ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 14, Nomor 2, Tahun 2025, Halaman 423 - 432
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

FUZZY POSSIBILISTIC C-MEANS (FPCM) *CLUSTERING* UNTUK IDENTIFIKASI KELUHAN UTAMA PELANGGAN INDIHOME PADA DATA *TWEETS*

Taufik Aji Putra[1*], Iut Tri Utami[2], Ardiana Alifatus Sa'adah[3]
[1,2,3] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail**: taufikajiputra84@gmail.com

DOI: 10.14710/j.gauss.14.2.423-432

Article Info:
Received: 2024-10-27
Accepted: 2025-10-31
Available Online: 2025-11-07

Keywords:
IndiHome; Tweets; Clustering; Fuzzy Possibilistic C-Means; Extended Xie-Beni Index; Word Cloud

Abstract: Customer complaints are issues faced by customers regarding a company's products or services. Many companies use Twitter as a platform to interact with customers, making handling complaints through social media crucial for building a positive image and maintaining customer loyalty. IndiHome Regional 4 faces challenges in identifying main complaints due to a high volume of complaints on Twitter. Cluster analysis groups similar complaints, aiding the identification process. Text mining converts textual data into numerical format, streamlining complaint processing. Fuzzy Possibilistic C-Means Clustering, a fuzzy-based method, enables data membership across clusters with varying degrees of membership. By adopting relative (fuzzy) and absolute (possibilistic) membership, more accurate data placement is achieved. Data consists of IndiHome Regional 4 customer complaint tweets received via the Twitter channel "IndiHomeCare" from January to December 2022. The clustering process formed 4 clusters based on the smallest Extended Xie-Beni Index value, tested with different cluster numbers (3-7). Witel (also known as "Wilayah Telekomunikasi") Yogyakarta had the highest members and complaints in each cluster, while Witel Kudus had the lowest. Word Cloud analysis revealed main complaints in each cluster, including WiFi-related subscription costs, internet disruptions, customer service issues, and slow connections.

1. PENDAHULUAN

Keluhan pelanggan mencerminkan masalah dalam produk atau layanan perusahaan. Penanganan keluhan pelanggan di media sosial menjadi semakin penting karena memungkinkan pelanggan untuk menyampaikan keluhan dengan cepat dan mudah, yang dapat berdampak pada reputasi perusahaan di mata publik (Sigurdsson, *et al*., 2021). Contoh media sosial yang digunakan adalah Twitter, dimana salah satu perusahaan yang memanfaatkan Twitter untuk layanan pelanggan adalah IndiHome yang menggunakan akun Twitter @IndiHome dan @IndiHomeCare untuk layanan pelanggan.

IndiHome adalah layanan yang memiliki akses internet, telepon rumah, dan tayangan TV interaktif (IndiHome, 2023). Cakupan wilayah IndiHome berada secara luas di Indonesia, salah satunya adalah IndiHome Regional 4 yang berbasis di Provinsi Jawa Tengah dan Daerah Istimewa Yogyakarta. IndiHome Regional 4 merasa kesulitan untuk mengidentifikasi masalah utama dari keluhan pelanggan di Twitter. Dalam mengelompokkan data keluhan, diperlukan metode untuk mengidentifikasi keluhan utama pelanggan yang serupa, sehingga proses identifikasi dapat mudah dilakukan. (Aulia, 2023).

Fuzzy Possibilistic C-Means Clustering (FPCM) ialah algoritma *clustering* dengan logika *fuzzy* dimana memiliki konsep tingkat keanggotaan setiap data yang bervariasi antara 0 hingga 1, sehingga memungkinkan setiap data menjadi anggota dari semua *cluster*. FPCM

dapat memperbaiki dan menghasilkan penempatan data pada *cluster* yang lebih akurat karena menggunakan dua jenis keanggotaan, *fuzzy* dan *possibilistic*.

