

## **KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto)**

**Lalu Mutawalli<sup>1</sup>, Mohammad Taufan Asri Zaen<sup>2</sup>, Wire Bagye<sup>3</sup>**

<sup>1</sup>Pusat Studi Data Sains, STMIK Lombok, Jl. Basuki Rahmat No.105 Praya Lombok Tengah

<sup>2</sup>Sistem Informasi, STMIK Lombok, Jl. Basuki Rahmat No.105 Praya Lombok Tengah

<sup>3</sup>Teknik Informatika, STMIK Lombok, Jl. Basuki Rahmat No.105 Praya Lombok Tengah

<sup>1</sup>laluallistilo@gmail.com, <sup>2</sup>opanzain@gmail.com, <sup>3</sup>wirestmik@gmail.com

### **Abstract**

*In the era of technological disruption of mass communication, social media became a reference in absorbing public opinion. The digitalization of data is very rapidly produced by social media users because it is an attempt to represent the feelings of the audience. Data production in question is the user posts the status and comments on social media. Data production by the public in social media raises a very large set of data or can be referred to as big data. Big data is a collection of data sets in very large numbers, complex, has a relatively fast appearance time, so that makes it difficult to handle. Analysis of big data with data mining methods to get knowledge patterns in it. This study analyzes the sentiments of netizens on Twitter social media on Mr. Wiranto stabbing case. The results of the sentiment analysis showed 41% gave positive comments, 29% commented neutrally, and 29% commented negatively on events. Besides, modeling of the data is carried out using a support vector machine algorithm to create a system capable of classifying positive, neutral, and negative connotations. The classification model that has been made is then tested using the confusion matrix technique with each result is a precision value of 83%, a recall value of 80%, and finally, as much as 80% obtained in testing the accuracy.*

**Keywords:** Model, SVM, Classification, Text, Social Media

### **Abstrak**

Pada era teknologi distrupsi komunikasi masa, sosial media menjadi acuan dalam menyerap opini publik. Digitalisasi data sangat pesat diproduksi oleh pengguna sosial media karena merupakan upaya untuk merepresentasikan perasaan khalayak. Produksi data yang dimaksud adalah pengguna melakukan postingan status dan komentar di sosial media. Produksi data oleh khalayak di sosial media tersebut memunculkan sekumpulan data yang sangat besar atau bisa disebut dengan istilah big data. Big data adalah merupakan kumpulan himpunan data dalam jumlah yang sangat besar, kompleks, memiliki waktu kemunculan relatif cepat, sehingga menjadikanya sulit untuk ditangani. Analisa terhadap data yang besar dengan metode *data mining* untuk mendapatkan pola pengetahuan di dalamnya. Penelitian ini melakukan analisa terhadap sentimen netizen di media sosial twitter terhadap kasus penusukan Wiranto. Hasil sentimen analisis menunjukkan 41% memberikan komentar positif, 29% berkomentar secara netral, dan 29% berkomentar secara negatif terhadap kejadian. Selain itu, pemodelan terhadap data dilakukan dengan menggunakan algoritma *support vector machine* agar dapat membuat sistem yang mampu melakukan klasifikasi terhadap komentar berkonotasi positif, netral, dan negatif. Model klasifikasi yang telah dibuat kemudian diuji dengan menggunakan teknik confusion matrix dengan masing-masing hasil adalah nilai presisi sebesar 83%, nilai recall sebesar 80%, dan terakhir sebesar 80% yang didapatkan pada pengujian akurasi.

Kata kunci: Model, SVM, Klasifikasi, Teks, Media Sosial

## 1. Pendahuluan

Pada era teknologi distrupsi komunikasi masa di sosial media menjadi acuan dalam menyerap opini publik. Digitalisasi data sangat pesat dan masif di produksi oleh pengguna sosial media, hal tersebut merupakan upaya dalam merepresentasikan perasaan mereka. Produksi data yang dimaksud adalah bilamana pengguna sosial media memposting status dan komentar dalam bentuk text, foto, audio, video dan lain sebagainya. Proses produksi tersebut memunculkan sekumpulan data yang sangat besar kemudian disebut dengan istilah big data. Big data adalah istilah umum untuk kumpulan himpunan data dalam jumlah yang sangat besar, kompleks, memiliki waktu kemunculan relatif cepat, sehingga menjadikanya sulit untuk ditangani atau di proses jika hanya menggunakan manajemen basis data atau aplikasi pemeroses data tradisional[1]. Perkembangan data di sosial media setiap menit bahkan dalam hitungan detik mengalami penambahan. Pengolahan dan analisa data sosial media sangat penting dilakukan guna mendapatkan informasi maupun pola pengetahuan yang termudah dari data-data tersebut.

