

## MODEL KLASTERING SKM3 (*SUBCONTROLLED K-MEANS MAX-MIN*) DAN APLIKASINYA DALAM MENGHITUNG ELEKTABILITAS PASANGAN CALON KEPALA DAERAH

Patuan P. Tampubolon<sup>1</sup>, Tesdiq Prigel Kaloka<sup>2</sup>, Olivia Swasti<sup>3</sup>,  
Widya Fajar Mustika<sup>4</sup>, Alhadi Bustamam<sup>5</sup>

<sup>1,2,3,4,5</sup>Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Indonesia

**Abstract:** Indonesia is a legal state that chooses a leader based on the results of general elections, such as the election of presidents and regional leaders. Electability is statistical data for each pair of candidates who show public interest to choose the candidate. Electability data is usually obtained from the results of questionnaires or interviews with constituents. The data search process is carried out by a survey institution. Most people discuss voluntarily in social media related to the candidate that they will choose. This study uses discussion data from social media to calculate the electability of each pair of candidates by using cluster method. The cluster method is K-Means. K-Means employs euclidean distance to determine the cluster of each data, while the number of cluster can be determined by the user. This study proposes SKM3 model (Subcontrolled K-Means Max-Min), which applies the minimum and maximum average values to decide the cluster of each data. SKM3 cluster is controlled by K-Means method that uses Euclidian distance. SKM3 model is processed using news data from *detik.com* site for the election of regional leader of West Java, Central Java, and East Java. The error value of SKM3 model is calculated through RMSE (Root Mean Square Error). The error value of West Java is 0.0452, the error value of Central Java up to 0.0343, and the error value of East Java is 0.2382. Based on the error values of each electoral region, it shows that SKM3 model has a small error value, so it can be concluded that SKM3 model is good for calculating the electability of the leader by using clustering method.

**Keywords:** *Electability, Clustering, K-Means, SKM3.*

### PENDAHULUAN

Klasterisasi sudah banyak digunakan dalam berbagai bidang dengan beragam aplikasi vital yang sangat penting, diantaranya adalah: riset pasar, dimana klasterisasi digunakan untuk segmentasi dan *profiling* pelanggan yang membantu dalam merancang strategi-strategi produk, harga, tempat dan promosi. Klasterisasi juga digunakan untuk mengimplementasikan *customer relationship management* (CRM) yang efektif; sistem rekomendasi produk dalam sistem jual beli *daring*; *business intelligence*; sistem keamanan; mesin pencarian di internet (*search engine*), dan sebagainya (Suyanto.2017).

Salah satu metode klasterisasi yang sudah terbukti sangat baik adalah algoritma *K-Means*. Metode ini menentukan beberapa fitur menjadi pusat klaster kemudian menghitung jarak antar fitur yang lain pada dataset terhadap pusat, kemudian mengelompokkan anggota-anggota yang memiliki kesamaan, yaitu yang memiliki jarak terdekat pada satu pusat yang ada. Metode perhitungan jarak yang paling sering digunakan adalah jarak *Euclidean*, sedangkan metode jarak yang lain adalah metode jarak *Cosinus*.

Penelitian ini menggunakan algoritma *K-means Max-min* yang diusulkan oleh Visalakshi dan Suguna (2009). Algoritma ini dikontrol oleh metode *K-means* yang menggunakan jarak *Euclidean*. Adapun model matematika yang diusulkan ini diberi nama *Sub-controlled K-Mean Max-min (SKM3)*. Untuk mengukur performa dari model matematika SKM3, akan ditunjukkan aplikasi pada perhitungan elektabilitas calon gubernur dan wakil gubernur Jawa Barat, Jawa Tengah dan Jawa Timur untuk masa jabatan 2018-2023. Data yang digunakan merujuk pada situs <https://pilkada.detik.com/> dengan mengamati berita yang terbit dari bulan Januari 2018 sampai Mei 2018. Daerah Jawa Barat diambil 80 berita dengan setiap pasangan mempunyai pemberitaan sebanyak 20 berita. Untuk daerah Jawa Tengah dan Jawa Timur yang masing-masing daerah diambil 102 berita.