Penelitian ini bertujuan untuk mengidentifikasi keluhan utama dari pelanggan IndiHome dengan membentuk *cluster* dari keluhan yang diterima oleh IndiHome Regional 4 yang telah dilakukan proses FPCM dan mengidentifikasi keluhan utama pelanggan IndiHome Regional 4 (Provinsi Jawa Tengah dan Daerah Istimewa Yogyakarta) berdasarkan persebaran Wilayah Telekomunikasi (Witel) dan *word cloud* dari hasil *cluster* FPCM.

2. TINJAUAN PUSTAKA

Pelanggan IndiHome memiliki beberapa opsi untuk mengirimkan keluhan, seperti mengadukan masalah, memberikan usulan, atau meminta informasi melalui kanal pengaduan seperti *call center* 147, *social media* IndiHome (Whatsapp, Instagram, Facebook, dan Twitter), serta aplikasi myIndiHome (IndiHome, 2023). Pelanggan memiliki kebebasan untuk mengirimkan pengaduan melalui berbagai saluran, salah satunya melalui Twitter. Pelanggan dapat mengirimkan keluhan atau permintaan dalam bentuk *tweet* dengan menggunakan kata kunci IndiHome atau melakukan *tagging* pada akun terkait seperti @IndiHome atau @IndiHomeCare. Sebelum dilakukan proses *clustering*, data *tweets* keluhan pelanggan IndiHome Regional 4 yang berbentuk teks diproses dengan metode *text mining*.

Text mining ialah metode untuk mendapatkan informasi yang bermutu dari dokumen teks menggunakan berbagai alat analitik untuk menemukan dan mengeksplorasi tren data yang menarik (Feldman dan Sanger, 2007). *Text mining* menghubungkan dan membandingkan informasi dokumen untuk mengidentifikasi kata-kata yang merepresentasikan teks, dengan mengungkap hubungan dan menghasilkan informasi baru.

Text preprocessing dalam t*ext mining* merupakan cara penting untuk mengolah data tidak terstruktur menjadi data yang terstruktur, melalui penghapusan, pembersihan, dan transformasi data, agar siap digunakan pada tahap analisis selanjutnya. (Feldman dan Sanger, 2007). Langkah-langkah yang dilakukan dalam *text preprocessing* yaitu: *case folding, removing* (*url, mention, number, punctuation, emoticon, strip white space*), pembakuan kata, *stopwords removal*, *stemming*, serta *tokenizing*.

Text Representation merupakan proses mengolah teks menjadi format lebih sederhana untuk dilakukan analisis. *Term Frequency-Inverse Document Frequency* (TF-IDF) ialah salah satu teknik pembobotan dalam pendekatan *text representation*. TF-IDF menggunakan dua konsep berupa *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Perhitungan TF-IDF bisa dihitung dengan persamaan (1) (Upendra dan Babu, 2016):

$$w_{j,\mathrm{i}} = \frac{b_{\mathrm{j,i}}}{\sum_{j=1}^{m} b_{j,i}} \cdot \log_2 \frac{\mathrm{D}}{\mathrm{D_j}} \tag{1}$$

Teknik analisis data yang mengelompokkan objek data pada kelompok-kelompok dengan perbedaan karakteristik, yang memiliki sifat relatif homogen di dalam setiap kelompok, salah satunya adalah metode *Clustering* (Talakua, *et al*., 2017). Dalam analisis *cluster*, pendekatan umum untuk mengukur kemiripan antara objek adalah dengan menggunakan konsep ukuran jarak. (Alwi dan Hasrul, 2018). Salah satu metode perhitungan jarak kedekatan ialah *Euclidean*. 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* yakni (Vijaymeena dan Kavitha, 2016):

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

Fuzzy Possibilistic C-Means Clustering (FPCM) ialah salah satu metode *clustering* non-hierarki yang menggunakan logika *fuzzy*, dengan setiap data dapat memiliki tingkat keanggotaan antara 0 dan 1 dalam semua *cluster*. (Kusumadewi dan Purnomo, 2013). Pada teori himpunan *fuzzy*, peranan derajat keanggotaan sebagai penentu keberadaan elemen dalam suatu himpunan termasuk penting. Nilai keanggotaan atau derajat keanggotaan atau *membership function* menjadi ciri utama dari penalaran dengan logika *fuzzy* tersebut (Kusumadewi dan Purnomo, 2013). FPCM menggunakan dua jenis keanggotaan, yakni pemodelan keanggotaan *fuzzy* pada *Fuzzy C-Means* (FCM) dan keanggotaan *possibilistic* pada *Possibilistic C-Means* (PCM), yang dapat mengatasi kelemahan FCM yang sensitif terhadap pemilihan parameter awal dan waktu komputasi yang lama, serta PCM yang sensitif terhadap inisialisasi karena matriks kekhasan absolut (matriks keanggotaan PCM) saling independen pada baris dan kolom (Suganya dan Shanthi, 2012).

