

Analisis Sentimen pada Twitter Tentang Calon Presiden 2019 Menggunakan Metode SVM (*Support Vector Machine*)

Okky Zoellanda A.Tane, Kemas Muslim Lhaksana,
Fhira Nhita

Fakultas Informatika, Universitas Telkom, Bandung
okkyzoellandaatane@students.telkomuniversity.ac.id, kemasmuslim@telkomuniversity.ac.id,
fhiranhita@telkomuniversity.ac.id,
Prodi S1 Ilmu Komputasi, Fakultas Informatika, Universitas Telkom

Abstrak

Twitter mempunyai banyak fitur yang bisa digunakan. Salah satu fitur tersebut adalah user bisa membagikan tweet atau berita berupa tulisan, foto, atau video. Data pada informasi yang bermanfaat salah satunya yaitu analisis sentimen. Untuk mengetahui seberapa besar tanggapan dari masyarakat terkait dengan pemilihan umum presiden Indonesia 2019, apakah lebih condong kepada sentimen positif atau sentimen negatif. Salah satu metode analisis sentimen adalah SVM (*Support Vector Machine*). Analisis sentimen pada media sosial twitter terhadap calon presiden pada pemilu 2019 menggunakan metode SVM dilakukan pengujian terhadap beberapa persentase data yang beragam. Dari pengujian didapatkan nilai akurasi sebesar 91.5 %. Beberapa faktor yang memengaruhi tingkat nilai akurasi tersebut adalah persentase jumlah data training dan data testing, proses pre-processing data yang dilakukan, serta persentase jumlah data positif dan data negatif.

Kata kunci: twitter, SVM(*support vector machine*), media sosial, pemilu

Abstract

Twitter has many features that can be used. One of these features is that users can share tweets or news in the form of text, photos or videos. One of the useful data on information is sentiment analysis. To find out how much the response from the public is related to the 2019 Indonesian presidential election, is it more inclined to positive sentiment or negative sentiment. One method of sentiment analysis is SVM (*Support Vector Machine*). The sentiment analysis on social media twitter on presidential candidates in the 2019 election using the SVM method was tested on various percentage of data. From the test, the accuracy value was 91.5%. Several factors that influence the level of accuracy value are the percentage of training data and testing data, the data pre-processing process performed, and the percentage of positive data and negative data.

Keyword: twitter, social media, SVM(*Support Vector Machine*), presidential election

1. Pendahuluan

1.1 Latar belakang

Media sosial menjadi media referensi bagi masyarakat pada saat ini[1]. Internet dan jejaring sosial seperti Twitter kini telah menjadi medium penting sebagai arena promosi dan komersialisasi produk dan budaya. Pola promosi tampak kian massif melalui situs online hingga akun jejaring sosial baik yang dilakukan oleh perseorangan, ataupun kelompok pemerintah, dan juga korporasi. Tidak hanya itu, dalam proses komunikasi politik pun media sosial telah dilirik oleh para elit politik. Berbagai kepala Negara dan menteri serta anggota parlemen mulai membuat akun dan bekicau di Twitter.

Pertarungan informasi politik dan pesan pesan politik di internet terutama sosial media seperti Facebook dan Twitter sudah tidak dapat dihindari lagi. Sudah saatnya penelitian mengenai perilaku di Twitter kini berfokus pada konten apa yang mereka keluarkan (posting). Pentingnya meneliti konten dalam komunikasi adalah karena konten dipercaya memiliki efek. Beberapa penelitian sebelumnya yang meneliti tweet yang berasal dari politikus ketika menggunakan Twitter kebanyakan lebih berfokus pada pemimpin partai atau perdana menteri atau presiden sbagai objek yang diteliti “[Glassma et al,2010:219-233; Grant et al 2010:579-604]” dibandingkan politikus yang mencalonkan atau diajukan ke pemilihan umum berlangsung, dan bahaya beberapa yang menggunakan topic apa saja yang dikeluarkan oleh para politikus di Twitter mereka[2]. Supervised Learning adalah pembelajaran yang menggunakan dua jenis data training dataset dan testing dataset. Supervised learning mempelajari data training sehingga mendapatkan hasil yang paling baik di data testing[3]. Pada penelitian ini supervised learning yang digunakan adalah *Support Vector Machine* (SVM)

1.2 Tujuan

Adapun tujuan Tugas Akhir ini adalah sebagai berikut :

1. Bagaimana masyarakat menanggapi pemilihan umum presiden Indonesia 2019 melihat dari sentiment analisisnya
2. Menerapkan metode SVM (*Support Vector Machine*) pada data twitter
3. Mengetahui tingkat akurasi pada metode SVM (*Support Vector Machine*), dan hasil perhitungan yg didapatkan.

