

ANALISIS SENTIMEN PENGGUNAAN KOTAK SUARA KARDUS MENGUNAKAN SVM *LINEARLY SEPARABLE DATA*

¹Anastasia Dhika Prasmitasari Wardhani, ²Isnandar Slamet, ³Sugiyanto

^{1,2,3} Program Studi Statistika, FMIPA Universitas Sebelas Maret, (0271)669376

e-mail: anastasiadpw@gmail.com

Abstrak

Pemilihan umum merupakan sarana bagi warga negara untuk memilih wakil-wakilnya yang akan menyalurkan aspirasinya. Komisi Pemilihan Umum sebagai penyelenggara pemilu memutuskan menggunakan kotak suara kardus dalam Pemilihan Umum 2019. Keputusan tersebut menimbulkan berbagai opini dalam masyarakat yang disampaikan melalui media sosial, salah satunya *Twitter*. Opini tersebut dijadikan bahan analisis sentimen. Penelitian ini bertujuan untuk mengetahui dan mengklasifikasikan sentimen masyarakat terhadap penggunaan kotak suara kardus. Opini masyarakat diklasifikasikan menjadi positif atau negatif. Tahapan dalam melakukan analisis sentimen adalah *scraping* data, *preprocessing* data, pelabelan data, klasifikasi, dan visualisasi dengan *word cloud*. Algoritma klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine Linearly Separable Data*. Berdasarkan pembahasan, dapat disimpulkan bahwa sentimen positif terhadap penggunaan kotak suara kardus adalah aman, hemat, kedap, tahan, didukung, dan disetujui, sedangkan sentimen negatifnya adalah rawan, curang, percaya, rusak, polemik, dan gila. Selain itu, diperoleh 79,24% data dapat diklasifikasikan dengan benar.

Kata Kunci: Analisis Sentimen, *Twitter*, Kotak Suara Kardus, *SVM Linearly Separable Data*.

Abstract

General election is a momentum for citizens to elect their representatives who will deliver their aspirations. General Election Commission as the election organizer decides to use a cardboard ballot box in the 2019 General Election. The decision causes various opinions in the public that are conveyed through social media, one of them is *Twitter*. This opinion is used as a material for sentiment analysis. This research aims to find out and classify public's sentiment towards the use of cardboard ballot box. Public opinion is classified as positive or negative. The steps to do sentiment analysis are *scraping* data, *data preprocessing*, *data labeling*, *classification*, and *visualization* with *word cloud*. The classification algorithm that used in this research is *Support Vector Machine Linearly Separable Data*. Based on the discussion, it can be concluded that the positive sentiment towards the use of cardboard ballot boxes is safe, economical, impermeable, resistant, supported, and approved, while the negative sentiments are vulnerable, cheating, believing, broken, polemic, and crazy. In addition, 79.24% of data can be classified correctly.

Keywords: Sentiment Analysis, *Twitter*, Cardboard Ballot Box, *SVM Linearly Separable Data*.

PENDAHULUAN

Pemilu atau pemilihan umum merupakan sarana dalam suatu negara yang menganut asas demokrasi yang memberi kesempatan warga negaranya untuk memilih wakil-wakilnya yang akan menyalurkan aspirasinya. Lembaga yang bertanggungjawab dalam pelaksanaan pemilu adalah Komisi Pemilihan Umum (KPU). Salah satu komponen penting dalam pemilu adalah kotak suara. Kotak suara yang digunakan dalam pemilu 2019 adalah kotak suara berbahan karton kedap air atau biasa disebut dengan kotak suara kardus. Keputusan KPU ini menimbulkan berbagai opini, baik pro maupun kontra. Opini tersebut disampaikan melalui berbagai cara, misalnya melalui media sosial. Salah satu media sosial yang banyak digunakan adalah *Twitter*. Media sosial *Twitter* berada di peringkat ke-4 *platform* aktif media sosial, dengan presentase 52% dari total pengguna aktif media sosial sebanyak 150 juta orang (datareportal.com). Data yang diperoleh dari media sosial berupa data teks yang merupakan data tidak terstruktur sehingga diperlukan algoritma yang memadai untuk mengekstraksi pengetahuan dari data tersebut. Metode tradisional memadai untuk analisis data terstruktur tetapi tidak sesuai untuk volume besar data tidak terstruktur untuk mengekstraksi pengetahuan (Kanimozhi dan Venkatesan, 2015).

