

Menentukan Popularitas Calon Presiden dan Tren pada Pilpres 2019 menggunakan Algoritma DBSCAN

Rifaldi Saputra¹, Mustakim², Muhammad Ridwan³, Okfalisa⁴

Puzzle Research Data Technology (Predatech) Fakultas Sains dan Teknologi^{1,2,3}

Program Studi Sistem Informasi Fakultas Sains dan Teknologi^{1,2,3}

Jurusan Teknik Informatika Fakultas Sains dan Teknologi⁴

Universitas Islam Negeri Sultan Syarif Kasim Riau^{1,2,3,4}

Jl.HR. Soebrantas No. 155 Simpang Baru, Tampan, Pekanbaru, Riau - Indonesia 28293

e-mail: rifaldisaputra11@gmail.com¹; mustakim@uin-suska.ac.id²; muhammad.ridwan02@student.uin-suska.ac.id³; okfalisa@gmail.com⁴.

Abstrak

Indonesia merupakan negara demokrasi terbesar ketiga di dunia. Salah satu cerminan demokrasi tersebut adalah pemilihan presiden. Seorang tokoh politik yang ingin maju menjadikan opini masyarakat yang sekarang disampaikan melalui media sosial sebagai pertimbangan. Peranan media sosial menjadi sangat penting, karena mampu mendongkrak suara secara signifikan bahkan dijadikan senjata baru bagi banyak bidang terutama kampanye politik. Salah satu media sosial yang sangat populer adalah Twitter. Twitter menggunakan tweet yang mengandung data apabila diolah dapat menjadi informasi. Data dari tweet dapat dijadikan bahan untuk mencari opini masyarakat terhadap calon presiden dan pola yang terbentuk dan pengetahuan pada pemilihan presiden 2019. Untuk menangani tweet yang berbentuk data tekstual dapat dilakukan dengan menggunakan text mining. Metode yang digunakan adalah algoritma partitioning clustering yaitu Density-Based Spatial Clustering of Application with Noise (DBSCAN). Hasil dari penelitian ini adalah DBSCAN menjadi metode terbaik karena mempunyai validitas silhouette index (SI) sebesar 0.8094 dan waktu eksekusi di RapidMiner 2.5676 detik. Frekuensi nama Joko Widodo mendominasi kategori positif, negatif dan netral. Hasil penelitian ini dapat digunakan untuk orang, organisasi dan proses bisnis yang berkaitan erat dengan Pilpres 2019.

Kata kunci: Calon Presiden, DBSCAN, Text Mining, Twitter dan Pemilihan Presiden 2019

Abstract

Indonesia Indonesia is the third largest democracy in the world. One reflection of that democracy is the presidential election. A political figure who wants to move forward makes public opinion now conveyed through social media a consideration. The role of social media is very important, because it is able to jack up the sound significantly and even become a new weapon in many fields, especially political campaigns. One very popular social media is Twitter. Twitter uses tweets that contain data when it is processed into information. Data from tweets can be used as material to seek public opinion on candidates of presidential and patterns formed and knowledge in the 2019 presidential election. To handle tweets in the form of textual data can be done using text mining. The method used is partitioning clustering algorithm, namely Density-Based Spatial Clustering of Application with Noise (DBSCAN). The results of this study are DBSCAN to be the best method because it has a silhouette index (SI) validity of 0.8094 and an execution time on RapidMiner 2.5676 seconds. The frequency of the name Joko Widodo dominates the positive, negative and neutral categories. The results of this study can be used for people, organizations and business processes that are closely related to the Presidential Election 2019.