FPCM menggunakan konsep dasar dari FCM dimana pusat *cluster* dan matriks kekhasan relatif (matriks keanggotaan) hasil FCM digunakan untuk menghitung matriks kekhasan pada FPCM. Dalam FPCM, perbaikan dilakukan pada pusat *cluster* dan tingkat keanggotaan setiap data demi mencapai lokasi yang tepat. Proses ini berulang dengan tujuan meminimalkan fungsi objektif yang menceminkan bahwa tingkat keanggotaan mempengaruhi jarak antara setiap data dan pusat *cluster* (Putri, *et al*., 2022).

Sebelum melakukan pengelompokan dengan FPCM, tahap awal adalah melakukan perhitungan FCM. Berikut langkah-langkah untuk FCM (Nayak, *et al*., 2015):

1. Memasukkan data pada matriks $\boldsymbol{X}$ ukuran $n \times m$, dimana n ialah jumlah data, sedangkan m ialah variabel data.
$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$
2. Mendefinisikan beberapa parameter awal, seperti jumlah *cluster* (c), pembobot (w), maksimal iterasi (*MaxIter*), *error* (ξ), fungsi objektif awal (P_0), dan iterasi awal (t).
3. Menghasilkan bilangan acak μ_{ik} untuk anggota matriks kekhasan relatif awal **U**, dengan i ialah jumlah data dan k ialah urutan *cluster*.
$$\boldsymbol{U} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1c} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2c} \\ \vdots & \vdots & \dots & \vdots \\ \mu_{n1} & \mu_{n2} & \dots & \mu_{nc} \end{bmatrix}$$
dengan syarat:
$$\mu_{ik} \in [0,1] \,;\, (i = 1,2,\dots,n);\, (k = 1,2,\dots,c)$$
$$\sum_{k=1}^{c} \mu_{ik} = 1 \,;\, (k = 1,2,\dots,c)$$
4. Menentukan pusat *cluster* $\boldsymbol{V}_k$, dimana $\boldsymbol{V}_k = (v_{k1}, v_{k2}, \dots, v_{kj})$ 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. Menentukan fungsi objektif pada iterasi ke-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. Mengubah matriks kekhasan relatif **U**:

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

7. Meninjau keadaaan iterasi berakhir
   Apabila: ($|P_t - P_{t-1}| < \xi$) atau ($t \geq MaxIter$), iterasi tidak dilanjutkan;
   Apabila: ($|P_t - P_{t-1}| > \xi$) atau ($t < MaxIter$), $t = t + 1$, ulangi tahapan ke-4

Adapun untuk tahapan FPCM yakni (Boudouda, *et al*., 2005); (Putri, *et al*., 2022):

1. Memasukkan data (X_{ij}) pada matriks $\boldsymbol{X}$ ukuran $n \times m$, dengan n ialah jumlah data, sedangkan m ialah variabel data.

