

IMPLEMENTASI *SUPPORT VECTOR MACHINE* PADA ANALISA SENTIMEN TWITTER BERDASARKAN WAKTU

Faisal Rahutomo¹, Imam Fahrur Rozi², Haris Setiyono³

^{1,2,3}Prodi Teknik Informatika Politeknik Negeri Malang

^{1,2,3}Jl. Soekarno Hatta No.9, Jatimulyo, Kec. Lowokwaru, Kota Malang, Jawa Timur 65141

E-mail : faisal@polinema.ac.id¹ , imam.rozi@polinema.ac.id² , harissetiyono@gmail.com³

Abstrak

Analisis sentimen merupakan salah satu cabang ilmu dari *data mining* yang bertujuan untuk menganalisis, memahami, mengolah, dan mengekstrak data tekstual yang berupa opini terhadap entitas seperti produk, servis, organisasi, individu, dan topik tertentu. Dalam menentukan kategori positif, negatif atau netral suatu tanggapan masyarakat di twitter dapat dilakukan dengan manual dengan cara membaca setiap *tweet*. Hal ini tentu membutuhkan banyak waktu dan menghabiskan banyak tenaga. Pada penelitian ini menggunakan algoritma klasifikasi *Support Vector Machine* untuk melakukan klasifikasi data tweet menjadi sentimen positif, negatif atau netral. Analisa dilakukan berdasarkan rentang waktu tertentu, karena setiap waktu dapat memiliki topik pembahasan yang berbeda dan dari hasil data tersebut dapat dilihat perkembangan tren sentimen dan dapat dilihat bagaimana tanggapan masyarakat terhadap topik tersebut. Data tweet didapatkan dengan cara *crawling* secara periodik dengan target kata kunci nama capres dan cawapres pada pemilu 2019. *Dataset* yang digunakan dalam penelitian ini menggunakan 600 tweet. Pada pengujian hasil klasifikasi menggunakan *k-fold cross validation* dengan membagi menjadi 10 bagian data, nilai rata-rata akurasi 66%, presisi 67% dan *recall* 66%.

Keywords: Analisis sentimen twitter, pemilu 2019, support vector machine

Abstract

Sentiment analysis is one branch of science from data mining that aims to analyze, understand, process, and extract textual data in the form of opinions on entities such as products, services, organizations, individuals, and certain topics. In determining positive, negative or neutral categories, a public response on twitter can be done manually by reading each tweet. This certainly requires a lot of time and takes a lot of energy. In this study using the Support Vector Machine classification algorithm to classify tweet data into positive, negative or neutral sentiments. Analysis is carried out based on a certain time span, because each time can have a different topic of discussion and from the results of these data can be seen the development of sentiment trends and can be seen how the public response to a particular topic. The tweet data is obtained by crawling periodically with the target keywords of the names of candidates and vice president in the 2019 election. The dataset used in this study uses 600 tweets. In testing the classification using k-fold cross validation by dividing into 10 data parts, average value of 66% accuracy, 67% precision and 66% recall.

Keywords: Twitter sentiment analysis, 2019 election, support vector machine

I. PENDAHULUAN

Perkembangan teknologi informasi yang sangat cepat sangat berpengaruh dengan kehidupan sosial manusia sekarang ini. Salah satunya adalah dengan hadirnya media sosial seperti facebook, twitter, instagram yang dimanfaatkan untuk memudahkan komunikasi dan berbagi informasi. Twitter merupakan salah satu media sosial yang banyak digunakan oleh masyarakat.