Keywords: Candidate of Presidential, DBSCAN, Text Mining, Twitter dan Presidential Election 2019

1. Pendahuluan

Indonesia merupakan negara yang menganut sistem demokrasi terbesar ketiga di dunia setelah Amerika Serikat dan India [3]. Nilai demokrasi tercermin melalui kebebasan dan keterlibatan partai politik atau gabungan partai politik peserta pemilu untuk mengusung calon presiden dan wakil presiden sepanjang memenuhi persyaratan yang ditentukan dalam peraturan perundang-undangan [13]. Pemilihan presiden akan diadakan pada tahun 2019. Seorang tokoh politik yang akan maju sebagai calon presiden tentu menjadikan popularitas mereka berdasarkan opini masyarakat sebagai bahan pertimbangan [18].

Pada zaman sekarang opini cenderung disampaikan melalui media sosial [18]. Media menjadi kekuatan keempat setelah pemerintah, legislatif dan masyarakat. Media dianggap sebagai *watch dog* yang menjaga keberlangsungan demokrasi dan kepentingan publik dalam demokrasi [3]. Media sosial bahkan dijadikan senjata baru bagi banyak bidang bahkan kampanye politik pada pemilu 2014 termasuk salah satunya [9]. Media sosial merupakan salah satu *channel* yang efektif dalam menguasai komunikasi publik sebagai salah satu kunci guna memenangkan kompetisi pada

dunia politik [2]. Salah satu media sosial yang sangat populer adalah *Twitter* [1]. Indonesia adalah negara dengan pengguna aktif *Twitter* ketiga terbesar di dunia [7]. Pada akhir tahun 2016 menyebutkan bahwa 77 persen pengguna *Twitter* Indonesia adalah pengguna aktif [4]. *Twitter* merupakan aplikasi yang sebagian besar menggunakan tweet untuk mengekspresikan perasaan pengguna [17].

Data dari *tweet* ini dapat dijadikan bahan untuk mencari opini masyarakat terhadap calon presiden dan wakil presiden dan pola yang menarik terkait pemilihan presiden 2019 [6]. Untuk menarik kesimpulan hasil agregasi dari *tweet* menggunakan text mining [20]. Text mining merupakan proses mengekstrak pola yang menarik dan signifikan untuk mengeksplorasi pengetahuan dari sumber data tekstual [19]. Pada penelitian ini dilakukan dengan melakukan *clustering*. *Clustering* adalah teknik analisis data yang populer dan dapat diterapkan secara luas di berbagai bidang termasuk: pengenalan pola, *machine learning*, pemrosesan gambar, pencarian informasi dan sebagainya [18].

DBSCAN merupakan metode pengklusteran *partitioning* [5]. Metode pengklusteran *partitioning* lebih cepat dibandingkan metode hierarki dan lebih menguntungkan apabila dalam jumlah objek yang besar [16]. DBSCAN adalah algoritma yang melakukan cluster berdasarkan parameter yang diberikan radius lingkungan, *eps* dan jumlah poin minimum, *MinPts* [12].

Penelitian tentang text mining dan metode *clustering* yang digunakan telah diterapkan, diantaranya adalah Pencarian trend Pilkada Pekanbaru dengan menggunakan Algoritma DBSCAN telah dilakukan oleh Reza (2017) menggunakan teknik text mining berdasarkan data *Twitter*. Hasil dari penelitian ini ternyata sesuai dengan hasil Pilkada Pekanbaru dimana pemenangnya adalah Firdaus dan Ayat menjadi Walikota dan wakil walikota [14]. Penelitian dari Nurhuda dkk (2013) tentang Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari *Twitter* menggunakan Metode Naïve Bayes Classifier. Hasil dari penelitian ini adalah popularitas sentiment masyarakat terhadap pasangan Prabowo Subianto – Hatta Rajasa mendapatkan sentiment positif sebesar 47,7 %, sentiment negatif 26,4 % dan sentiment netral 25,9%. Sedangkan pasangan Joko Widodo – Jusuf Kalla mendapatkan total sentiment positif 37,6%, sentiment negatif 34,4 dan sentiment netral 27,9 % [18]. Penelitian dari Lukmana dkk (2019) tentang Analisis Sentimen Terhadap Calon Presiden 2019 Dengan Support Vector Machine Di *Twitter*. Dari penelitian tersebut didapatkan hasil bahwa algoritma klasifikasi SVM menggunakan fungsi Kernel dengan melakukan klasifikasi dengan kata kunci Jokowi mendapatkan akurasi sebesar 86,82% dan kata kunci Prabowo dengan mendapatkan akurasi sebesar 86,27%. Untuk kata kunci Jokowi mendapatkan sentiment positif dari tweet tentang pembangunan infrastruktur di Indonesia, sedangkan sentiment negatif didapatkan dari tweet mengenai pernyataan Jokowi tentang kebakaran hutan yang tidak sesuai data pada debat kedua. Untuk kata kunci Prabowo mendapatkan sentiment positif tentang elektabilitas Prabowo sebagai oposisi yang berkualitas termasuk sedangkan sentiment negatif didapatkan dari pernyataan Prabowo tentang data luas daerah dan unicorn yang tidak sesuai data pada debat kedua.

