengguna *twitter* terhadap pelayanan Telkom serta Biznet. Pengujian memanfaatkan *Confussion Matrix* dan *K- Fold Cross Validation* diperuntukan untuk membagi informasi *training* serta informasi *testing*. *K- Fold Cross Validation* serta *Confusion Matrix* membagikan hasil nilai akurasi 79, 6%, *precision* 76, 5%, *recall* 72, 8%, serta *F1- score* 74, 6% untuk Telkom, dan akurasi

83,2%, *precision* 78, 8% *recall* 71, 6%, serta *F1- score* 75% untuk biznet.

Penelitian oleh (Taufiq H.R, Wardoyo & Pratama, 2018), mempelajari tentang Analisis sentimen pada *twitter* terhadap kinerja Komisi Pemberantasan Korupsi (KPK) di Indonesia dengan prosedur *Naïve Bayes*. Dengan tujuan untuk megakomodasi permasalahan terhadap kinerja KPK berdasarkan data *twitter*. Hasil pada penelitian ini menunjuka bahwa sebanyak 50,96% akun *twitter* memiliki sentimen positif dan 49,03% memiliki sentimen negatif. Sedangkan hasil analisis pada data uji menunjukkan tingkat akurasi 65,51%, tingkat presisi 51,35%, dan *recall* 90,47%.

Penelitian oleh (Tane, Lhaksana & Nhita, 2019), Mempelajari tentang Analisis Sentimen pada *Twitter* tentang Calon Presiden 2019 Memanfaatkan Tata cara *SVM( Support Vector Machine)*. Dengan tujuan untuk mengenali besar asumsi warga terpaut dengan pemilihan universal presiden Indonesia pada tahun 2019. Pengujian mendapatkan nilai akurasi sebesar 91.5%.

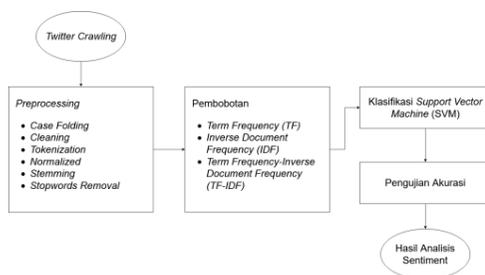
Penelitian oleh (Rahmawati, Marjuni & Zeniarja, 2017), meneliti tentang Analisis Sentimen Publik Pada Media Sosisal *Twitter* Terhadap Pelaksanaan PILKADA Serentak Menggunakan Algoritma *Support Vector Machine*. Proses klasifikasi menghasilkan 2 kelas *tweet* kategori yaitu positif dan negatif. Dengan

3000 dataset *tweet* bahasa Indonesia pemberian label dilakukan dengan metode clustering, dengan membagi 3000 data 2700 data *training* dan 300 data *testing*. Hasil klasifikasi mendapatkan akurasi tertinggi 91% di banding clustering l-means dengan hasil 82%.

Penelitian ini bertujuan untuk memberikan rekomendasi tentang opini masyarakat tentang kinerja KPK berdasarkan data yang diambil dari *twitter* kemudian diklasifikasikan menggunakan algoritma SVM dengan label klasifikasi yaitu positif, negatif dan netral.

## METODE PENELITIAN

### Kerangka penelitian



**Gambar 1.** Kerangka penelitian (Debby, Fernando, & Sulistiani, 2020)

Gambar 1 menyajikan alur dari kerangka penelitian dengan penjelasan sebagai berikut :

#### 1. *Twitter Crawling*

Proses *crawling* ini merupakan proses awal untuk pengumpulan data, proses ini memanfaatkan fasilitas *twitterscraper*.

#### 2. *Preprocessing*

*Preprocessing* merupakan tahap awal dalam pemrosesan untuk penyeleksian kata pada data teks komentar *tweet*, sehingga menghasilkan kata-kata yang lebih ringkas yang berisi sentimen dengan menyeleksi dan membuang kata-kata yang tidak diperlukan, sehingga perlu dilakukan penghapusan beberapa komponen dari data teks komentar *tweet* untuk menyeleksi *tweet*. Beberapa dari tahapan *preprocessing* yaitu: *case folding*, *cleansing*, *tokenizing*, *normalize*, *stemming*, dan *stopword removal*.

