

ANALISIS KECENDERUNGAN LAPORAN MASYARAKAT PADA “LAPORGUB..!” PROVINSI JAWA TENGAH MENGGUNAKAN *TEXT MINING* DENGAN *FUZZY C-MEANS CLUSTERING*

Ratna Kurniasari^{1*}, Rukun Santoso², Alan Prahutama³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
ratnakurniasari55@gmail.com

ABSTRACT

Effective communication between the government and society is essential to achieve good governance. The government makes an effort to provide a means of public complaints through an online aspiration and complaint service called “LaporGub..!”. To group incoming reports easier, the topic of the report is searched by using clustering. Text Mining is used to convert text data into numeric data so that it can be processed further. Clustering is classified as soft clustering (fuzzy) and hard clustering. Hard clustering will divide data into clusters strictly without any overlapping membership with other clusters. Soft clustering can enter data into several clusters with a certain degree of membership value. Different membership values make fuzzy grouping have more natural results than hard clustering because objects at the boundary between several classes are not forced to fully fit into one class but each object is assigned a degree of membership. Fuzzy c-means has an advantage in terms of having a more precise placement of the cluster center compared to other cluster methods, by improving the cluster center repeatedly. The formation of the best number of clusters is seen based on the maximum silhouette coefficient. Wordcloud is used to determine the dominant topic in each cluster. Word cloud is a form of text data visualization. The results show that the maximum silhouette coefficient value for fuzzy c-means clustering is shown by the three clusters. The first cluster produces a word cloud regarding road conditions as many as 449 reports, the second cluster produces a word cloud regarding covid assistance as many as 964 reports, and the third cluster produces a word cloud regarding farmers fertilizers as many as 176 reports. The topic of the report regarding covid assistance is the cluster with the most number of members.

Keywords: LaporGub, Text Mining, Clustering, Fuzzy C-Means, Silhouette Coefficient, Wordcloud

1. PENDAHULUAN

Pesatnya perkembangan teknologi informasi telah menjadikan internet sebagai teknologi komunikasi yang menjadi sumber kebutuhan manusia. Kehadiran internet sebagai media informasi dan komunikasi semakin diterima dan dibutuhkan oleh masyarakat dunia seiring meningkatnya jumlah pengguna internet (Kristina, 2015).

Terwujudnya pelayanan publik yang berkualitas merupakan salah satu ciri tata kelola pemerintahan yang baik (*good governance*) (Handoyo, 2014). Komunikasi yang efektif antara pemerintah dan rakyat sangat penting diperlukan untuk mencapai tata kelola pemerintahan yang baik. Layanan pengaduan masyarakat dengan memanfaatkan teknologi informasi menjadi jawaban yang tepat untuk memenuhi keperluan tersebut. Pemerintah Provinsi Jawa Tengah memanfaatkan jumlah pengguna internet di Provinsi Jawa Tengah untuk meningkatkan kinerja pemerintah. Pemerintah mengupayakan penyediaan pengaduan masyarakat melalui suatu layanan aspirasi dan pegaduan secara online bernama “LaporGub..!”. Masyarakat dapat menggunakan fasilitas “LaporGub..!” untuk menyampaikan permasalahan yang ada di tengah masyarakat agar dapat langsung melaporkannya (Sinaga & Waluyo, 2016).

Semakin bertambahnya jumlah laporan yang masuk maka semakin meningkatkan jumlah data teks yang dapat diolah menjadi suatu informasi. Untuk memudahkan memperoleh laporan masuk yang dominan, maka dicarilah topik dari laporan masuk dengan menggunakan *clustering*. Pada penelitian ini, metode analisis kecenderungan topik laporan yang digunakan yaitu *text mining* dengan *fuzzy c-means clustering*. *Text Mining* digunakan untuk mengubah data teks menjadi data numerik agar dapat diolah lebih lanjut. *Fuzzy c-means clustering* adalah suatu teknik pengelompokan data dengan keberadaan tiap titik

data dalam suatu *cluster* ditentukan oleh derajat keanggotaan (Jaya et al., 2014). Pembentukan banyaknya cluster terbaik dilihat berdasarkan *Silhouette coefficient* yang maksimum. *Silhouette coefficient* bekerja dengan cara mengevaluasi penempatan setiap objek dalam setiap cluster dengan membandingkan jarak rata-rata objek dalam satu *cluster* dan jarak antara objek dengan *cluster* yang berbeda (Aini, 2014).