Di Indonesia pengguna media sosial aktif mencapai 130 juta pengguna atau 49% dari populasi masyarakat di Indonesia [1]. Indonesia adalah negara yang menganut sistem demokrasi. Sebagai masyarakat dalam menyampaikan opini atau gagasan terhadap suatu masalah yang dihadapi dalam negara biasa dilakukan melalui media social. Ini dilakukan karena kemudahan akses dan kemungkinan besar apa yang disampaikan bisa dibaca langsung oleh yang bersangkutan. Selain itu negara demokrasi ditandai dengan adanya program pemilihan umum terhadap

presiden dan wakil presiden yang diselenggarakan secara periodic. Di Indonesia sendiri dilakukan dalam waktu 5 tahun sekali. Sebagai tokoh calon presiden dan wakil presiden tentu akan mempertimbangkan elektabilitas mereka berdasarkan tanggapan dari masyarakat langsung, mengingat masyarakat yang akan menentukan suara terhadap pilihan mereka. Dahulu untuk menganalisa tanggapan masyarakat lebih sulit karena keterbatasan informasi yang diperoleh. Namun di era sekarang ini data respon masyarakat akan lebih mudah didapatkan melalui media sosial twitter mengingat Indonesia memiliki pengguna yang cukup banyak.

Dalam menentukan kategori positif, negatif atau netral suatu komentar dapat dilakukan dengan manual dengan cara membaca setiap tweet. Hal ini tentu membutuhkan banyak waktu dan menghabiskan banyak tenaga karena terkadang antusias masyarakat yang begitu besar yang menyebabkan akan semakin banyak opini yang harus dibaca untuk

mengklasifikasikan apakah termasuk dalam kategori positif, negatif atau netral. Pada penelitian sebelumnya perancangan klasifikasi tweet berdasarkan sentimen dan fitur calon gubernur DKI Jakarta 2017 yaitu peneliti melakukan klasifikasi sentimen berdasarkan tweet dari masyarakat [2]. Pada penelitian Implementasi twitter *sentiment analysis* untuk *review* film menggunakan algoritma *support vector machine* yang dilakukan oleh Faisal Rahutomo, Pramana Yoga S., Miftahul Agtamas Fidyawan yaitu melakukan klasifikasi dengan menggunakan metode *support vector machine* dengan hasil dengan rata-ratanya adalah 76,06 persen, 76,83 persen, 81,07 persen, dan 83,3 persen [3]. Dari penelitian sebelumnya terkait sentimen twitter masih belum melakukan klasifikasi terhadap data secara berkala. Analisa dilakukan berdasarkan rentang waktu tertentu, karena setiap waktu dapat memiliki topik pembahasan yang berbeda dan dari hasil data tersebut dapat dilihat perkembangan tren sentimen dan dapat dilihat bagaimana tanggapan masyarakat terhadap topik tertentu. Oleh karena itu dalam penelitian ini mengusulkan studi kasus yang digunakan adalah Implementasi *support vector machine* pada analisa sentimen twitter terhadap tokoh calon presiden dan wakil presiden pada pemilihan umum 2019 berdasarkan waktu.

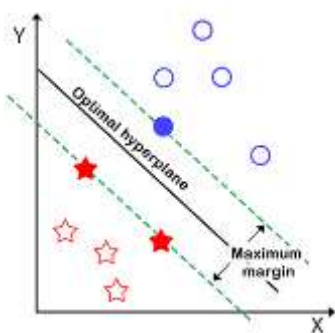
II. LANDASAN TEORI

2.1. Analisa Sentimen

Analisa sentimen adalah sistem komputasi yang mempelajari pendapat, sikap dan emosi terhadap suatu entitas yang dapat mewakili individu, peristiwa atau topik dari sekumpulan teks [4].

2.2. Support Vector Machine

Support Vector Machine pertama kali diperkenalkan oleh Vapnik pada tahun 1992. *support vector machine* adalah algoritma *machine learning* yang bekerja atas prinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah class pada *input space* [5].



Gambar 1. contoh *hyperplane optimal* [3]

Dapat dilihat pada Gambar 1 bahwa terdapat 3 simbol (bintang dan lingkaran) yang terlintas garis putus-putus merupakan *support vector*, proses untuk mencari *hyperplane* terbaik terhadap *support vector* adalah inti dari *support vector machine* ini.

Konsep dari *support vector machine non-linear* adalah mengubah data x yang dipetakan oleh fungsi $\Phi(x)$ ke ruang vektor yang memiliki dimensi lebih tinggi. Pemetaan ini bertujuan untuk merepresentasikan data pada ruang vektor baru [6].