Penelitian ini akan membahas mengenai analisis data tidak terstruktur, yaitu *tweet* pada media sosial *Twitter* tentang penggunaan kotak suara kardus dalam Pemilu 2019. Penelitian ini bertujuan untuk

mengetahui dan mengklasifikasikan sentimen masyarakat terhadap penggunaan kotak suara kardus tersebut. Data tersebut kemudian diklasifikasikan secara manual menjadi dua sentimen, yaitu positif atau negatif. Novantirani (2015) melakukan analisis sentimen terhadap penggunaan transportasi umum darat dalam kota dengan menggunakan SVM dan mendapat tingkat akurasi sebesar 78,12%. Susilowati (2015) juga menggunakan SVM untuk mengklasifikasikan kemacetan lalu lintas pada *Twitter* dengan tingkat akurasi di atas 90%. Rajeswary dan Divya (2015) memperoleh tingkat akurasi 100% menggunakan SVM tingkat lanjut dalam mengklasifikasikan hama berdasarkan fitur-fiturnya. Latifah (2018) membandingkan algoritma SVM dan Naïve Bayes untuk mengklasifikasikan tanggapan masyarakat mengenai Traveloka melalui *Twitter*, hasilnya tingkat akurasi SVM lebih tinggi yaitu sebesar 79,14% dibandingkan dengan Naïve Bayes yaitu 75,61%. Karena algoritma SVM memiliki performa yang baik dalam banyak penelitian dengan memberikan tingkat akurasi yang cukup tinggi maka penelitian ini menggunakan algoritma SVM, khususnya SVM *Linearly Separable Data*.

METODE PENELITIAN

Pemilu

Pemilihan umum atau disebut Pemilu adalah sarana kedaulatan rakyat untuk memilih anggota Dewan Perwakilan Rakyat, anggota Dewan Perwakilan Daerah, Presiden dan Wakil Presiden, dan untuk memilih anggota Dewan Perwakilan Rakyat Daerah, yang dilaksanakan secara langsung, umum, bebas, rahasia, jujur, dan adil dalam Negara Kesatuan Republik Indonesia berdasarkan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945. Lembaga yang menyelenggarakan pemilu adalah Komisi Pemilihan Umum (UU No. 7 Tahun 2017). Salah satu komponen penting dalam pemilu adalah kotak suara. Berdasarkan Peraturan Komisi Pemilihan Umum (PKPU) Republik Indonesia Nomor 15 Tahun 2018, pasal 7, spesifikasi kotak suara yang digunakan adalah: terbuat dari bahan karton kedap air yang pada satu sisinya bersifat transparan; barang habis pakai; berbentuk kotak yang kokoh pada setiap sisinya, dengan ukuran panjang 40 cm, lebar 40 cm, dan tinggi 60 cm; dan berwarna putih.

Twitter

Twitter adalah situs web yang dimiliki dan dioperasikan oleh Twitter, Inc., yang menawarkan jaringan sosial berupa *microblog*. *Microblog* memungkinkan penggunanya mengirim dan membaca pesan *blog* seperti pada umumnya namun terbatas hanya sejumlah 280 karakter yang ditampilkan pada halaman profil pengguna. *Twitter* memiliki karakteristik dan format penulisan yang unik dengan simbol ataupun aturan khusus. Pesan dalam *Twitter* dikenal dengan istilah *tweet* (Novantirani, 2015).

Data Mining

Data mining adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan, dan *machine learning* untuk mengekstraksi serta mengidentifikasi informasi yang bermanfaat untuk pengetahuan yang terkait dari berbagai *database* besar (Turban *et al*, 2005).