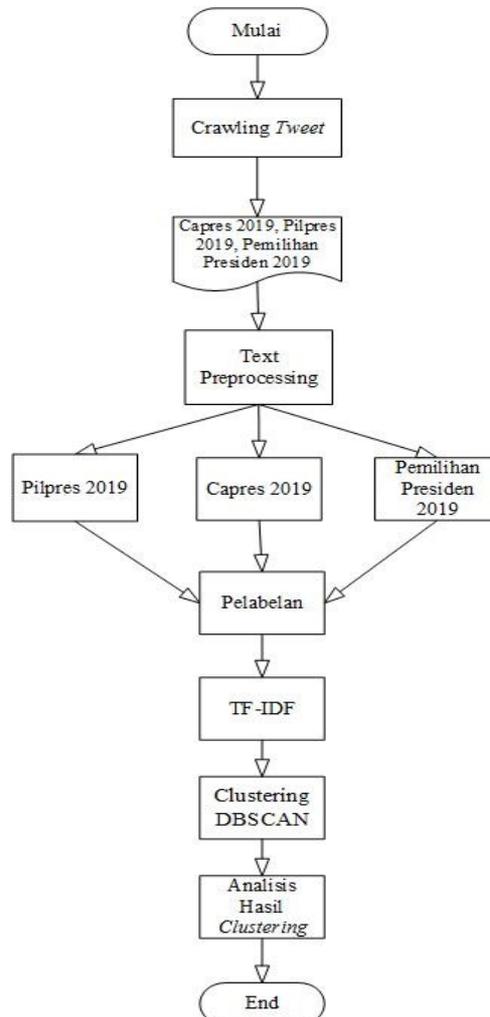
Penelitian ini bertujuan untuk membandingkan hasil calon terkuat baik sebelum ditetapkan secara resmi untuk maju menjadi calon presiden yang diambil dari lembaga survei di Indonesia dengan setelah penetapan resmi oleh Komisi Pemilihan Umum (KPU) baik secara positif, negatif dan netral. Selain itu, untuk mengetahui tren dari Pemilihan Presiden 2019.

2. Metode Penelitian

Penelitian ini diawali dengan pengambilan dataset yang bersumber dari media sosial *twitter*, dengan menggunakan teknik pengambilan data yaitu *crawling*. Data yang telah terkumpul akan diberi label secara otomatis. Tahapan selanjutnya yaitu tahap preprocessing serta pembobotan kata dengan metode TF-IDF, setelah tahapan ini selesai maka selanjutnya adalah mencari pola pengelompokan/*clustering* keterkaitan antar data menggunakan algoritma DBSCAN, jika pola telah ditentukan maka pola akan dianalisis untuk mendapatkan suatu pengetahuan. Adapun langkah-langkah yang dilakukan dalam penelitian ini terdapat pada Gambar 1.