2. Studi Terkait

2.1 Media Sosial Twitter

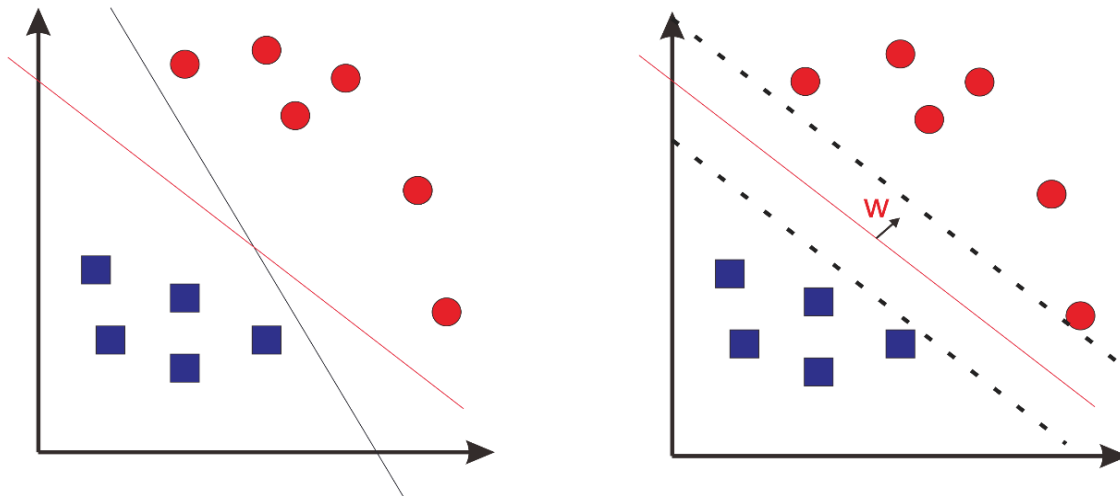
Pada dasarnya media sosial merupakan perkembangan mutakhir dari teknologi-teknologi web baru berbasis internet, yang memudahkan semua orang untuk dapat berkomunikasi, berpartisipasi, saling berbagi dan membentuk sebuah jaringan secara online, sehingga dapat menyebarluaskan konten mereka sendiri. Post di blog, tweet, atau video YouTube dapat direproduksi dan dapat dilihat secara langsung oleh jutaan orang secara gratis[4]. Media sosial mempunyai banyak bentuk, diantaranya yang paling populer yaitu microblogging (Twitter), facebook, dan blog. Twitter adalah suatu situs web yang merupakan layanan dari microblog, yaitu suatu bentuk blog yang membatasi ukuran setiap post-nya, yang memberikan fasilitas bagi pengguna untuk dapat menuliskan pesan dalam twitter update hanya berisi 140 karakter. Twitter merupakan salah satu jejaring sosial yang paling mudah digunakan, karena hanya memerlukan waktu yang singkat tetapi informasi yang disampaikan dapat langsung menyebar secara luas[5].

Twitter adalah sebuah situs web yang dimiliki dan dioperasikan oleh TwitterInc, yang menawarkan jaringan sosial berupa microblog sehingga memungkinkan pengguna untuk mengirim dan membaca pesan tweets[6]. Tweet adalah penulisan teks yang memiliki 140 karakter yang berada di halaman public twitter, pengguna bias membatasi pengiriman atau pesan yang hanya dapat dilihat oleh teman-teman atau pengikut (followers). Pengguna dapat memeriksa orang-orang yang membatalkan pertemanan (unfollowing) melalui layanan yang ada di dalam twitter. Pengguna dapat menulis pesan dan mengelompokannya berdasarkan topic atau jenis pesan dengan menggunakan tagar “#” (hashtag). Pengguna juga dapat menghubungkan pesan tweet mereka dengan orang lain menggunakan tanda “@” yang diikuti dengan nama pengguna yang akan dihubungkan untuk mengirim atau membalas pengguna lain[7].

