

Sentimen Analisis Politik Berita Media Online Dalam Pemilihan Presiden 2019 Menggunakan Metode *Support Vector Machine*

Aditya Gifhari Soenarya¹, Kemas Muslim L.²

^{1,2} Fakultas Informatika, Universitas Telkom, Bandung

¹adityags@students.telkomuniversity.ac.id, ²kemasmuslim@telkomuniversity.ac.id

Abstrak

Penyedia layanan berita dalam menerbitkan berita seharusnya tidak boleh memiliki keberpihakan atau sentimen dalam hal politik untuk menjaga kualitas berita. Semakin dekatnya Pemilu Presiden dan wakil Presiden 2019 membuat gencarnya berita tentang Jokowi-Ma'ruf dan Prabowo-Sandi. Permasalahannya adalah bagaimana cara menentukan suatu teks berita memiliki sentimen positif atau negatif kepada salah satu kubu Jokowi-Ma'ruf atau Prabowo-Sandi dengan akurasi tinggi. Pada penelitian tugas akhir ini, penulis membuat dua sistem klasifikasi teks berita yang pertama untuk kubu Jokowi-Ma'ruf dan kedua untuk kubu Prabowo-Sandi menggunakan metode *Support Vector Machine* dengan pembobotan menggunakan metode *TF-IDF*. Penelitian ini menggunakan metode pembobotan *TF-IDF* pada semua kata yang telah dilakukan preprocessing lalu hasil pembobotan diolah menggunakan *Support Vector Machine* menggunakan tiga kernel yaitu *radial basis function*, linear, dan polynomial. Dari hasil percobaan pada model *Support Vector Machine* untuk kubu Jokowi-Ma'ruf, hasil akurasi terbaik diperoleh dari skema data latih 90% dan data uji 10% menggunakan kernel linear dengan akurasi 93.33%. Untuk hasil percobaan model *Support Vector Machine* kubu Prabowo-Sandi, hasil akurasi terbaik diperoleh dari skema data latih 90% dan data uji 10% menggunakan kernel linear dengan akurasi 83.33%.

Kata kunci : SVM, TF-IDF, kernel, *radial basis function*, polynomial, linear

Abstract

News service providers in publishing news should not have a bias or sentiment in terms of politics to keep news quality. The closer the 2019 President and Vice President Election makes the news about Jokowi-Ma'ruf and Prabowo-Sandi incessant. The problem is how to determine a news text that have positive or negative sentiments to one of Jokowi-Ma'ruf's or Prabowo-Sandi's camps with high accuracy. In this final task research, the authors made two news text classification systems, first for the Jokowi-Ma'ruf camp and the second for the Prabowo-Sandi camp using the *Support Vector Machine* method with weighting using the *TF-IDF* method. This study uses the *TF-IDF* weighting method on all preprocessing words and the weighting results are processed using *Support Vector Machine* using three kernels, *radial basis function*, linear, and polynomial. From the results of experiments on the *Support Vector Machine* model for the Jokowi-Ma'ruf camp, the best accuracy results were obtained from the 90% training data scheme and the 10% test data using a linear kernel with an accuracy of 93.33%. For the experimental results of the *Support Vector Machine* model of the Prabowo-Sandi camp, the best accuracy results were obtained from the 90% training data scheme and the 10% test data using a linear kernel with an accuracy of 83.33%.

Keywords: SVM, TF-IDF, kernel, *radial basis function*, polynomial, linear

1. Pendahuluan

1.1. Latar Belakang

Berita adalah sebuah informasi tentang apa yang sedang atau telah terjadi di suatu tempat baik berupa cetak, siaran, internet atau dari mulut ke mulut. Selain sarana penyampai informasi, berita juga merupakan penggiring opini publik yang cukup efektif. Semakin dekatnya pemilihan presiden dan wakil presiden 2019 membuat masyarakat mulai terpecah kedalam 2 kubu yaitu kubu Jokowi dan Prabowo. Penyedia layanan berita yang harusnya bersifat netral, beberapa malah menjadi senjata dalam menggiring opini publik ke suatu kubu.