Machine Learning

Machine learning adalah bidang yang mempelajari pengembangan algoritma komputer untuk mengubah data menjadi informasi. *Machine learning* menggunakan teknik-teknik data mining dan algoritma pembelajaran lainnya untuk membuat model dari hal-hal yang terjadi pada data untuk memprediksi kejadian selanjutnya. *Machine learning* dibedakan menjadi dua tipe, yaitu *supervised learning* dan *unsupervised learning*. Terdapat beberapa algoritma dalam *supervised learning*, yaitu K-Nearest Neighbors, Naïve Bayes, Support Vector Machine, Decision Trees, Linear Regression, dan Neural Network (Faizal, 2016).

Klasifikasi

Salah satu metode data mining adalah klasifikasi. Klasifikasi digunakan untuk memprediksi kategori dari sekumpulan data dengan berbagai atribut (Fanani, 2017). Sebelum melakukan prediksi, terlebih dahulu dilakukan proses pembelajaran. Data yang digunakan dalam proses pembelajaran

disebut *data training* (data latih), sedangkan data yang digunakan pada proses prediksi disebut *data testing* (data uji) (Faizal, 2016).

Text Mining

Text mining adalah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur, sedangkan *data mining* mengolah data yang terstruktur (Han *et al.*, 2012).

Analisis Sentimen

Analisis sentimen merupakan salah satu cabang ilmu dari *text mining*, *natural language program*, dan *artificial intelligence*. Proses yang dilakukan oleh analisis sentimen untuk memahami, mengekstrak, dan mengolah data teks secara otomatis sehingga menjadi suatu informasi yang bermanfaat (Luqyana, 2018).

SVM Linearly Separable Data

Support Vector Machine (SVM) adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik (Christianini & Taylor, 2000). Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari garis (*hyperplane*) terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space* (Octaviani, 2014). *Linearly separable data* berarti data yang dapat dipisahkan secara *linear*. Misalnya terdapat dua kelas, positif dan negatif. *Hyperplane* pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. Persamaan untuk dua kelas yang dapat dipisahkan oleh bidang pembatas yang sejajar adalah (Fikria, 2018)

$$\begin{aligned}x_i \cdot w + b &\geq +1, \text{ untuk } y_i = +1 \\x_i \cdot w + b &\leq -1, \text{ untuk } y_i = -1\end{aligned}\quad (1)$$

Fungsi pemisah linier didefinisikan sebagai (Latifah, 2018)

$$g(x) := \text{sgn}(f(x)) \quad (2)$$

Dengan $f(x) = wx + b$, dimana $x, w \in R^n$ dan $w \in R$.

Secara matematis, formulasi masalah SVM *Linearly Separable Data* di dalam *primal space* adalah

$$\min_w (w) = \frac{1}{2} \|w\|^2 \quad (3)$$

$$y_i(w \cdot x_i + b) \geq 1 \quad (4)$$

dimana, x_i adalah data *input*, y_i adalah *output* dari data x_i , w dan b adalah parameter yang dicari nilainya. Pada (3) dan (4) bertujuan untuk meminimalkan fungsi tujuan (3) dengan memperhatikan (4). Setelah mendapatkan fungsi pemisah, dengan (2) tersebut semua nilai $f(x) < 0$ masuk ke dalam kelas -1 (negatif) dan nilai $f(x) > 0$ masuk ke dalam kelas +1 (positif).

Tahap Analisis

Data yang digunakan dalam penelitian ini adalah *tweets* di *user timeline* yang berkaitan dengan penggunaan kotak suara kardus dalam Pemilu 2019. Metode yang digunakan dalam penelitian ini adalah metode klasifikasi dengan algoritma *Support Vector Machine Linearly Separable Data*. Tahap awal yang dilakukan dalam penelitian ini adalah *crawling data*, yaitu pengumpulan data dari *Twitter* melalui *Application Programming Interface (API) Twitter*. Kemudian dilakukan pelabelan data secara manual menjadi positif atau negatif. Selanjutnya dilakukan *preprocessing data*, yaitu proses membersihkan data agar data yang tidak terstruktur menjadi data terstruktur.