Tujuan penelitian ini adalah mengetahui *cluster* laporan masyarakat yang terbentuk menggunakan *text mining* dengan *fuzzy c-means clustering* serta mengetahui kecenderungan laporan masyarakat menggunakan *wordcloud* dari laporan masyarakat yang masuk melalui kanal website “LaporGub..!” Provinsi Jawa Tengah.

2. TINJAUAN PUSTAKA

2.1. “LaporGub..!” Provinsi Jawa Tengah

“LaporGub..!” merupakan portal pengaduan online tentang pemerintah Provinsi Jawa Tengah yang dikembangkan oleh Dinas Komunikasi dan Informatika Provinsi Jawa Tengah. Bentuk pengaduan masyarakat dapat berupa keluhan, saran dan masukan. Laporan dari masyarakat yang masuk ke “LaporGub..!” diproses dan diteruskan untuk ditindaklanjuti oleh pihak terkait sesuai isi laporan. Sistem “LaporGub..!” menyediakan berbagai bentuk pilihan kanal antara lain *Website*, aplikasi berbasis android, (*Short Message Service*) SMS, Instagram, Facebook, Twitter, WhasApp, *Hotline* serta aduan secara langsung.

2.2. *Text Mining*

Text mining adalah bidang baru pada cabang ilmu komputer yang mempelajari cara menangani krisis informasi yang berlebihan dengan menggabungkan beberapa teknik dari data mining, *machine learning*, *natural language processing*, pengambilan informasi (*informasi retrieval*) dan pengelolaan ilmu pengetahuan (*knowledge management*) (Feldman & Sanger, 2007). Menurut Feldman dan Sanger (2007) *text mining* melibatkan beberapa tahap, diantaranya *text preprocessing*, *text representation*, teknik analisis (seperti analisis distribusi, pengelompokan, analisis tren, dan aturan asosiasi) serta visualisasi hasil.

Text Preprocessing adalah tahap awal pada operasi *knowledge discovery* dari *text mining*, meliputi berbagai jenis teknik yang diadaptasi dari pengambilan informasi dan ekstraksi informasi dengan cara mengubah format tidak terstruktur menjadi terstruktur kemudian dapat diolah pada tahap berikutnya (Feldman & Sanger, 2007). Terdapat beberapa tahapan *text preprocessing* yang dilakukan berdasarkan data yang dimiliki. Penelitian ini melakukan tahapan-tahapan sebagai berikut:

1. *Case Folding*, yaitu menyeragamkan semua bentuk huruf menjadi *lowercase*
2. *Removing*, yaitu tahapan menghilangkan karakter yang dianggap tidak memberikan pengaruh pada proses analisis. Proses *removing* yang digunakan sebagai berikut:
 - *Remove URL* yaitu menghilangkan *URL* yang termuat dalam dokumen. *URL* merupakan serangkaian karakter berupa huruf, angka dan simbol yang diawali protokol “https” dan digunakan mengakses website, contohnya <https://laporgub.jatengprov.go.id>.
 - *Remove Mention* yaitu menghapus nama pengguna yang diawali simbol “@”.
 - *Remove Hastag* yaitu menghilangkan kata berawalan simbol “#”. *Hastag* merupakan penunjuk sebuah kata yang dibicarakan dan memiliki simbol “#”.
 - *Remove Punctuation* yaitu menghilangkan karakter yang bukan merupakan *alphabet*. *Punctuation* merupakan tanda baca yang digunakan dalam teks.
 - *Remove Number* yaitu menghilangkan karakter berupa angka
3. *Strip Whitescape*, yaitu menghapus spasi yang berlebih pada dokumen yang akan dianalisis karena spasi digunakan sebagai pemisah berjumlah satu spasi