Indonesia termasuk negara ketiga yang menganut sistem demokrasi terbesar setelah Amerika Serikat dan India yang dipimpin oleh kepala negara yang disebut dengan presiden. Secara periodik pemilihan presiden ini diadakan sekali dalam 5 tahun [3], [18]. Menurut survei, yang dilakukan PolMark 10 tokoh yaitu Joko Widodo, Prabowo Subianto, Agus Harimurti Yudhoyono, Anies Baswedan, Hary Tanoesoedibjo, Gatot Nurmantyo, Jusuf Kalla, Megawati Soekarno, Rhoma Irama dan Mahfud MD. Berdasarkan Peraturan (PKPU) Nomor 5 tahun 2018 tentang Tahapan, Program, dan Jadwal Penyelenggaraan Pemilu 2019, jadwal pendaftaran berlangsung selama tujuh hari, yaitu 4-10 Agustus 2018. Berdasarkan hasil rapat Pleno dari

pendaftaran tersebut hanya menghasilkan 2 pasangan calon presiden dan wakil presiden yaitu Joko Widodo dengan KH Ma'aruf Amin dan Prabowo Subianto dengan Sandiaga Uno ditetapkan sebagai calon presiden dan wakil presiden 2019 [10], [11].



Gambar 1. Metode Penelitian

Pengklastran dan analisis data sentimen dari *tweet-tweet* masyarakat Indonesia akan dilakukan dengan menggunakan algoritma DBSCAN. Pengolahan data dengan algoritma ini dilakukan berdasarkan beberapa parameter yaitu radius lingkungan, *Eps* dan jumlah poin, *MinPts* [12]. Alur ini dapat dilihat pada Gambar 2.