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

2. Menetapkan beberapa parameter awal, seperti banyak *cluster* (c), pembobot (w) dan (η), maksimal iterasi (*MaxIter*), *error* (ξ), fungsi objektif awal (P_0), dan iterasi awal (l).
3. Mengembalikan hasil optimal yakni matriks kekhasan relatif (**U**) dan pusat *cluster* ($\mathbf{V}_k$) pada hasil FCM, untuk menentukan matriks kekhasan absolut awal (**T**)

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

   dengan syarat:

$$t_{ik} \in [0,1]\,;(i = 1,2,\dots,n);(k = 1,2,\dots,c)$$

$$\sum_{i=1}^{n} t_{ik} = 1\,;(i = 1,2,\dots,n)$$

   Elemen matriks **T** awal dapat dihitung dengan persamaan berikut:

$$t_{ik}{}^{(0)} = \frac{\left[\sum_{j=1}^{m}\left(x_{ij}-v_{kj}\right)^2\right]^{\frac{-1}{\eta-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}\left(x_{ij}-v_{kj}\right)^2\right]^{\frac{-1}{\eta-1}}} \quad (6)$$

4. Memperbaiki pusat *cluster* ($\boldsymbol{V}_k$)

$$v_{kj} = \frac{\sum_{i=1}^{n}(\mu_{ik}{}^{w} + t_{ik}{}^{\eta})\, x_{ij}}{\sum_{i=1}^{n}(\mu_{ik}{}^{w} + t_{ik}{}^{\eta})} \quad (7)$$

5. Menentukan fungsi objektif pada iterasi ke-l

$$P_l = \sum_{i=1}^{n}\sum_{k=1}^{c}\left((\mu_{ik}{}^{w} + t_{ik}{}^{\eta})\left[\sum_{j=1}^{m}(x_{ij}-v_{kj})^2\right]\right) \quad (8)$$

6. Mengubah matriks kekhasan relatif **U**

$$\mu_{ik}{}^{(l)} = \frac{\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{w-1}}} \quad (9)$$

7. Mengkalkulasi perbaikan matriks kekhasan absolut **T**

$$t_{ik}{}^{(l)} = \frac{\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{\eta-1}}}{\sum_{i=1}^{n}\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{\eta-1}}} \quad (10)$$

8. Memeriksa keadaan iterasi berakhir
   Apabila: ($|P_l - P_{l-1}| < \xi$) atau ($l \geq MaxIter$), iterasi tidak dilanjutkan;
   Apabila: ($|P_l - P_{l-1}| > \xi$) atau ($l < MaxIter$), $l = l + 1$, mengulangi langkah ke-4

Indeks *Extended Xie-Beni* digunakan untuk menguji validitas hasil *cluster*. Berdasarkan penelitian Zhou (2008), Indeks *Extended Xie-Beni* adalah modifikasi dari *Xie-Beni* yang berpatokan pada nilai *Compactness* berupa jarak *intracluster* untuk mengukur kedekatan objek data dalam setiap *cluster*, serta *Separation* yakni jarak *intercluster* untuk mengukur perbedaan antara *cluster* satu dengan *cluster* lainnya (Muchsin dan Sudarma, 2015). Semakin rendah nilai Indeks *Extended Xie-Beni*, semakin optimal hasil *cluster*. Persamaan Indeks *Extended Xie-Beni* yakni sebagai berikut:

$$EXB = \frac{\sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left(\mu_{ik}{}^{w} + t_{ik}{}^{\eta}\right)\left[\sum_{j=1}^{m}(x_{ij} - v_{kj})^2\right]\right)}{n \min\limits_{h \neq k}\sum_{j=1}^{m}(v_{hj} - v_{kj})^2} \quad (11)$$

Word cloud ialah gambar yang memperlihatkan kata-kata dari teks dengan ukuran berbeda untuk menonjolkan kata-kata yang sering termuat pada suatu kumpulan data teks (Ivosights, 2023), sehingga memberikan gambaran visual topik utama pada teks.

3. METODE PENELITIAN

Data yang dipakai ialah data kualitatif berbentuk teks keluhan terhadap layanan IndiHome yang dituliskan oleh pengguna Twitter yang berlokasi di wilayah Regional 4, termasuk dalam data sekunder dan didapatkan di IndiHome Regional 4 selama periode Januari 2022 sampai Desember 2022 sebanyak 2.500 data. Pengerjaan analisis data menggunakan aplikasi RStudio dan Microsoft Excel.