4. Normalisasi Kata, yaitu mengkonversi kata yang tidak baku menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).
5. *Stopword Removal*, yaitu menghilangkan kata-kata umum yang sering muncul namun dianggap tidak bermakna dalam dokumen.
6. *Stemming*, yaitu proses menemukan kata dasar dari sebuah kata dengan menghapus semua imbuhan (*affixes*) baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) serta kombinasi dari awalan dan akhiran (*confixes*) pada kata turunan (Febriyanto, 2019)
7. *Tokenizing*, yaitu adalah proses pemecahan konten teks menjadi kata-kata, istilah, simbol, atau beberapa elemen bermakna lainnya yang disebut token (Vijayarani & Janani, 2016).

Text representation adalah tahap mengubah data teks menjadi representasi yang lebih mudah untuk diproses. Salah satu pendekatan untuk *text representation* adalah menggunakan matriks dokumen atau yang biasa disebut *Document Term Matrix* (Putri et al., 2019). Baris pada matriks mewakili dokumen, sedangkan kolom pada matriks berisi kata-kata, frase atau unit hasil indeks lainnya dalam dokumen yang digunakan untuk mencari konteks dari dokumen tersebut (*term*).

Setiap kata memiliki tingkat kepentingan yang berbeda dalam dokumen, sehingga perlu dilakukan pembobotan pada setiap kata. TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah teknik untuk memberikan bobot hubungan suatu kata atau *term* terhadap suatu dokumen (Robertson, 2004). Pembobotan TF-IDF adalah hasil perkalian dari pembobotan *Term Frequency* dan *Inverse Document Frequency* dari suatu kata dengan formula sebagai berikut (Salton & Buckley, 1988):

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

dengan:

$W_{j,i}$ = Pembobotan TF-IDF untuk term ke-j pada dokumen ke-i

$n_{j,i}$ = Banyaknya kemunculan *term* ke-j pada dokumen ke-i

$\sum_k n_{k,i}$ = Banyaknya kemunculan seluruh term pada dokumen ke-i

D = Banyaknya semua dokumen dalam koleksi

d_j = Banyaknya dokumen yang mengandung term ke-j.

Text visualization merupakan teknik yang digunakan untuk menampilkan visualisasi data dalam bentuk teks salah satunya yaitu *wordcloud*. *Wordcloud* memberikan informasi yang relevan dengan menampilkan istilah penting dalam teks (Tessem et al., 2015). Besar kecilnya ukuran teks akan menyesuaikan dengan frekuensi kata, semakin banyak frekuensi kata yang digunakan maka semakin besar ukuran kata yang ditampilkan (Pradana, 2020).

2.3. Distance Space

Distance space atau yang biasa disebut dengan pengukuran jarak merupakan langkah perhitungan jarak setiap objek dengan objek lainnya (Rofiqi, 2017). Pasangan objek dengan jarak lebih pendek akan mirip satu sama lain, dibandingkan dengan pasangan objek yang memiliki jarak lebih panjang. Formula pengukuran jarak Euclidean dua vektor $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ dan $\mathbf{y} = (y_1, y_2, \dots, y_p)'$ sebagai berikut (Rencher, 2002):

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

dengan:

- $d(x, y)$ = Jarak Euclidean objek x dan y
 p = Banyaknya parameter
 x_j = Nilai pengamatan vektor x objek ke- j
 y_j = Nilai pengamatan vektor y objek ke- j

2.4. Fuzzy C-Means Clustering

Clustering adalah proses membagi sekelompok objek data menjadi beberapa kelompok, sehingga objek-objek dalam suatu kelompok memiliki banyak kesamaan, tetapi memiliki banyak perbedaan dengan objek-objek dalam kelompok lain (Han et al., 2012). Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota kelompok dan meminimalkan kesamaan antar *cluster*. Pengelompokan yang baik harus mengelompokkan objek yang serupa dan memisahkan objek yang berbeda.