Sosial media analisis adalah kegiatan untuk mengumpulkan data-data sosial media seperti facebook, twitter, instagram, dan sosial media lainnya kemudian di analisa untuk keperluan bisnis, selain itu, dapat pula untuk mengukur sentimen terhadap suatu produk atau layanan[2]. Analisa sosial media dalam pada saat ini begitu marak digunakan oleh banyak institusi pemerintah dan perusahaan untuk mendapatkan suatu inputan pengetahuan tentang suatu kebijakan atau produk yang dibuat. Pada penelitian ini bertujuan untuk melakukan pemodelan teks yang didapat melalui postingan pengguna twitter dalam kasus penusukan Wiranto yang merupakan menteri kordinator bidang politik hukum dan keamanan republik Indonesia. Wiranto ditusuk pada tanggal 10 Oktober 2019 di Pandeglang Banten[3]. Penusukan pada Wiranto mendapatkan membuat perbincangan netizen khususnya di *twitter* buming dibicarakan. Pemberitaan dan komentar netizen terkait dengan kasus penusukan tersebut menjadi konsep penelitian ini untuk mendapatkan informasi tentang sentimen netizen pada seseorang figur. Pada penelitian terdahulu menerangkan bahwa, analisa sosial media mengembangkan kerangka kerja untuk memfasilitasi pembelajaran organisasi melalui

text media sosial untuk meningkatkan kualitas layanan publik oleh pemerintah, melakukan analisa terhadap komentar masyarakat terkait dengan informasi yang di publikasikan oleh pemerintah[4]. Analisis teks media sosial *twitter* dapat memberikan pemahaman ilmiah tentang sentimen terhadap publik figur[5]. Penelitian terdahulu menunjukkan bahwa analisa text sosial media dapat memberikan *secondary opinion* terhadap institusi pemerintah dan figur secara personal.

Pada penelitian ini juga akan melakukan klasifikasi terhadap khususnya pada teks postingan dan komentar netizen tentang kasus figure Wiranto dalam kasus penusukannya. Metode *machine learning* yang digunakan untuk klasifikasi teks adalah *support vector machine*. Penelitian terdahulu yang menggunakan *support vector machine (SVM)* untuk klasifikasi teks, studi perbandingan tentang efektifitas dalam konteks klasifikasi teks yang tidak seimbang menggunakan SVM, hasil penelitian tingkat akurasi menggunakan SVM sebanyak 86% menunjukkan metode ini cukup baik dalam melakukan klasifikasi pada teks[6]. Pemodelan pendapat pengguna pada sosial media twitter menggunakan metode SVM untuk mendapatkan pengetahuan tentang angka kejahatan yang terjadi di Malaysia, Singapura dan Myanmar. Pada penelitian terdahulu menunjukkan bahwa metode SVM cukup baik dalam melakukan klasifikasi teks sosial media, sehingga pada penelitian ini mengusulkan metode SVM sebagai metode memodelkan text komenter netizen di sosial media twitter dalam kasus penusukan Wiranto.

## 2. Tinjauan Pustaka Dan Teori

### 2.1. Tinjauan Pustaka

Penelitian yang dilakukan oleh peneliti merujuk dari penelitian-penelitian terdahulu yang melakukan penelitian terkait klasifikasi teks. Berikut adalah uraian mengenai penelitian yang telah dilakukan sebelumnya. Pembuatan model klasifikasi teks dengan menkombinasikan metode CNN dan RNN kemudian disebut dengan istilah C-LTSM. Hasil penelitian menunjukkan bahwa model yang dibangun cukup baik dalam melakukan klasifikasi[7]. Pembuatan model klasifikasi dengan metode *convolutional network*, dengan *convolutional network* didapatkan hasil klasifikasi yang mutakhir dan kompetitif[8]. Implementasi *latent selection augmented naïve* untuk pemilihan fitur yang lebih baik pada data

sehingga hasil pemodelan yang dibagun akan lebih baik[9]. Pembuatan model klasifikasi teks dengan *hebb rule based feature selection (HRFS)*. Hasil eksperimen menunjukkan bahwa efektof untuk membuat model dengan kinerja lebih baik dalam melakukan tesk[10]. Beberapa penelitian sebelumnya belum terdapat pengolahan teks dengan Bahasa Indonesia. Penelitian selanjutnya akan dibuat model klasifikasi dengan menggunakan algoritma *support vector machine* dengan memanfaatkan library sastrawi pada python untuk melakukan pembersihan data.

## 2.2. Klasifikasi Teks

Pengklasifikasian teks sangat dibutuhkan dalam era digitalisasi, jumlah produksi data teks saat ini sangat besar dan pertumbuhannya sangat cepat. Data teks yang tersedia saat ini sangat besar khususnya pertumbuhan dalam sosial media, perlunya melakukan analisa untuk melakukan klasifikasi terhadap data teks tersebut untuk mendapatkan pengetahuan yang termuat dalam data. Klasifikasi adalah suatu proses pengklasifikasian dokumen-dokumen kedalam satu atau lebih kategori yang telah didefinisikan sebelumnya atau kedalam kelas-kelas dari dokumen-dokumen yang sama[11]. Gambar 1. menunjukkan pengklasifikasian pada teks.



