

## PENERAPAN *TEXT MINING* DAN *FUZZY C-MEANS CLUSTERING* UNTUK IDENTIFIKASI KELUHAN UTAMA PELANGGAN PDAM TIRTA MOEDAL KOTA SEMARANG

Genisia Pramestiloka Aulia<sup>1\*</sup>, Tatik Widiharih<sup>2</sup>, Iut Tri Utami<sup>3</sup>

<sup>1,2,3</sup>Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

\*e-mail : [genispramesti6990@gmail.com](mailto:genispramesti6990@gmail.com)

DOI: 10.14710/j.gauss.12.1.126-135

### Article Info:

Received: 2022-10-24

Accepted: 2022-12-30

Available Online: 2023-05-04

### Keywords:

Text Mining; Fuzzy C-Means Clustering; Xie-Beni Index; Word cloud.

**Abstract:** Customer complaints can be handled effectively by identifying the main complaints that cause customers to be dissatisfied. Many customer complaints cause difficulty for PDAM Tirta Moedal Semarang to identify problems, which are frequently the primary complaints of customers. Grouping complaints that have similarities using Fuzzy C-Means Clustering will make the identification of the main customer complaints easier. Fuzzy C-Means uses fuzzy models, allows data to be a member of all formed clusters with membership level between 0-1. Fuzzy C-Means Clustering can also introduce more flexible patterns and show results in more accurate cluster placement. Text mining is used to convert textual data into numerical data. Customer complaints received through all contacts in PDAM Tirta Moedal Semarang from October–December 2021 were used as data. The clustering process forms 6 clusters, with the number of clusters tried being 3, 4, 5, and 6, which are seen by the smallest Xie-Beni Index. The main complaints from PDAM Tirta Moedal Semarang customer that seen through Word cloud in each cluster are that the water stops running in clusters 1 and 6 and the pipes leak in clusters 4 and 5. Complaints in clusters 2 and 3 are complaints related to water meters and water flow.

## 1. PENDAHULUAN

Kota Semarang memiliki Badan Usaha Milik Daerah (BUMD) yang menyediakan layanan air bersih untuk masyarakat Kota Semarang yaitu PDAM Tirta Moedal Kota Semarang (Perumda Air Minum Tirta Moedal Kota Semarang, 2019). PDAM Tirta Moedal Kota Semarang mencatat bahwa keluhan pelanggan semakin meningkat setiap tahunnya dan akan terus melakukan kegiatan evaluasi untuk mengetahui masalah-masalah yang sering dikeluhkan pelanggan (Prayoga, 2020). Menurut Ghoniyah (2012), proses penanganan keluhan yang baik dimulai dengan mengidentifikasi sumber ketidakpuasan pelanggan. Banyaknya keluhan yang diperoleh membuat PDAM Tirta Moedal Kota Semarang merasa kesulitan untuk mengidentifikasi masalah yang sering menjadi keluhan utama pelanggan.

Pengelompokan pada keluhan yang memiliki kemiripan dibutuhkan untuk memudahkan identifikasi keluhan utama pelanggan. Penelitian ini menerapkan *Text Mining* untuk mempersiapkan keluhan sebelum dikelompokkan dan *Fuzzy C-Means Clustering* untuk mengelompokkan keluhan pelanggan yang memiliki kemiripan. *Fuzzy C-Means Clustering* menggunakan model pengelompokan *fuzzy* yaitu pengelompokan yang menggunakan logika *fuzzy*.

Logika *fuzzy* merupakan logika bernilai banyak (*multivalued logic*) yang dapat mendefinisikan nilai diantara logika konvensional, seperti ya atau tidak dan benar atau salah yang dinyatakan dalam derajat keanggotaan dengan rentang nilai 0-1. Model pengelompokan *fuzzy* menyebabkan data dapat menjadi anggota dari semua *cluster* yang

terbentuk. *Fuzzy C-Means Clustering* juga dapat mengenalkan pola yang lebih fleksibel dan memudahkan pemecahan perhitungan.