Untuk mengetahui kecondongan sebuah penyedia layanan berita *online*, maka dibutuhkan metode analisis sentimen terhadap berita-berita yang di terbitkan di website penyedia layanan berita *online* tersebut. Seperti pada contoh berita "Ke Pasar Ikan di Aceh, Sandiaga: Ekonomi Sepi, Pedagang Butuh Modal" yang diterbitkan detik.com, berita tersebut akan di analisis judul dan isi beritanya apakah memiliki sentimen positif atau negatif pada salah satu calon presiden atau calon wakil presiden. Karena berita-berita yang di terbitkan pada suatu penyedia layanan berita *online* sangat banyak dan akan sulit jika di lakukan secara manual, maka dibutuhkan metode analisis sentimen yang yang dapat berjalan secara otomatis untuk menentukan kecondongan politik suatu penyedia layanan berita *online*.

Pada Tugas Akhir ini, dilakukan sentimen analisi menggunakan metode *Support Vector Machine* (SVM) untuk melakukan klasifikasi teks berita berbahasa Indonesia. *Support Vector Machine* (SVM) merupakan salah satu metode klasifikasi untuk menentukan fungsi pemisah yang optimal. Keunggulan SVM adalah dapat diimplementasikan relatif mudah karena SVM dapat di rumuskan dalam masalah Quadratic Programming

1.2. Topik dan Batasannya

Banyaknya berita yang beredar tentang salah satu kubu capres atau cawapres membuat sulitnya pengidentifikasian suatu penyedia layanan berita memiliki sentimen positif atau negatif kepada salahsatu kubu calon presiden dan calon wakil presiden, oleh karna itu dibutuhkan metode pengotomatisan dalam mengklasifikasi suatu berita memiliki sentimen positif atau negatif pada salah kubu dan masalah pengotomatisan itu akan di selesaikan dengan menggunakan metode *Support Vector Machine*. Beberapa batasan masalah diantaranya, teks yang di digunakan sebagai data hanya teks berita berbahasa Indonesia, teks berita yang digunakan hanya berasal dari laman penyedia layanan berita online kompas.com, berita yang di ambil adalah berita yang di terbitkan pada rentang waktu 1 Desember 2018 sampai dengan 28 Februari 2019, dan teks berita yang di olah hanya teks berita yang memiliki tag calon presiden dan calon wakil presiden yang terkait.

1.3. Tujuan

Tujuan pembuatan Tugas Akhir ini adalah.

1. Menghasilkan sistem yang dapat menentukan kumpulan teks berita memiliki sentimen positif atau negatif kepada salah satu kubu calon presiden dan wakil presiden Jokowi-Ma'ruf atau Prabowo-Sandi menggunakan metode SVM
2. Menguji dan menganalisis performansi klasifikasi SVM untuk menentukan kumpulan berita mempunyai sentimen positif atau negatif ke kubu calon presiden dan wakil presiden Jokowi-Ma'ruf atau kubu Prabowo-Sandi.

2. Studi Terkait

2.1. Sentimen Analysis

Sentiment Analysis merupakan proses mengolah data untuk mendapatkan teks yang terdapat dalam di dalam suatu kalimat. *Sentiment Analysis* dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif [1].

2.2. Preprocessing

Sebelum mengolah teks berita yang telah diambil dibutuhkan beberapa proses agar teks yang diambil mudah untuk diolah. *Preprocessing* adalah suatu proses mengubah dari bentuk teks yang tidak beraturan ke bentuk yang terstruktur sesuai dengan kebutuhan. Beberapa tahapan yang di gunakan dalam *preprocessing* teks berita diantaranya.

- a. Case Folding
Case Folding berfungsi untuk mengubah semua huruf pada teks berita menjadi huruf kecil.
- b. Tokenizing
Tokenizing berfungsi untuk pemecahan teks menjadi berbentuk kata dengan melakukan penghilangan angka dan tanda baca.
- c. Filtering
Filtering merupakan proses untuk membuang kata yang dianggap tidak penting seperti kata :yang, tetapi, atau, ke, di dan sebagainya.
- d. Steaming
Proses *steaming* adalah proses mengubah kata menjadi kata dasar agar membuat proses pengolahan teks berita menjadi lebih optimal [2].

2.3. Term Weighting

Term Weighting adalah pemberian bobot atau suatu nilai pada kata dalam suatu teks. Bobot atau nilai menentukan klasifikasi suatu teks. Ada beberapa metode pembobotan diantaranya *TF*, *TF-IDF*.