Penelitian tentang bidang politik yang menggunakan data dari sosial media telah banyak dilakukan, antara lain, Schroeder dkk (2015) melakukan analisa efek berita pada *Fox News* terhadap hasil pemungutan suara. Kemudian Shim dkk (2015) menggunakan metode *Structural Equation Modeling (SEM) Analysis* dimana setiap parameter diestimasi dengan metode *Maximum Likelihood* dalam menelusuri motif pada penggunaan berita seluler. Panagiotopoulos dkk (2017) melakukan analisa kebijakan-kebijakan yang berkaitan dengan pertanian di Britania Raya melalui media daring *twitter*. Selain itu, Yacub dkk (2017) juga menggunakan media daring *twitter* dalam menganalisa pidato politik pada pemilihan presiden US tahun 2016. Choi dkk (2017) melakukan analisa aktivitas-aktivitas di mediasosial yang berhubungan dengan partisipasi ketika masyarakat melakukan pembicaraan politik melalui sosial media dibanding yang tidak dengan tingkah laku di dunia nyata. Kim (2018) menganalisa cara penggunaan berita pada media online *facebook* di Korea Selatan.

## **METODE PENELITIAN**

Penelitian ini menggabungkan proses teori dan praktek, dengan proses teori meliputi studi literatur yang menjadi dasar penelitian ini. Proses praktikal meliputi proses pengumpulan data dan proses penarikan kesimpulan dengan bantuan program R. Berikut disajikan hal-hal yang berkaitan dengan penelitian ini.

### **A. Metode Klasterisasi**

Klasterisasi adalah proses pengelompokkan himpunan data kedalam beberapa grup atau klaster sedemikian sehingga objek-objek dalam grup tersebut memiliki kemiripan yang tinggi, namun sangat berbeda terhadap objek-objek di grup yang lain (J Han et al.2012). Metode klasterisasi yang telah ada sangat beragam, metode-metode tersebut dapat dikelompokkan kedalam empat kategori:

### 1. Metode Partisi

Bekerja dengan cara membagi data kedalam sejumlah kelompok. Metode ini juga dikenal dengan metode berbasis pusat atau metode berbasis representatif (zaki et al.2013) karena bekerja dengan menentukan pusat-pusat klaster, dimana pusat klaster bisa berupa rata-rata, modus, atau sebuah objek representatif dari semua objek dalam suatu klaster berdasarkan suatu ukuran tertentu. Algoritma-algoritma yang termasuk kedalam metode berbasis partisi adalah *k-means*, *k-modes*, *k-medoids*, *fuzzy c-means*, dll.

### 2. Metode Hirarki

Bekerja dengan cara mengelompokkan objek-objek data kedalam sebuah hirarki klaster. Hirarki disini hanya untuk merangkum dan merepresentasikan data secara ringkas agar mudah dalam visualisasi. Algoritma-algoritma yang termasuk kedalam *hierarchical clustering* dapat dikelompokkan kedalam tiga kategori, yaitu: metode algoritmik, metode probabilistik dan metode Bayesian.

### 3. Metode Berbasis Kepadatan

Metode ini digunakan untuk mengatasi permasalahan kompleksitas yang terdapat pada metode hirarki. Algoritma yang menggunakan metode berbasis kepadatan, yaitu OPTICS (Ankerst et al.1999) dan DENCLUE (Hinneburg et al. 1998).

### 4. *Grid-based methods*

Metode ini digunakan untuk mengatasi masalah kompleksitas komputasi yang tinggi yang terdapat pada *partitioning methods*, *hierarchical methods*, dan *density-based method*. Keunggulan dari metode ini yaitu komputasi yang sangat cepat dan tidak bergantung pada jumlah objek data.