Hasil kelompok keluhan yang telah terbentuk tersebut digunakan untuk mengidentifikasi keluhan utama. Tujuan dari penelitian ini adalah memperoleh *cluster* dari keluhan pelanggan PDAM Tirta Moedal Kota Semarang yang telah dilakukan proses *Text Mining* dengan menggunakan *Fuzzy C-Means Clustering* dan mengidentifikasi keluhan utama pelanggan PDAM Tirta Moedal Kota Semarang menggunakan *Word cloud* berdasarkan *cluster* yang terbentuk.

## 2. TINJAUAN PUSTAKA

Pelanggan PDAM Tirta Moedal Kota Semarang dapat melakukan pengaduan seperti melaporkan keluhan, memberikan saran atau memperoleh informasi dengan menggunakan beberapa pilihan kanal pengaduan, yaitu: *call center*, sosial media PDAM Tirta Moedal Kota Semarang (*Whatsapp, Instagram, Facebook, dan Twitter*), serta aplikasi SI AGAN (Perumda Air Minum Tirta Moedal Kota Semarang, 2021). Keluhan pelanggan PDAM Tirta Moedal Kota Semarang masih berbentuk data teks, sehingga proses *text mining* dibutuhkan sebelum dilakukan proses *clustering*.

*Text mining* dapat didefinisikan sebagai proses menemukan informasi berkualitas dengan mengidentifikasi dan mengeksplorasi pola data dari dokumen teks menggunakan berbagai alat analisis (Vijayarani dan Nithya, 2015). *Text mining* memberikan penyelesaian untuk masalah yang sering dihadapi dalam pemrosesan, pengorganisasian, dan analisis teks tidak terstruktur dalam jumlah besar. Sumber data yang diolah pada *text mining* berasal dari teks yang relative tidak terstruktur dan berjumlah besar, sehingga dibutuhkan *text preprocessing*.

*Text Preprocessing* merupakan proses pembersihan, penghilangan, perubahan sumber data yang dianggap tidak perlu untuk digunakan pada proses analisis selanjutnya, baik berupa karakter non alfabet maupun kata-kata. Berikut merupakan tahapan *Text Preprocessing* yang akan digunakan dalam penelitian ini:

1. *Case folding*, digunakan untuk mengubah seluruh kata menjadi huruf kecil pada dokumen teks.
2. *Remove number*, digunakan untuk menghapus angka pada dokumen teks.
3. *Remove punctuation*, digunakan untuk menghilangkan semua karakter selain huruf, seperti tanda baca atau symbol.
4. *Strip whitespace*, yaitu proses penghapusan spasi berlebih pada kalimat.
5. Pembakuan kata, yaitu proses merubah kata tidak baku ke kata baku.
6. *Stopword removal*, yaitu proses penghilangan kata umum yang sering muncul tetapi dianggap tidak berarti. Contoh: dengan, pada, yang, di.
7. *Stemming*, yaitu proses merubah kata berimbuhan ke kata dasarnya (Putri dan Setiadi, 2014)
8. *Tokenizing*, yaitu proses pemotongan rangkaian kata dalam kalimat menjadi bagian-bagian kata tunggal.

Tingkat kepentingan yang dimiliki setiap kata dalam sebuah dokumen teks berbeda satu sama lain, sehingga setelah dilakukan *text preprocessing* perlu dilakukan pembobotan kata. Pembobotan kata yang digunakan adalah pembobotan TF-IDF, pembobotan TF-IDF menggunakan penggabungan terhadap dua konsep yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF).

*Term Frequency* merupakan frekuensi sebuah kata atau *term* muncul dalam dokumen dan *Inverse Document Frequency* merupakan pembobotan kata yang didasarkan pada

jumlah dokumen yang memuat kata-kata tertentu (Prasetyo, 2012). Menurut Upendra dan Babu (2016), perhitungan TF-IDF dapat dihitung menggunakan persamaan (1):

$$W_{i,j} = \frac{n_{j,i}}{\sum_k n_{k,i}} \cdot \log_2 \frac{D}{D_j} \quad (1)$$

dengan  $i$  dokumen,  $j$  term,  $W_{i,j}$  pembobotan TF-IDF untuk term ke- $j$  di dokumen ke- $i$ ,  $n_{j,i}$  jumlah kemunculan term ke- $j$  di dokumen ke- $i$ ,  $\sum_k n_{k,i}$  banyaknya kemunculan semua term di dokumen ke- $i$ ,  $D$  banyaknya semua dokumen yang digunakan dan  $D_j$  banyaknya dokumen yang mengandung term ke- $j$ .