Proses pembelajaran pada *support vector machine* saat menemukan *support vector* hanya bergantung dari *dot product* data yang telah ditransformasi pada ruang baru. Nilai *dot product* dapat dihitung tanpa mengetahui proses transformasi data Φ . Fungsi *kernel* memberikan kemudahan dalam proses pembelajaran *support vector machine* untuk menentukan *support vector* [7]. Fungsi kernel dalam metode ini sering disebut dengan *kernel trick*. *Kernel trick* adalah fungsi yang mengelompokkan data dari dimensi rendah kedalam dimensi tinggi. Salah satu fungsi kernel yang dipakai pada sebuah aplikasi untuk mengatasi masalah metode *support vector machine non-linear* adalah *Kernel Gaussian (radial basis function, RBF)* [8].

$$K(x, x_k) = \exp(-\|x - x_k\|^2 / \sigma^2) \quad (1)$$

Sequential Training support vector machine merupakan salah satu algoritma yang dapat digunakan untuk melakukan proses *training*. Selain algoritma *Sequential Training* dalam metode *support vector machine*, terdapat algoritma lain seperti proses *Quadratic Programming (QP)* dan *Sequential Minimal Optimization (SMO)* [9].

2.3 Support Vector Machine Multiclass

Pada awalnya, *support vector machine* digunakan untuk klasifikasi data ke dalam dua kelas. Dalam perkembangannya, *support vector machine* dapat diperluas untuk klasifikasi *multi* kelas dengan sejumlah skema. Skema multi kelas diantaranya adalah *One-Against-All*, *One-Against-One*, *Decision Direct Acyclic Graph* dan *Adaptive Directed Acyclic Graph*. *One-Against-One (OAO)* merupakan sebuah pendekatan untuk menjawab permasalahan pada *multi-class* yang ada dalam metode *Support Vector Machine*. Dalam penelitian ini pendekatan yang digunakan adalah *One-Against-One*, pendekatan ini bekerja dengan cara menjadikan kelas dari data ke- i bernilai positif dan untuk data ke- j bernilai negatif. Pada Tabel 1 akan dijelaskan contoh permasalahan klasifikasi dengan menggunakan tiga kelas.

Tabel 1 Pendekatan *One-Against-One*

$y_i = 1$	$y_j = -1$	Hipotesis Kernel
Kelas 1	Kelas 2	$f^{1.2}(x) = (w^{1.2})x + b^{1.2}$
Kelas 2	Kelas 3	$f^{2.3}(x) = (w^{2.3})x + b^{2.3}$
Kelas 3	Kelas 1	$f^{3.1}(x) = (w^{3.1})x + b^{3.1}$

Proses klasifikasi menggunakan algoritma *support vector machine*, *support vector machine* mewakili dokumen sebagai points dalam ruang vektor. *support vector machine* berusaha untuk menemukan *hyperplane* paling optimum, sebuah *hyperplane* dikatakan paling optimum jika berada tepat ditengah-tengah kedua kelas sehingga memiliki jarak paling jauh ke data-data terluar di kedua kelas.

Dengan kata lain *hyperplane* tersebut memiliki *margin* maksimum.

Setelah mendapatkan nilai *TF-IDF* proses perhitungan *support vector machine* dapat dilakukan dengan menambahkan hasil *TF-IDF* dengan label, contoh disini memberikan label 1, 0, -1 dimana 1 mewakili sentimen positif, 0 mewakili netral dan -1 mewakili sentimen negatif. Dalam percobaan dalam *dataset* akan dibagi menjadi 2 data yaitu data latih dan data uji. Adapun langkah-langkah perhitungan sebagai berikut :