Gambar 1. Ilustrasi klasifikasi teks

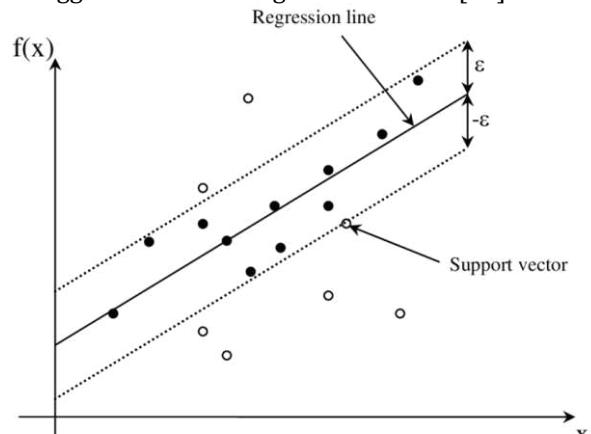
## 2.3. Data Mining

Data mining adalah sebuah proses pencarian secara automatis informasi yang berguna dalam tempat penyimpanan berukuran besar. Fungsi data mining adalah mengidentifikasi fakta-fakta atau kesimpulan yang disarankan berdasarkan penyaringan data untuk menemukan pola yang terdapat pada data[12]. Proses dalam data mining dalam

konteks penelitian ini berupa, melakukan data collecting melakukan pengumpulan data melalui sosial media hususnya twitter. Data cleaning adalah fase dimana data-data dibersihkan, data yang telah bersih dan relevan dapat digunakan untuk proses untuk penggalian pengetahuan. Data integration menapatkan data dari berbagai macam postingan dan komentar di sosial media. Selection atau pemilihan data yang relevan dapat diterima dari koleksi data yang ada. Transformasi data kedalam bentuk yang cocok dengan cara melakukan normalisasi. Data mining merupakan inti pada penlitian ini adalah menerapkan salah satu algoritma data mining yaitu *support vector machine* (SVM) untuk mengekstrak pola-pola potensial yang berguna. Pattern evaluation melakukan evaluasi terhadap model yang telah dibuat dengan SVM dengan metode evaluasi seperti pengujian confusion matrix. Knowledge representation berupa melakukan visualisasi pada data untuk membantu user dalam mengerti dan menginterpretasi hasil data mining.

## 2.4. Support vector machine

*Support vector machine* (SVM) merupakan salah satu metode regresi atau pengklasifikasi data berdasarkan data-data sebelumnya dan pemodelanya disupervisi terlebih dahulu. SVM termasuk kedalam jenis klasifikator yang biner, linier dan non probabilistik[13]. Svm menggunakan decision boundary (batas keputusan) yang akan menentukan klasifikasi dari data-data pelatihan, sehingga dapat dibentuk sebuah model linier atau *hyperplane* yang paling optimal untuk mengklasifikasi data. Gambar 2. Merupakan gambaran pengklasifikasian menggunakan SVM dengan model linier[14].



Gambar 2. SVM Model linier

SVM menggunakan model linier sebagai decision boundary dalam bentuk persamaan dapat dijelaskan sebagai berikut:

$$y(x) = \mathbf{w}^t \varphi(x) + b \quad \dots \quad (1)$$

Dimana  $x$  adalah vector input,  $w$  adalah parameter bobot,  $\varphi(x)$  adalah fungsi basis, dan  $b$  adalah suatu bias. Sehingga, sebuah hyperplane  $y(x)$  dengan maksimumkan margin didapatkan fungsinya. Apabila data terkласifikasi secara linier maka data  $x$  ke  $n$  dapat dicari jaraknya dengan.

$$\frac{t_n(x_n)}{\|\mathbf{w}\|} = \frac{t_n(\mathbf{w}^t \varphi(x_n) + b)}{\|\mathbf{w}\|} \quad \dots \quad (2)$$