Langkah-langkah dalam penelitian:

1. Memasukkan data *tweets* keluhan pelanggan IndiHome Regional 4.
2. Menjalankan *text preprocessing* data.
3. Menjalankan *text representation* dengan pembobotan kata menggunakan TF-IDF.
4. Menerapkan FCM:
   a. Memilih banyak *cluster* (c = 3-7), nilai pembobot (w = 2), maksimal iterasi (*MaxIter*=1.000), *error* yang kecil ($\xi = 10^{-9}$), fungsi objektif awal ($P_0 = 0$), dan iterasi awal (t =1).
   b. Menghasilkan bilangan acak μ_{ik} untuk anggota matriks kekhasan relatif awal **U**
   c. Menentukan pusat *cluster* ($\boldsymbol{V}_k$)
   d. Menentukan fungsi objektif (P_t)
   e. Menentukan perbaikan matriks kekhasan relatif **U**
   f. Meninjau kondisi tidak berlanjut, apabila ($|P_t - P_{t-1}| < \xi$) atau ($t \geq MaxIter$), iterasi berakhir, namun apabila: $|P_t - P_{t-1}| > \xi$) dan ($t < MaxIter$) akan $t=t+1$ lalu mengulangi tahapan (c).

5. Menerapkan FPCM:
   a. Memilih banyak *cluster* (c = 3-7), nilai pembobot (w = 2) dan (η = 2), maksimal iterasi (*MaxIter* = 1.000), *error* yang kecil (ξ =10^{-9}), fungsi objektif awal (P_0 = 0), dan iterasi awal (l =1).
   b. Mengembalikan nilai akhir matriks kekhasan relatif (**U**) dan pusat *cluster* ($\boldsymbol{V}_k$) dari hasil FCM untuk menentukan matriks kekhasan absolut **T**
   c. Menentukan perbaikan pusat *cluster* ($\boldsymbol{V}_k$)
   d. Menentukan fungsi objektif iterasi ke-l (P_l)
   e. Menentukan perbaikan matriks kekhasan relatif **U**
   f. Melakukan perbaikan matriks kekhasan absolut **T**
   g. Meninjau kondisi tidak berlanjut, apabila diperoleh ($|P_l - P_{l-1}| < \xi$) atau ($l \geq MaxIter$) maka iterasi berakhir, jika: $|P_l - P_{l-1}| > \xi$) dan ($l < MaxIter$) maka l=l+1 dan mengulang kembali langkah (c).
   h. Menentukan anggota pada setiap *cluster* dengan matriks kekhasan absolut **T**.
6. Menentukan *cluster* optimum menggunakan metode *Extended Xie-Beni*, yakni memilih indeks nilai terkecil.
7. Melakukan *profiling* semua *cluster* dengan *word cloud.*
8. Intepretasi *output profiling* dari semua *cluster* untuk menyimpulkan keluhan utama.

4. HASIL DAN PEMBAHASAN

Text preprocessing mengolah data teks yang tidak terorganisir menjadi lebih terstruktur. *Text preprocessing* yang dilakukan yakni *case folding,* penghapusan (*url, mention,* angka*, punctuation,* emoji*, strip whitespace*), pembakuan kata, penghapusan *stopwords*, *stemming*, dan *tokenizing*. Hasil *Text Preprocessing* dengan contoh data nomor 20, 915, 1.647 dan 2.398 dapat dilihat dalam Tabel 1.

Tabel 1. Contoh *Output Text Preprocessing*

Nomor	Data *tweets*	Sesudah *Text Preprocessing*
20	Ppp lag bgt @IndiHomeCare	ganggu
915	masih error, mohon segera perbaiki min thanks	error baik admin
1.647	gangguan kah wifi signal pagi ini?	ganggu wifi sinyal pagi
2.398	Internet saya gangguan tolong dilayani	internet ganggu tolong layan

Setelah *text preprocessing* selesai, akan dilakukan pembobotan kata dengan TF-IDF. Nilai bobot yang diperoleh dari pembobotan TF-IDF akan dipakai pada langkah awal FCM kemudian juga pada tahap awal FPCM. Hasil TF-IDF pada data keluhan di IndiHome Regional 4 dengan contoh keluhan nomor 10, 377, 926, dan 2.249 pada Tabel 2.