1. Term Frequency (TF)

Term Frequency adalah salah satu metode yang paling sederhana untuk pemberian bobot. Caranya dengan memeberi bobot pada kata yang sering muncul pada sebuah teks. Dalam metode ini, setiap kata diasumsikan mempunyai kepentingan yang sesuai dengan jumlah munculnya kata tersebut di dalam teks. Persamaanya adalah [3].

$$W(d, t) = TF(d, t) \quad (2.1)$$

$TF(d,t)$ = frekuensi kemunculan kata t pada teks d .

2. Term Frequency – Inverse Document Frequency

Adalah metode pembobotan berdasarkan inverse *Inverse Document Frequency*, semakin sedikit suatu kata muncul maka semakin tinggi bobotnya [4] [5].

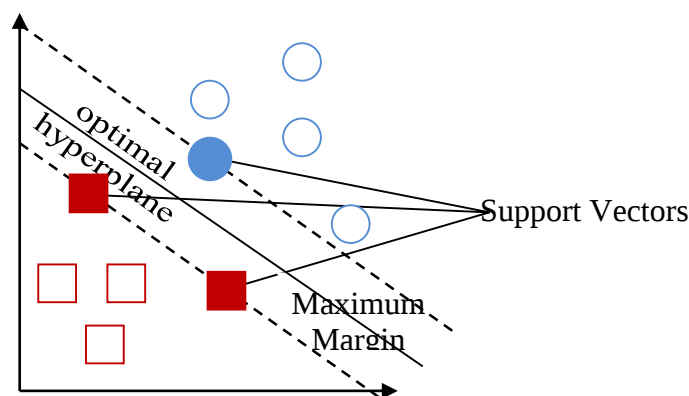
$$idf = \log\left(\frac{n}{df}\right) \quad (2.2)$$

n = jumlah teks

df = jumlah teks yang mengandung satu kata

2.4. Support Vector Machine

SVM adalah metode learning yang bekerja dengan menggunakan prinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan *hyperplane* terbaik dalam input space [6]. seperti pada gambar 2.1 *hyperplane* terbaik dari dua data adalah *hyperplane* yang memiliki margin yang maksimal. Margin merupakan jarak antara *hyperplane* dengan pattern terdekat masing masing kelas. Patten yang paling dekat disebut *support vector*.



Gambar 2.1 Ilustrasi SVM [7].

Data yang tersedia dinotasikan sebagai $\vec{x}_i \in \mathbb{R}^d$. Untuk label dinotasikan $y_i \in \{+1, -1\}$ untuk $i = 1, 2, \dots, l$, dimana l merupakan banyaknya data. Diasumsikan kelas $+1$ dan -1 dapat terpisah sempurna oleh *hyperplane* berdimensi d , yang didefinisikan

$$\vec{w} \vec{x}_i + b = 0 \quad (2.6)$$

Pattern $\vec{x}_i$ yang termasuk kelas $+1$ memenuhi pertidaksamaan

$$\vec{w} \vec{x}_i + b \geq +1 \quad (2.7)$$

Pattern $\vec{x}_i$ yang termasuk kelas -1 memenuhi pertidaksamaan

$$\vec{w} \vec{x}_i + b \leq -1 \quad (2.8)$$

Margin maksimum ditentukan dengan memaksimalkan jarak antara *hyperplane* dan titik terdekatnya (*support vector*), yaitu $\frac{1}{\|\vec{w}\|}$ hal ini dapat di rumuskan sebagai Quadratic Programming (QP), yaitu mencari titik minimal persamaan (2.9), dengan *constraint* persamaan (2.10).

$$\min_{\vec{w}} \tau(\vec{w}) = \frac{1}{2} \|\vec{w}\|^2 \quad (2.9)$$

$$y_i(\vec{x}_i \vec{w} + b) - 1 \geq 0, \forall i \quad (2.10)$$

Problem ini dapat di selesaikan dengan berbagai teknik komputasi, salah satunya *Lagrange Multiplier* seperti pada persamaan berikut.