## B. Ukuran Jarak

Terdapat beberapa ukuran jarak menurut Soman K.P. et al(2006) dan Wunsch X. (2005), yaitu:

1. *Euclidean distances* biasanya dihitung dari *raw data* dan bukan dari standarisasi data.

$$d(t_i, t_j) = \sqrt{\sum_{k=1}^d (t_{ik} - t_{jk})^2} \quad 1)$$

2. *Cosine coefficient* berhubungan dengan overlap dengan rata-rata geometrik pada dua himpunan.

$$d(t_i, t_j) = \frac{\sum_{h=1}^d t_{ih} t_{jh}}{\sqrt{\sum_{h=1}^d t_{ih}^2 \sum_{h=1}^d t_{jh}^2}} \quad 2)$$

3. *Max-min distance measure* dapat dicari dengan cara perhitungan operasi *min* dan *max* pada setiap pasangan dalam data objek.

$$d(t_i, t_j) = \frac{\sum_{k=1}^d \min_i(t_{ik}, t_{jk})}{\sum_{k=1}^d \max_i(t_{ik}, t_{jk})} \quad 3)$$

Pastikan seluruh singkatan dan akronim didefinisikan pada saat penggunaan pertama kali untuk menghindari salah interpretasi. Singkatan yang lazim seperti SD atau SMP tidak perlu didefinisikan.

### C. Algoritma K-means

Metode *k-means* dapat digunakan dengan mudah diimplementasikan dalam berbagai aplikasi kecil hingga menengah. Riset tentang *k-means* telah banyak dilakukan dari beragam ilmu sejak tahun 1950-an, diantaranya adalah Lloyd (1957,1982) Forgey (1965), Friedman dan Rubin (1967) serta MacQueen (1967). Ide dasar dari algoritma *k-means* sangatlah sederhana, yaitu meminimalkan *Sum of Squared Error (SSE)* antara objek-objek data dengan sejumlah *k* sentroid. Algoritma *K-means* dapat dilihat pada Tabel 1.

**Tabel 1. Algoritma K-Means Clustering**

<p>Algoritma <i>k-means clustering</i></p> <p><b>K-means (D,k)</b>                  Pilih sejumlah <i>k</i> objek secara acak dari himpunan data <i>D</i> sebagai sentroid awal</p> <p><b>Repeat</b>                  for semua objek di dalam <i>D</i>                      Masukkan setiap objek yang bukan sentroid ke kluster yang paling dekat diantara <i>k</i> kluster yang ada</p> <p><b>End</b>                  Perbarui setiap sentroid dengan menghitung rata-rata dari semua objek yang berada didalam kluster tersebut</p> <p><b>until</b> tidak ada perubahan sentroid</p>
---

Fungsi objektif pada *k-means clustering* ini berdasarkan pada jarak Euclidean antara objek *x* pada grup *j* dan kluster yang bersesuaian dengan sentroid *C<sub>j</sub>*,m dapat didefinisikan sebagai:

$$J = \sum_{j=1}^K \sum_{i=1}^n \|x_i - C_j\|^2 \quad 4)$$

### D. K-means Max-Min Clustering

Secara umum, algoritma *k-means clustering* menggunakan jarak Euclidean sebagai ukuran jarak untuk menghitung *similarity* antara objek dengan sentroidnya. Pada paper

(Visalakshi N., Suguna C. 2009) ukuran jarak *max-min* dapat mengganti ukuran jarak Euclidean. Ukuran jarak *max-min* memerlukan nilai dalam rentang [0,1]. Prosedur normalisasi *min-max* (Shalabi et al.2006) diikuti sebagai langkah *pre-processing* yang diusulkan untuk algoritma *k-means*. Algoritma *K-means Max-Min* dapat dilihat pada Tabel 2.