1. Menghitung *matrix kernel* dengan *kernel gaussian*

$$K(x, x_k) = \exp(-\|x - x_k\|^2 / \sigma^2) \quad (1)$$

2. Melakukan perhitungan nilai *matrix hessian* dengan rumus:

$$D_{ij} = y_i y_j (K(x_i x_j) + \lambda^2) \quad (2)$$

Keterangan :

x_i = Data yang ke- *i*

x_j = Data yang ke- *j*

y_i = Kelas yang ada pada data ke-*i*

y_j = Kelas yang ada pada data ke-*j*

3. Menghitung nilai γ baru dengan membagi nilai inisial γ dengan nilai maksimal dari *matrik hessian*.

4. Menghitung persamaan dibawah ini untuk memperbarui nilai *E (Epsilon)*.

$$E_i = \sum_i^n \alpha_i D_{ij} \quad (3)$$

$$\delta\alpha_i = \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\} \quad (4)$$

$$\alpha_i = \alpha_i + \delta\alpha_i \quad (5)$$

Keterangan :

α_i = Nilai parameter *alpha*

D_{ij} = Hasil nilai *matriks Hessian*

C = Nilai untuk parameter *C*

$\delta\alpha_i$ = Parameter delta alpha untuk ke- *i*

Untuk menentukan nilai *Epsilon (E)* harus melakukan perhitungan iterasi, dimana *alpha (α)* ke-0 = 0.

5. Melakukan langkah 4 hingga iterasi maksimum atau $\max(|\delta\alpha|) < \epsilon$. Setelah proses diatas selesai, maka akan didapatkan α dan *support vector*.

6. Tahap selanjutnya adalah menghitung nilai bias (*b*) Dalam proses mencari nilai bias diharuskan mencari nilai *Weight (W)* dari data latih.

$$b = -\frac{1}{2} (\sum_{i=0}^n \alpha_i y_i K(x_i, x^-) + \sum_{i=0}^n \alpha_i y_i K(x_i, x^+)) \quad (6)$$

7. Menghitung fungsi $f(x)$ pada masing-masing perbandingan kelas.

$$f(x) = \sum_{i=0}^n \alpha_i y_i K(x, x_i) + b \quad (7)$$

8. Setelah ditentukan nilai $f(x)$ maka dilakukan perhitungan dengan rumus *Sign*.

9. Menentukan hasil akhir klasifikasi menggunakan data perbandingan antar kelas. Kemudian menghitung total kemunculan label yang paling banyak.

III. ANALISA SENTIMEN BERDASARKAN WAKTU

Sistem yang dirancang dan dibangun dalam penelitian ini adalah sistem yang mengimplementasikan metode *support vector machine* untuk mengklasifikasikan sentimen pada twitter dengan menargetkan kata kunci tertentu. sistem akan dapat menjalankan *crawling* data, pelabelan data latih, *preprocessing* data, pembobotan menggunakan *TF-IDF*, pembagian data latih dan data uji, klasifikasi dan pengujian klasifikasi. Grafik sentimen akan menampilkan persentase jumlah tiap-tiap kategori sentimen berdasarkan kata kunci terkait dan juga menampilkan tren waktu dengan grafik tren yang dapat dikelompokkan data perhari dalam rentang waktu tertentu.



Gambar 2. Diagram alir sistem

3.1. Pengambilan data

Dalam penelitian ini data didapatkan dari media sosial Twitter dengan memanfaatkan fasilitas *application programming interface (API)* yang disediakan oleh twitter dengan menggunakan fitur *search* pada twitter dengan kata kunci tertentu. Penggunaan *twitter API* diharuskan menggunakan *key* dan *token* rahasia yang diberikan untuk melakukan *authentication*, jika *key* dan *token valid* maka dilakukan langkah selanjutnya sesuai dengan kebutuhan data.

Pengambilan data dilakukan secara berkala setiap menit dengan target 8 kata kunci, yaitu sandiuno, Sandianga Uno, Ma'ruf Amin, MarufAminNU, Jokowi, Joko Widodo, Prabowo atau Prabowo subianto. Pengambilan data secara periodik ini menggunakan fitur *scheduler* pada *framework* Laravel yang dijalankan setiap menit dalam *crawling* data twitter yang kemudian akan disimpan dalam *database*.

