

CLUSTERING DATA CUTI SAKIT MENGGUNAKAN ALGORITMA AFFINITY PROPAGATION (STUDI KASUS: PERUSAHAAN TELEKOMUNIKASI DI JAKARTA)

¹Tika Purnama Putri, ²Febriani

^{1,2} Magister Manajemen Sistem Informasi Universitas Gunadarma
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat lamat institusi
¹tikapurnamaputri@gmail.com, ²febriani@staff.gunadarma.ac.id

Abstrak

Konsep data mining dapat diimplementasikan pada berbagai data, salah satunya adalah data cuti karyawan. Penerapan data mining pada data cuti sakit karyawan berfungsi untuk menghasilkan analisa pola penyakit sehingga dapat dijadikan referensi untuk tindakan preventif dan rehabilitatif terkait cuti sakit dan kesehatan karyawan. Penelitian ini menggunakan teknik Clustering dengan Algoritma Affinity Propagation pada data cuti sakit Perusahaan Telekomunikasi dengan atribut berupa, departemen karyawan, usia dan penyakit. Tahapan penelitian terdiri dari pengumpulan data, pre-processing, implementasi Affinity Propagation, pembuatan laporan dan visualisasi. Hasil percobaan menunjukkan bahwa Algoritma Affinity Propagation dipengaruhi oleh preferences, damping factor dan iterasi. Hasil percobaan juga menunjukkan bahwa untuk menghasilkan kualitas cluster terbaik, penelitian ini menggunakan nilai preference median, damping factor 0,9 dan iterasi 100 kali, sehingga menghasilkan silhouette coefficient sebesar 0,426, running time 0,56 detik dan menghasilkan 2 cluster. Cluster pertama didominasi oleh departemen Cellular Operation, Permit Relation dan IT dengan dominasi penyakit Gastritis, Infeksi Saluran Pernafasan bagian atas (Influenza, Tonsilitis, Pharyngitis, Laryngitis, Sinusitis, Rhinitis) serta Diare dan Disentri dengan usia berkisar 23 hingga 64 tahun. Cluster kedua didominasi oleh departemen NOC, Cellular Operation dan Human Resources dengan dominasi penyakit Infeksi Saluran Pernafasan lain (Bronchitis, Pleuritis, Pneumonia, Asma dan lainnya), Dermatitis Alergi dan Pulpitis Gingivitis dengan range usia berkisar 24 hingga 49 tahun.

Kata Kunci: Affinity Propagation, Analisa data cuti sakit, Clustering

Abstract

The concept of data mining can be implemented on various data, such as employee leave data. The application of data mining in employee sick leave data serves to produce an analysis of disease data patterns that can be used as a reference to take preventive and rehabilitative measures related to sick leave and employee health. The study used clustering techniques with affinity propagation algorithms on telecommunications company sick leave data with attributes in the form of, employee department, age and employee disease. The research phase consists of collection of sick leave data, pre-processing, affinity propagation implementation, generate analysis report and visualization. The results showed that clustering of affinity propagation algorithms is influenced by preferences, damping factors and iterations. To produce the best cluster quality, the study used a quartile preference value of 3, a damping factor of 0.5 and an iteration of 100 times, resulting in a silhouette coefficient value of 0.426, a running time of 0.56 seconds and resulting in 2 clusters. The first cluster has 632 itemset dominated by the departments of Cellular Operation, Permit Relation and IT with the dominance of Gastritis, Throat disease, upper Respiratory Tract Infections including Influenza, Tonsillitis, Pharyngitis, Laryngitis, Sinusitis, Rhinitis and Diarrhea and Dysentery with an age range ranging from 23 to 64 years. The second cluster has 56 itemset dominated

by the NOC, Cellular Operations and Human Resources departments with the dominance of other Respiratory Tract Infections including Bronchitis, Pleurisy, Pneumonia, Asthma and others, Allergic Dermatitis and Gingivitis Pulpitis with an age range ranging from 24 to 49 years.

Keywords: *Affinity Propagation, Analysis of sick leave data, Clustering*

PENDAHULUAN

Pada era industri 4.0 beberapa teknologi berkembang pesat, diantaranya *data science*, *artificial intelligent* (AI), *internet of think* (IoT) serta *big data*. Pemanfaatan *big data* dapat memberikan beragam manfaat bagi suatu perusahaan. Pengolahan *big data* untuk menghasilkan suatu informasi dapat menggunakan teknik yang disebut *data mining*. *Data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar [1]. Konsep *data mining* pada perusahaan dapat diimplementasikan pada berbagai data dan salah satunya adalah data cuti karyawan.