Preprocessing data terdiri dari *cleansing data* yaitu menghapus *Uniform Resource Locator (URL)*, emoji, *mention*, *username*, *hashtag*, tanda baca, angka, RT, dan *strip white space*; *case folding* yaitu tahap mengubah teks menjadi huruf kecil; *stemming* yaitu tahap mengubah kata-kata dalam kalimat menjadi kata dasar; *remove stopword* yaitu tahap menghapus kata-kata yang tidak memiliki

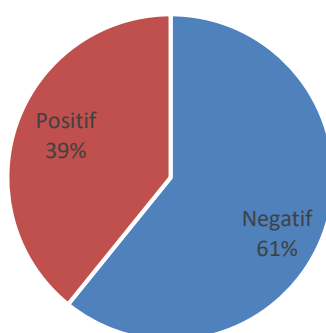
makna; dan *tokenizing* yaitu tahap pemotongan *string input* dalam suatu kalimat berdasarkan kata-kata yang menyusunnya.

Setelah data bersih, dilakukan pembobotan karena masing-masing kata yang berbeda dalam suatu dokumen memiliki tingkat kepentingan berbeda. Selanjutnya data dibagi menjadi data latih dan data uji. Data latih digunakan untuk membentuk model klasifikasi sedangkan data uji digunakan untuk mengetahui tingkat akurasi model tersebut. Kemudian dilakukan klasifikasi menggunakan algoritma *Support Vector Machine Linearly Separable Data* dan dapat diketahui tingkat akurasi penggunaan algoritma tersebut. Kemudian dilakukan visualisasi data menggunakan *word cloud* untuk mengetahui apa saja yang termasuk ke dalam sentimen positif dan negatif tentang penggunaan kotak suara kardus dalam Pemilu 2019.

HASIL DAN PEMBAHASAN

Tahap awal yang dilakukan adalah melihat karakteristik data. Analisis statistika deskriptif digunakan untuk melihat eksplorasi data atau melihat gambaran umum dari data. Data yang diambil berupa *tweet* mengenai penggunaan kotak suara kardus dalam Pemilu 2019 dengan kata kunci “kotak suara kardus”. *Crawling data* yang dilakukan memperoleh 10.000 *tweets*.

Kemudian dilakukan pelabelan secara manual terhadap *tweet* tersebut menjadi dua kelas, yaitu positif atau negatif. Data *tweet* yang tidak berkaitan dengan penggunaan kotak suara kardus tidak digunakan dan diperoleh 3145 *tweets* hasil pelabelan sentimen positif atau negatif. Gambar 1 menunjukkan *pie chart* mengenai proporsi jumlah sentimen positif dan negatif penggunaan kotak suara kardus.



Gambar 1. *Pie Chart* Proporsi Sentimen

Berdasarkan Gambar 1, dapat dilihat bahwa *tweet* yang mengandung sentiment positif sebesar 39% atau 1234 *tweets* sedangkan yang mengandung sentimen negatif sebesar 61% atau 1911 *tweets*. Selanjutnya, Tabel 1 menunjukkan contoh *tweet* yang mengandung sentimen positif dan negatif mengenai penggunaan kotak suara kardus dalam Pemilu 2019.

Tabel 1. Contoh *Tweet*

No.	Tweet	Class
1.	RT @Susilo35 Bawaslu Gunungkidul menguji kekuatan kotak suara berbahan karton dupleks. Setelah 3 kali pengujian, bagaimana hasilnya? Aman dan kuat :) https://t.co/OzNTZe4IMC	Positif
2.	RT @Fahri: Kotak Suara 'Kardus' Dipikul Naik-Turun Gunung Tak akan Selamat :(#2019GagalPemilu - https://t.co/COutDn7UUw	Negatif

Preprocessing Data

Preprocessing data dilakukan untuk menghilangkan *noise*, memperjelas fitur, meningkatkan akurasi dan mengkonversi data asli agar dapat diolah sesuai kebutuhan. Tahap awal yang dilakukan dalam *preprocessing* adalah *cleansing* yang terdiri dari menghapus URL, *mention*, *username*, RT, *hashtag*, emoji, angka, tanda baca, dan *stripe white spac* seperti dalam Tabel 2.