3.2. Pengolahan data teks

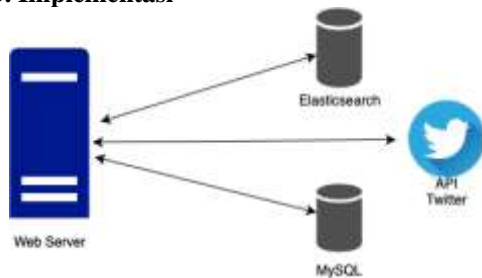
Tahap analisa data dilakukan dengan melalui beberapa proses diantaranya proses *preprocessing*, pembobotan dan klasifikasi sentimen. *Preprocessing* adalah pengolahan data yang bertujuan untuk meminimalisir data / kata yang tidak memiliki nilai dalam objek penelitian dan data yang tidak konsisten. Tahap *preprocessing* dibagi menjadi beberapa proses yaitu *case folding*, *cleaning*, *tokenizing*, *filtering* dan *stemming* [10][11].

- *Case folding* adalah Proses mengubah semua teks dalam dokumen menjadi konsisten, bisa berupa *lowercase* atau *uppercase*.
- *Cleaning* adalah proses penghapusan *url*, *mention* dan *hashtag*.

- *Tokenizing* adalah proses pemotongan kalimat menjadi tiap kata.
- *Filtering* adalah proses hasil *tokenizing* dilakukan penghapusan kata-kata yang kurang memiliki arti dalam tweet yang ada dalam *list stopwords*.
- *Stemming* adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya.

Data hasil dari *preprocessing* kemudian dilakukan pembobotan tiap kata, tujuannya adalah untuk mengetahui jumlah kata yang sering muncul di setiap dokumen atau biasa disebut *term frequency*. Setelah mendapatkan nilai *term frequency* pada dokumen kemudian mencari nilai *inverse document frequency (IDF)*. *IDF* menunjukkan hubungan ketersediaan sebuah *term* dalam seluruh dokumen. Semakin sedikit jumlah dokumen yang mengandung *term* yang dimaksud, maka nilai *IDF* semakin besar.

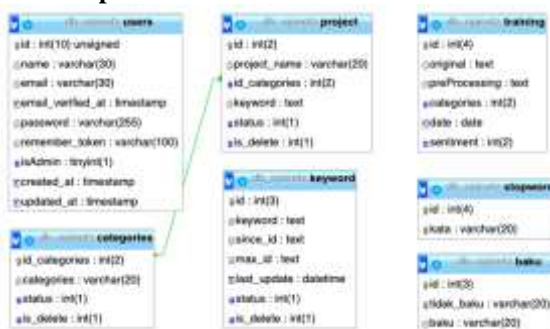
3.3. Implementasi



Gambar 3. Arsitektur sistem

Arsitektur sistem menggunakan 2 *database*, yaitu *database mysql* untuk menyimpan data *project* yang meliputi kata kunci target dan *database elasticsearch* untuk menyimpan data hasil *crawling* twitter. *Web server* menggunakan Apache dan PHP versi 7.2.

3.3.1. Implementasi Database



Gambar 4. Desain database

Implementasi *database* pada sistem ini menggunakan 7 tabel yaitu tabel *users*, *project*, *training*, *categories*, *keyword*, *stopwod* dan baku. Pada gambar 4 merupakan *database MySQL* yang digunakan dalam sistem untuk mengelola data *project*, *data training* dan *keyword target* pada *crawling* twitter.

Sedangkan data hasil *crawling* twitter disimpan dalam *database* yang berbeda menggunakan *database engine elasticsearch* dengan skema sesuai yang diberikan twitter. *Elasticsearch* merupakan *database* yang difokuskan pada *search engine* yang dapat memproses teks lebih cepat dan memiliki *query*

yang lebih mudah dalam melakukan pencarian serta menggunakan basis *RESTful*. *Elasticsearch* memiliki keunggulan yang memiliki performa yang cepat dan mampu menangani jumlah data yang besar.