#### 3. Pembobotan

Pada tahapan ini setiap dokumen mendapat perlakuan perhitungan pembobotan atau perhitungan *scoring* dengan menggunakan kamus KBBI atau kamus bahasa slang yang berisikan *score* yang telah dibuat untuk dicocokkan dengan kata yang ada pada dokumen. Kemudian kata tersebut direpresentasikan kedalam bentuk vektor dan TF-IDF dan dibuat tabel informasi dokumen yang berisi *term frequency* (TF), *dokumen frequency* (DF), dan IDF dari masing-masing *term*, lalu mengalikan TF dan IDF (Styawati & Mustofa, 2019). Tujuannya adalah untuk mendapatkan label/sentimen dari setiap *term*/kata yang ada pada dokumen.

#### 4. Klasifikasi *Support Vector Machine* (SVM)

SVM merupakan model yang berasal dari teori pembelajaran statistika yang akan memberikan hasil yang lebih baik dibandingkan metode lainnya. (Sulistiani, 2018). Oleh karena itu bagian ini digunakan dalam pengelompokan label yang telah didapatkan dari proses pembobotan dan hasil labelling. Pengelompokan dalam tiga kelas yaitu, positif, negatif, dan netral.

5. Pengujian Akurasi  
Digunakan untuk menghasilkan seberapa akuratnya hasil penelitian.
6. Hasil Analisis Sentimen  
Hasil dari keseluruhan penelitian yang telah dilaksanakan.

## HASIL PENELITIAN DAN PEMBAHASAN

### 1. Twitter Crawling

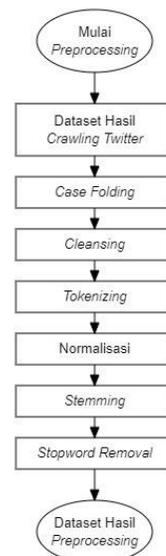
Proses *crawling* menghasilkan 2000 data komentar *tweet twitter* KPK RI yang diambil dengan rentang waktu tahun 2018-2019 dengan memanfaatkan fasilitas *twitterscraper* dengan menginstal *pip* dan *tools python 3.7.6* cara ini dapat memudahkan proses *crawling*, karena dalam proses *crawling* tidak ada limitasi *tweet* dan hasilnya dapat langsung disimpan dalam bentuk *file .json* atau *.csv* lalu disimpan dalam bentuk *file excel*. Contoh dokumen hasil *crawling* dapat dilihat pada Tabel 1.

**Tabel 1.** Contoh Dokumen Hasil *Crawling*

Dokumen	<i>Tweet</i>
Dokumen 1	Lah baru tau.. kalo korupsi 2triliun trs lapor lumayan dong 2M balik lagi. Gitu ?
Dokumen 2	Betul Pak Mahfud kalau pelapor diberi fee artinya mereka yg melapor hanya mengharapkan itu dan tdk ada peduli dan pemberantasan korupsi secara ikhlas. Laporkan koruptor secara ikhlas dan peduli pd NKRI
Dokumen 3	Skenario. Ini akan terus terulang...

### 2. Preprocessing

*Preprocessing* merupakan tahap awal pada proses pengolahan data teks hasil ekstraksi sebelum diolah lebih lanjut hingga sampai ke proses klasifikasi. Tahapan *preprocessing* disajikan pada Gambar 2.



**Gambar 2.** Tahapan *Preprocessing* (Isnain, Sihabuddin, & Suyanto, 2020)

**Gambar 2** merupakan tahapan *preprocessing* yang terdiri dari :

1. *Case folding* dan *Cleansing*

Proses ini merupakan proses Poses menyeragamkan huruf- huruf kapital menjadi huruf kecil dan pembersihan dokumen dari kata- kata yang tidak dibutuhkan untuk mengurangi noise. Kata yang dihilangkan merupakan suatu karakter HTML, kata kunci, *emoticon*, *hastag* (#), *mention username* (@username), angka dan *url*. Dapat dilihat pada contoh yang di ambil dari dokumen pertama pada Tabel 2.