*Word cloud* adalah teknik visualisasi yang sering digunakan pada *text mining* untuk memberikan gambaran umum sebuah dokumen teks dengan menampilkan kata-kata yang sering muncul secara visual. Kata-kata tersebut divisualisasikan dalam bentuk yang menarik namun tetap informatif, sehingga lebih mudah untuk dipahami.

Analisis *cluster* merupakan proses dimana objek data dibagi ke dalam beberapa kelompok yang disebut *cluster*. Pendekatan yang biasa digunakan untuk mengukur kemiripan antar objek-objek dalam analisis *cluster* adalah ukuran jarak (Mulekar dan Brown, 2017). Salah satu pengukuran jarak kedekatan adalah jarak *Euclidean*. Menurut Vijaymeena dan Kavitha (2016), fungsi pengukuran jarak antara dua buah objek  $x = (x_1, x_2, \dots, x_p)$  dan  $y = (y_1, y_2, \dots, y_p)$  menggunakan *Euclidean* didefinisikan sebagai berikut:

$$d(x, y) = \sqrt{\sum_{j=1}^p (x_j - y_j)^2} \quad (2)$$

dengan  $d(x, y)$  merupakan jarak *Euclidean* objek  $x$  dan  $y$ ,  $p$  jumlah atribut yang diamati,  $x_j$  nilai pengamatan objek  $x$  atribut ke- $j$ , dan  $y_j$  nilai pengamatan objek  $y$  atribut ke- $j$ .

Salah satu metode dari analisis *cluster* adalah *Fuzzy C-Means Clustering*. *Fuzzy C-Means Clustering* menggunakan model pengelompokan *fuzzy*, menyebabkan data mampu menjadi anggota dari seluruh *cluster* yang telah terbentuk. Keberadaan titik data pada suatu *cluster* dinyatakan menggunakan derajat keanggotaan antara 0-1 (Rahakbauw, 2017). Menurut Tanatavikorn dan Yamashita (2015), konsep dasar dari *Fuzzy C-Means Clustering* yaitu menentukan pusat *cluster*, dimana pusat *cluster* tersebut belum tepat sehingga perbaikan pusat *cluster* dan derajat keanggotaan dilakukan secara berulang pada setiap data. Proses perulangan tersebut menyebabkan pusat *cluster* berpindah mengarah ke titik yang tepat, dimana proses ini didasarkan pada minimasi fungsi objektif.

Menurut Nayak dkk., (2015) tahapan-tahapan *Fuzzy C-Means Clustering* adalah sebagai berikut:

1. Memasukkan data yang akan di *cluster* ke dalam sebuah matriks  $X$  berukuran  $n \times m$ , dengan  $n$  adalah jumlah data dan  $m$  adalah atribut pada setiap data.  $X_{ij}$  = data ke- $i$  ( $i = 1, 2, \dots, n$ ), atribut ke- $j$  ( $j = 1, 2, \dots, m$ )

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

2. Menentukan beberapa parameter awal, berikut merupakan parameter awal yang ditentukan:
  - a. Banyaknya *cluster* yang akan dibentuk ( $c$ ) =  $2 \leq c < (n)$ ; dengan  $n$  adalah banyaknya data
  - b. Jumlah nilai pembobot ( $w$ ) =  $w \geq 1$ ; menurut Klawonn dan Höppner (2003)  $w=2$  sering digunakan untuk pembobot karena dianggap paling halus.

- c. Jumlah maksimum pengulangan iterasi = MaksIter;
- d. Jumlah error terkecil yang diharapkan ( $\xi$ ) =  $\xi > 0$ ;
- e. Fungsi objektif awal =  $P_0 = 0$ ;
- f. Iterasi awal =  $t = 1$
3. Membangkitkan bilangan acak  $\mu_{ik}$  dengan  $i$  ( $i = 1, 2, \dots, n$ ) adalah jumlah data dan  $k$  ( $k = 1, 2, \dots, c$ ) adalah jumlah *cluster* yang merupakan elemen dari matriks partisi awal **U**.