3.3.2. Implementasi Crawling

Proses *crawling* merupakan langkah awal pada sistem ini, melakukan pengambilan data twitter dengan memanfaatkan *API Search* twitter. *Crawling* ini dijalankan setiap menit untuk melakukan pengambilan data, dengan limit 150 kali *request* data dengan masing-masing hasil maksimal 100 tweet. Data yang berhasil diambil akan disimpan pada *database elasticsearch* dengan *schema* sesuai dengan hasil *response* dari *API* twitter.

3.3.3. Implementasi Klasifikasi

Tahap analisa klasifikasi data dilakukan dengan melalui beberapa proses diantaranya proses *preprocessing*, pembobotan dan klasifikasi sentimen. *Preprocessing* adalah pengolahan data yang bertujuan untuk meminimalisir data / kata yang tidak memiliki nilai dalam objek penelitian dan data yang tidak konsisten. Proses klasifikasi *support vector machine* pada sistem ini menggunakan *library php machine learning*.

Tren sentimen diklasifikasikan menjadi 3 kategori, yaitu sentimen positif, negatif dan netral. Sentimen yang ditampilkan berdasarkan rentang waktu mulai 1 januari 2019 sampai 31 maret 2019. Dengan total keseluruhan data tweet sejumlah 3.642.407 tweet.



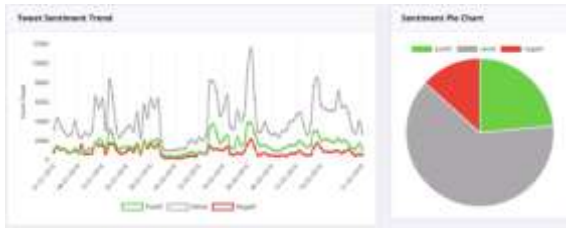
Gambar 5. Tren sentimen Jokowi

Pada gambar 5 merupakan data sentimen dengan kata kunci Jokowi dan joko widodo, dengan jumlah 1.648.220 tweet dan klasifikasi sentimen positif 33%, negatif 11% dan netral 54%.



Gambar 6. Tren sentimen Prabowo

Pada gambar 6 merupakan data sentimen dengan kata kunci prabowo dan prabowo subianto, dengan jumlah 1.209.802 tweet dan klasifikasi sentimen positif 26%, negatif 16% dan netral 58%.



Gambar 7. Tren sentimen Sandianga Uno
 Pada gambar 7 merupakan data sentimen dengan kata kunci sandiuno dan sandianga uno, dengan jumlah 581.826 tweet dan klasifikasi sentimen positif 24%, negatif 13% dan netral 63%.



Gambar 8. Tren sentimen Ma'rif Amin
 Pada gambar 8 merupakan data sentimen dengan kata kunci MarufAminNU dan ma'rif amin, dengan jumlah 202.559 tweet dan klasifikasi sentimen positif 28%, negatif 3% dan netral 69%.

3.4. Pengujian

Untuk mengevaluasi suatu model klasifikasi, memerlukan himpunan data uji. Pengujian dilakukan dengan tujuan untuk mendapatkan nilai akurasi, presisi, *recall* dan *k-fold cross-validation*. Evaluasi performansi dilakukan untuk menguji hasil dari klasifikasi dengan mengukur nilai performansi dari sistem yang telah dibuat, parameter pengujian.

IV. PEMBAHASAN

4.1 Validasi

Validasi adalah salah satu proses pengujian algoritma, pada umumnya pengujian dilakukan menggunakan data secara acak beberapa kali pengujian hingga dihasilkan hasil rata-rata dari data tersebut. Pada validasi menggunakan *10 fold cross validation*, yaitu dengan membagi *dataset* menjadi 2 bagian, data latih dan data uji dengan presentase pembagian yang sama pada setiap k (jumlah iterasi).