2.2 Algoritma SVM (Support Vector Machine)

Pada dijelaskan bahwa SVM pertama kali diperkenalkan oleh Guyon Boser dan Vapnik. SVM adalah salah satu cara untuk melakukan klasifikasi dengan menggunakan teknik regresi. Dengan kata lain, SVM adalah alat prediksi klasifikasi dan regresi yang menggunakan teori dari *machine learning* untuk memaksimalkan akurasi dengan cara menghindari *overfit* pada data secara otomatis. *Support Vector Machine* dapat juga didefinisikan sebagai sistem yang menggunakan ruang hipotesis dari fungsi linier yang didapat dari algoritma yang diturunkan dari *statistical learning*[12].

Ide dari SVM ini adalah membagi dua buah data sehingga terbentuk dua buah kelas. Tetapi untuk membagi data menjadi dua kelas *hyperplane* yang memungkinkan cukup banyak dan harus ditentukan *hyperplane* yang terbaik untuk membagi kelas tersebut. SVM adalah salah satu algoritma klasifikasi terfavorit karena selain mempunyai dasar teori yang kuat SVM juga mengklasifikasi lebih akurat dari algoritma lainnya di banyak bidang seperti klasifikasi *web page* dan bioinformatika. Dalam gambar 1[8].



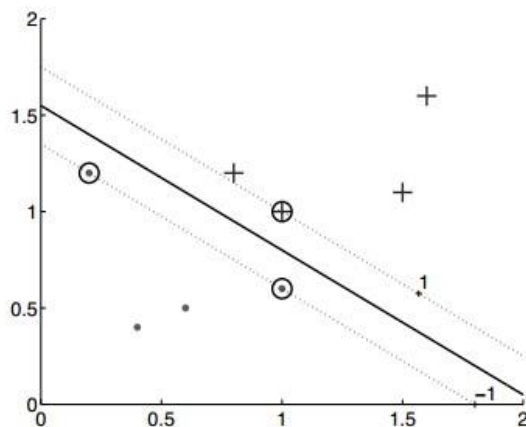
Gambar 1. Ilustrasi cara kerja SVM

Hyperplane terbaik dicari dengan mengukur jarak margin antara dua kelas. Dalam hal ini, kelas dinotasikan sebagai $y_i \in \{-1, +1\}$ yaitu $i = 1, 2, \dots, l$ dengan l adalah banyaknya data. Jarak antara data terdekat dengan margin disebut *halfspace*, dinotasikan sebagai b . Diasumsikan bahwa kedua kelas terpisah sempurna oleh *hyperplane* yang berdimensi d . Fungsi *hyperplane* dapat dicari dengan rumus sebagai berikut

$$f: \vec{w} \cdot \vec{x} + b = 0 \tag{1.1}$$

Keterangan: 1

- \vec{w} = Paramater *hyperplane* yang dicari (1,0)
- \vec{x} = data input SVM
- b = nilai bias sebagai parameter *hyperplane*



Gambar 2 Ilustrasi cara kerja SVM

Hyperplane yang paling optimal adalah yang memiliki margin terbesar. Dengan margin yang besar, maka kita dapat menghindari kesalahan dalam pengklasifikasian data [11]. Contohnya, jika terdapat sebuah data baru pada kelas “+” yang posisinya berada sedikit lebih dekat ke kelas “•”, maka data tersebut masih dapat diklasifikasikan ke kelas yang tepat.

Jika kita memiliki suatu data $X = \{x_t, r_t\}$ dengan 2 kelas berbeda, yaitu C_1 dan C_2 , di mana $r_t = +1$ jika $x_t \in C_1$ dan $r_t = -1$ jika $x_t \in C_2$, persamaan garis *hyperplane* dapat dinyatakan sebagai berikut:

$$w^T x + w_0 = 0 \quad (1.2)$$

Dengan w dan w_0 adalah parameter dari model. Nilai w dapat dinyatakan dengan rumus berikut:

$$w = \sum_{t=1}^N \alpha_t r_t x_t \quad (2.2) \quad t = 1, 2, \dots, N$$

Keterangan:

α_t : Lagrange multiplier

r_t : Label kelas, bernilai 1 untuk $x_t \in C_1$ atau -1 jika $x_t \in C_2$

x_t : Fitur dari data

N : jumlah data

Nilai α_t dapat dicari dengan memaksimalkan persamaan berikut:

$$L_d = -\frac{1}{2} \|w\|^2 - \sum_t \alpha_t r_t (w^T x_t) + \sum_t \alpha_t \quad (1.3)$$

$$\text{s.t. } \sum_t \alpha_t r_t = 0 \text{ dan } \alpha_t \geq 0, \forall t$$

Setelah semua nilai α_t didapatkan, beberapa akan memiliki nilai $\alpha_t = 0$ dan hanya sebagian kecil yang memiliki nilai $\alpha_t > 0$. Kumpulan x_t dengan nilai $\alpha_t > 0$ adalah yang disebut dengan support vector. Kemudian nilai w_0 bisa didapatkan dengan rumus berikut:

$$w_0 = r_t - w^T x_t \quad (1.4)$$

Untuk menguji kelas data *testing* dapat menggunakan rumus berikut:

$$g(x) = \text{sign}(w^T x + w_0) \quad (1.5)$$

jika $g(x) = 1$ pilih C_1 , atau jika $g(x) = -1$ pilih C_2 .

2.3 Ekstraksi Fitur.

Feature Extraction atau ekstraksi fitur merupakan suatu pengambilan ciri/feature dari suatu bentuk yang nantinya nilai yang didapatkan akan dianalisis untuk proses selanjutnya. Ekstraksi fitur didapat dari melihat bentuk-bentuk umum yang ada di kalimat sitasi dan pengambilan dari penelitian sebelumnya, sehingga didapatkan fitur-fitur sebagai berikut[9].

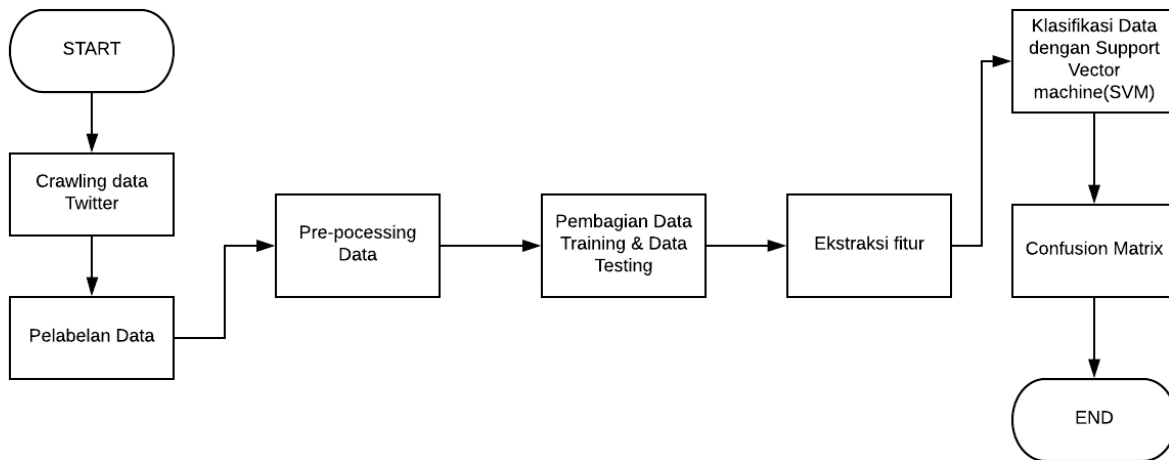
'adil'
'agama'
'allah'
'ancam'
berbahaya'
'berita'
buta'
cebong'
'curang'
'damai'

2.4 Analisis Sentimen

Sentiment Analysis adalah penambangan kontekstual teks yang mengidentifikasi dan mengekstrak informasi subjektif dalam sumber, dan membantu para pembisnis untuk memahami sentimen sosial dari merek, produk atau layanan mereka saat memantau percakapan online[10].