3. Analisa dan Hasil

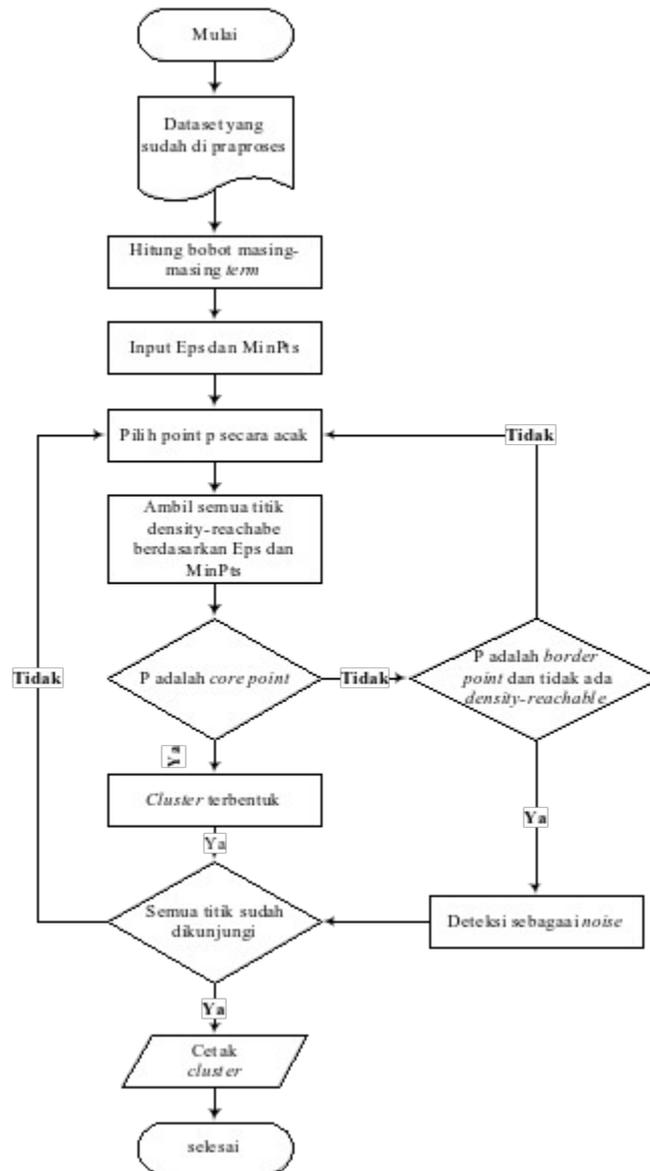
3.1. Analisis Data

Data yang digunakan berupa data sekunder karena diambil dari *Twitter*. Data ini diperoleh dengan dua tahap. Tahap pertama adalah sebelum pencalonan dimulai dari tanggal 25 Februari sampai 3 Agustus 2018 (160 hari) dan tahap kedua adalah setelah pencalonan dimulai dari 4 Agustus sampai 12 September 2018 (40 hari atau 25% dari tahap sebelum pencalonan). Data sebelum pencalonan didapatkan sebanyak 8.194 data dan data setelah pencalonan didapatkan sebanyak 5.369. Data yang telah didapat akan dilakukan *duplicate removal* sehingga data sebelum pencalonan menjadi 6.838 data dan setelah pencalonan menjadi 4.374 data. Atribut yang digunakan adalah isi *tweet* atau text dari data *Twitter*.

3.2. Text Preprocessing

Data dilakukan proses *cleaning* untuk merubah teks menjadi huruf kecil, menghapus semua karakter selain huruf, *username* pengguna, URL, *retweet* dan *hashtag*. Data hasil *cleaning*

akan diproses pada tahap *Text Preprocessing*. Tahap ini terdiri atas 3 tahap yaitu filtering, tokenizing dan stemming. Hasil *Text Preprocessing* dapat dilihat pada Tabel 1 dan 2.



Gambar 2. Alur DBSCAN

Tabel 1. Hasil *Text Preprocessing* Sebelum Pencalonan

No	Teks
1	partai rakyat dukung jokowi capres
2	anies peluang aja ngambil ilusi
...	...
6696	temu rawan sulsel

Tabel 2. Hasil *Text Preprocessing* Setelah Pencalonan

No	Teks
1	pagi sob pasti pilih capres gandeng ya
2	arogan kuasa media asai lebih lemah capres prabowo lemah
...	...
78	hujan batu ya kancapolitik com hujan batu ket ika kampung narkoba

3.3. Pelabelan dan Pembobotan

Data akan dilakukan pelabelan menjadi 6 kategori yaitu data sebelum pencalonan label negatif, data sebelum pencalonan label netral, data sebelum pencalonan label positif, data setelah

pencalonan label negatif, data setelah pencalonan label netral dan data setelah pencalonan label positif. positif, negatif dan netral. 6 data yang telah dilabeli akan dilakukan pembobotan menggunakan Term Frequency-Inverse Document Frequency. Hasil pembobotan tersebut dapat dilihat pada Tabel 3, 4, 5, 6, 7 dan 8.

Tabel 3. Pembobotan TF-IDF Sebelum Pencalonan Label Negatif

No	Kata						
	aja	berita	calon	capres	...	wakil	ya
	1	2	3	4	...	38	39
1	1.000	0.000	0.000	0.000	...	0.000	0.000
2	0.000	0.000	0.000	0.000	...	0.000	0.000
...
825	0.000	0.000	0.000	0.000	...	0.000	0.000

Tabel 4. Pembobotan TF-IDF Sebelum Pencalonan Label Netral

No	Kata						
	calon	capres	cawapres	daftar	...	ulama	usung
	1	2	3	4	...	23	24
1	0.000	0.336	0.000	0.000	...	0.000	0.000
2	0.000	0.612	0.791	0.000	...	0.000	0.000
...
3238	0.000	0.000	0.000	0.000	...	0.000	0.000

Tabel 5. Pembobotan TF-IDF Sebelum Pencalonan Label Positif

No	Kata						
	ahy	baik	berita	calon	...	ulama	widodo
	1	2	3	4	...	39	40
1	0.000	0.000	0.000	0.000	...	0.000	0.000
2	0.000	0.000	0.000	0.000	...	0.000	0.000
...
1495	0.000	0.000	0.000	0.000	...	0.000	0.000