$$U = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1c} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2c} \\ \vdots & \vdots & \dots & \vdots \\ \mu_{n1} & \mu_{n2} & \dots & \mu_{nc} \end{bmatrix}$$

dengan syarat:

$$\mu_{ik} \in [0, 1]; (1 \leq i \leq n; 1 \leq k \leq c)$$

dan

$$\sum_{k=1}^c \mu_{ik} = 1; 1 \leq i \leq n$$

4. Menghitung pusat *cluster* ke- $k$ :  $V_k$ .  $V_k = (V_{k1}, V_{k2}, \dots, V_{kj})$  dengan  $k = 1, 2, \dots, c$  dan  $j = 1, 2, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w X_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

5. Menghitung fungsi objektif pada iterasi ke- $t$ ,  $P_t$ :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \mu_{ik}^w \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] \right) \quad (4)$$

6. Memperbaiki matriks partisi **U**

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-1}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-1}} \quad (5)$$

dengan  $i = 1, 2, \dots, n$  dan  $k = 1, 2, \dots, c$

7. Mengecek kondisi berhenti

- Jika: ( $|P_t - P_{t-1}| < \xi$ ) atau ( $t > \text{MaksIter}$ ) maka proses berakhir;
- Jika: ( $|P_t - P_{t-1}| > \xi$ ) atau ( $t < \text{MaksIter}$ ) maka  $t=t+1$  dan melanjutkan kembali langkah ke-4.

Ukuran yang digunakan untuk menguji validitas hasil *cluster* adalah Indeks *Xie-Beni*. Indeks *Xie-Beni* didasarkan pada jarak *intracluster* yaitu kedekatan antar anggota di setiap *cluster* dan *intercluster* yaitu perbedaan antara suatu *cluster* dengan *cluster* lain (Muranishi dkk., 2014). Semakin kecil nilai *Xie-Beni*, semakin baik hasil *cluster*. Indeks *Xie-Beni* dapat dinyatakan dalam persamaan (6):

$$XB = \frac{\sum_{i=1}^n \sum_{k=1}^c (\mu_{ik}^w [\sum_{j=1}^m (X_{ij} - V_{kj})^2])}{n \min_{i \neq k} \sum_{j=1}^m (V_{ij} - V_{kj})^2} \quad (6)$$

dengan  $X_{ij}$  objek data ke- $i$  pada atribut ke- $j$ ,  $V_{kj}$  nilai pusat *cluster* ke- $k$  pada atribut ke- $j$ ,  $V_{ij}$  nilai pusat *cluster* ke- $i$  pada atribut ke- $j$ ,  $\mu_{ik}$  nilai keanggotaan objek data ke- $i$  dengan pusat *cluster* ke- $k$ ,  $n$  banyaknya objek data,  $c$  jumlah *cluster*,  $w$  jumlah nilai pembobot.

### 3. METODE PENELITIAN

Data keluhan pelanggan yang digunakan adalah data dengan jenis kualitatif, yaitu berupa data teks keluhan pelanggan PDAM Tirta Moedal Kota Semarang yang masuk pada semua kontak pengaduan, yaitu: *call center, Whatsapp, Instagram, Facebook, Twitter* dan SI AGAN yang termasuk dalam data sekunder dan didapatkan di PDAM Tirta Moedal Kota Semarang selama periode Oktober 2021 sampai Desember 2021 sebanyak 3.577 data. *Software RStudio* akan digunakan untuk membantu mengolah data penelitian.

Langkah-langkah analisis data:

1. Memasukkan data teks keluhan PDAM Tirta Moedal Kota Semarang.
2. Melakukan *Text Preprocessing* data.
3. Melakukan pembobotan kata menggunakan pembobotan TF-IDF.
4. Penerapan *Fuzzy C-Means Clustering*:
  - a. Menentukan banyaknya *cluster* ( $c$ ), nilai bobot ( $w$ ), maksimal iterasi (MaksIter), error terkecil ( $\xi$ ), fungsi objektif awal ( $P_0=0$ ), dan iterasi awal ( $t=1$ ).
  - b. Membangkitkan bilangan acak  $\mu$  sebagai elemen-elemen matriks partisi awal  $U$
  - c. Menghitung pusat *cluster* ( $V_k$ )
  - d. Menghitung fungsi objektif ( $P_t$ )
  - e. Memperbaiki matriks partisi  $U$
  - f. Jika diperoleh ( $|P_t - P_{t-1}| < \xi$ ) atau ( $t > \text{MaksIter}$ ) maka proses berakhir, jika:  $|P_t - P_{t-1}| > \xi$  dan ( $t < \text{MaksIter}$ ) maka  $t=t+1$  kemudian melakukan kembali langkah (c).
5. Menghitung nilai Indeks Xie-Beni pada *cluster* yang terbentuk.
6. Menentukan *cluster* optimum dari *cluster* yang terbentuk berdasarkan nilai Indeks Xie-Beni.
7. *Profiling* pada setiap *cluster* menggunakan bantuan *Word cloud*.
8. Analisis hasil *cluster* untuk mengidentifikasi keluhan utama pelanggan.

#### 4. HASIL DAN PEMBAHASAN

*Text Preprocessing* mengubah data keluhan yang berbentuk tidak teratur ke bentuk data keluhan yang lebih teratur. Tahapan *text preprocessing* yang diperlukan yakni *case folding, removing number, removing punctuation, pembakuan kata, stopwords removal, stemming* dan *tokenizing*. Contoh hasil *Text Preprocessing* menggunakan data keluhan nomor 16, 549, 683 dan 802 disajikan pada Tabel 1.

Tabel 1. Contoh Hasil *Text Preprocessing*

No	Sebelum <i>Text Preprocessing</i>	Setelah <i>Text Preprocessing</i>
16	informasi dari pelanggan pipa bocor di depan rumah nomor pelanggan : 08150277	pipa bocor
549	AIR MATI KURANG LEBIH 10 HARI KKTH	air mati kkth
683	meteran bureng mohon bisa ganti meteran	meter buram ganti meter
802	Tagihan 2 bulan ini melonjak	tagih lonjak

Pembobotan pada data setelah melalui tahap *text preprocessing* akan dilakukan dengan menggunakan pembobotan TF-IDF. Hasil yang didapatkan dari pembobotan ini diperlukan untuk proses *Fuzzy C-Means Clustering*. Hasil pembobotan TF-IDF pada keluhan pelanggan PDAM Tirta Moedal Kota Semarang dengan contoh keluhan nomor 6, 13, 292, 494 dan 3047 disajikan pada Tabel 2.

Tabel 2. Contoh Hasil Pembobotan TF-IDF

No		air	alir	bocor	...	rusak	tutup
6	curi air	0,633	0	0	...	0	0
13	air alir kecil kko	0,316	0,920	0	...	0	0
292	ganti meter masalah	0	0	0	...	0	0
494	air nyala kecil bocor	0,316	0	0,232	...	0	0
3047	pipa bocor rusak paving bocor tutup gentong	0	0	0,265	...	1,16E+14	1,29+E14

Banyaknya *cluster* yang akan di coba pada proses *Fuzzy C-Means Clustering* adalah 3,4,5 dan 6. Jumlah pembobot yang digunakan  $w = 2$ , jumlah maksimum pengulangan iterasi sebanyak 1000, dan jumlah error terkecil yang diharapkan  $(\xi) = 10^{-9}$  sesuai dengan bawaan *Software R-Studio*.

Indeks *Xie-Beni* akan digunakan untuk menentukan jumlah *cluster* yang optimum dari hasil proses *clustering*. Nilai Indeks *Xie-Beni* pada jumlah cluster  $k = 3,4,5$  dan 6 disajikan pada Tabel 3.

Tabel 3. Hasil Nilai Indeks *Xie-Beni*

k	Indeks <i>Xie-Beni</i>
3	0,05802351
4	0,05701159
5	0,05649308
6	0,05617681

Berdasarkan Tabel 3, terlihat bahwa jumlah *cluster*  $c = 6$  memiliki nilai Indeks *Xie-Beni* terkecil dibanding jumlah *cluster* lainnya, yaitu sebesar 0,05617681. Jumlah *cluster* digunakan untuk penelitian ini yaitu  $c=6$ . Hasil *clustering* dengan jumlah *cluster* sebanyak 6 secara ringkas disajikan pada Tabel 4.