**Tabel 2. Algoritma K-Means Berdasarkan Jarak Max-Min**

<b>Input</b>	: Dataset dari $n$ objek dengan banyak $d$ fitur dan nilai $K$
<b>Prosedur</b>	:
<b>Step 1</b>	: Menormalisasikan data input ke dalam jangkauan [0,1] menggunakan prosedur Normalisasi <i>Max-min</i>
<b>Step 2</b>	: Mendeklarasikan keanggotaan matrix $U$ yang berdimensi $n \times K$
<b>Step 3</b>	: Membangun kluster sebanyak $K$ secara random dari data sebagai pusat kluster (sentroid) awal. Misal sentroid awal adalah $C_1, C_2, \dots, C_K$
<b>Step 4</b>	: Menghitung ukuran jarak $d_{ij}$ menggunakan Similaritas Ukuran <i>Max-Min</i> $d_{ij} = \frac{\sum_{k=1}^d \min(x_{ik}, C_{jk})}{\sum_{k=1}^d \max(x_{ik}, C_{jk})}$ untuk semua <i>sentroid</i> $C_j, j = 1, 2, \dots, K$ dan objek-objek data $x_i, i = 1, 2, \dots, n$
<b>Step 5</b>	: Menghitung matriks keanggotaan $U$ $U_{ij} = \begin{cases} 1; & d_{ij} \leq d_{il}, j \neq l \\ 0; & \text{untuk sebaliknya} \end{cases}$ $i = 1, 2, \dots, n; j = 1, 2, \dots, K$
<b>Step 6</b>	: Menentukan sentroid yang baru $C_j = \frac{\sum_{i=1}^n (U_{ij})x_i}{\sum_{i=1}^n (U_{ij})}$ untuk $j = 1, 2, \dots, K$
<b>Step 7</b>	: Ulangi langkah 4 sampai 6 sampai konvergen
<b>Output</b>	: Partisi sebanyak $K$ kluster dari data yang diinput

E. Perhitungan Presentase

Presentasi suatu data dapat dihitung dengan menggunakan formula sebagai berikut.

$$Presentase = \frac{\text{jumlah data target}}{\text{jumlah keseluruhan data}} \times 100\% \tag{5}$$

F. Root Mean Square Error (RMSE)

*Root Mean Square Error* (RMSE) adalah fungsi eror yang digunakan sebagai metode statistik dasar untuk mengukur performa model dalam meteorologi, dan kualitas udara.

Perhitungan RMSE untuk dataset sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \tag{6}$$

dengan

$n$  = banyaknya sampel

$e_i$  = model eror

(Chai T., & Draxler R.R. 2014)

Jangkauan nilai RMSE berada diantara 0 dan 1, jika nilai RMSE yang dihasilkan semakin mendekati 0 maka model yang digunakan semakin bagus.

## HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian. Hasil penelitian dapat dilengkapi dengan tabel, grafik (gambar), dan/atau bagan. Bagian pembahasan memaparkan hasil pengolahan data, menginterpretasikan penemuan secara logis, dan mengaitkan dengan sumber rujukan yang relevan.

### A. Model Subcontrolled K-means Max-Min (SKM3)

Hasil yang dikeluarkan dari model klasterisasi adalah dataset dibentuk menjadi beberapa kelompok berdasarkan suatu kesamaan yang terdefinisi. Pada algoritma *K-means*, kesamaan yang dimiliki dari satu kelompok adalah jarak data yang terdekat dari suatu sentroid. Algoritma *K-means Max-Min* memilih sentroid sebanyak  $K$  sentroid secara acak, dimana setiap sentroid memiliki beberapa anggota. Algoritma tersebut sangat baik dipilih menjadi dasar model matematika yang diusulkan, karena dalam bidang statistik data yang diambil untuk diteliti adalah data sampel yang dipilih secara acak dengan aturan tertentu yang diharapkan sangat menggambarkan data dari populasi. Sebagai gambaran, pada pemilihan presiden, lembaga survei akan mengambil sampel masyarakat dari berbagai provinsi, berbagai profesi, berbagai status pendidikan, dan aturan-aturan lainnya.

Berdasarkan konsep statistik, algoritma *K-means Max-min* sendiri masih memiliki kekurangan, yaitu, walaupun sudah memilih sentroid secara acak, perlu adanya aturan yang jelas agar sentroid tersebut menggambarkan dataset yang ada. Ide yang digunakan untuk menangani kekurangan tersebut adalah dengan menggunakan algoritma *K-means* untuk mengontrol kualitas dari banyaknya anggota setiap sentroid yang kemudian akan menghasilkan algoritma *K-means Max-Min*, karena algoritma *K-means* dapat menghasilkan sentroid yang lebih konsisten dari pada *K-means Max-Min*.