**Tabel 2.** Proses *Casefolding* dan *Cleansing*

lah baru tau kalo korupsi t trs lapor lumayan dong m balik lagi gitu
---

2. Proses Tokenisasi

Merupakan proses menyeleksi, memecah dan pemotongan kata pada dokumen menjadi *term-term* berdasarkan *spasi*. Contoh hasil tokenisasi dapat dilihat pada Tabel 3.

**Tabel 3.** Proses Tokenisasi

['lah', 'baru', 'tau', 'kalo', 'korupsi', 't', 'trs', 'lap or', 'lumayan', 'dong', 'm', 'balik', 'lagi', 'gitu']
---

3. Proses Normalisasi

Tahap ini mengubah kalimat tidak baku/slang menjadi kalimat baku yang sesuai dengan KBBI seperti contoh yang disajikan pada Tabel 4.

**Tabel 4.**Proses Normalisasi

['lah', 'baru', 'tahu', 'kalau', 'korupsi', 'triliun', 'terus', 'lapor', 'lumayan', 'dong', 'miliar', 'balik', 'lagi', 'begitu']
--

4. Proses *Steaming*

Merupakan proses untuk menghilangkan awalan atau akhiran kata yang terdapat kata sambung, kata depan, kata ganti, menjadi kata dasar yang sesuai dengan KBBI. Hasil dari proses *stemming* disajikan pada Tabel 5.

**Tabel 5.** Proses *Stemming*

['lah', 'baru', 'tahu', 'kalau', 'korupsi', 'triliun', 'terus', 'lapor', 'lumayan', 'dong', 'miliar', 'balik', 'lagi', 'begitu']
--

5. Proses *Stopwords Removal*

Merupakan proses penyaringan kata pada dokumen yang tidak ada hubungannya dalam analisis sentimen seperti yang disajikan pada Table 6.

**Tabel 6.**Proses *Stopwords Removal*

['korupsi', 'triliun', 'lapor', 'lumayan', 'miliar']
--

3. Pembobotan

Pada tahapan ini kata tersebut direpresentasikan kedalam bentuk vektor dan TF-IDF. Penggunaan metode TF-IDF dalam proses pembobotan dapat menghasilkan sebuah vektor dengan banyak *term* sehingga dapat dikenali tiap kata yang dihitung sebagai satu fitur.

1. *Term Frecuency* (TF)

Proses ini akan menghitung jumlah kemunculan kata pada *dataset*. Untuk menentukan bobot

dari masing-masing *term/kata* dalam sebuah dokumen yang ada pada *dataset* menggunakan *augmented TF* berdasarkan persamaan (1).

$$tf = 0,5 + 0,5 \frac{tf}{\max(tf)} \quad (1)$$

Keterangan:

tf = banyaknya kata yang muncul pada sebuah dokumen

max(t f) = panjang kata dari sebuah dokumen itu sendiri.

Sebagai contoh, misalkan *Term Frecuency* (TF) diambil dari dokumen ke-1 sebagai berikut :

['korupsi', 'triliun', 'lapor', 'lumayan', 'miliar'].

Jika dokumen tersebut dimasukkan ke dalam persamaan (1) maka:

$$\text{korupsi} = 0,5 + 0,5 \left(\frac{1}{5}\right) = 0,2$$

Karena setiap kata hanya muncul 1 kali pada dokumen maka hasilnya akan sama, kemudian bobot TF pada dokumen diurutkan sesuai dengan abjad sehingga hasilnya sebagai berikut:

['korupsi' = 0,2, 'lapor' = 0,2, 'lumayan' = 0,2, 'miliar' = 0,2, 'triliun' = 0,2]

2. *Inverse Document Frecuency* (IDF)

Proses ini merupakan jumlah dokumen yang berisikan term yang dicari dalam dokumen *dataset* menggunakan persamaan (2).

$$idf = \ln \frac{N}{df} + 1 \quad (2)$$

Keterangan:

ln : Logarima Natural

N : Jumlah semua dokumen

df :Jumlah term/ kata pada dokumen

Contoh *Inverse Document Frecuency* (IDF) yang diambil dari dokumen ke-1:

['korupsi', 'triliun', 'lapor', 'lumayan', 'miliar']

Langkah pertama mencari nilai DF, untuk perhitungan manual, sedangkan pada program menggunakan *library scikit-learn*. Hasil perhitungan DF disajikan pada Tabel 7.