Tabel 4. Hasil *Clustering* dengan *Fuzzy C-Means*

Cluster ke-	Nomor Keluhan	Jumlah Anggota
1	1789, 1790, 1791, 1792, ..., 2400,2401,2402,2403	615
2	2404,2405,2406,2407,2408, ..., 3007,3008,3009,3010	607
3	1175,1176,1177,1178,1179, ..., 1785,1786,1787,1788	614
4	3011,3012,3013,3014,3015, ..., 3574,3575,3576,3577	567
5	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15, ..., 563,564,565,567	567
6	568,569,570,571,572,573, ..., 1171,1172,1173,1174	607

Keluhan utama pelanggan pada masing-masing *cluster* yang terbentuk akan diidentifikasi menggunakan *Word cloud* untuk melihat kata yang sering muncul di setiap *cluster*. *Word cloud* yang dihasilkan pada setiap *cluster* memiliki kemiripan pada kata yang ada pada masing-masing *cluster*, dengan kata “bocor”, “pipa”, “air” dan “mati” yang memiliki ukuran terbesar.

Berdasarkan hal tersebut, akan digunakan 10 kata teratas yang sering muncul di setiap *cluster* untuk melihat kata yang memiliki frekuensi kemunculan tertinggi atau paling banyak dibandingkan dengan *cluster* lainnya yang ditampilkan pada Tabel 5.

Tabel 5. Kata dengan Frekuensi Kemunculan Terbanyak pada Setiap *Cluster* Dibandingkan dengan *Cluster* Lainnya

<i>Cluster</i>	Kata	Frekuensi Kemunculan
1	tangki	90
	bantu	88
2	meter	37
3	alir	63
4	sm (sekitar meteran)	34
5	kko (kanan kiri oke)	50
	sbm (sebelum meteran)	39
6	kks (kanan kiri sama)	83
	stopkran	75
	on	72
	kecil	40

Hal ini membantu untuk menganalisis lebih lanjut terkait keluhan yang ada di setiap *cluster*. Berikut merupakan hasil analisis pada setiap *cluster* dengan menggunakan *Word cloud*:

a. *Cluster* 1

Gambar 1 memperlihatkan hasil *Word cloud* untuk *cluster* 1. Terlihat bahwa kata “bocor”, “air”, “pipa” dan “mati” memiliki ukuran yang terbesar.



Gambar 1. *Word Cloud Cluster* 1

Berdasarkan *Word cloud* pada Gambar 1, terdapat pula kata “bantu” dan “tangki” yang memiliki ukuran cukup besar, kata tersebut juga memiliki frekuensi kemunculan terbanyak dibandingkan dengan *cluster* lainnya. Kata ini berhubungan dengan keluhan air mati dimana keluhan air mati sudah terjadi terlalu lama dan pelanggan meminta bantaun air tangki karena sudah kehabisan air.

b. *Cluster* 2

Hasil *Word cloud* untuk *cluster* 2 ditunjukkan pada Gambar 2. Terlihat bahwa kata “bocor”, “pipa” memiliki ukuran terbesar, selain itu kata “air” dan “mati” juga memiliki ukuran cukup besar.



Gambar 2. *Word Cloud Cluster* 2

Berdasarkan Tabel 5, kata “meter” di *cluster* 2 merupakan kata dengan frekuensi kemunculan terbanyak dibandingkan *cluster* lainnya. Kata “meter” menunjukkan keluhan terkait kondisi meteran seperti meteran mati atau rusak, meteran burem, permohonan normalisasi meteran dan penggantian meteran.

c. *Cluster* 3

Gambar 3 memperlihatkan hasil *Word cloud* untuk *cluster* 3. Terlihat bahwa kata “air” memiliki ukuran terbesar, kata “bocor”, “mati” dan “pipa” juga memiliki ukuran yang cukup besar.



Gambar 3. *Word Cloud Cluster* 3

Tabel 5 memperlihatkan bahwa kata “alir” pada *cluster* 3 merupakan kata dengan frekuensi kemunculannya terbanyak dibandingkan *cluster* lainnya. Kata “alir” menunjukkan keluhan terkait kondisi aliran air, seperti air mengalir kecil dan terkadang mati, serta air yang mengalir tidak sesuai jadwal aliran.

d. *Cluster* 4

Hasil *Word cloud* untuk *cluster* 4 ditunjukkan pada Gambar 4. Terlihat bahwa kata “bocor” dan “pipa” memiliki ukuran terbesar, selain itu juga terlihat kata “air” dan “mati” yang memiliki ukuran cukup besar.