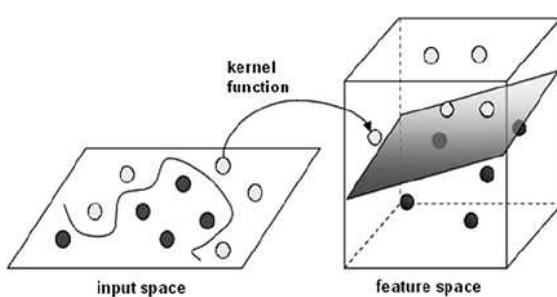
Margin jarak *decision boundary* dengan data terdekat oleh karena itu maksimumkan margin dapat menggunakan:

$$\arg\max \quad \dots \quad (3)$$

Pada dunia nyata data-data pengukuran umumnya tidak dapat diklasifikasi dengan menggunakan metode SVM linier, khususnya dalam konteks penelitian ini adalah klasifikasi teks yang memiliki kompleksitas pada data yang digunakan yaitu berupa teks, namun dapat diklasifikasi dengan lebih baik dengan menggunakan SVM non linier. Oleh karena itu SVM dimodifikasi dengan memasukkan fungsi kernel. Fungsi kernel adalah fungsi  $k$  yang mana untuk semua vektor input  $x$  dan  $z$  akan memenuhi kondisi:

$$k(x, z) = \varphi(x)^t \varphi(z) \quad \dots \quad (4)$$

Dimana  $\varphi$  adalah fungsi pemetaan dari ruang input ke ruang fitur. Gambar 3. Menunjukkan ilustrasi transformasi data ruang input ke ruang fitur. Dalam SVM, kernel trick saat *learning* model sangat membantu dalam mengatasi masalah *feature space*.



Gambar 3. Transformasi data ruang input ke ruang fitur

Pemilihan kernel berpengaruh terhadap akurasi yang dihasilkan[15]. Pada tabel 1. Merupakan Kernel-kernel yang dapat dipakai pada SVM.

Tabel 1. Kernel yang umum dipakai dalam SVM

Jenis Kernel	Definisi
Polynomial	$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^t \mathbf{x}_j + 1)^p$
Gausian	$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2}\right)$
Sigmoid	$k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(a\mathbf{x}_i^t \mathbf{x}_j + \beta)$

## 2.5. Pengujian Dengan Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Terdapat empat istilah sebagai representasi hasil klasifikasi. Keempat istilah tersebut adalah True Positif (TP) merupakan nilai positif yang dideteksi dengan benar, True Negatif (TN) merupakan jumlah data negatif yang dideteksi dengan benar, False Postif (FP) merupakan data negatif namun terdeteksi secara positif dan False Negatif (FN) merupakan data negatif terdeteksi sebagai data negatif.

Tabel 2. Confusion Matrix

		TRUE VALUES	
		TRUE	FALSE
PREDICTION	TRUE	TP <i>Correct Result</i>	FP <i>Unexpected Result</i>
	FALSE	FN <i>Missing Result</i>	TN <i>Correct absence of result</i>

Presisi adalah data yang diambil berdasarkan informasi yang kurang. Presisi dapat dibuat dengan nilai prediksi positif. Berikut adalah aturan presisi:

$$Precision = \left( \frac{TP}{TP + FP} \right) * 100\% \quad \dots \quad (5)$$

Recall adalah data penghapusan yang berhasil diambil dari data yang relevan dengan kueri. Recall juga dikenal sebagai sensitifitas. Munculnya data relevan dapat dilihat dengan recall, berikut adalah persamaan recall:

$$Recall = \left( \frac{TP}{TP + FN} \right) * 100\% \quad \dots \quad (6)$$

Akurasi adalah persentase dari total data yang diidentifikasi dan dinilai. Berikut ini adalah aturan akurasi:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100\% \quad (7)$$