Fuzzy C-Means Clustering adalah suatu teknik pengelompokan data dengan keberadaan tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan (Jaya et al., 2014). *Fuzzy C-Means clustering* menggunakan model pengelompokan *fuzzy* sehingga data dapat menjadi anggota dari semua *cluster* yang dibentuk oleh derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1, tingkat keberadaan data dalam suatu *cluster* ditentukan berdasarkan derajat keanggotaannya (Luthfi, 2007). Konsep dasar *fuzzy c-means clustering* adalah menentukan pusat *cluster* yang akan menandai lokasi rata-rata untuk tiap *cluster*. Setiap titik data memiliki derajat keanggotaan untuk tiap *cluster* yang terbentuk. Pada kondisi awal pusat *cluster* belum akurat, oleh karena itu dilakukan perbaikan pusat *cluster* dan derajat keanggotaan tiap-tiap titik data secara berulang hingga berada pada titik yang tepat. Perulangan yang dilakukan berdasarkan pada minimasi fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut, perulangan ini akan membuat pusat *cluster* bergerak menuju ke lokasi yang tepat (Kusumadewi & Purnomo, 2010). Tahapan-tahapan *fuzzy c-means clustering* sebagai berikut (Kusumadewi & Purnomo, 2010) :

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

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

2. Menentukan:
 - a. Banyaknya *cluster* = c ;
 - b. Pangkat/pembobot = w ;
 - c. Maksimum iterasi = MaksIter ;
 - d. Error yang diharapkan = ξ ;
 - e. Fungsi Objektif awal = $P_0 = 0$;
 - f. Iterasi awal = $t=1$;
3. Membangkitkan bilangan acak μ_{ik} dengan i ($i=1,2,\dots,n$) merupakan banyak data dan k ($k=1,2,\dots,c$) merupakan banyak *cluster* sebagai elemen-elemen matriks partisi awal U .

$$U^{(0)} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1c} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2c} \\ \vdots & \vdots & \cdots & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nc} \end{bmatrix}$$

dengan syarat :

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

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

4. Menghitung pusat *cluster* ke- k : 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 :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^w d_{ik}^2(X_{ij}, V_{kj}) \quad (4)$$

dengan $d_{ik}(X_{ij}, V_{kj})$ adalah jarak Euclidean antara data ke- i dengan pusat *cluster* ke- k , sehingga diperoleh sebagai berikut:

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

6. Menghitung perubahan derajat keanggotaan setiap data pada setiap *cluster* (memperbaiki matriks partisi U) dengan :

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

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

7. Menyelidiki kondisi berhenti

- Jika : $(|P_t - P_{t-1}| < \xi)$ atau $(t > \text{MaksIter})$ maka berhenti;
- Jika : $(|P_t - P_{t-1}| > \xi)$ dan $(t < \text{MaksIter})$ maka $t=t+1$ kemudian mengulangi langkah 4

2.5. Silhouette Coefficient

Silhouette bekerja dengan cara mengevaluasi penempatan setiap objek yang berada pada setiap *cluster* dengan membandingkan jarak rata-rata objek dalam satu *cluster* dan jarak antara objek dengan *cluster* lainnya (Aini, 2014). Langkah-langkah metode *silhouette coefficient* sebagai berikut (Struyf et al., 1997):

1. Menghitung rata-rata jarak objek ke- i dengan semua objek yang berada didalam satu *cluster*.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (7)$$

dengan :

$a(i)$ = rata-rata jarak objek ke- i dengan semua objek dalam satu *cluster* A

$d(i, j)$ = jarak antara objek i dan j

$|A|$ = banyaknya objek pada *cluster* A

2. Menghitung rata-rata jarak objek ke- i dengan semua objek pada *cluster* lain.

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (8)$$

dengan:

$d(i, C)$ = jarak rata-rata objek i dengan semua objek pada *cluster* C dengan $C \neq A$