3. Perancangan Sistem

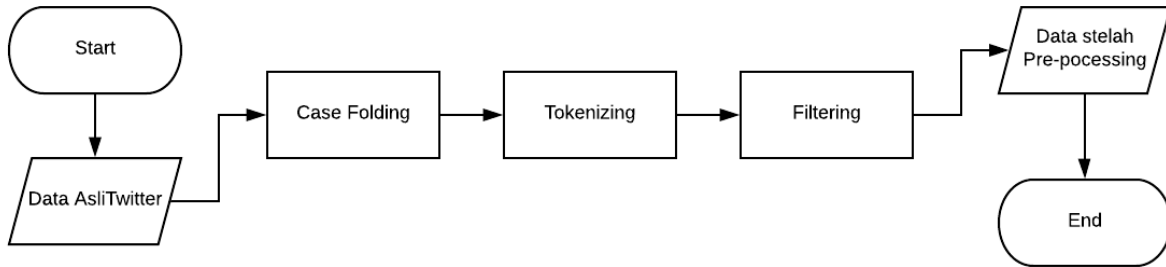
3.1 Pembangunan model



Gambar 3. Alur Pembangunan Model

Pada proses pembangunan model pada tugas akhir ini dapat dilihat pada gambar 2 di atas. Proses pertama kali dilakukan dalam pembangunan model ini adalah crawling data twitter bias disebut pengumpulan data dari medio sosial twitter, setelah didapat dari hasil crawling data selanjutnya dilakukan pelabelan secara manual dengan masing masing kelas sentiment nya yaitu positif dan negatif.Data yang telah dilakukan pelabelan berjumlah 476 data. Setelah proses pelabelan dilakukan porses Pre-pocessing. Preprocessing

dilakukan untuk membuat data menjadi lebih baik adri hasil sebelumnya untuk dilakukan klasifikasi.Karena data yang di peroleh dari crawling data adalah data *noise* atau bentuk data yang tidak terstruktur secara rapi.



Gambar 4. Alur Pada Proses Pre-processing

Pada gambar diatas dapat dilihat alur pada Pre-processing untuk mendapatkan hasil data yang terstruktur rapi

3.3 Skenario Pengujian Support Vector Machine (SVM)

Data set yang berjumlah 476 dibagi ke dalam dua kategori yaitu data training dan data testing. Skenario pengujian dilakukan untuk melihat perbandingan jumlah data yang akan dimasukkan pada dua kategori yaitu data training dan data testing. Berikut adalah tabel untuk skenario pengujian dalam bentuk persentase.

Tabel 1. Jumlah data berdasarkan label positif dan negatif

No.	Label	Jumlah Tweet
1	Positif	253
2	Negatif	223

Jumlah porsi data training dan data testing untuk skenario pengujian dapat dilihat pada tabel 2 berikut :

Tabel 2. Jumlah porsi data training dan data testing

No.	Data Training	Data Testing
1	55 % (262 data training)	45 % (214 data testing)
2	60 % (286 data training)	40 % (190 data testing)
3	70 % (333 data training)	30 % (143 data testing)
4	80 % (381 data training)	20 % (95 data testing)

3.4 Confusion Matrix

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya confusion matrix mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya

- TN : merupakan data negatif yang terdeteksi dengan benar
- FP : merupakan data negatif namun terdeteksi sebagai data positif
- TP : merupakan data positif terdeteksi benar
- FN : merupakan data positif terdeteksi sebagai data negative

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{2.1}$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\% \tag{2.2}$$

$$\text{Recall} = \frac{TP}{FN+TP} \times 100\% \tag{2.3}$$

4. Hasil dan Analisis

4.1 Skenario Pengujian

4.1.1 Skenario Pengujian 1 (55% : 45 %)

Perbandingan data training dan data testing : 262 data training dan 214 data testing. Untuk jumlah data skenario pengujian 1 bisa dilihat pada tabel 3 berikut :

Tabel 3. Skenario pengujian 1

Kelas Aktual	Kelas klasifikasi	
	Positif	Negatif
Positif	115	0
Negatif	21	78

4.1.2 Skenario Pengujian 2 (60% : 40 %)

Perbandingan data training dan data testing : 286 data training dan 190 data testing. Untuk jumlah data skenario pengujian 2 bisa dilihat pada tabel 4 berikut :

Tabel 4. Skenario pengujian 2

Kelas Aktual	Kelas klasifikasi	
	Positif	Negatif
Positif	98	0
Negatif	23	69

4.1.3 Skenario Pengujian 3 (70% : 30%)

Perbandingan data training dan data testing : 333 data training dan 143 data testing. Untuk jumlah data skenario pengujian 3 bisa dilihat pada tabel 5 berikut :

Tabel 5. Skenario pengujian 3

Kelas Aktual	Kelas klasifikasi	
	Positif	Negatif
Positif	86	0
Negatif	14	43

4.1.4 Skenario Pengujian 4 (80% : 20%)

Perbandingan data training dan data testing : 381 data training dan 95 data testing. Untuk jumlah data skenario pengujian 4 bisa dilihat pada tabel 6 berikut :