### 3. Metodologi

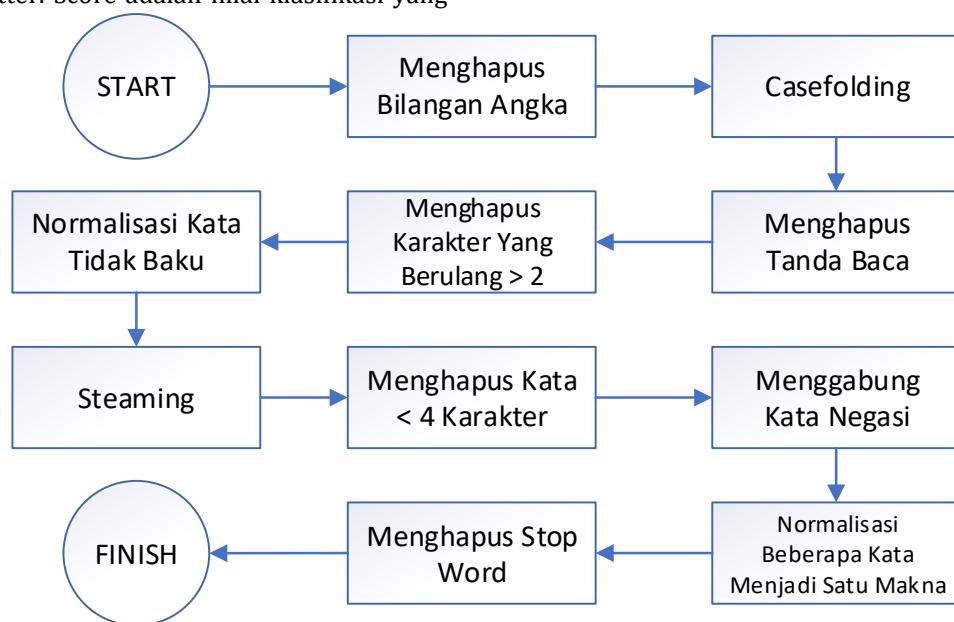
#### 3.1 Pengumpulan Data

Pengumpulan data dengan memanfaatkan *Application Programming Interface (API)* yang telah disediakan oleh twitter. Data yang terkumpul merupakan data teks berdasarkan postingan dan komentar pengguna twitter. Data dikumpulkan mulai dari tanggal 10 Oktober 2019 sampai dengan tanggal 12 Oktober 2019, jumlah data yang terkumul selama tiga hari sebanyak 17.148 baris, pada tabel 1 adalah contoh 5 baris dataset yang dikumpulkan. Variabel yang digunakan pada penelitian ini, teks adalah tweets atau status maupun komentar pengguna sosial media twitter. Score adalah nilai klasifikasi yang

sudah disaring berdasarkan kamus positif, netral, dan negatif. Klasifikasi adalah setiap tweet yang dibagi menjadi tiga kategori yaitu, sentimen positif, netral dan sentimen negatif. Sentimen positif adalah tweet yang memuat konotasi kebaikan dan rasa empati terhadap peristiwa. Sentimen netral tidak memiliki tweets yang tidak memberikan opini berkonotasi positif dan negatif. Sentimen negatif adalah memuat opini negatif.

#### 3.2 Data Preprocessing

Preprocessing merupakan langkah untuk menghasilkan data yang siap untuk dilakukan pelebelan dan pemodelan data. Tujuan dari tahap ini adalah untuk membersihkan kata-kata yang tidak perlu atau kata-kata yang tidak memiliki makna. Pada gambar 1. merupakan tahapan yang dilakukan dalam pembersihan data.



Gambar 4. Tahapan-tahapan preprocessing data

1. Menghapus bilangan angka  
Munculnya bilangan angka pada postingan status maupun komentar seperti nomor "hp" dan "tanggal" dan lain-lain tidak memiliki konotasi yang merepresentasikan suatu pesan yang akan dianalisa, sehingga perlu untuk melakukan penghapusan.
2. Case folding  
Case folding adalah proses untuk menaamaratakan bentuk semua text menjadi huruf kecil semua atau besar semua.
3. Menghapus tanda baca atau simbol  
Menghapus tanda baca atau simbol seperti "/@#\$^&\*()", dan lain-lain.
4. Menghapus karakter berulang lebih dari dua kali

- Dalam Bahasa indonesia jarang menemukan adanya kata yang memiliki karakter berulang lebih dari 2 karakter, oleh sebab itu pada tahap ini dilakukan penghapusan pada karakter yang berulang lebih dari dua kali.
5. Normalisasi kata tidak baku  
Perkataan yang tidak baku seperti penyingkatan kata seperti, "nma", "ulng", "jht", dan lain-lain pada tahapan ini dikumpulkan kata-kata tidak baku secara manusia untuk kemudian diubah menjadi kata baku. Nma menjadi nama, ulng menjadi ulang, jht menjadi jahat.
  6. Steaming  
Tahapan ini dilakukan dengan menggunakan tokenization terlebih dahulu. Tokenization adalah tahap memecah kalimat menjadi bagian-bagian dinamakan token. Sebuah token dianggap sebagai suatu bentuk kata, frasa, atau suatu elemen yang berarti. Setelah tokenization dilakukan dilanjutkan dengan proses steaming. Steamming adalah tahap mengubah sebuah kata ke dalam bentuk kata dasarnya dengan menghapus kata imbuhan "di" maupun imbuhan di belakang kata
  7. Menghapus kata kurang dari empat karakter  
Kata yang kurang dari empat karakter biasanya merupakan kata yang tidak memiliki makna sangat berarti. Oleh karena itu, dilakukan proses pelepasan. Contohnya seperti: ya, ok, iya, dan sebagainnya.
  8. Menggabung kata negasi  
Tahapan ini dilakukan untuk menggabungkan kalimat yang terdapat kata "tidak" dengan kalimat setelanya. Karena "tidak" merupakan kalimat negasi yang bermakna berlawanan dengan makna sebenarnya.
  9. Normalisasi beberapa kata menjadi satu makna

Banyak terdapat kata yang berbeda namun memiliki makna yang sama. Agar kata tersebut dianggap berbeda maka diperlukan normalisasi beberapa kata menjadi satu makna. Oleh karena itu, perlu dilakukan pengumpulan kata yang sama untuk digunakan sebagai *base dictionary*.