Tabel 2. Contoh Tweet Setelah Cleansing

No.	Sebelum	Sesudah
1.	RT @Susilo35 Bawaslu Gunungkidul menguji kekuatan kotak suara berbahan karton dupleks. Setelah 3 kali pengujian, bagaimana hasilnya? Aman dan kuat :) https://t.co/OzNTZe4IMC	Bawaslu Gunungkidul menguji kekuatan kotak suara berbahan karton dupleks Setelah kali pengujian bagaimana hasilnya Aman dan kuat
2.	RT @Fahri: Kotak Suara 'Kardus' Dipikul Naik-Turun Gunung Tak akan Selamat :(#2019GagalPemilu https://t.co/COutDn7UUw	Kotak Suara Kardus Dipikul Naik Turun Gunung Tak akan Selamat

Langkah selanjutnya yang dilakukan adalah *case folding* atau mengubah huruf dalam *tweet* menjadi huruf non kapital seperti dalam Tabel 3.

Tabel 3. Contoh Tweet Setelah Case Folding

No.	Sebelum	Sesudah
1.	Bawaslu Gunungkidul menguji kekuatan kotak suara berbahan karton dupleks Setelah kali pengujian bagaimana hasilnya Aman dan kuat	bawaslu gunungkidul menguji kekuatan kotak suara berbahan karton dupleks setelah kali pengujian bagaimana hasilnya aman dan kuat
2.	Kotak Suara Kardus Dipikul Naik-Turun Gunung Tak akan Selamat	kotak suara kardus dipikul naik turun gunung tak akan selamat

Langkah selanjutnya yang dilakukan adalah *stemming*, yaitu mengubah kata-kata dalam *tweet* menjadi kata dasar seperti dalam Tabel 4.

Tabel 4. Contoh Tweet Setelah Stemming

No.	Sebelum	Sesudah
1.	bawaslu gunungkidul menguji kekuatan kotak suara berbahan karton dupleks Setelah kali pengujian bagaimana hasilnya aman dan kuat	bawaslu gunungkidul uji kuat kotak suara bahan karton dupleks setelah kali uji bagaimana hasil aman dan kuat
2.	kotak suara kardus dipikul naik turun gunung tak akan selamat	kotak suara kardus pikul naik turun gunung tak akan selamat

Langkah selanjutnya adalah menghapus *stopword* seperti dalam Tabel 5. Fungsi dari penghapusan *stopword* adalah mengurangi jumlah kata yang akan diproses dan menghilangkan *noise*.

Tabel 5. Contoh Tweet Setelah Menghapus Stopword

No.	Sebelum	Sesudah
1.	bawaslu gunungkidul uji kuat kotak suara bahan karton dupleks setelah kali uji bagaimana hasil aman dan kuat	bawaslu gunungkidul uji kuat kotak suara bahan karton dupleks kali uji hasil aman kuat
2.	kotak suara kardus pikul naik turun gunung tak akan selamat	kotak suara kardus pikul naik turun gunung selamat

Langkah selanjutnya yang dilakukan adalah *tokenization* yaitu proses membagi teks berupa kalimat menjadi token-token atau kata-kata seperti dalam Tabel 6.

Tabel 6. Contoh Tweet Setelah Tokenization

No.	Output
1.	"bawaslu", "gunungkidul", "uji", "kuat", "kotak", "suara", "bahan", "karton", "dupleks", "kali", "uji", "hasil", "aman", "kuat"
2.	"kotak", "suara", "kardus", "pikul", "naik", "turun", "gunung", "selamat"

Term-Weighting TF-IDF

Evaluasi untuk mengetahui seberapa penting suatu kata atau *term* pada koleksi dokumen atau *corpus* juga merupakan salah satu tahap dalam *text mining* dan *information retrieval*. Hal tersebut dititikberatkan pada semakin sering kata dalam dokumen muncul di dalam *corpus*. Teknik reduksi

variabel yang dilakukan pada penelitian ini adalah *term frequency – inverse document frequency*, atau disebut dengan TF-IDF. Nilai TF-IDF meningkat secara proporsional sesuai dengan berapa kali suatu kata muncul pada dokumen. Hal ini bertujuan untuk menyesuaikan fakta bahwa ada beberapa kata yang muncul lebih sering di dalam corpus tersebut secara umum. Setelah dilakukan pembobotan, terdapat 96 kata yang berhubungan dengan kotak suara kardus di Pemilu Indonesia 2019.