- $|C|$ = banyaknya objek pada *cluster* C
3. Menghitung $d(i, C)$ untuk semua C kemudian menentukan nilai minimumnya, $b(i)$ menunjukkan perbedaan rata-rata objek i untuk *cluster* terdekat dengan tetangganya.
- $$b(i) = \min d(i, C) \quad (9)$$

dengan :

$b(i)$ = nilai minimum dari jarak rata-rata objek i dengan semua objek pada *cluster* lain C

4. Menghitung nilai *silhouette*

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (10)$$

dengan :

$s(i)$ = nilai *silhouette*, nilai berada pada kisaran -1 hingga 1

Kriteria subjektif yang digunakan untuk melihat pengukuran baik atau tidaknya pengelompokan berdasarkan *silhouette coefficient* sebagai berikut (Kaufman et al., 1990):

Tabel 1. Interpretasi Nilai *Silhouette Coefficient*

Nilai <i>Silhouette coefficient</i>	Interpretasi
0.71 - 1.00	Memiliki persebaran <i>cluster</i> yang kuat
0.51 - 0.70	Memiliki persebaran <i>cluster</i> yang standar
0.26 - 0.50	Memiliki persebaran <i>cluster</i> yang lemah
≤ 0.25	Memiliki persebaran <i>cluster</i> yang tidak terstruktur

3. METODE PENELITIAN

3.1. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data teks laporan masyarakat pada kanal website “LaporGub..!” Provinsi Jawa Tengah yang diambil melalui Dinas Komunikasi dan Informatika Provinsi Jawa Tengah sebanyak 1.589 laporan selama periode Oktober – Desember 2020.

3.2. Langkah-langkah Analisis Data

Analisis kecenderungan laporan dilakukan menggunakan metode *text mining* dengan metode lanjutan berupa *fuzzy c-means clustering*. Data penelitian diproses menggunakan perangkat lunak RStudio. Adapun Langkah-langkah analisis yang dilakukan sebagai berikut:

1. Menyiapkan data teks laporan teks.
2. Melakukan *preprocessing* data untuk merubah data teks menjadi data numerik.
3. Data laporan diubah menjadi sebuah matriks dengan pembototan TF-IDF.
4. Penerapan *Fuzzy C-Means Clustering*
 - a. Menginisiasi banyaknya *cluster* (c), nilai bobot (w), Maksimal iterasi (MaksIter), Error terkecil (ξ) dan fungsi objektif awal (P_t)
 - b. Membangkitkan bilangan acak μ
 - c. Menghitung pusat *cluster* (V)
 - 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 berhenti, jika : ($|P_t - P_{t-1}| > \xi$) dan ($t < \text{MaksIter}$) maka $t = t + 1$ kemudian mengulangi langkah (c)
5. Menghitung indeks validitas *cluster* menggunakan *silhouette coefficient*.
6. *Profiling* masing-masing *cluster* dengan bantuan *wordcloud*.

4. HASIL DAN PEMBAHASAN

4.1. Website “LaporGub..!” Jawa Tengah

Kanal website “LaporGub..!” merupakan salah satu kanal pengaduan laporan masyarakat secara online yang dikembangkan oleh Dinas Komunikasi dan Informatika Provinsi Jawa Tengah. Laporan yang disampaikan pada kanal website dapat diakses melalui alamat <https://laporgub.jatengprov.go.id/>. Penelitian ini menggunakan data laporan yang masuk pada kanal website “LaporGub..!” selama periode Oktober 2020 – Desember 2020 dengan jumlah laporan sebanyak 1.589.

Tabel 2. Contoh Laporan Masuk pada Kanal Website “LaporGub..!”