### 3.3 Pelabelan Data

Data yang didapatkan belum memiliki kelas, oleh karena itu pada tahap ini pemberian label untuk mengetahui kelas dari setiap teks. Label memiliki tiga kelas antara lain, positif, netral, dan negatif.

### 3.4 Analisa Data

Analisa data dilakukan dengan tiga tahapan pendekatan yaitu, analisa deskriptif. Pendeskripsi data dilakukan untuk mengetahui pola yang terdapat didalam kumpulan data. Melakukan scoring dengan visualisasi untuk mengetahui dari teks berapa persentase yang memiliki komentar maupun status yang berkonotasi positif, netral, dan negatif. Klasifikasi teks dilakukan dengan menerapkan metode *support vector machine* (SVM) dengan tahapan sebagai berikut, melakukan indexing, weighting, training, dan testing. Melakukan pengujian kinerja algoritma. Pengukuran terhadap kinerja sistem klasifikasi merupakan hal yang penting, kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasi data. Confusion matrix mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan sistem dengan hasil klasifikasi yang seharusnya[16].

## 4. Analisa Hasil Dan Pembahasan

Pada penelitian ini dilakukan pengujian analisa deskriptif, pengujian klasifikasi, dan pengujian terhadap hasil klasifikasi. Penegujian deskriptif dengan cara visualisasi sentimen yang terbagi menjadi tiga kelompok yaitu, sentimen positif, sentimen netral, dan sentimen negatif. Selanjutnya, melakukan uji klasifikasi dengan *support vector machine* untuk mendapatkan model learning. Pada tahap terakhir melakukan pengujian terhadap model klasifikasi yang telah dibuat.

#### 4.1 Skenario Pengujian

Data yang digunakan pada penelitian ini adalah bersumber dari sosial media twitter dengan melakukan proses ekstraksi data melalui twitter API. Gambar 5. *source code* untuk melakukan data *collecting*.

```
import tweepy
from tweepy import OAuthHandler
import csv

consumer_key = 'rFXtLHcp100PD9q8knwT0Mdf4'
consumer_secret = '4rNUdsyqlrfwgWVYo67jkq3IgUH667MrAQRXDDaX8XzuGxX1k'
access_token = '1141908353683116032-Zz8ulepzkhdsjeqAvilrtrunFcNS'
access_secret = 'T0jwLyZxSANKxF7bqfIx0Vh2t14RGMzHMWmmWtd164GmD'
auth = OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)
api = tweepy.API(auth, wait_on_rate_limit=True)
csvFile = open('#data.csv', 'a')
csvWriter = csv.writer(csvFile)
for tweet in tweepy.Cursor(api.search,
                           q = '#wiranto',
                           count=50,
                           lang="id",
                           since = "2019-01-01").items():
    print(tweet.created_at, tweet.text)
    csvWriter.writerow([tweet.created_at, tweet.text.encode('utf-8')])
```

Gambar 5. Pengumpulan Data

Didapatkan data sebanyak 10000 baris postingan dan komentar, Tabel 3. Data Sebelum melakukan pembersihan adalah merupakan bentuk data yang didapatkan dengan menampilkan lima baris data ditampilkan.

Tabel 3. Dataset sebelum melakukan pembersihan

No	Text
1	2019-10-12 01:06:45 Kalau takmir masjid ditusuk, penyerang disebut orang gila.\n\nKalau pejabat negara ditusuk, apkh penyerang disebut orang\ufe0f\x80\xaf"
2	2019-10-11 17:20:42 Entah apa yang merasuki mu\nInnalilahi, kuyang, Wiranto\n#Innalillahi #Wiranto #kuyang Menkopolhukam Pak Wiranto <a href="https://t.c/xe2\x80\xaf">https://t.c/xe2\x80\xaf</a> "
3	2019-10-11 17:04:00 Semoga Wiranto tidak menjadi beban ""negara"". Seperti kata dia ke Saudara saudara di Ambon.\n\n#Wiranto \n#Wirantoditusuk <a href="https://\xe2\x80\xaf">https://\xe2\x80\xaf</a> "
4	2019-10-11 16:49:41 Tidak suka pada Wiranto, scr pribadi maupun kebijakannya, ya terserah saja.\n\nMembela teroris, menutupinya dg segala dalih\ufe0f\x80\xaf"
5	2019-10-11 14:33:00 Selebgram Karin Novilda alias Awkarin turut bersimpati atas musibah yang menimpa Wiranto. #KAMUHARUSTAU #awkarin\ufe0f\x80\xaf" <a href="https://t.co/OT1tHbB11D">https://t.co/OT1tHbB11D</a>