Perlu menjadi perhatian bahwa algoritma *K-means* dan *K-means Max-Min* memiliki kekurangan yang sama, yaitu pengguna harus menentukan berapa jumlah sentroid yang akan digunakan. Hal ini merupakan kekurangan, karena pengguna belum tahu berapa jumlah sentroid yang baik untuk semua jenis dataset. Jika dataset diubah,

maka jumlah sentroid yang digunakan kemungkinan juga akan berubah. Untuk kekurangan yang satu ini, jumlah sentroid yang dimasukkan berupa deret aritmatika dengan beda satu. Tidak disarankan untuk memasukkan angka satu sebagai jumlah sentroid, karena satu sentroid merupakan populasi dari dataset itu sendiri. Dari sekian banyak sentroid yang dimasukkan akan dipilih satu sentroid yang paling baik. Pemilihan satu sentroid terbaik adalah dengan menghitung selisih banyaknya anggota dari sekian banyak kluster dari algoritma *K-means* dan *K-means Max-Min* yang paling minimal. Ini dilakukan untuk menghindari kompleksitas perhitungan yang semakin tinggi, dan untuk mempersingkat waktu pengerjaan.

Selanjutnya, satu sentroid yang optimal kemudian menghasilkan satu kluster yang merepresentasikan satu aturan. Untuk itu, model matematika yang akan dibuat harus dilakukan beberapa kali untuk mendapatkan beberapa kluster-kluster yang merepresentasikan beberapa aturan-aturan yang mirip dengan proses statistik.

Model *Subcontrolled K-means Max-Min* (SKM3) dapat lebih jelas dilihat pada Tabel 3 berikut.

**Tabel 3. Algoritma Subcontrolled K-Means Max-Min**

<b>Input</b>	Dataset dengan $n$ objek dan $d$ fitur
Normalisasi Min-max	
$M =$ matriks ukuran $(n \times p)$	
Untuk $i$ dari 1 sampai $p$	
Untuk $j$ dari $a$ sampai $b$	Gunakan Algoritma <i>K-means Max-Min</i> (jumlah sentroid = $j$ ) Gunakan Algoritma <i>K-means</i> (jumlah sentroid = $j$ ) $K_{optimal} = \min  k(km3)_j - k(km)_j $
Sudah $j$	
Simpan pada matriks $M$ kolom ke $i$ .	
Sudah $i$	
$V = \sum_{i=1}^p m_{np}$ dimana $\forall m_{np} \in M$	
<b>Output</b>	Kluster optimal $V$
Parameter SKM3	

*Subcontrolled K-means Max-min* memiliki beberapa parameter yang dapat ditentukan oleh pengguna, yaitu, banyaknya percobaan, dilambangkan dengan  $(p)$ , dan banyaknya jumlah sentroid yang merupakan deret aritmatika dari  $a$  sampai  $b$  dimana  $a > 1$  dan  $b \neq a$ .

Untuk penggunaan model matematika yang diusulkan, pada perhitungan elektabilitas pemilihan kepala daerah nilai parameter yang digunakan adalah  $p = 100$ ,  $a = 3$  dan  $b = 5$ . Parameter tersebut sangat baik pada dataset pilkada dibuktikan dengan nilai RMSE (Pers. 6) yang sangat kecil yang akan ditunjukkan pada bagian C.

## B. Perhitungan Elektabilitas

Algoritma SKM3 menghasilkan suatu vektor  $V$  yang merupakan kluster optimal. Vektor ini akan digunakan untuk menghitung elektabilitas. Jika data berupa sebuah vektor kolom yang sudah diatur sedemikian sehingga data yang akan dihitung merupakan perhitungan persentase. Bila urutan dari baris ke  $g$  sampai ke  $h$  dari total  $n$  baris adalah milik dari pasangan calon tertentu, maka elektabilitas dari pasangan calon dapat diformulasikan seperti berikut.

$$\text{Elektabilitas} = \frac{\sum_{w=g}^h v_w}{\sum_1^n v_n} \times 100\% \quad 7)$$

## C. Elektabilitas Pemilihan Kepala Daerah

Untuk mendukung penelitian ini, metode SKM3 digunakan untuk menghitung elektabilitas pasangan calon Gubernur dan Wakil Gubernur tahun 2018 berdasarkan dataset yang berisi bobot pada diskusi publik warganet berupa komentar di media sosial. Hal ini dilakukan karena dari diskusi tersebut dapat diambil informasi tentang pilihan warganet atas pasangan calon tertentu yang berdampak pada pemilihan kepala daerah tahun 2018.