$$L(\vec{w}, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i(\vec{x}_i \vec{w} + b) - 1) \quad (2.11)$$

$$(i = 1, 2, \dots, l)$$

α_i adalah *Lagrange Multipliers*, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan (2.11) adalah minimal L terhadap $\vec{w}$ dan b , dan maksimum L terhadap α_i memperhatikan sifat bahwa titik optimal gradient $L = 0$, persamaan (2.11) dapat dimodifikasi sebagai memaksimalkan problem yang hanya mengandung α_i , sebagai persamaan (2.12) yaitu:

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \vec{x}_i \vec{x}_j \quad (2.12)$$

s.t

$$\alpha_i \geq 0 (i = 1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0$$

Dari hasil perhitungan ini diperoleh α_i yang kebanyakan bernilai positif, data yang berkorelasi dengan α_i positif lah yang disebut *support vector*.

Untuk menyelesaikan problem non linier, SVM dapat dimodifikasi dengan menggunakan fungsi *Kernel*. Awalnya data $\vec{x}$ dipetakan oleh fungsi $\Phi(\vec{x})$ ke ruang vektor yang berdimensi lebih tinggi. Pada ruang vektor yang baru, *hyperplane* dapat dibangun untuk memisahkan kedua kelas ini. Notasi matematika dari mapping ini adalah.

$$\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^q \quad d < q \quad (2.13)$$

Pemetaan ini dilakukan dengan menjaga topologi data, seperti dua data yang berjarak jauh pada input space akan berjarak jauh juga pada feature space dan sebaliknya. Karna umumnya transformasi Φ tidak diketahui dan sulit di pahami maka perhitungan dot product diganti dengan fungsi *Kernel* $K(\vec{x}_i, \vec{x}_j)$ yang mendefinisikan transformasi Φ secara implisit, yang disebut dengan *Kernel Trick* dengan rumus

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) \quad (2.14)$$

Kernel memberikan berbagai kemudahan, pada SVM kita hanya cukup mengetahui fungsi *Kernel* yang dipakai dan tidak perlu mengetahui wujud fungsi non linear Φ

Selanjutnya hasil klasifikasi dari data $\vec{x}$ diperoleh dari persamaan berikut:

$$\begin{aligned} f(\Phi(\vec{x})) &= \vec{w} \cdot \Phi(\vec{x}) + b \\ &= \sum_{i=1, \vec{x}_i \in SV}^n \alpha_i y_i \Phi(\vec{x}) \cdot \Phi(\vec{x}_i) + b \\ &= \sum_{i=1, \vec{x}_i \in SV}^n \alpha_i y_i K(\vec{x}, \vec{x}_i) + b \end{aligned} \quad (2.15)$$

SV di atas dimaksudkan pada subset training set yang terpilih sebagai *support vector*.

Ada beberapa jenis kernel yaitu, *radial basis function* dan *polynomial*.

Radial basis function kernel atau *RBF* kernel merupakan fungsi kernel yang populer digunakan pada algoritma pembelajaran, khususnya dalam *Support Vector Machine* [8]. *RBF* kernel didefinisikan sebagai [9]

$$K(x, y) = \exp\left(\frac{-\|x - y\|^2}{2\sigma^2}\right)$$

Polynomial kernel dengan degree-d di definisikan sebagai

$$K(x, y) = (x^T y + c)^d$$

2.5. Evaluasi

Evaluasi algoritma akan dilakukan dengan *confusion matrix* menggunakan nilai akurasi, *precision*, dan *recall*. *confusion matrix* dapat dilihat pada table 2.1 berikut

Tabel 2.1 confusion matrix

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	<i>True Positive</i>	<i>False Negative</i>
Negatif	<i>False Positive</i>	<i>True Negative</i>

Dimana:

True Negative (TN) : Kelas yang diprediksi negatif, dan faktanya adalah negatif.

True Positive (TP) : Kelas yang diprediksi positif dan faktanya adalah positif.

False Negative (FN) : Kelas yang diprediksi negatif dan faktanya adalah kelas positif.

False Positive (FP) : Kelas yang diprediksi positif dan faktanya adalah kelas negatif.

1. Akurasi

Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya. Berikut persamaan untuk menghitung akurasi [10].