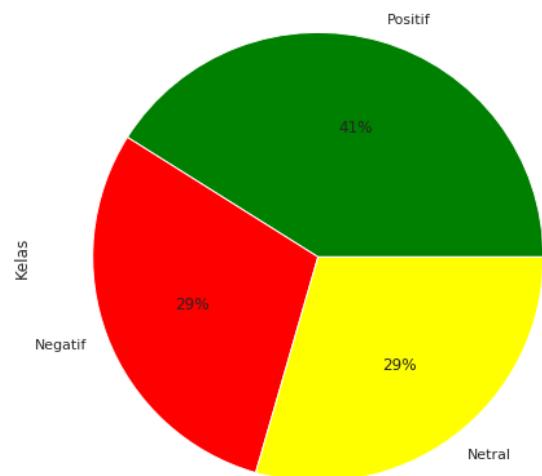
Dataset yang didapatkan kemudian dibersihkan dengan teknik preprosesing agar data lebih mudah untuk diberikan kategorisasi. Kategori dibagi menjadi tiga kelas yaitu, positif,

netral, dan negatif. Tabel 4. Merupakan hasil data yang sudah dibersihkan dan kategorisasi.

Tabel 4. Dataset setelah dibersihkan

No	Text	Label
1	kalau takmir masjid ditusuk penyerang disebut orang gila kalau pejabat negara ditusuk, apkh penyerang disebut orang	Negatif
2	entah apa yang merasuki mu inalilah kuyang wiranto inalilah wiranto kuyang menkopolhukam pak wiranto	Negatif
3	semoga wiranto tidak menjadi beban negara seperti kata dia kepada saudara saudara ambon	Negatif
4	tidak suka pada wiranto pribadi maupun kebijakannya terserah saja membela teroris menutupi segala dalih	Negatif
5	seleb karin novilda alias awkarin turut bersimpati atas musibah menimpa wiranto	Positif

Dataset yang telah melalui pase preprosesing kemudian divisualisasi dengan melakukan scoring pada atribut label. Hasil visualisasi pada gambar 4. Visualisasi sentimen analisis pada teks yang di dapatkan. Dari hasil ini menunjukkan bahwa 41% merasa simpati terhadap kejadian, dengan memberikan postingan dan komentar-komentar yang berkonotasi positif. Berdasarkan hasil pengujian sentimen netral sebanyak 29% dengan komentar berkonotasi netral dan 29% sentimen masa berkonotasi negatif.



Gambar 6. Visualisasi sentimen netizen

#### 4.2 Skenario pemodelan

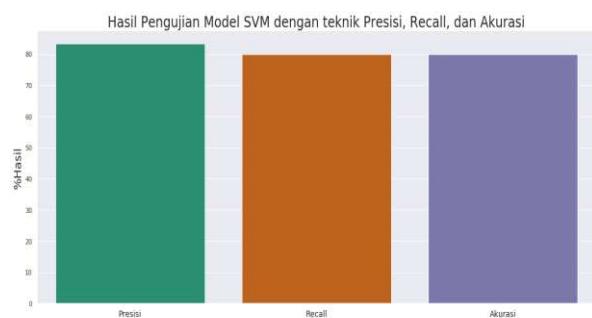
Pada tahapan ini menjelaskan pemodelan pelatihan data untuk membuat model klasifikasi. Data dibagi menjadi dua bagian yaitu, data training dan data testing. Dari total jumlah data yaitu 10000 data sebanyak 70% atau sebanyak 7000 digunakan sebagai data latih atau data training. Kemudian 30% sisanya atau sebanyak 3000 data yang digunakan sebagai data pengujian atau data testing.

#### 4.3 Pengujian model

Model yang telah dibuat menggunakan algoritma svm kemudian dilakukan tahap evaluasi untuk mengetahui hasil seberapa baik metodologi, algoritma dan teknik yang digunakan pada penelitian ini. Pengujian model dilakukan dengan menggunakan teknik pengujian dengan confusion matrix dengan melihat nilai presisi, recall, dan nilai akurasi. Hasil pengujian pada model yang telah dibuat dengan algoritma SVM adalah sebagai berikut, nilai presisi 83%, nilai recall sebanyak 80%, dan nilai akurasi 80%, gambar 7. menujukkan visualisasi hasil pengujian model.

	precision	recall	f1-score	support
Negatif	0.97	0.83	0.89	867
Netral	0.63	0.88	0.74	919
Positif	0.89	0.72	0.79	1214
accuracy			0.80	3000
macro avg	0.83	0.81	0.81	3000
weighted avg	0.83	0.80	0.80	3000
Total Akurasi	0.8			

Gambar 7. Hasil pengujian model klasifikasi



Gambar 8. Bar Chart Hasil pengujian model klasifikasi

#### 5. Kesimpulan Dan Saran

##### 1. Kesimpulan

Pada penelitian ini, telah berhasil dibangun model untuk melakukan klasifikasi pada text. Model dibangun dengan menggunakan algoritma *support vector machine* dengan metode nonlinier. Model yang telah dibuat kemudian diuji dengan cara menghitung nilai presisi, recall, dan nilai akurasi. Hasil evaluasi menunjukkan bahwa model yang dibangun cukup baik dalam melakukan klasifikasi dengan masing-masing diperoleh nilai presisi sebesar 83%, nilai recall sebesar 80%, dan didapatkan nilai uji presisi sebesar 80%.