Gambar 4. *Word Cloud Cluster* 4

Berdasarkan Tabel 5, kata “sm” atau “sekitar meteran” pada *cluster* 4 memiliki frekuensi kemunculan yang paling banyak dibandingkan *cluster* lainnya. Kata “sm” atau “setelah meteran” berkaitan dengan keluhan pipa bocor, dimana keluhan pipa bocor pada *cluster* 4 sering terjadi pada daerah sekitar meteran.

e. *Cluster* 5

Gambar 5 memperlihatkan hasil *Word cloud* untuk *cluster* 5. Terlihat bahwa kata “bocor” dan “pipa” memiliki ukuran terbesar, selain itu kata “air” dan “mati” juga memiliki ukuran yang cukup besar.



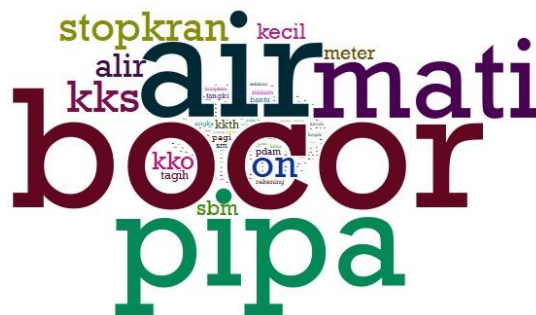


Gambar 5. Word Cloud Cluster 5

Berdasarkan Tabel 5, kata “sbm” atau “sebelum meteran” dan kata “kko” atau “kanan kiri oke” pada *cluster* 5 merupakan kata dengan frekuensi kemunculan terbanyak dibandingkan *cluster* lainnya. Kata “sbm” atau “sebelum meteran” berhubungan pada keluhan pipa bocor, dimana keluhan pipa bocor pada *cluster* 5 sering terjadi sebelum instalasi meteran. Kata “kko” atau “kanan kiri oke” berhubungan pada keluhan air mati dimana banyak pelanggan yang mengeluhkan air mati tetapi tetangga tidak mengalami hal yang sama atau air tidak mati.

f. *Cluster* 6

Hasil *Word cloud* untuk *cluster* 6 ditunjukkan pada Gambar 6. Terlihat bahwa kata “bocor”, “air”, “mati” dan “pipa” memiliki ukuran yang terbesar.



Gambar 6. Word Cloud Cluster 6

Berdasarkan Tabel 5, kata “kks” atau “kanan kiri sama”, “stopkran”, “on” dan “kecil” merupakan kata dengan frekuensi kemunculan terbanyak dibandingkan *cluster* lainnya. Kata ini merupakan kondisi dari keluhan air mati, dimana pada *cluster* 6 keluhan air mati sering terjadi dengan kondisi seperti: tetangga mengalami keluhan yang sama, stopkran sudah dalam posisi on, dan air mengalir kecil.

## 5. KESIMPULAN

Proses *clustering* pada 3.577 keluhan pelanggan PDAM Tirta Moedal Kota Semarang yang telah dilakukan proses *Text mining* dengan menggunakan *Fuzzy C-Means Clustering* terbentuk 6 *cluster*. Jumlah *cluster* dilihat berdasarkan nilai Indeks *Xie-Beni* yang terkecil yaitu sebesar 0,05617681. Berdasarkan *Word cloud* yang terbentuk pada 6 *cluster* yang telah dihasilkan, keluhan utama pelanggan PDAM Tirta Moedal Kota Semarang adalah keluhan air mati dan pipa bocor.