Pada permasalahan ini, metode SKM3 melakukan proses pengambilan sampel dengan mengelompokkan berita-berita menjadi beberapa kluster yang kemudian dilakukan analisa jarak antara berita dengan pusat kluster, dimana pusat kluster dipilih secara acak. Kemudian anggota setiap kluster adalah berita yang memiliki jarak terpendek ke pusat kluster tertentu. Setelah itu, pusat kluster optimum yang dihasilkan oleh SKM3 dijadikan acuan menjadi elektabilitas setiap pasangan calon.

Data yang dihimpun untuk penelitian ini adalah komentar pembaca pada portal media daring <http://pilkada.detik.com/>. Proses pengumpulan data dapat dilihat sebagaiberikut

### 1. Pengumpulan dan pengelompokkan komentar setiap berita

Proses pengumpulan komentar warganet pada setiap pemberitaan terhadap suatu pasangan calon yang kemudian dijadikan sebuah tabel. Kemudian komentar dikelompokkan kedalam 3 kategori, yaitu pro, netral, dan kontra. Pengelompokkan tersebut memerlukan analisa terhadap indikator emosi dalam sebuah komentar. Contoh, jika dalam sebuah komentar terdapat kata “mantap”, “pasti menang”, “setuju”, “sukses selalu”, dan lain-lain, maka komentar tersebut termasuk kedalam kategori pro. Jika dalam sebuah komentar terdapat kata yang merupakan katanya, membahas permasalahan yang tidak berkaitan dengan daerah pemilihan berita yang diambil, maka komentar tersebut termasuk kedalam kategori netral. Jika dalam



sebuah komentar terdapat kata “bahaya”, “hadeh”, “pencitraan doang”, “jangan pilih”, “omdo”, dan lain-lain, maka komentar tersebut termasuk kedalam kategori kontra. Vontoh hasil pengelompokkan komentar dapat dilihat pada Tabel 3.

**Tabel 4. Contoh Pengelompokkan Komentar**

No	Komentar	Pro/Kontra/Netral
1	Hanura yg ini yg punya kekuatan ril, basis massa nya jelas.	Pro
2	perahu kecil apa yg dibanggakan	Kontra
3	kl kursinya kecil dan hanya jadi pengembira saja, sebaiknya tidak gaya-gaya-an ganti dukungan segala, suaranya gak berpengaruh apa2	Kontra
4	rame klw udah kubu2an sih	Netral

2. Pembobotan komentar dan pembuatan dataset

Data berbentuk komentar tidak dapat diproses dalam perhitungan matematika, sehingga perlu adanya proses pembobotan yang dapat merubah komentar tersebut menjadi angka. Pembobotan dilakukan dengan pemberian nilai 1 pada setiap komentar, kemudian nilai-nilai setiap komentar pada setiap berita diakumulasi. Setelah semua komentar diberikan bobot, kemudian dibuat suatu dataset. Pembuatan dataset penting, karena dari data ini yang kemudian dapat dilakukan perhitungan menggunakan program. Dataset dibentuk dari bobot setiap komentar, judul pemberitaan. Nomor urut pasangan calon, dan waktu pemberitaan. Contoh dataset dapat dilihat pada Tabel 5.

**Tabel 5. Contoh Dataset**

No	Waktu berita	Judul berita	Paslon	Pro	Netral	Kontra
1	Kamis 26 April 2018, 19:26 WIB	Tim Demiz-Dedi Laporkan Paguyuban Paranormal ke Bawaslu Jabar	4	0	0	1
2	Kamis 26 April 2018, 12:40 WIB	Ridwan Kamil Bikin Mobil Avatar Untuk Kampane Interaktif	1	0	0	0
3	Rabu 25 April 2018, 13:07 WIB	Pasangan Deddy-Dedi Janji Perluas Akses Air Bersih dan Irigasi	4	1	0	2
4	Kamis 08 Februari 2018, 23:07 WIB	Ridwan Kamil-Uu Larang Pendukungnya Gelar Kampanye Hitam	1	4	1	1

Hasil perhitungan elektabilitas yang dilakukan dengan model SKM3 dapat dilihat pada Tabel 6, Tabel 7, dan Tabel 8. Perhitungan dilakukan dengan menggunakan bantuan aplikasi R.