Klasifikasi dengan SVM *Linearly Separable Data*

Data yang telah melalui tahap *preprocessing* kemudian di partisi menjadi data *training* dan data *testing* menggunakan metode *5 fold cross validation*. Data *training* digunakan untuk membentuk model klasifikasi sementara data *testing* digunakan untuk menguji model klasifikasi yang telah terbentuk sehingga dapat dievaluasi nilai ketepatan klasifikasinya. Total data adalah 3.145 *tweet*, 4 *fold* berjumlah 2.516 data untuk *training* dan satu *fold* berjumlah 629 data untuk *testing*.

Langkah selanjutnya adalah membangun model klasifikasi menggunakan metode *support vector machine* pada data *training*. Fungsi kernel yang digunakan adalah *linear*. Model SVM yang didapatkan kemudian diuji pada data *testing* untuk melihat ketepatan klasifikasi model. Ketepatan klasifikasi dapat diketahui dengan dilakukannya beberapa pengukuran evaluasi kinerja model seperti nilai akurasi, *specificity*, dan *sensitivity*. Tabel 7 menunjukkan hasil evaluasi model.

Tabel 7. Evaluasi Model

Evaluasi Model	Linear	
	Training	Testing
Akurasi	0,8027	0,7924
Specificity	0,9275	0,9180
Sensitivity	0,6090	0,5918

Berdasarkan Tabel 7, dapat dilihat bahwa nilai akurasi untuk model linier sebesar 0,7924 yang artinya bahwa presentase model SVM dengan fungsi *Kernel Linear* benar melakukan prediksi sebesar 79,24%. Model linier tidak mengalami *overfitting* maupun *underfitting* karena nilai akurasi, *specificity*, dan *sensitivity* antara model *training* dan *testing* tidak terpaut jauh. Nilai *specificity* sebesar 0,9180 artinya presentase kelas sentimen negatif yang berhasil diprediksi sebagai kelas negatif sebesar 91,8%. Nilai *sensitivity* sebesar 0,5918 berarti presentase kelas sentimen positif yang berhasil diprediksi sebagai kelas positif sebesar 59,18%.

Visualisasi Hasil

Visualisasi yang dilakukan menggunakan *word cloud*. Fungsi dari *word cloud* ini adalah untuk mengetahui kata-kata yang sering muncul yang berpengaruh dalam pengerjaan model klasifikasi. Dengan menggunakan *wordcloud* dapat mengetahui frekuensi kata yang sering muncul di kedua kategori yaitu positif dan negatif seperti dalam Gambar 2 dan Gambar 3.



Gambar 2. *Word Cloud* Sentimen Positif



Gambar 3. *Word Cloud* Sentimen Negatif

Berdasarkan Gambar 2, dapat dilihat bahwa sentimen positif terhadap penggunaan kotak suara kardus adalah: aman dan jamin (terjamin aman untuk digunakan); sepakat (disepakati oleh berbagai pihak dalam pemerintahan); tahan dan duduk (tahan beban saat diduduki walaupun hanya sebuah karton); kedap dan air (kedap air meskipun berbahan karton); hemat (menghemat anggaran pemerintah dalam penyelenggaraan pemilu); dukung (didukung oleh berbagai pihak); dan sesuai (sesuai dengan peraturan perundang-undangan).

Berdasarkan Gambar 3, dapat dilihat bahwa sentimen negatif terhadap penggunaan kotak suara kardus adalah: curang dan rawan (rawan kecurangan); polemik (penggunaannya menimbulkan polemik); rusak (kotak suara rawan rusak karena berbahan karton); gila (ide penggunaan kotak suara kardus merupakan ide gila); percaya (kurangnya kepercayaan masyarakat terhadap kotak suara tersebut); dan kawal (menimbulkan kecurangan masyarakat sehingga perlu adanya pengawalan kotak suara tersebut).