Tabel 2. *10 fold cross validation*

	Dataset dibagi menjadi 10 bagian									
	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
Percobaan 1	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji
Percobaan 2	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji
Percobaan 3	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji
Percobaan 4	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji
Percobaan 5	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji
Percobaan 6	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji
Percobaan 7	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji
Percobaan 8	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji
Percobaan 9	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji
Percobaan 10	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji

Keterangan :

- = Data latih
- = Data uji

Seperti contoh yang akan diterapkan adalah 90 % merupakan data latih, 10% menjadi data uji dan

membagi menjadi 10 bagian. Proses ini diulang sebanyak 10 kali hingga semua data mendapatkan bagian menjadi data uji. Data yang akan diujikan sebanyak 600 data, pada setiap perulangan 60 data menjadi data uji dan 540 menjadi data latih. Tabel 2 merupakan representasi dari *10 fold cross validation*.

4.2 Pengujian Algoritma

Pengujian dilakukan untuk mengukur kinerja sistem atau metode dalam menentukan klasifikasi data tweet berdasarkan data latih dan data uji yang telah disediakan dan mendapatkan tingkat nilai akurasi, presisi dan *recall* pada data hasil klasifikasi.

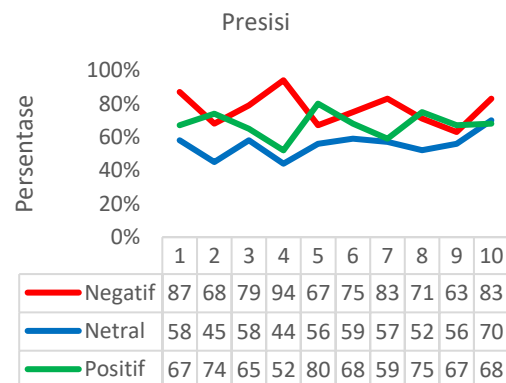
Tabel 3. Hasil pengujian akurasi algoritma

Kfold	Presisi			Recall			Rata-rata		Akurasi
	-1	0	1	-1	0	1	presisi	recall	
1	87%	58%	67%	65%	70%	70%	71%	68%	68%
2	68%	45%	74%	65%	50%	70%	62%	62%	62%
3	79%	58%	65%	55%	75%	65%	67%	65%	65%
4	94%	44%	52%	80%	35%	70%	63%	62%	62%
5	67%	56%	80%	80%	45%	80%	68%	68%	68%
6	75%	59%	68%	75%	59%	68%	67%	67%	67%
7	83%	57%	59%	50%	60%	80%	66%	63%	63%
8	71%	52%	75%	60%	60%	75%	66%	65%	65%
9	63%	56%	67%	75%	50%	60%	62%	62%	62%
10	83%	70%	68%	75%	70%	75%	74%	73%	73%
Avg	77%	56%	68%	68%	57%	71%	67%	66%	66%

Keterangan :

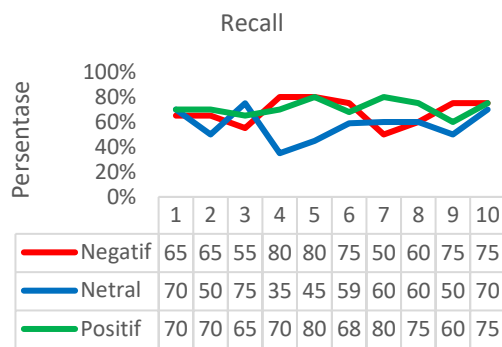
- 1 : label negatif
- 0 : label netral
- 1 : label positif

Berdasarkan Tabel 3 didapatkan hasil pada *10 fold cross validation* rata-rata presisi dengan nilai minimal 62% dan maksimal 74%, *recall* minimal 62% dan maksimal 73%, untuk nilai akurasi adalah minimal 62% dan maksimal memiliki nilai 73%. Sedangkan untuk rata-rata pada presisi sebesar 67%, *recall* 66% dan akurasi sebesar 66%.