Tabel 6. Pembobotan TF-IDF Setelah Pencalonan Label Negatif

No	Term						
	aja	amin	calon	capres	...	wakil	ya
	1	2	3	4	...	43	44
1	0.000	0.000	0.000	0.472	...	0.000	0.000
2	0.000	0.000	0.000	0.472	...	0.000	0.000
...
746	0.000	0.000	0.000	0.000	...	0.000	0.000

Tabel 7. Pembobotan TF-IDF Setelah Pencalonan Label Netral

No	Term						
	aja	amin	calon	capres	...	widodo	ya
	1	2	3	4	...	41	42
1	0.000	0.000	0.553	0.246	...	0.000	0.000
2	0.000	0.000	0.000	1.000	...	0.000	0.000
...
831	0.000	0.000	0.000	0.000	...	0.000	0.000

Tabel 8. Pembobotan TF-IDF Setelah Pencalonan Label Positif

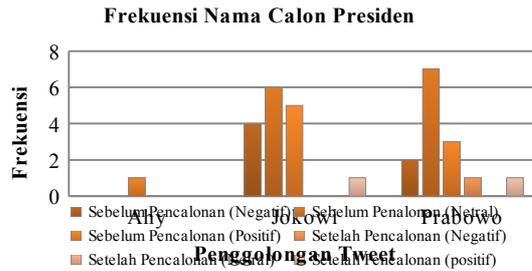
No	Term						
	amin	calon	capres	cawapres	...	widodo	ya
	1	2	3	4	...	23	24
1	0.000	0.000	0.390	0.000	...	0.000	0.814
2	0.000	0.000	0.000	0.000	...	0.000	0.000
...
2142	0.000	0.000	0.000	0.000	...	0.000	0.741

3.4. Algoritma Density-Based Spatial Clustering of Application with Noise (DBSCAN)

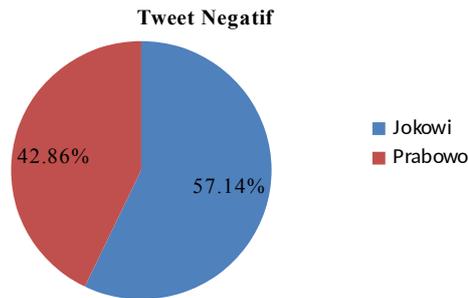
Algoritma DBSCAN diolah menggunakan *Rapidminer* dengan beberapa parameter *Eps* dan *MinPts* yang berbeda. Adapun *Eps* yang digunakan yaitu: 0.3, 0.4, 0.5 dan 0.6 sedangkan

MinPts yang digunakan yaitu: 5, 6, 7, 8, 9 dan 10. Percobaan dilakukan sebanyak 24 kali untuk setiap kategori sehingga berjumlah 144 kali percobaan. Waktu eksekusi dari rata-rata 144 kali percobaan tersebut adalah sebesar 2.5764 detik. Berdasarkan analisis 6 data dengan DBSCAN didapatkan hasil bahwa Frekuensi nama calon presiden sebelum dan setelah pencalonan terdapat pada Gambar 3.

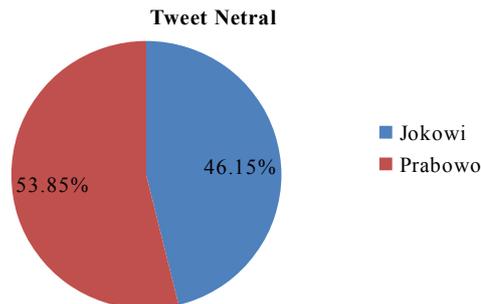
Berdasarkan *tweet-tweet* untuk nama calon presiden baik sebelum dan setelah pencalonan yang telah *dicluster* didapatkan 3 label. Persentase 3 label dapat dilihat pada Gambar 4, 5 dan 6.