Penerapan *data mining* pada data cuti sakit karyawan ini dapat menjadi solusi bagi perusahaan untuk menghasilkan analisa pola data penyakit yang menjadi referensi dan menghasilkan informasi secara maksimal, sehingga dapat dilakukan tindakan preventif dan rehabilitatif terkait cuti sakit dan kesehatan karyawan. Pengimplementasian *data mining* untuk data cuti sakit dapat menerapkan Teknik *Clustering* dengan algoritma *Affinity Propagation*.

Clustering adalah pengelompokan data ke dalam cluster (*group*) sehingga setiap *cluster* akan berisi data yang semirip mungkin

dan berbeda dengan objek dalam *cluster* yang lainnya [2]. Sedangkan Algoritma *Affinity Propagation* adalah algoritma *clustering* dengan penerapan algoritma *message-passing* dimana *clustering* berdasarkan pesan yang dikirim dan diterima di antara setiap titik data. Setiap titik data akan mengirim pesan ke titik data lain tentang seberapa baik titik data menjadi pusat *cluster* (Eksemplar) dan seberapa baik titik data untuk menjadi anggota *cluster* [3].

Penelitian sebelumnya terkait *clustering* dengan algoritma *Affinity Propagation* adalah penelitian terhadap karakteristik kekerasan seksual terhadap anak. Penelitian ini menggunakan parameter jenis kekerasan, *range* usia dan faktor pemicu [4]. Hasil dari penelitian tersebut memperoleh tiga *cluster*. *Cluster* yang mendominasi adalah *cluster* pertama dengan kriteria jenis kekerasan seksual pencabulan pada anak di range usia tiga sampai dengan enam belas tahun dengan faktor pemicu terbesar adalah kesempatan.

Penelitian lainnya adalah melakukan *clustering* kemampuan kognitif penderita penyakit Parkinson berdasarkan skor *Mini Mental State Examination* [5]. Hasil dari penelitian tersebut menghasilkan empat kelompok pasien di mana dua kelompok (68 dan 30 peserta) menunjukkan kemampuan kognitif yang hampir baik, satu kelompok

memiliki gangguan kognitif sedang (34 peserta), dan kelompok terakhir memiliki defisit kognitif yang lebih tinggi (8 peserta).

Kedua penelitian diatas menunjukkan bahwa penerapan *data mining* khususnya teknik *clustering* dapat menunjukkan pola suatu data sehingga dapat memberikan informasi yang lebih maksimal. Namun, masih banyak perusahaan yang belum menerapkan *data mining* dalam pengolahan data. Salah satunya adalah pengolahan data cuti sakit pada Perusahaan Telekomunikasi di Jakarta. Data cuti sakit pada perusahaan ini masih dianalisa secara manual menggunakan *Microsoft Excel*. Informasi yang diperoleh dari pengelolaan data tersebut hanya terbatas untuk melihat jumlah penyakit terbanyak yang menyebabkan cuti serta departemen terbanyak yang mengajukan cuti. Hal ini dikarenakan Perusahaan Telekomunikasi ini masih belum melakukan teknik *data mining* secara maksimal.

Berdasarkan uraian masalah di atas, penelitian ini dilakukan dengan mengimplementasi-kan teknik *data mining* untuk pengelompokan cuti sakit dengan data berskala besar dengan parameter departemen karyawan, rentang usia dan tipe penyakit agar dapat menemukan suatu pola kelompok yang dibutuhkan oleh Perusahaan Telekomunikasi tersebut. Pemilihan *clustering* menggunakan tiga parameter tersebut dikarenakan permintaan Perusahaan Telekomunikasi. Hasil analisis data cuti sakit dapat menjadi referensi kebijakan

Perusahaan Telekomunikasi dalam melakukan pencegahan dan perhatian tertentu terhadap suatu penyakit untuk meminimalisir cuti sakit.

METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan yaitu, mengumpulkan data cuti sakit, melakukan pre-processing, implementasi *Affinity Propagation*, generate analysis report dan visualisasi. Tahapan metode penelitian tersebut dapat dilihat melalui bagan pada Gambar 1.