##### 2. Saran

Pada penelitian ini masih perlu untuk dilanjutkan agar mendapatkan model yang lebih baik. Penelitian selanjutnya agar mempertimbangkan jenis kernel yang digunakan. Selain itu, data preprosesing diperlukan untuk membuat dataset yang digunakan lebih baik agar terhindar dari kalimat-kalimat yang ambigu. Pada penelitian selanjutnya dapat menerapkan metode-metode klasifikasi yang lebih baik dalam melakukan klasifikasi pada data yang berukuran besar.

#### Daftar Pustaka

- [1] Oracle, "The Definition of Big Data," 2014. [Online]. Available: <https://www.oracle.com/big-data/guide/what-is-big-data.html>.
- [2] G. Szabo, G. Polatkan, O. Boykin, and A. Chalkiopoulos, *Social Media Data Mining and Analytics*. Indianapolis: Wiley, 2019.
- [3] Ri. S. I. Nissa and S. Anggreni, "Kasus Penusukan Wiranto, Butuh Berapa Lama Luka Tusuk Sembuh Total," *Suara*, 2019. [Online]. Available: <https://www.suara.com/health/2019/10/1/142753/kasus-penusukan-wiranto-butuh-berapa-lama-luka-tusuk-sembuh-total>.
- [4] C. G. Reddick, A. Takeoka, and A. Ojo, "A social media text analytics framework for double-loop learning for citizen-centric public services: A case study of a local government Facebook use," *Gov. Inf. Q.*, vol. 34, no. 1, pp. 110–125, 2017.
- [5] T. Aleti, J. I. Pallant, A. Tuan, and T. Van Laer, "ScienceDirect Tweeting with the Stars: Automated Text Analysis of the Effect of Celebrity Social Media Communications on Consumer Word of Mouth," *J. Interact. Mark.*, vol. 48, no. 1, pp. 17–32, 2019.

- [6] A. Sun, E. Lim, and Y. Liu, "On strategies for imbalanced text classification using SVM : A comparative study," *Decis. Support Syst.*, vol. 48, no. 1, pp. 191–201, 2009.
- [7] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM Neural Network for Text Classification," *Smantic Scholar*, 2015. [Online]. Available: <https://www.semanticscholar.org/paper/A-C-LSTM-Neural-Network-for-Text-Classification-Zhou-Sun/10f62af29c3fc5e2572baddca559ffbfd6be8787>. [Accessed: 26-Nov-2019].
- [8] X. Zhang, J. Zhao, and Y. Lecun, "Character-level Convolutional Networks for Text," *Computer Science Machine Learning*, 2015. [Online]. Available: <https://arxiv.org/abs/1509.01626>. [Accessed: 10-Nov-2019].
- [9] G. Feng, J. Guo, B. Jing, and T. Sun, "Feature subset selection using naive Bayes for text classification," *Pattern Recognit. Lett.*, vol. 65, no. 1, pp. 109–115, 2015.
- [10] W. Heyong and H. Ming, "Supervised Hebb rule based feature selection for text classification," *Inf. Process. Manag.*, vol. 56, no. October 2018, pp. 167–191, 2019.
- [11] Y. Data and D. Sarkar, *Text Analytics with Python*. Bangalore: apress, 2016.
- [12] M. J. . Betty and G. S. Linoff, *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management*, Second Edi. Indianapolis: Wiley Publishing, 2004.
- [13] J. Han and M. Kamber, *Data Mining Concept and Technique*. San Fransico: Morgan Kaufman, 2006.
- [14] E. T. Qasthari, "Teknik Pengukuran : Metode Klasifikasi *Support vector machine* ( SVM ) pada Data Pengukuran," *github*, 2017. [Online]. Available: <https://eufat.github.io/docs/teknik-pengukuran-2.pdf>. [Accessed: 20-Nov-2019].
- [15] R. Diani, U. N. Wisesty, and A. Aditsania, "Analisis Pengaruh Kernel *Support vector machine* ( SVM ) pada Klasifikasi Data Microarray untuk Deteksi Kanker," *J. Comput.*, vol. 2, no. 1, pp. 109–118, 2017.
- [16] E. Prasetyo, *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi, 2012.
- [17] Imtihan, K., Hadawiyah, R., & Lombok, H. A. S. (2018). Sistem Informasi Penggajian Guru Honorer Menggunakan Konsep Agile Software Development dengan Metodologi Extreme Programming (XP) pada SMK Bangun Bangsa. *IJNS-Indonesian Journal on Networking and Security*, 7(2).