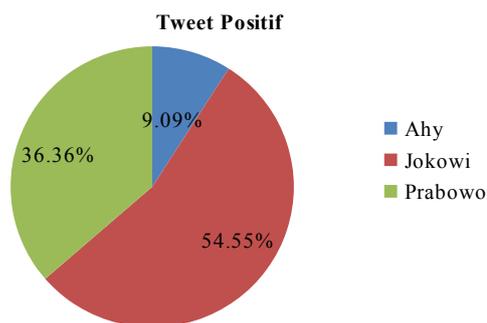
Gambar 3. Frekuensi Nama Calon Presiden Sebelum dan Setelah Pencalonan



Gambar 4. Tweet Negatif Calon Presiden



Gambar 5. Tweet Netral Calon Presiden



Gambar 6. Tweet Positif Calon Presiden

3.5 Silhouette Index (SI)

Silhouette index digunakan untuk menentukan *cluster* terbaik. Penelitian ini terdiri dari kategori sebelum dan setelah pencalonan dengan 3 label yaitu negatif, positif, dan netral pada setiap kategori. SI keseluruhan diambil dari rata-rata SI semua data. SI data sebelum pencalonan label negatif adalah 0.7762, SI data sebelum pencalonan label netral adalah 0.8618, SI data sebelum pencalonan label positif adalah 0.8158, SI data setelah pencalonan label negatif adalah 0.7623, SI data setelah pencalonan label netral adalah 0.7913 dan SI data setelah pencalonan label positif adalah 0.8487. Sehingga rata-rata SI keseluruhan yaitu 0.8094.

4. Kesimpulan

Berdasarkan hasil dari penelitian dan pembahasan yang telah dilakukan. Hasil frekuensi kata nama calon presiden pada metode DBSCAN untuk waktu sebelum pencalonan dengan label negatif dan positif didominasi oleh Joko Widodo dan netral didominasi oleh Prabowo. Untuk waktu setelah pencalonan dengan label positif didominasi oleh Joko Widodo dan negatif didominasi oleh Prabowo sedangkan netral didominasi oleh Prabowo dan Joko Widodo. Pola yang terbentuk pada pemilihan presiden 2019 didapatkan dari analisis semua cluster. Pada Algoritma DBSCAN untuk sebelum pencalonan tweet negatif tentang berita tentang pengusungan calon presiden, tweet netral politik yang terjadi dalam pemilihan presiden, tweet positif adalah dukungan dari berbagai kalangan untuk calon yang maju dalam pemilihan presiden dan untuk setelah pencalonan tweet negatif tentang gerakan yang dilakukan rakyat dalam memberikan dukungan untuk masing-masing calon, tweet netral tentang sebagian rakyat ingin mengganti presiden dan lainnya tidak dan tweet positif tentang kampanye pasangan calon presiden dan wakil presiden. Selain itu algoritma DBSCAN mempunyai *Silhouette Index* yang tinggi sebesar 0.8094.

Daftar Pustaka

- [1] Adarsh, M. J. dan Ravikumar, P. *Survey: Twitter Data Analysis Using Opinion Mining, International Journal of Computer Applications*. Vol. 128, No.5. Oktober 2015.
- [2] Ardha, Berliani. *Social Media sebagai Media Kampanye Partai Politik 2014 di Indonesia*, Jurnal Visi Komunikasi. Vol. 13, No. 1, Hal. 105-120. Mei 2014.
- [3] Bachtiar, F. R. *Pemilu Indonesia: Kiblat Negara Demokrasi dari Berbagai Representasi*, Jurnal Politik Profetik. Vol 3, No. 1. 2014.
- [4] Beritasatu.com. *Indonesia Masuk Lima Besar Pengguna Twitter*. 3 Mei 2017. [Online] Available. <http://www.beritasatu.com/digital-life/428591-indonesia-masuk-lima-besar-pengguna-twitter.html>, diakses 20 Maret 2018.
- [5] Budiman, S.A.D., Safitri, D. dan Ispriyanti, D. *Perbandingan Metode K-Means dan Metode DBSCAN pada Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang Semarang*, Jurnal Gaussian. Vol. 5, No. 4, Hal. 757-762. 2016.
- [6] Hidayatullah, A.F., dan Ma'arif, M.R. *Penerapan Text Mining dalam Klasifikasi Judul Skripsi*, Seminar Nasional Teknologi Informasi (SNATI). Yogyakarta, Indonesia. Agustus 2016.
- [7] Katadata.co.id. *Indonesia Pengguna Twitter besar Ketiga di Dunia*. 22 November 2016. [Online] Available. <https://databoks.katadata.co.id/datapublish/2016/11/22/indonesia-pengguna-twitter-terbesar-ketiga-di-dunia>, diakses 20 Maret 2018.