Tabel 6. Skenario pengujian 4

Kelas Aktual	Kelas klasifikasi	
	Positif	Negatif
Positif	41	0
Negatif	13	41

4.2 Hasil Confusion Matrix

Hasil perhitungan confusion matrix bisa dilihat pada tabel 7 berikut :

Tabel 7. Tabel Hasil Confusion Matrix

Confusion Matrix	Porsi Data Training dan Data Testing			
	55 % : 45% (262 : 214)	60 % : 10 % (286 : 190)	70 % : 30% (333 : 143)	80 % : 20% (381 : 95)
Akurasi	91,5	90	86,7	86,3
Presisi	100	100	100	100
Recall	87,8	83,7	79,1	75,9

4.3 Analisis

Dari hasil pengujian dari data diatas, dapat disimpulkan system ini dapat menghasilkan tingkat akurasi yang cukup baik dengan akurasi sebesar 91,5T% menggunakan confusion matrix. Pada proses Pre-processing data yg sudah diolah berpengaruh terhadap proses yang dilakukan untuk menentukan beberapa ekstraksi fitur yang membantu menentukan kriteria nilai dari data sebelumnya.Pada proses yang telah dilakukan dalam tahap skenario penguji didapatkan semakin banyak data testing yang dipakai semakin besar tingkat akurasinya. Dapat dilihat pada table 7 Hasil Confusion Matrix di atas.

5. Kesimpulan

Pada penelitian ini, dibangun sistem untuk mengimplementasikan Algoritma SVM untuk analisis sentimen pada data tweet tentang PASLON PILPRES 2019 dengan keyword “jokowi” dan “prabowo”.Hal ini diperlihatkan dengan hasil akurasi menggunakan data twitter sebanyak 476 data,dimana dalam penelitian ini menggunakan ekstraksi fitur untuk pengambilan suatu ciri dari suatu bentuk yang nantinya nilai yang didapatkan akan dianalisis untuk proses selanjutnya.

Hasil dari pengamatan sentimen masyarakat terhadap rangkaian proses pemilu dan terhadap PASLON PILPRES 2019 adalah sebesar 253 data menunjukkan sentimen positif dan 223menunjukkan sentiment negatif. Dalam perhitungan metode Support Vector Machine (SVM) dan dilanjutkan dengan metode *Confusion Matrix* menghasilkan 91,5% hasil akurasi yang di dapat. Hal ini menunjukkan perhitungan proses sentiment analisi menggunakan Support Vector Machine (SVM) menghasilkan akurasi yang cukup baik.

DAFTAR PUSTAKA

- [1] M. e. a. Lister, A Critical Introduction, Second Edition, New York, 2009, .
- [2] E. Qualman, "Socialnomics: How Social Media Transforms the Way We Live and Do Business.," New Jersey, Wiley, John & Sons, Incorporated, 2009, .
- [3] T. M. B. ., K. H. & G. v. ' . H. Graham, "Between Broadcasting Political Messages And Interacting With Voters. Journal of Information, Communication & Society," 2013, .
- [4] E. Learned-Miller, " "Introduction to Supervised Learning", " 2014.
- [5] D. Zarella, The Social Media Marketing Book, Jakarta: PT Serambi Ilmu Semesta Anggota IKAPI, 2010, .
- [6] D. Zarella, The Social Media Marketing Book, Jakarta: PT Serambi Ilmu Semesta Anggota IKAPI, 2010.
- [7] Twitter, "Support Twitter", 2013.
- [8] Ammar dan Shadiq, "*Keoptimalan Naïve Bayes Dalam Klasifikasi*" .
- [9] Junita, Beby Dwi (2017) Ekstraksi fitur dan klasifikasi menggunakan metode glcm dan svm pada citra mammogram untuk identifikasi kanker Payudara. Teknologi Rekaya.
- [10] Burhanudin, Rahmat (2018,9 May).*Mengenal Sentiment Analisis* .Dikutip 21 Juli 2019 Mamat Programer: <https://mamat.co/mengenal-sentiment-analysis/>
- [11] Ghifar. (2015, 5 Juni). Support Vector Machine: Penjelasan matematis dan intuitif. Dikutip 21 Juli 2019 <https://ghifar.wordpress.com/2015/06/05/support-vector-machines-penjelasan-matematis-dan-intuitif/>
- [12] R. WP, A. Novianty S.T., M.T, and C. Setianingsih S.T., M.T, "Analisis Sentimen Menggunakan Support Vector Machine dan Maximum Entropy,"