No	Laporan	Waktu
1	Harga pupuk tak sebanding harga jual jagung pak..tolong kami..kamu rakyat kecil	01-10-2020 00:38
2	Mohon maaf sebelumnya atas pengaduan saya pada tgl 26 sep 2020 tentng bansos itu keliru krn trnyata istri saya sudh mendptkn bansos dd mohon maaf ts kekeliruan saya seetajg saya ucapkn terimakasih ts perhatiannya ...matusuwun	01-10-2020 01:48

4.2. Text Preprocessing

Text preprocessing mengubah format seluruh laporan menjadi data teks dengan format yang sama. Pada tahap ini proses yang dilakukan antara lain *case folding*, *removing* (*remove URL*, *remove mention*, *remove hastag*, *remove punctuation*, *remove number*), *strip whitespace*, normalisasi kata, *stopword removal*, *stemming* dan *tokenizing*.

Tabel 3. Contoh Hasil Proses *Text Preprocessing*

No	Laporan
1	harga pupuk sebanding harga jual jagung tolong rakyat
2	pengaduan september bansos keliru istri bansos dana desa kekeliruan

4.3. Text Representation

Pada proses ini, dilakukan perubahan data teks yang telah melalui proses *text preprocessing* menjadi matriks yang berisi keunculan kata (TF) pada setiap dokumen, kemudian menghitung pembobotan kata dengan menggunakan pembobotan (TF-IDF). Tabel 4 merupakan hasil pembobotan kata menggunakan TF-IDF.

Tabel 4. Hasil Pembobotan TF-IDF

No	aandri	...	adu	...	ketua	...	pemerintah	...	rakyat	...	solusi	...	zumrodin
1	0	...	0	...	0	...	0	...	0,518	...	0	...	0
2	0	...	0,505	...	0	...	0	...	0	...	0	...	0
3	0	...	0	...	0,520	...	0	...	0	...	0	...	0
4	0	...	0	...	0	...	0,125	...	0	...	0	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1589	0	...	0	...	0	...	0	...	0	...	0,141	...	0

4.4. Pengaplikasian *Fuzzy C-Means Clustering*

Proses *clustering* dengan *fuzzy c-means* dilakukan dengan bantuan Rstudio. Data yang digunakan adalah data dari hasil pembobotan TF-IDF. Pangkat fuzzyfier yang digunakan yaitu $w=2$, dengan batas error terkecil (ξ) = 10^{-9} dan maksimum iterasi sebanyak 1000. Proses *clustering* dilakukan sebanyak 5 kali dimulai dengan banyak 3 *cluster* sampai dengan

7 cluster. Banyaknya cluster optimum dari proses clustering ditentukan menggunakan validasi *silhouette coefficient*.

Tabel 5. Nilai *Silhouette Coefficient*

Banyak cluster	<i>Silhouette Coefficient</i>
3	0,05046042
4	0,01169989
5	0,01758649
6	-0,05830728
7	-0,193728

Berdasarkan Tabel 5 dapat dilihat bahwa banyaknya cluster yang paling optimum yaitu pada banyaknya 3 cluster, karena memiliki nilai *silhouette coefficient* terbesar yaitu 0,05046042. Sehingga banyaknya cluster yang akan dianalisis pada penelitian ini yaitu sebanyak 3 cluster. Hasil analisis cluster dengan banyak 3 cluster sebagai berikut:

Tabel 6. Hasil Analisis Cluster pada Cluster Optimum

Cluster ke-	Nomor Anggota Laporan	Jumlah Anggota
1	3, 9, 10, 11, 14, 21, 23, 24, 26, 28, 33, 34, 43, 44, ..., 1518	449
2	2, 4, 5, 6, 7, 8, 12, 13, 15, 16, 17, 18, 19, 20, ..., 1589	964
3	1, 36, 39, 60, 67, 100, 101, 109, 114, 121, 132, ..., 1587	176

Berdasarkan Tabel 6 didapatkan hasil jumlah anggota dari masing-masing cluster yaitu cluster 1 beranggotakan 449 laporan, cluster 2 beranggotakan 964 laporan, dan cluster 3 beranggotakan 176 laporan.