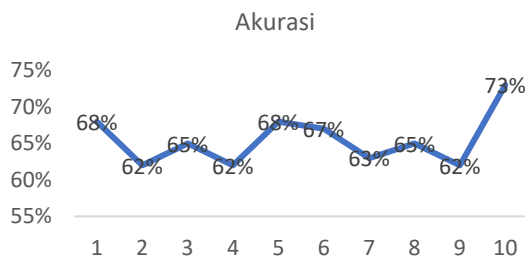
Gambar 9. Persentase pengujian presisi

Dari percobaan uji presisi pada gambar 9 didapatkan nilai tertinggi pada ketiga label klasifikasi terdapat pada label negatif dengan nilai 94% dan sedangkan nilai terendah adalah 44% pada label netral.



Gambar 10. Persentase pengujian recall

Selain itu pada gambar 10 yaitu pengujian recall pada label netral memiliki nilai terendah sebesar 35%. Label positif dan negatif memiliki nilai tertinggi yang sama yaitu 80%.



Gambar 11. Persentase pengujian akurasi

Pada gambar 11 merupakan hasil persentase pengujian akurasi. Dari 10 fold cross validation didapatkan nilai terendah 62%, nilai tertinggi 73% dan rata-rata keseluruhan 66%.

V. KESIMPULAN

Kesimpulan yang dapat diambil dari hasil penelitian yang dilakukan tentang Implementasi Support Vector Machine pada Analisa Sentimen Twitter Berdasarkan Waktu adalah crawling data twitter menggunakan API twitter secara periodik berjalan dengan baik, dengan memperhatikan jumlah kata kunci yang ditargetkan karena memiliki batas penggunaan layanan API twitter. Dalam preprocessing teks pada data twitter berjalan dengan baik dan memerlukan beberapa proses untuk menjadikan tweet lebih optimal untuk selanjutnya diproses pengklasifikasian. Implementasi algoritma support vector machine memiliki persentase akurasi rata-rata sebesar 66% yang dibagi menjadi 3 klasifikasi dengan validasi data menggunakan k-fold cross validation sebanyak 10 bagian data.

Berdasarkan penelitian, ada beberapa hal yang disarankan untuk pengembangan atau penelitian berikutnya adalah pada data stopword disarankan lebih banyak data, lebih banyak kosa kata yang digunakan semakin baik pada pemrosesan teks. Sistem dapat dilakukan optimasi sehingga dapat menghasilkan tingkat kinerja sistem dan akurasi yang lebih baik.

Daftar Pustaka

- [1] S. Kemp, "Digital in 2018 in Southeast Asia," *We Are Soc.*, p. 362, 2018.
- [2] S. Adi, "Perancangan Klasifikasi Tweet Berdasarkan Sentimen Dan Fitur Calon Gubernur DKI Jakarta 2017," *J. Inform. Pelita Nusant.*, vol. 3, no. 1, pp. 10–16, 2018.
- [3] F. Rahutomo, P. Y. Saputra, and M. A. Fidyawan, "Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine," *J. Inform. Polinema*, pp. 93–100, 2016.
- [4] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [5] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeding Eng.*, vol. 2, no. 1, pp. 1–7, 2015.
- [6] Suyanto, *Machine Learning Tingkat Dasar dan Lanjut*. Bandung: Informatika, 2018.
- [7] D. J. Haryanto, L. Muflikhah, and M. A. Fauzi, "Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2909–2916, 2018.
- [8] C. Rahmad, F. Rahutomo, M. A. Gustalika, and I. F. Rahmah, "Identification of sugarcane maturity scale based on RGB, Gabor feature extraction and Support Vector Machine," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 434, no. 1, 2018.
- [9] J. Mase, M. T. Furqon, and B. Rahayudi, "Penerapan Algoritme Support Vector Machine (SVM) Pada Pengklasifikasian Penyakit Kucing," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 3648–3654, 2018.
- [10] F. Rahutomo, I. Yanuar, R. Pratiwi, and D. M. Ramadhani, "Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia Naïve Bayes ' s Experiment On Hoax News Detection In Indonesian Language," vol. 23, no. 1, pp. 1–15, 2019.
- [11] E. Hardiyanto and F. Rahutomo, "Studi Awal Klasifikasi Artikel Wikipedia Bahasa Indonesia Dengan Menggunakan Metoda K Nearest Neighbor," vol. 9035, pp. 334–338, 2016.