Tabel 2. Contoh *Output* Pembobotan TF-IDF

Nomor		aba	⋯	directmessage	⋯	zte
10	bantu restart modem lambat	0	⋯	0	⋯	0
377	cek directmessage wifi ganggu	0	⋯	0,944	⋯	0
926	admin bantu followup keluh via directmessage	0	⋯	0,629	⋯	0
2.249	kemarin directmessage directmessage aba	2,577	⋯	1,888	⋯	0

Banyaknya *cluster* yang ditentukan dalam proses FPCM *Clustering* ialah 3, 4, 5, 6 dan 7. Nilai pembobot sebesar w = 2 & $\eta = 2$, nilai *MaxIter* sejumlah 1.000, dan nilai *error* terendah yang diinginkan (ξ) = 10^{-9} sama dengan nilai *default* pada RStudio. Indeks *Extended Xie-Beni* berguna untuk mencari jumlah *cluster* yang paling optimal pada FPCM.

Angka hasil Indeks Extended *Xie-Beni* dengan jumlah *cluster k* = 3, 4, 5, 6, dan 7 ditampilkan pada Tabel 3.

Tabel 3. *Output* Indeks *Extended Xie-Beni*

k	Nilai *Extended Xie-Beni*
3	2,436567E+26
4	8,287821E+25
5	3,296357E+29
6	2,691483E+27
7	6,550994E+27

Dilihat dari Tabel 3, jumlah *cluster c* = 4 memperoleh angka Indeks *Extended Xie-Beni* yang paling rendah dibandingkan dengan jumlah *cluster* lainnya, yakni sebesar 8,287821E+25. Hasil *clustering* dengan jumlah *cluster c* = 4 ditampilkan pada Tabel 4.

Tabel 4. Jumlah Anggota Hasil *Cluster* FPCM

Cluster ke-	Nomor *Tweet*	Jumlah Anggota
1	7, 26, 27, 28, …, 2.472,2.479,2.488,2.499	482
2	2, 5, 6, 8, 9, …, 2.489, 2.490, 2.492, 2.496	1.078
3	3, 4, 60, 92, 108, …., 2.396, 2.410, 2.491, 2493	87
4	1, 10, 12, 15, 22, …, 2.495, 2.497, 2.498, 2.500	853

Keluhan utama pelanggan dari setiap *cluster* akan diidentifikasi berdasarkan persebaran Wilayah Telekomunikasi (Witel) PT Telkom Indonesia Regional 4, serta dengan menggunakan *Word cloud* untuk memeriksa kata-kata dengan frekuensi kemunculan terbanyak di setiap *cluster*. Hasil persebaran keluhan setiap Witel berdasarkan *cluster* dilihat pada Tabel 5.

Tabel 5. Persebaran Anggota *Cluster* FPCM berdasarkan Witel

Witel	Jumlah Anggota *Cluster*				Jumlah Keluhan
	Cluster 1	*Cluster* 2	*Cluster* 3	*Cluster* 4	
Semarang	84	195	14	156	449
Kudus	18	38	1	24	81
Pekalongan	26	68	8	66	168
Magelang	26	59	5	52	142
Purwokerto	32	103	9	51	195
Solo	52	129	10	96	287
Yogyakarta	244	486	40	408	1.178
Jumlah	482	1.078	87	853	2.500

Word cloud pada setiap *cluster* menunjukkan beberapa kesamaan pada kata-kata yang ada di setiap *cluster*, seperti “wifi” dan “internet”, sehingga 10 kata teratas yang sering muncul di semua *cluster,* kecuali kata-kata yang mempunyai kesamaan antar *cluster,* akan digunakan untuk melihat kata-kata lain dengan frekuensi kemunculan tertinggi yang berbeda dengan *cluster* lainnya yang dapat dilihat pada Tabel 6.