4.5. Penentuan Topik Laporan Menggunakan Wordcloud

Analisis lanjut untuk mengetahui kecenderungan topik laporan pada setiap cluster menggunakan bantuan pembentukan *wordcloud* pada masing-masing cluster. Pembuatan *wordcloud* masing-masing cluster dilakukan dengan bantuan *software* R Studio. *Wordcloud* pada Gambar 1 terlihat bahwa cluster 1 dengan kata frekuensi muncul paling tinggi yaitu kata “jalan” diikuti dengan kata yang lain yaitu desa, warga, lapor, tindak, rusak, tolong, bangun, dan lain-lain. Setelah ditinjau kembali dengan mencari laporan yang berisi kata “jalan” dapat disimpulkan bahwa laporan yang berada pada cluster 1 merupakan laporan yang berisi tentang keluhan mengenai kondisi jalan rusak yang banyak terjadi di daerah desa, oleh karena itu warga meminta pihak pemerintah untuk dilakukan tindakan perbaikan jalan. Sehingga cluster 1 merupakan cluster dengan anggota laporan mengenai permasalahan yang terjadi dengan kondisi jalan.



Gambar 1. *Wordcloud Cluster 1*

Wordcloud pada 2 terlihat bahwa cluster 2 dengan kata frekuensi muncul paling tinggi adalah kata “bantu” diikuti dengan kata yang lain yaitu covid, keluarga, desa, rumah, orang, dan lain-lain. Setelah ditinjau kembali dengan mencari laporan yang berisi kata “bantu” dapat disimpulkan bahwa laporan yang berada pada cluster 2 merupakan laporan yang berisi

tentang keluhan banyak masyarakat yang belum menerima bantuan covid, selain itu banyak masyarakat pula yang tidak mendapatkan bantuan covid, karena dirasa bantuan covid diberikan pada sasaran yang tidak sesuai. Sehingga *cluster 2* merupakan *cluster* dengan anggota laporan mengenai banyaknya masyarakat yang meminta bantuan kepada pemerintah akibat dampak wabah covid.



Gambar 2. Wordcloud Cluster 2

Wordcloud pada Gambar 3 terlihat bahwa *cluster 3* dengan kata frekuensi muncul paling tinggi adalah kata “pupuk” diikuti dengan kata yang lain yaitu tani, kartu, subsidi, tolong, harga, bantu dan lain-lain. Setelah ditinjau kembali dengan mencari laporan yang berisi kata “pupuk” dapat disimpulkan bahwa laporan yang berada pada *cluster 3* merupakan laporan yang berisi mengenai keluhan para petani saat pembelian pupuk dengan menggunakan kartu tani. Walaupun para petani telah mempunyai kartu tani namun para petani mengalami kesulitan untuk membeli pupuk bersubsidi. Sehingga *cluster 3* merupakan *cluster* dengan anggota laporan mengenai permasalahan yang terjadi pada pembelian pupuk yang dialami oleh petani.



Gambar 3. Wordcloud Cluster 3

5. KESIMPULAN

Penentuan pembentukan banyaknya *cluster* terbaik yang terbentuk dari 1.589 laporan masuk pada “LaporGub..!” Provinsi Jawa Tengah dilihat berdasarkan nilai validasi *silhouette coefficient* paling maksimum. Nilai *silhouette coefficient* terbesar untuk *fuzzy c-means clustering* yaitu 0,05046042 pada banyaknya 3 *cluster*. Kecenderungan topik laporan masyarakat pada “LaporGub..!” dapat dilihat dari terbentuknya *wordcloud* pada masing-masing *cluster*. Penerapan *fuzzy c-means clustering* menghasilkan 3 *cluster* laporan dengan topik mengenai kondisi jalan, bantuan covid, serta mengenai pupuk petani. *Cluster* dengan topik bantuan covid merupakan *cluster* dengan jumlah anggota paling banyak.