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.16)$$

2. Precision

Dalam data mining *precision* adalah jumlah teks yang benar diklasifikasi dalam sebuah kelas dibagi total teks dalam kelas tersebut [11]. Dengan persamaan [10].

$$Precision(P) = \frac{TP}{TP+FP} \quad (2.17)$$

3. Recall

Dalam data mining *recall* adalah jumlah teks yang benar diklasifikasikan dalam sebuah kelas dibagi jumlah semua teks yang berlabel kelas tersebut [11]. dengan persamaan [10].

$$Recall(R) = \frac{TP}{TP+FN} \quad (2.18)$$

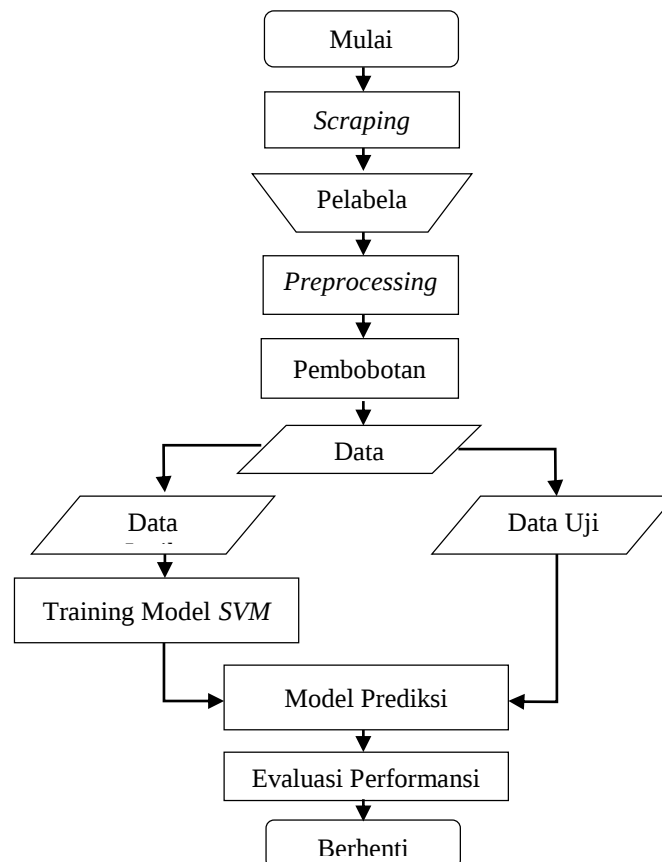
4. F-Measure

F-Measure digunakan untuk melakukan evaluasi rata-rata nilai *Precision* dan *Recall*. Semakin besar *F-Measure*, maka performansi model semakin baik [11]. Dengan persamaan [10].

$$F - Measure = \frac{2RP}{R+P} \quad (2.19)$$

3. Sistem yang Dibangun

Rancangan sistem dapat dilihat pada gambar 3.1 yang terdiri dari *scraping*, pelabelan, *preprocessing*, pembobotan, pemisahan data latih dan data uji, training model SVM, model prediksi dan evaluasi performansi.



Gambar 3.1.

4. Evaluasi

Bagian ini berisi dua sub-bagian, yaitu Hasil Pengujian dan Analisis Hasil Pengujian. Pengujian dan analisis yang dilakukan selaras dengan tujuan TA sebagaimana dinyatakan dalam Pendahuluan.

4.1. Hasil Pengujian

Bagian ini berisi pelabelan dan hasil pengujian model SVM berupa *confusion matrix* dan akurasi.

4.1.1. Pelabelan

Pelabelan dilakukan dengan mengelompokkan teks berita ke dalam kelas positif dan negatif kepada salah satu kubu Jokowi-Ma'ruf atau Prabowo-Sandi. Untuk kriteria teks berita berlabel positif yaitu:

1. Berita membahas kampanye berupa pertemuan dengan warga dan pemaparan janji-janji bila terpilih menjadi presiden
2. Berita membahas deklarasi dukunga suatu ormas/kelompok/masyarakat suatu daerah kepada calon tersebut

3. Berita membahas keunggulan visi dan misi atau pencapaian baik calon tersebut
4. Berita membahas bantahan/klarifikasi isu negatif calon atau kubu calon dengan menyertakan argumen yang logis atau bukti data yang kuat

Sementara untuk kriteria teks berita berlabel negatif yaitu:

1. Berita membahas isu negatif calon atau kubu calon tanpa adanya pembelaan dari calon atau kubu calon yang di isukan
2. Berita membahas sanggahan tentang pernyataan calon atau pihak calon dengan bukti data yang kuat atau argumen yang logis tanpa pembelaan dari calon atau kubu calon bersangkutan

4.1.2. Hasil klasifikasi SVM

Hasil pengujian model SVM dengan tiga kali percobaan *random* untuk kubu Jokowi-Ma'ruf pada **tabel 4.3** dan untuk kubu Prabowo-Sandi pada tabel 4.4

Tabel 4.3 Hasil Percobaan Model SVM Jokowi-Ma'ruf

No	Data Set	Kernel	Akurasi 1	Akurasi 2	Akurasi 3	Rata-Rata Akurasi
1	90:10	linear	93.33%	93.33%	93.33%	93.33%
2	90:10	polynomial	83.33%	90%	90%	87.78%
3	90:10	RBF	83.33%	90%	90%	87.78%
4	80:20	linear	88.33%	81.67%	86.67%	85.56%
5	80:20	polynomial	83.33%	81.67%	85%	83.33%
6	80:20	RBF	83.33%	81.67%	85%	83.33%
7	70:30	linear	87.91%	81.32%	87.91%	85.71%
8	70:30	polynomial	84.62%	80.22%	84.62%	83.15%
9	70:30	RBF	84.62%	80.22%	84.62%	83.15%
10	60:40	linear	89.17%	82.50%	89.17%	86.95%
11	60:40	polynomial	86.67%	80%	86.67%	84.45%
12	60:40	RBF	86.67%	80%	86.67%	84.45%
13	50:50	linear	87.33%	83.33%	86.67%	85.78%
14	50:50	polynomial	85.33%	81.33%	83.33%	83.33%
15	50:50	RBF	85.33%	81.33%	83.33%	83.33%

Tabel 4.4 Hasil Percobaan Model SVM Prabowo-Sandi

No	Data Set	Kernel	Akurasi 1	Akurasi 2	Akurasi 3	Rata-Rata Akurasi
1	90:10	linear	93.33%	76.67%	80.00%	83.33%
2	90:10	polynomial	76.66%	66.67%	80.00%	74.44%
3	90:10	RBF	76.66%	66.67%	80.00%	74.44%
4	80:20	linear	73.33%	70.00%	80.00%	74.44%
5	80:20	polynomial	68.33%	65.00%	80.00%	71.11%
6	80:20	RBF	68.33%	65.00%	80.00%	71.11%
7	70:30	linear	76.92%	73.63%	81.32%	77.29%
8	70:30	polynomial	72.53%	70.33%	80.22%	74.36%
9	70:30	RBF	72.53%	70.33%	80.22%	74.36%
10	60:40	linear	74.17%	72.50%	81.67%	76.11%
11	60:40	polynomial	72.50%	70.00%	80.00%	74.17%
12	60:40	RBF	72.50%	70.00%	80.00%	74.17%
13	50:50	linear	71.33%	74.67%	80.00%	75.33%
14	50:50	polynomial	71.33%	72.67%	77.33%	73.78%
15	50:50	RBF	71.33%	72.67%	77.33%	73.78%

Pada tabel 4.3 skema yang menghasilkan akurasi paling tinggi pada model Jokowi-Ma'ruf adalah dengan perbandingan data latih 90% dan data uji 10% dan menggunakan kernel *linear* yang menghasilkan akurasi 93.33%. Untuk model Prabowo-Sandi pada tabel 4.4 akurasi paling tinggi dihasilkan pada skema perbandingan data latih 90% dan data uji 10% dengan menggunakan kernel linier yang menghasilkan akurasi 83.33%.

Pada Tabel 4.5 yang terdapat pada lampiran dilakukan perhitungan rata-rata *precision*, *recall* dan *F- Measure* pada setiap skema kernel dan rasio data pada model SVM Jokowi-Ma'ruf

Pada Tabel 4.6 yang terdapat pada lampiran dilakukan perhitungan rata-rata *precision*, *recall* dan *F- Measure* untuk setiap skema kernel dan rasio data pada model SVM Prabowo-Sandi

4.2 Analisis Hasil Pengujian

Dari percobaan beberapa skema rasio data dan kernel yang berbeda-beda, akurasi terbaik pada model SVM Jokowi-Ma'ruf pada tabel 4.3 adalah 93.33% dengan skema data latih 90% dan data uji 10% dengan menggunakan kernel *linier*. Untuk model Prabowo-Sandi pada tabel 4.4 akurasi paling tinggi dihasilkan pada skema perbandingan data latih 90% dan data uji 10% dengan menggunakan kernel linier yang menghasilkan akurasi 83.33%.