**Tabel 7.** Perhitungan DF pada Dokumen

Term/kata	DF
korupsi	187
triliun	34
lapor	65
lumayan	6
miliar	25

Jika dokumen pada Tabel 7 dimasukkan ke dalam persamaan (2) maka :

$$\begin{aligned} \text{korupsi} &= \ln \left(\frac{1890}{187}\right) + 1 \\ &= 3.3132234911991016 \end{aligned}$$

$$\begin{aligned} \text{triliun} &= \ln \left(\frac{1890}{34}\right) + 1 \\ &= 5.017971583437527 \end{aligned}$$

$$\begin{aligned} \text{lapor} &= \ln \left(\frac{1890}{65}\right) + 1 \\ &= 4.369944838158051 \end{aligned}$$

$$\begin{aligned} \text{lumayan} &= \ln \left(\frac{1890}{6}\right) + 1 \\ &= 6.752572638825633 \end{aligned}$$

$$\begin{aligned} \text{miliar} &= \ln\left(\frac{1890}{25}\right) + 1 \\ &= 5.3254562831854875 \end{aligned}$$

Kemudian bobot IDF pada dokumen diurutkan sesuai dengan abjad. sehingga hasilnya sebagai berikut:

['korupsi'=3.3132234911991016,  
'lapor'=4.369944838158051,  
'lumayan'=6.752572638825633,  
'miliar'=5.3254562831854875,  
'triliun'=5.017971583437527]

	45250488,	
	1.20787460	
	42298067]	
['skenario']	[8.54433210	Netral
	8053688]	

### 3. TF IDF

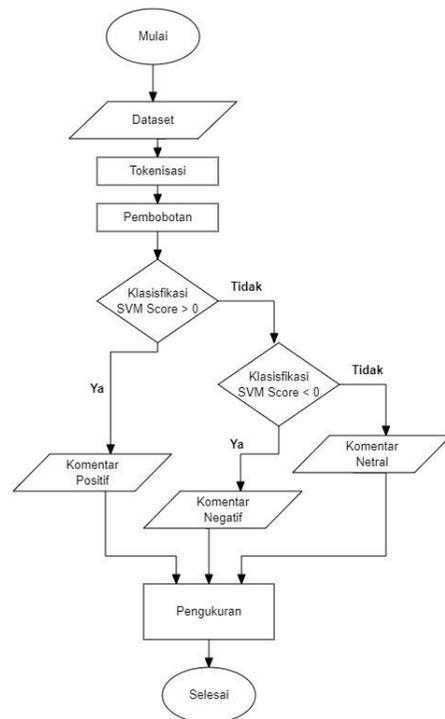
Cara menghitungnya adalah dengan mengalikan TF dan IDF dengan hasil yang disajikan pada Tabel 8.

**Tabel 8.** TF IDF

Dokumen	TF*IDF	Label
['korupsi', 'triliun', 'lapor', 'lumayan', 'miliar']	[0.66264469 82398204, 0.87398896 76316102, 1.35051452 77651267, 1.06509125 66370976, 1.00359431 66875054]	Negatif
['mahfud', 'lapor', 'lapor', 'harap', 'peduli', 'berantas', 'kosupsi', 'ikhlas', 'lapor', 'koruptor', 'ikhlas', 'peduli', 'nkri']	[0.41279063 67465956, 0.43931682 79998056, 1.14549535 68285506, 0.25486334 547685396, 0.34625237 232454914, 1.00844880 8805704, 0.57274767 84142753, 0.51942866	Negatif

### 4. Klasifikasi SVM

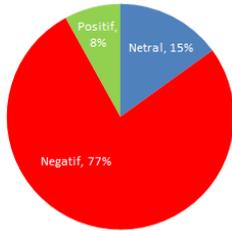
Penelitian ini menggunakan *skit-learn machine learning* pada *python* versi 3.7.6 dengan *Support Vector machine*. Alur dari proses klasifikasi disajikan pada Gambar 3.