**Tabel 6. Elektabilitas Pasangan Calon Daerah Jawa Barat**

Nomor Urut Pasangan Calon	Elektabilitas SKM3	Elektabilitas Dataset
1	20,67%	18,48%
2	11,39%	7,88%
3	5,02%	5,98%
4	4,33%	4,08%
Belum Pasti	58,59%	36,42%

**Tabel 7. Elektabilitas Pasangan Calon Daerah Jawa Tengah**

Nomor Urut Pasangan Calon	Elektabilitas SKM3	Elektabilitas Dataset
1	14%	13,98%
2	37,29%	34,69%
Belum Pasti	48,71%	48,67%

**Tabel 8. Elektabilitas Pasangan Calon Daerah Jawa Timur**

Nomor Urut Pasangan Calon	Elektabilitas SKM3	Elektabilitas Dataset
1	12,1%	10,31%
2	56,01%	51,13%
Belum Pasti	38,89%	61,44%

Tabel 6, Tabel 7, dan Tabel 8 memberikan informasi tentang nilai elektabilitas setiap pasangan calon pada daerah masing-masing, yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur. Terdapat 2 jenis nilai elektabilitas, yaitu elektabilitas dataset dan elektabilitas SKM3. Elektabilitas dataset merupakan elektabilitas dari dataset tanpa melalui proses model SKM3, elektabilitas dataset diperoleh dengan memasukkan dataset ke Pers. 7. Sedangkan elektabilitas SKM3 diperoleh dengan melakukan proses pada Tabel 3.

Perhitungan dengan menggunakan SKM3 juga mendapatkan hasil lain selain elektabilitas, yaitu kecenderungan warganet yang belum menentukan pilihan atau belum mendukung salah satu pasangan. Berdasarkan hasil perhitungan, sebanyak 58,59% warganet pada Tabel 6, 48,71% warganet pada Tabel 7, dan 38,89% warganet pada Tabel 8 belum menentukan pilihan ataupun kecenderungan untuk mendukung salah satu pasangan.

Hasil perhitungan elektabilitas dengan model SKM3 harus dievaluasi, agar diperoleh hasil yang meyakinkan. Evaluasi model SKM3 dilakukan dengan menghitung nilai erornya. Nilai eror dihitung dengan membandingkan hasil perhitungan SKM3 dengan hasil perhitungan dataset. Perhitungan nilai eror penelitian ini menggunakan fungsi eror RMSE. Dengan menggunakan Pers. 6, diperoleh nilai eror pada Tabel 6 sebesar 0,0452, Tabel 7 sebesar 0,0343, dan Tabel

8 sebesar 0,2382. Nilai eror yang diperoleh pada Tabel 6, Tabel 7, dan Tabel 8 kurang dari 0,5, bahkan cenderung mendekati 0. Semakin kecil nilai eror suatu model, maka model tersebut dikatakan semakin bagus. Sehingga, model SKM3 sangat baik jika digunakan untuk menghitung elektabilitas menggunakan dataset berupa komentar warganet di media sosial.

## **SIMPULAN DAN SARAN**

Model SKM3 yang diusulkan pada penelitian ini dapat digunakan untuk mencari elektabilitas suatu pasangan calon berdasarkan data komentar warganet di media sosial karena model SKM3 memiliki nilai eror yang kecil. Dibuktikan dengan penghitungan nilai eror dengan fungsi eror RMSE yaitu kurang dari 0,5.

Penelitian ini masih memiliki beberapa kekurangan, berikut saran yang dapat digunakan untuk melakukan penelitian lebih lanjut.