- [8] Lukmana, D.T, Subanti, S. dan Susanti, Y. *Analisis Sentimen Terhadap Calon Presiden 2019 dengan Support Vector Machine di Twitter*, Seminar Nasional Penelitian Pendidikan Matematika (SNP2M). Tangerang, Indonesia. Juni 2019.
- [9] Mulawarman dan Nurfitri, A.D. *Perilaku Pengguna Media Sosial beserta Implikasinya ditinjau dari Perspektif Psikologi Sosial Terapan*, Buletin Psikologi. Vol. 25, No. 1, Hal. 36-44. 2017. 15.
- [10] Nasional.kompas.com. *10 Capres dengan Elektabilitas Tertinggi Menurut Survey PolMark. 19 Oktober 2017*. [Online] Available <https://nasional.kompas.com/read/2017/10/22/17325561/10-capres-dengan-elektabilitas-tertinggi-menurut-survei-polmark>, diakses 10 Agustus 2018.
- [11] News.detik.com. *KPU Tetapkan Jokowi dan Prabowo Capres 2019*. 20 September 2018. [Online] Available. <https://news.detik.com/berita/4221168/kpu-tetapkan-jokowi-dan-prabowo-capres-2019>, diakses 10 Oktober 2018.
- [12] Nurhuda, F., Sihwi, S. W. dan Doewes, A. "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter menggunakan Metode Naïve Bayes Classifier," JURNAL ITSMART. Vol. 2, No. 2. Desember 2013.
- [13] Ozkok, F.O., dan Celik, M. *A New Approach to Determine Eps Parameter of DBSCAN Algorithm*, *International Journal of Intelligent Systems and Applications in Engineering (IJSAE)*. Vol. 5(4), Hal. 247-251. Desember 2017.
- [14] Rauta, Umbu. *Menggagas Pemilihan Presiden*, Jurnal Konstitusi. Vol. 11, No. 3. September 2014.
- [15] Reza, N. G. *Penerapan Algoritma DBSCAN untuk Pencarian Trend Topik Pilkada Pekanbaru 2017 pada Twitter*. [Skripsi] UIN SUSKA Riau. 2017.
- [16] Sangga, V. A. P. *Perbandingan Algoritma K-Means dan Algoritma K-Medoids dalam Pengelompokan Komoditas Peternakan di Provinsi Jawa Tengah Tahun 2015*. [Skripsi] Universitas Islam Indonesia. 2018.
- [17] Sari, Y.A., dkk. *User Emotion Identification in Twitter Using Specific Features: Hashtag, Emoji, Emoticon, and Adjective Term*, *Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)*. Vol. 7, Issue. 1, Hal. 18-23. Februari 2014.
- [18] Suthar, N. dkk. *A Technical Survey on DBSCAN Clustering Algorithm*, *International Journal of Scientific & Engineering Research*. Vol. 4, Issue 5. Mei 2013.
- [19] Talib, R. dkk. *Text Mining: Techniques, Applications and Issues*, *International Journal of Advanced Computer Science and Applications (IJACSA)*. Vol. 7, No. 11. 2016.
- [20] Yogapreethi, N. dan Maheswari, S. *A Review on Text Mining in Data Mining*, *International Journal on Soft Computing (IJSC)*. Vol. 7, No. 2/3. Agustus 2016.