**Gambar 3.** Klasifikasi *Support Vector Machine*

Gambar 3 menjelaskan pengolahan data yang telah melalui proses pembobotan dan kemudian akan diklasifikasikan menjadi tiga kelas, yaitu kelas positif, kelas negatif, dan kelas netral. Dengan

ketentuan apabila bobot/ *score* kata >0 adalah positif, <0 adalah negatif dan selain itu adalah netral. Dari hasil pengklasifikasian mesin dengan algoritma SVM didapatkan presentase kelas yaitu kelas positif, kelas negatif, dan kelas netral yang disajikan pada Gambar 4.



**Gambar 4.** Grafik presentase

Gambar 4 menunjukkan hasil presentase dari klasifikasi, dapat dijelaskan bahwa hasil sentimen lebih condong kepada sentimen negatif, yaitu sebesar 77%, untuk netral sebesar 15% dan positif hanya 8%. Data hasil klasifikasi ini diperoleh dengan membagi data sebagai informasi pelatihan serta pengujian, dengan ketentuan jumlah data pelatihan sebanyak 80% dan data pengujian sebanyak 20%.

**5. Pengujian dan Evaluasi**

Pengujian dilakukan menggunakan *confussion matrix* melalui *library SVM*, dari model *training* dan *testing* yang telah diklasifikasikan sebelumnya. Hingga menghasilkan matriks dengan ukuran 3x3 sebagai representatif kelas aktual dan kelas prediksi. Hasil dari model training dengan menggunakan data baru yang belum dilakukan training

sebelumnya. Tabel *confussion matrix* disajikan Tabel 9.

**Tabel 9.** *Confussion Matrix*

Data Aktual	Data Prediksi		
	Negatif	Netral	Positif
Negatif	TNg	NgN	FN
Netral	NNg	TN	NP
Positif	FP	PN	TP

Keterangan:

TP = Kelas kata terprediksi benar bernilai positif

PN = Kelas positif terprediksi netral

FP = Kelas positif terprediksi negatif  
 NP = Kelas netral terprediksi sebagai kelas kata positif

TN = Kelas terprediksi netral

NNg= Kelas netral terprediksi negatif

FN = Kelas negatif terprediksi positif

NgN= Kelas negatif terprediksi netral

TNg=Kelas Negatif terprediksi sebagai negatif

Berdasarkan hasil pengujian yang dilakukan, penelitian ini menghasilkan *confussion matrix* sebagai berikut:

**Tabel 10.** Hasil *Confussion Matrix*

Data Aktual	Data Prediksi		
	Negatif	Netral	Positif
Negatif	252	31	4
Netral	10	53	0
Positif	19	3	6

Pengujian ini menghasilkan akurasi, *recall*, *percision* serta *F1-Score/F-Measure* pada *consol spyder* yang disajikan pada Gambar 5.

	precision	recall	f1-score	support
Negatif	0.90	0.88	0.89	287
Netral	0.61	0.84	0.71	63
Positif	0.60	0.21	0.32	28
accuracy			0.82	378
macro avg	0.70	0.64	0.64	378
weighted avg	0.83	0.82	0.81	378

**Gambar 5.** Hasil Akurasi, Recall, Precision, dan F-1 Score

Gambar 5 menunjukkan bahwa hasil pengujian untuk metode klasifikasi *Support Vector Machine* mendapatkan akurasi sebesar 82%.