Tabel 6. Frekuensi Kemunculan Kata Utama pada Setiap *Cluster* dibandingkan dengan *Cluster* Lain

Cluster ke-	Kata	Frekuensi
1	bayar	104
	pakai	67
2	admin	435
	ganggu	327
	los	162
3	sinyal	10
	hilang	9
	mbps (*megabit per second*)	9
4	lambat	401
	jaring	97

Tabel 6 berguna sebagai acuan untuk melakukan analisis lebih mendalam terkait keluhan yang terdapat di setiap *cluster*, yang dapat memberikan gambaran tentang kata-kata utama yang paling sering muncul pada setiap *cluster*, sehingga dapat membantu dalam mengidentifikasi karakteristik utama dan permasalahan yang relevan pada setiap *cluster* yang terbentuk. Berikut adalah hasil intepretasi *Word cloud* setiap *cluster* yang terbentuk:

a. *Cluster* 1

Gambar 1. *Word Cloud Cluster* 1

Dapat dilihat pada Gambar 1, kata “bayar” dan “internet” mendominasi, diikuti oleh kata “pakai” dan “koneksi”. Kata-kata ini berhubungan dengan keluhan mengenai pembayaran biaya berlangganan WiFi dan pemakaian WiFi yang bermasalah seperti koneksi atau jaringan internet yang mengalami *error*.

b. *Cluster* 2

Gambar 2. *Word Cloud Cluster* 2

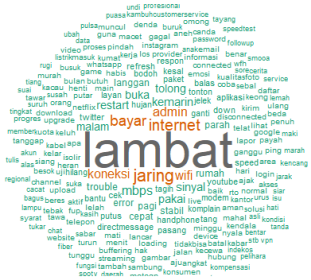
Dapat dilihat pada Gambar 2, kata “admin”, “internet, dan “ganggu” mendominasi, diikuti oleh kata “tolong”, “wifi”, dan “los”. Kata-kata ini berhubungan dengan keluhan mengenai gangguan internet mati, dilihat dari lampu indikator “los” yang menyala pada router WiFi, kemudian terdapat beberapa keluhan pelanggan yang meminta tolong kepada *admin* untuk dilakukan pengecekan dari *admin* dan atau teknisi IndiHome.

c. *Cluster* 3

Gambar 3. *Word Cloud Cluster* 3

Dapat dilihat pada Gambar 3, kata “wifi”, “bayar, dan “sinyal” mendominasi, diikuti oleh kata “hilang”, “mbps”, dan “admin”. Kata-kata ini berhubungan dengan keluhan mengenai sinyal jaringan internet yang hilang atau koneksi terputus, kemudian terdapat keluhan mengenai biaya yang dikeluarkan untuk pemakaian WiFi tidak sebanding dengan kualitas pelayanan dan kecepatan internet yang diperoleh pelanggan, hal tersebut dapat ditinjau dari frekuensi kata “mbps” yang cukup tinggi.

d. *Cluster* 4



Gambar 4. *Word Cloud Cluster* 4

Dapat dilihat pada Gambar 4, kata “lambat” mendominasi, diikuti oleh kata “jaring”, “internet”, dan “bayar”. Kata-kata ini berhubungan dengan mayoritas keluhan mengenai jaringan internet yang lambat, diikuti pembayaran biaya yang dikeluarkan untuk pemakaian WiFi tidak sebanding dengan kecepatan internet yang diperoleh pelanggan.

5. KESIMPULAN

Proses *clustering* 2.500 keluhan pelanggan IndiHome Regional 4 yang telah dilakukan proses FPCM menghasilkan sebanyak 4 *cluster*. Jumlah *cluster* optimal dilihat berpatokan pada angka Indeks *Extended Xie-Beni* yang terendah, yakni bernilai 8,287821E+25.

Ditinjau dari persebaran Witel (Wilayah Telekomunikasi) pada IndiHome Regional 4, Witel Yogyakarta memiliki jumlah anggota *cluster* dan keluhan tertinggi, sedangkan Witel Kudus yang terendah, kemudian Witel Semarang juga memiliki jumlah keluhan yang tinggi jika dibandingkan dengan Witel lainnya di provinsi Jawa Tengah, serta ditinjau berdasarkan *Word cloud* dari 4 *cluster* yang telah terbentuk, keluhan utama pelanggan IndiHome Regional 4 yakni keluhan tentang biaya berlangganan terhadap pemakaian WiFi, gangguan internet, pelayanan *Customer Service* (*admin*) dan internet lambat.