Untuk model SVM dengan menggunakan kernel RBF dan *polynomial* cenderung menghasilkan akurasi yang sama, pada tabel 4.3 dan tabel 4.4 akurasi yang di hasilkan pada model SVM Jokowi-Ma'ruf maupun model SVM Prabowo-Sandi selalu sama. Pada tabel 4.5 dan tabel 4.6 yang terdapat pada lampiran *confusion matrix* hasil tebakan model SVM menggunakan kernel RBF dan *polynomial* selalu sama sehingga menghasilkan nilai *precision*, *recall* dan *F- Measure* yang sama.

Model SVM yang dihasilkan oleh kernel linier memiliki akurasi yang paling baik pada setiap skema data seperti pada tabel 4.3 dan tabel 4.4 dimana model SVM menggunakan kernel linier selalu memiliki akurasi yang lebih tinggi daripada model SVM menggunakan kernel RBF atau *polynomial*. Untuk skema data terbaik pada tabel 4.3 dan tabel 4.4 adalah skema 90% data latih dan 10% data uji, SVM kernel linier, RBF dan *polynomial* mendapat akurasi terbaik pada skema data 90% data latih dan 10% data uji.

5. Kesimpulan

Pada penelitian ini dapat ditarik kesimpulan bahwa metode *Support Vector Machine* dapat digunakan untuk menentukan berita memiliki sentimen negatif atau positif pada salah satu kubu calon presiden dan wakil presiden. Dalam melakukan klasifikasi berita kedalam kategori positif atau negatif dengan menggunakan metode *Support Vector Machine* dapat dikategorikan cukup baik karena menghasilkan akurasi yang cukup tinggi yaitu 93.33% dengan menggunakan kernel liner dengan perbandingan data 90%:10% pada model Jokowi-Ma'ruf dan 83.33% untuk model Prabowo-Sandi dengan kernel dan perbandingan data yang sama. Untuk riset kedepannya disarankan untuk menambahkan kelas netral dan menambah jumlah data yang seimbang.

Daftar Pustaka

- [1] B. Pang, L. Lee dan S. Vaithyanathan, "Thumbs Up? Sentiment Classification using Machine Learning," 2002.
- [2] D. Wahyudi, T. Susyanto dan D. Nugroho, "Implementasi dan Analisis Algoritma Stemming Nazier & Adrian dan Porter pada Dokumen Berbahasa Indonesia," *Jurnal Ilmiah SINUS*.
- [3] T. Tokunaga dan I. Makoto, "Text Categorization Based on Weighted Inverse Document Frequency," 1996.
- [4] S. Menaka dan N. Radha, "Text Classification using Keyword Extraction Technique," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 12, 2013.
- [5] S. Charanyaa dan K. Sangeetha, "Term Frequency Based Sequence Generation Algorithm for Graph Based Data Anonymization," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 2, 2014.
- [6] A. S. Nugroho, A. B. Witarto dan D. Handoko, "Support Vector Machine –Teori dan Aplikasinya dalam Bioinformatika 1," 2003.
- [7] R. Aggiwal, "RPubs - Support Vector Machine," 7 February 2017. [Online]. Available: <http://rpubs.com/dimensionless/svm>. [Diakses 8 November 2018].

- [8] Y.-W. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard dan C.-J. Lin, "Training and Testing Low-degree Polynomial Data Mappings via Linear SVM," *Journal of Machine Learning Research* 11, 2010.
- [9] J.-P. Vert, K. Tsuda dan B. Schölkopf, A primer on kernel methods, 2004.
- [10] F. Guillet dan H. J. Hamilton, Quality Measures in Data Mining, Springer, 2007.
- [11] R. G. Rossellini, "Perbandingan Metode Pembobotan Term Menggunakan Term Frequency Chi-Square dan Term Frequency Inverse Document Frequency pada Text Mining," 2012.