Setelah ditinjau kembali untuk melihat isi laporan yang berada pada masing-masing *cluster* dapat disimpulkan bahwa *cluster 1* merupakan laporan yang berisi tentang keluhan mengenai kondisi jalan rusak yang banyak terjadi di daerah desa sehingga warga meminta pihak pemerintah untuk dilakukan tindakan perbaikan jalan. *Cluster 2* merupakan laporan yang berisi tentang keluhan banyak masyarakat yang belum menerima bantuan covid, selain itu banyak masyarakat pula yang tidak mendapatkan bantuan covid, karena dirasa bantuan covid diberikan pada sasaran yang tidak sesuai. Sedangkan *cluster 3* merupakan laporan yang berisi mengenai keluhan para petani saat pembelian pupuk dengan menggunakan kartu tani. Walaupun para petani telah mempunyai kartu tani namun para petani mengalami kesulitan untuk membeli pupuk bersubsidi.

DAFTAR PUSTAKA

- Aini, F. N. (2014). *Clustering Business Process Model Petri Net*. Jurnal Itsmart Vol. 3, No. 2: Hal. 47-51.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- Febriyanto, F. (2019). Sistem Penilaian Otomatis Jawaban Esai Dengan Menggunakan Metode Vector Space Model Pada Beberapa Perkuliahan Di Stmik Indonesia Banjarmasin. *Jurnal Teknologi Informasi* Vol XIV, No. 1: Hal. 53–68.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Data mining concepts and techniques, Third Edition*. Waltham: Morgan Kaufmann Publishers
- Jaya, T. S., Adi, K., & Noranita, B. (2014). *Sistem Pemilihan Perumahan dengan Metode Kombinasi Fuzzy C-Means Clustering dan Simple Additive Weighting*. Jurnal Sistem Informasi Bisnis, Vol. 1
- Kaufman, L., Rousseeuw, P. J., & Gentle, J. E. (1991). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons Inc.
- Kusumadewi, S., & Purnomo, H. (2010). *Aplikasi Logika Fuzzy untuk Pendukung Keputusan (Edisi 2)*. Yogyakarta: Graha Ilmu.
- Luthfi, E. T. (2007). *Fuzzy C-Means Untuk Clustering Data (Studi Kasus : Data Performance Mengajar Dosen)*. Seminar Nasional Teknologi 2007, Yogyakarta : 24 November 2007. Hal: 1-7.
- Pradana, M. G. (2020). *Penggunaan fitur wordcloud dan document term matrix dalam text mining*. Jurnal Ilmiah Informatika (JIF) Vol. 08, No. 1: Hal. 38-43.
- Putri, R. K., Warsito, B., & Mustafid, M. (2019). *Implementasi Algoritma Modified Gustafson-Kessel Untuk Clustering Tweets Pada Akun Twitter Lazada Indonesia*. Jurnal Gaussian Vol. 8, No. 3: Halm 285-295.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis. In Methods of Multivariate Analysis*. Amerika: John Wiley & Sons, Inc.
- Robertson, S. (2004). *Understanding inverse document frequency: On theoretical arguments for IDF*. Journal of Documentation Vol. 60, No. 5: Hal. 503–520.
- Rofiqi, A. Y. (2017). *Clustering Berita Olahraga Berbahasa Indonesia Menggunakan Metode K-Medoid Bersyarat*. Jurnal Simantec Vol. 6, No. 1: Hal. 25-32.
- Salton, G., & Buckley, C. (1988). *Term-weighting approaches in automatic text retrieval*. Information Processing & Management.
- Struyf, A., Hubert, M., & Rousseeuw, P. J. (1997). *Integrating robust clustering techniques in S-PLUS*. Computational Statistics and Data Analysis, Vol 26: No.1: Hal. 17–37.
- Tessem, B., Bjørnstad, S., Chen, W., & Nyre, L. (2015). *Word cloud visualisation of locative information*. Journal of Location Based Services Vol. 9, No. 4: Hal.254-272.
- Vijayarani, S., & Janani, R. (2016). *Text Mining: open Source Tokenization Tools – An Analysis*. Advanced Computational Intelligence: An International Journal (ACII) Vol. 3, No.1: Hal. 37–47.