Tahap pertama dalam metode penelitian ini adalah pengumpulan data. Tahap ini adalah tahap penarikan data cuti sakit dalam database menggunakan query dan disimpan dalam file .csv. Sumber data yang digunakan dalam penelitian ini adalah data cuti sakit pada Perusahaan Telekomunikasi. *Dataset* mentah sebelum dipindahkan ke data target memiliki atribut seperti pada Tabel 1.

Pada data target, periode yang digunakan dimulai dari Januari 2020 hingga Desember 2021 dan yang status pengajuan cuti sakitnya sudah disetujui. Atribut yang dipilih untuk data target adalah Dept, Date_birth dan Kategori Penyakit. Total data dalam jangka waktu tersebut untuk menjadi data target sebanyak 707 rows. Diambil 5 row data untuk contoh data target yang digunakan dalam penelitian ini dan dapat dilihat pada Tabel 2.

Tabel 1. Atribut data mentah cuti sakit

Nama Atribut	Keterangan
NIK	Atribut ini berisikan nomor induk kepegawaian yang mengajukan cuti sakit
Name	Atribut ini berisikan nama karyawan yang mengajukan cuti sakit
Dept	Atribut ini berisikan nama departemen dari karyawan yang mengajukan cuti sakit
Date_birth	Atribut ini berisikan tanggal lahir dari karyawan yang mengajukan cuti sakit
Status	Atribut ini berisikan status dari cuti sakit yang diajukan
Date_start	Atribut ini berisikan tanggal dimulainya cuti sakit
Date_end	Atribut ini berisikan tanggal berakhirnya cuti sakit
Kategori_Penyakit	Atribut ini berisi nama penyakit yang diajukan. Daftar nama penyakit ini telah di definisikan oleh HR berdasarkan undang-undang dan disesuaikan kembali sesuai peraturan perusahaan

Tabel 2. Contoh Data Target

Dept	Date_Birth	Kategori_Penyakit
Human Resources	11/07/1992	Infeksi saluran pernafasan bagian atas termasuk influenza, Tonsilitis, Pharingitis, Laringitis, Sinusitis, Rhinitis dan lain- lain.
Network Management	06/07/1989	Conjunctivitis, Keratitis, Skleritis. Katarak
NOC	01/08/1980	Diare, Disentri (Peradangan usus)
Business Support	15/07/1993	Dyspepsia (rasa tidak nyaman pada perut)
Data Center & Analysis	13/09/1992	Typus Abdominals, paratyphus

Tabel 3. Contoh Master data Dept

Deskripsi	ID DEPT
HUMAN RESOURCES	1
NETWORK MANAGEMENT	2
NOC	3
BUSINESS SUPPORT	4
DATA CENTER & ANALYSIS	5
PERMIT RELATION	6
RETAIL SALES	7

Selanjutnya, dilakukan pre-processing pada file tersebut yang meliputi, seleksi data, membuang duplikasi data, memeriksa data yang inkonsisten, memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi) serta melakukan transformasi data. Sebelum melakukan transformasi pada atribut Dept, dilakukan pembuatan master data Dept dengan cara pembersihan duplikasi data pada

variabel atribut ini dan di definisikan angka untuk menggantikan string pada data. Master data Dept dapat dilihat pada Tabel 3.

Selanjutnya melakukan perubahan tipe variabel pada atribut Dept yang semula string dengan ID_Dept pada Tabel 3 menjadi acuan dalam transformasi ini. Transformasi atribut Dept dapat dilihat pada Tabel 4.

Tabel 4. Transformasi Atribut Dept

Dept	Date_Birth	Kategori_Penyakit
1	11/07/1992	Infeksi saluran pernafasan bagian atas termasuk influenza, Tonsilitis, Pharingitis, Laringitis, Sinusitis, Rhinitis dan lain- lain.
2	06/07/1989	Conjungtivitis, Keratitis, Skleritis. Katarak
3	01/08/1980	Diare, Disentri (Peradangan usus)
4	15/07/1993	Dyspepsia (rasa tidak nyaman pada perut)
5	13/09/1992	Typus Abdominals, paratyphus

Tabel 5. Transformasi Atribut Date_Birth menjadi Atribut Age

Dept	Age	Kategori_Penyakit
1	30	Infeksi saluran pernafasan bagian atas termasuk influenza, Tonsilitis, Pharingitis, Laringitis, Sinusitis, Rhinitis dan lain- lain.
2	33	Conjungtivitis, Keratitis, Skleritis. Katarak
3	42	Diare, Disentri (Peradangan usus)
4	29	Dyspepsia (rasa tidak nyaman pada perut)
5	30	Typus Abdominals, paratyphus