1. Jenis media sosial yang menjadi penelitian ditambah agar data lebih beragam dan hasil lebih akurat dengan keadaan di dunia nyata.
2. Proses pemberian bobot komentar warganet akan lebih baik jika dilakukan oleh ahli bahasa atau menggunakan kecerdasan buatan (AI).

## **DAFTAR PUSTAKA**

- Ankerst, M., Breunig, M.M., Kriegel, H., & Sander, J. (1999). OPTICS: Ordering Points to Identify the Clustering Structure. *In ACM SIGMOD International Conference on Management of Data*, 28 (2), 49-60.
- Chai, T. & Draxler, R.R. (2014). Root Mean Square Error (RMSE) or mean absolute error (MAE)?-Arguments Against Avoiding RMSE in The Literature. *Geosci Model Development*, Vol 7, p. 1247 – 1250.
- Forgey, E.W. (1965). Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classifications. *Biometrics*, 21, 768–769.
- Friedman, H.P. & Rubin, J. (1967). On Some Invariant Criteria for Grouping Data. *Journal of the American Statistical Association*, 62, 1159-1178.
- Han, J., Kamber, M., & Peo, J. (2012). *Data Mining: Concepts and Techniques Third Edition*. Morgan Kaufmann, USA:Elsevier.
- Hinneburg, A. & Keim, D.A. (1998). An Efficient Approach to Clustering in Large Multimedia Databases with Noise. *American Association for Artificial Intelligence*, 58-65.
- Jihyang, C., Jae K.L., & Emily, T. M. (2017). Investigating Effects of Social Media News Sharing On the Relationship between Network Heterogeneity and Political Participation. *Computers in Human Behavior*, 75, 25-31.

- Kim, M.(2018). How Does Facebook News Use Lead to Actions in South Korea? The Role of Facebook Discussion Network Heterogeneity, Political Interest, and Conflict Avoidance in Predicting Political Participation. *Telematics and Informatics*, 35, 1373-1381.
- Lloyd, S.P. (1982). Least Squares Quantization in PCM. *IEEE Transactions On Information Theory*, 28(2), 129-137.
- MacQueen, J. (1967).Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281-297.
- Panagiotopoulos, P., Bowen, F., &Brooker, P. (2017). The Value of Social Media Data: Integrating Crowd Capabilities in Evidencebased Policy. *Government Information Quarterly*, 34, 601-612.
- Rui Xu. & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transaction on Neural Networks*, 16(3), 645-678.
- Sari, R.Y. (2017). Strategi Rakan Mualem dalam Meningkatkan Elektabilitas Muzakir Manaf sebagai Calon Gubernur Aceh pada Pilkada Tahun 2017. *Jurnal Ilmiah Mahasiswa FISIP UNSYIAH*, 2(3), 1-12.
- Schroeder, E. & Stone, D.F. (2015). Fox News and Political Knowledge. *Journal of Public Economics*, 126, 52-63.
- Shalabi. L.A., Shaaban, Z.,& Kassabeh, B. (2006). Data Mining A Preprocessing Engine.*Journal of Computer Science*, 2(9), 735-739.
- Shim, H., Kyung, H. Y., Jeong, K. L.,& Eun, G. (2015). Why Do People Access News with Mobile Devices? Exploring The Role of Suitability Perception and Motives On Mobile News Use. *Telematics and Informatics*, 32, 108-117.
- Soman,K.P., Diwakar, S.,& Ajay, V. (2006).*Insight into Data Mining Theory and Practice*. New Delhi, NJ: Prentice Hall of India Private.
- Timothy J.R. (2010).*Fuzzy Logic with Engineering Applications*. Newyork, NJ: Wiley.
- Visalakshi, N.K. & Suguna, J. (2009). *Conference: Fuzzy Information Processing Society*. Cincinnati: IEEE.
- Yaqub, U., Chun, S.A., Atluri, V., &Vaidya, J. (2017). Analysis of Political Discourse On Twitter in The Context of The 2016 US Presidential Elections. *Government Information Quarterly*, 34, 613-626.
- Zaki, M.J., & Meira, W. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. USA, NJ : Cambridge University Press.