Keluhan biaya berlangganan dalam *cluster* 1 dan 3 terkait dengan ketidakpuasan pelanggan terhadap penggunaan WiFi yang dimiliki. *Cluster* 2 mengenai keluhan gangguan internet ditandai dengan indikator "los" pada router WiFi, sedangkan *cluster* 3 dan 4 masing-masing seperti keluhan koneksi internet tidak stabil dan kecepatan internet yang lambat.

DAFTAR PUSTAKA

Alwi, W., dan Hasrul, M. 2018. Analisis Klaster untuk Pengelompokkan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Matematika Dan Statistika Serta Aplikasinya* (*Jurnal MSA*), Vol. 6, No. 1, Hal: 35-42.

Aulia, G. P., Widiharih, T., dan Utami, I. T. 2023. Penerapan Text Mining dan Fuzzy C-Means Clustering untuk Identifikasi Keluhan Utama Pelanggan PDAM Tirta Moedal Kota Semarang. *Jurnal Gaussian*, Vol. 12, No. 1, Hal: 126-135.

Boudouda, H., Seridi, H., dan Akdag, H. 2005. The Fuzzy Possibilistic C-Means Classifier. *Asian Journal of Information Technology,* Vol. 4, No. 11, Hal: 981-985.

Feldman, R. dan Sanger, J. 2007. *The Text Mining Handbook: Advances Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.

IndiHome. 2023. *Apa itu IndiHome?*. Tersedia: https://www.indihome.co.id/about-indihome. (diakses pada tanggal 28 Februari 2023).

Ivosights. 2023. *Apa itu Word Cloud? Kenali Fungsinya untuk Aplikasi Digital Monitoring*. Tersedia:https://ivosights.com/read/artikel/word-cloud-apa-itu-kenali-fungsinya-untuk-aplikasi-digital-monitoring (diakses pada tanggal 28 Februari 2023).

Klawonn, F. dan Höppner, F. 2003. What is Fuzzy about Fuzzy Clustering? Understanding and Improving the Concept of the Fuzzier. *5th International Symposium on Intelligent Data Analysis,* Berlin: 28-30 Agustus 2003. https://www.researchgate.net/publication/221460808_What_Is_Fuzzy_about_Fuzzy_Clustering_Understanding_and_Improving_the_Concept_of_the_Fuzzifier DOI:10.1007/978-3-540-45231-7_24

Kusumadewi, S. dan Purnomo, H. 2013. *Aplikasi Logika Fuzzy: Untuk Pendukung Keputusan*. Edisi ke-2. Yogyakarta: Graha Ilmu.

Muchsin, A.K., dan Sudarma, M. 2015. Penerapan Fuzzy C-Means Untuk Penentuan Besar Uang Kuliah Tunggal Mahasiswa Baru. *Lontar Komputer*, Vol. 6, No. 3, Hal: 175-184.

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.

Putri, G.N.S, Ispriyanti, D., dan Widiharih, T. 2022. Implementasi Algoritma Fuzzy C-Means dan Fuzzy Possibilistics C-Means untuk Klasterisasi Data Tweets pada Akun Twitter Tokopedia. *Jurnal Gaussian,* Vol. 11, No. 1, Hal: 86-98.

Suganya, R., dan Shanthi, R. 2012. Fuzzy C-Means Algorithm- A Review. *International Journal of Scientific and Research Publications (IJSRP)*, Vol 2, No. 11, Hal: 1-3.

Talakua, M. W., Leleury, Z. A., dan Talluta, A.W. 2017. Analisis *Cluster* dengan Menggunakan Metode K-Means untuk Pengelompokkan Kabupaten/Kota di Provinsi Maluku berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *Jurnal Ilmu Matematika dan Terapan*, Vol. 11, No. 2, Hal: 119-128.

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.

Zhou, Y., Martin, T., dan Zhang, C. 2008. A Validity Index for Fuzzy and Possibilistic C-means Algorithm. *Proceedings of IPMU'08*, Malaga: 22-27 Juni 2008.