## KESIMPULAN DAN SARAN

### Kesimpulan

1. Data yang digunakan pada penelitian ini didapat dari hasil data *crawling* sebanyak 2000 data dan dilakukan *Preprocessing* sehingga tersisa 1890 data yang dapat diolah. Data yang diambil merupakan data dari *twitter* KPK RI tahun 2018 dan 2019
2. Hasil dari klasifikasi menggunakan metode SVM dibagi menjadi tiga kelas, yaitu kelas positif sebanyak 8%, kelas negatif sebanyak 77%, dan kelas netral sebanyak 15%.
3. Berdasarkan hasil pengujian dan evaluasi yang dilakukan pada nilai akurasi, *precession*, *recall*, dan *F1-Score*, dapat disimpulkan bahwa sentimen masyarakat pengguna *twitter* mengenai kinerja KPK dengan presentase sangat kurang baik yaitu ditunjukkan dengan munculnya kecondongan sentimen negatif sebesar 77% dengan keakuratan hasil pengujian akurasi sebesar 82% dan pengujian *precision*

sebesar 90%, serta *recall* sebesar 88% dan *f1-score* sebesar 89%.

### Saran

Untuk meningkatkan hasil penelitian yang lebih maksimal, ada beberapa saran yang dapat dilakukan dan dapat direkomendasikan untuk pengembangan pada penelitian selanjutnya yaitu :

1. peneliti selanjutnya diharapkan dapat membuat kamus KBBI yang lebih lengkap demi terciptanya hasil akurasi yang lebih baik.
2. Peneliti selanjutnya disarankan memasukan dan menambahkan metode klaifikasi yang lain agar dapat membandingkan hasil akurasi yang lebih akurat.

### DAFTAR PUSTAKA

- Alita, D. (2017). Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter, *Journal of Information Systems Engineering and Business Intelligence*, Vol 5 (2), 100-109.
- Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter. *Jurnal Tekno Kompak*, 14(2), 86-91
- Haranto, F. F, & Sari, B. W. (2019). Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap. *Jurnal PILAR Nusa*

- Mandiri*, Vol 15(2), 171–176.
- Isnain, A. R., Sihabuddin, A., & Suyanto, Y. (2020). Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(2).
- Mulyanto, A., & Nurhuda, Y. A. (2017). Penyelesaian Kata Ambigu Pada Proses POS Tagging Menggunakan Algoritma Hidden Markov Model ( HMM ). *SEMINAR NASIONAL METODE KUANTITATIF*, p. 347.
- Rahmawati, A., Marjuni, A., & Zeniarja, J. (2017). Analisis Sentimen Publik Pada Media Sosial Twitter Terhadap Pelaksanaan Pilkada Serentak Menggunakan Algoritma Support Vector Machine. *CCIT Journal*, 10(2), 197–206.
- Rahutomo, F., Saputra, P. Y. & Fidyawan, M. A. (2018) Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine, *Jurnal Informatika Polinema*, 4(2), 93.
- Styawati, & Mustofa, K. (2019). A Support Vector Machine-Firefly Algorithm For Movie Opinion Data Classification. *13(3)*, 219–230.
- Sulistiani, H., & Muludi, K. (2018). Implementation of Dynamic Mutual Information and Support Vector Machine for Customer Loyalty Classification. *The 2nd International Conference on Applied Sciences, Mathematics and Informatics*, 1338(1), p. 12050.
- Surahman, A., Pasaribu, A. O., & Darwis, D. (2020). Ekstraksi Data Produk E-Marketplace Sebagai Strategi Pengolahan Segmentasi Pasar Menggunakan Web Crawler. *SISTEMASI: Jurnal Sistem Informasi*, 9(1), 73-81.
- Tane, O. Z. A., Lhaksana, K. M., & Nhita, F. (2019). Analisis Sentimen Pada Twitter Tentang Calon Presiden 2019 Menggunakan Metode Svm (Support Vector Machine). *Seminar Nasional Teknologi Fakultas Teknik Universitas Krisnadwipayana*, 1(1), 739–742.
- Taufiq H.R., R., Wardoyo, A. E., & Pratama, M. R. (2018). Analisis Sentimen Pada Twitter Terhadap Kinerja Komisi Pemberantasan Korupsi (Kpk) Di Indonesia Dengan Metode Naive Bayes. *3(1)*, 58–67.