Keluhan air mati yang terdapat pada *cluster* 1 terjadi dengan durasi yang lama, dimana mengakibatkan banyak pelanggan meminta bantuan air tangki. Keluhan yang terjadi pada *cluster* 2 yaitu keluhan terkait kondisi meteran, seperti meteran rusak, pergantian meteran dan normalisasi meteran. *Cluster* 3 memiliki keluhan yang berhubungan dengan kondisi

aliran air, dimana air mengalir kecil dan tidak sesuai jadwal aliran. Keluhan yang sering terjadi pada *cluster* 4 dan *cluster* 5 yaitu pada *cluster* 4 keluhan pipa bocor sering terjadi didaerah sekitar meteran, pada *cluster* 5 keluhan pipa bocor sering terjadi sebelum instalasi meteran dan terdapat keluhan air mati dengan kondisi tetangga tidak mengalami keluhan air mati. *Cluster* 6 memiliki keluhan air mati dengan kondisi seperti tetangga mengalami keluhan yang sama, stopkran sudah dalam posisi on, dan air mengalir kecil.

## DAFTAR PUSTAKA

- Ghonyah, N. 2012. Perilaku Komplain dan Pengaruhnya Terhadap Kepuasan dan Loyalitas Pelanggan Jasa. *Dharma Ekonomi*, Vol. 19, No.35, Hal: 1-12.
- Klawonn, F. dan Höppner, F. 2003. What is Fuzzy about Fuzzy Clustering? Understanding and Improving the Concept of the Fuzzier. *Internatioal Symposium on Intelligent Data Analysis*, Hal: 254-264
- Mulekar, M.S. dan Brown, C.S. 2017. *Distace and Similarity Measures in Encyclopedai of Social Network Analysis and Mining*. New York: Springer.
- Muranishi, M., Honda, K., dan Notsu, A. 2014. Xie-Beni Type Fuzzy Cluster Validation in Fuzzy Co-clustering of Documents and Keywords. *Soft Computing in Artificial Intelligence*, Vol. 270, Hal: 29-38.
- Nayak, J., Naik, B., dan Behera, H.S. 2015. Fuzzy C-Means (FCM) Clustering Algorithm: A Decade Review from 2000 to 2014. *Computational Intelligence in Data Mining*, Vol. 2, Hal: 133-149.
- Perumda Air Minum Tirta Moedal Kota Semarang. 2019. *Sejarah PDAM*. <https://www.pdamkotasmg.co.id/page/sejarah>. Diakses pada tanggal 18 April 2022.
- Perumda Air Minum Tirta Moedal Kota Semarang. 2021. *Beranda*. <https://www.pdamkotasmg.co.id/>. Diakses pada tanggal 20 April 2022.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta. Penerbit Andi
- Prayoga, M. A. 2020. Jumlah Aduan Pelanggan PDAM Meningkat 150 Orang/hari. *Suara Merdeka*. <https://www.suaramerdeka.com/semarang-raya/pr-04122840/jumlah-aduan-pelanggan-pdam-meningkat-150-oranghari?page=1>. Diakses pada tanggal 18 April 2022.
- Putri, E. K. dan Setiadi, T. 2014. Penerapan Text Mining pada Sistem Klasifikasi Email Spam Menggunakan Naïve Bayes. *Jurnal Sarjana Teknik Informatika*, Vol. 2, No. 3, Hal: 73-83.
- Rahakbauw, D.I., Ilwaru, V. Y., dan Hahury, M.H. 2017. Implementasi Fuzzy C-Means Clustering dalam Penentuan Beasiswa. *Jurnal Ilmu Matematika dan Terapan*. Vol. 11, No. 1, Hal: 1-11
- Tanatavikorn, H. dan Yamashita, Y. 2015. Improving Data Reliability for Process Monitoring with Fuzzy Outlier Detection. *Computer Aided Chemical Engineering*, Vol. 37, Hal: 1595-1600.
- Upendra, B., dan Babu, A.S. 2016. KNN TFIDF Based Named Entity Recognition. *International Journal of Scientific Development and Research (USDR)*, Vol. 1, No.12, Hal: 35-39.
- Vijaymeena, M.K. dan Kavitha, K. 2016. A Survey on Similiarity Measures in Text Mining. *Machine Learning and Applications*, Vol. 3, No.1, Hal: 19-28.
- Vijayarani, S. dan Nithya, N. 2015. Efficient Machine Learning Classifiers for Automatic Information Classification. *International Journal of Modern Trends in Engineering and Research*, Vol. 2, Hal: 685-694.