SIMPULAN DAN SARAN

Simpulan

Berdasarkan pembahasan di atas, dapat disimpulkan bahwa

1. Sentimen positif terhadap penggunaan kotak suara kardus adalah: aman dan jamin (terjamin aman untuk digunakan); sepakat (disepakati oleh berbagai pihak dalam pemerintahan); tahan dan duduk (tahan beban saat diduduki walaupun hanya sebuah karton); kedap dan air (kedap air meskipun berbahan karton); hemat (menghemat anggaran pemerintah dalam penyelenggaraan pemilu); dukung (didukung oleh berbagai pihak); dan sesuai (sesuai dengan peraturan perundang-undangan). Sedangkan sentimen negatif terhadap penggunaan kotak suara kardus adalah: curang dan rawan (rawan kecurangan); polemik (penggunaannya menimbulkan polemik); rusak (kotak suara rawan rusak karena berbahan karton); gila (penggunaan kotak suara kardus merupakan ide gila); percaya (kurangnya kepercayaan masyarakat terhadap kotak suara tersebut); dan kawal (menimbulkan kecurigaan masyarakat sehingga perlu adanya pengawalan kotak suara tersebut).
2. Hasil klasifikasi penggunaan kotak suara kardus memperoleh akurasi sebesar 0,7924 yang artinya bahwa presentase model SVM dengan fungsi *Kernel Linear* benar melakukan prediksi sebesar 79,24%. Model linier tidak mengalami *overfitting* maupun *underfitting* karena nilai akurasi, *specificity*, dan *sensitifity* antara model *training* dan *testing* tidak terpaut jauh. Nilai *specificity* sebesar 0,9180 artinya presentase kelas sentimen negatif yang berhasil diprediksi sebagai kelas negatif sebesar 91,8%. Nilai *sensitifity* sebesar 0,5918 berarti presentase kelas sentimen positif yang berhasil diprediksi sebagai kelas positif sebesar 59,18%.

Saran

Kepada para peneliti:

1. perlu dilakukan penelitian serupa dengan menggunakan algoritma lain sehingga dapat dibandingkan performanya;
2. data yang didapat merupakan hasil *crawling* sebelum terlaksananya Pemilu 2019 sehingga perlu dilakukan penelitian lagi setelah terlaksananya Pemilu 2019.

DAFTAR PUSTAKA

- Christianini, N. and Taylor, J.S. 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press.
- Faizal, M.R. 2016. *Seni Belajar Pemrograman Supervised Learning dengan R*. Banjarmasin: INDC Indonesia.
- Fikria, N. 2018. *Analisis Klasifikasi Sentimen Review Aplikasi E-Ticketing Menggunakan Metode Support Vector Machine dan Asosiasi*. Skripsi. Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
- Han, J., Kamber, M., and Pei, J. 2012. *Data Mining Concepts and Techniques Third Edition*. Waltham: Morgan Kaufmann.

- Kanimozhi, K.V. and Venkatesan, M. 2015. *Unstructured Data Analysis- A Survey*. International Journal of Advanced Research in Computer and Communication Engineering. Volume 4, Issue 3.
- Latifah, E.F.U. 2018. *Perbandingan Kinerja Machine Learning Berbasis Algoritma Support Vector Machine dan Naïve Bayes*. Skripsi. Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
- Luqyana, W.A. 2018. *Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine*. Skripsi. Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya.
- Novantirani, A. 2015. *Analisis Sentimen pada Twitter Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine*. Proceeding of Engineering. 2(1): 1177-1178.
- Octaviani, P.A. 2014. *Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang*. Jurnal Gaussian. 3(4): 813-814.
- Rajeswary, B. and Divya, S. 2015. *Identification and Classification of Pests in Greenhouse Using Advanced SVM in Image Processing*. International Journal of Scientific Engineering and Research. Volume 3, Issue 5.
- Susilowati, E. 2015. *Implementasi Metode Support Vector Machine untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter*. Proceeding of Engineering. 2(1): 1478.
- Turban, E., Aronson, J., and Liang, T. 2005. *Decision Support System And Intelligent System*. Upper Saddle River, New Jersey USA: Prentice Hall.

Website:

<https://datareportal.com/reports/digital-2019-indonesia> (diakses pada tanggal 12 Februari 2019)