Tabel 6. Contoh Master Kategori_Penyakit

Deskripsi	ID_Kategori_Penyakit
Infeksi saluran pernafasan bagian atas termasuk influenza, Tonsilitis, Pharingitis, Laringitis, Sinusitis, Rhinitis dan lain- lain.	1
Conjungtivitis, Keratitis, Skleritis. Katarak	2
Diare, Disentri (Peradangan usus)	3
Dyspepsia (rasa tidak nyaman pada perut)	4
Typus Abdominals, paratyphus	5

Atribut Date_birth akan dilakukan transformasi menjadi umur. Sehingga akan dilakukan pengurangan tahun saat ini, yaitu 2022 dengan tahun lahir pada atribut ini untuk menghasilkan atribut umur dengan bilangan bulat. Hasil transformasi atribut Dept menjadi Age dapat dilihat pada Tabel 5.

Atribut Kategori_Penyakit yang bertipe string akan dilakukan transformasi menjadi angka. Namun sebelumnya perlu

dilakukan pembuatan master Kategori_Penyakit dengan cara pembersihan duplikasi data pada variabel atribut ini dan di definisikan angka untuk dijadikan acuan transformasi Atribut Kategori_Penyakit. Mater Kategori_Penyakit dan dapat dilihat dari Tabel 6. Sedangkan hasil transformasi Atribut Atribut Kategori_Penyakit dapat dilihat pada Tabel 7. Hasil akhir transformasi ini akan digunakan ke tahap berikutnya, yaitu melakukan Normalisasi Data.

Tabel 7. Contoh Data Transformasi Akhir

Dept	Age	Kategori_Penyakit
1	30	1
2	33	2
3	42	3
4	29	4
5	30	5

Tabel 8. Contoh Normalisasi Data

Dept	Age	Kategori_Penyakit
-1,41421	-0,58435	-1,41421
-0,70711	0,041739	-0,70711
0	1,920003	0
0,707107	-0,79304	0,707107
1,414214	-0,58435	1,414214

Selanjutnya adalah normalisasi data. Normalisasi data dilakukan dengan mencari *Z-score*. *Z-score* sering disebut dengan nilai baku atau nilai standar. *Z-score* digunakan untuk membantu menentukan apakah sebuah

data bernilai ekstrem, atau outlier [6]. Perhitungan *Z-score* dilakukan dengan mengurangi variabel dengan rata-rata masing-masing kolom dan dibagi dengan standar deviasi masing-masing kolom sesuai dengan Persamaan (1).

$$z\ score = \frac{x - \mu}{\sigma} \quad (1)$$

Berdasarkan perhitungan *z-score* pada Persamaan (1) maka hasil dari normalisasi data dapat dilihat pada Tabel 8.

Data yang telah di normalisasi kemudian diimplementasikan Algoritma Affinity Propagation. Implementasi ini di-

lakukan menggunakan bahasa pemrograman Python beserta library-nya seperti Sklearn, Scipy, Pandas, Numpy, Matplotlib dan Xlsxwriter. Program pertama akan menghitung similarity menggunakan rumus *Negative Square Euclidean Distance* seperti Persamaan (2).

$$S_{(i,j)} = - \sum_{k=1}^n (x_{ik} - x_{jk})^2 \quad (2)$$

Matriks data target yang semula berdimensi 707 x 3 maka diubah menjadi matriks *similarity* yang berdimensi 707 x 707. Contoh perhitungan *similarity* pada Tabel 8

untuk elemen data pada baris 1 dan elemen data pada baris 2 sehingga menghasilkan elemen data (1,2) pada Tabel 9 dapat dilihat sebagai berikut:

$$S_{(1,2)} = \sum_{k=1}^3 (x_{1,k} - x_{2k})^2$$

$$S_{(1,2)} = -((-1,414 - (-0,707))^2 + (-0,584 - 0,041)^2 + (-1,4142 - (-0,707))^2)$$

$$S_{(1,2)} = -((-0,70710678)^2 + (-0,6260879)^2 + (-0,70710678)^2)$$

$$S_{(1,2)} = -(0,499999998321968 + 0,39198605852641 + 0,499999998321968)$$

$$S_{(1,2)} = -(1,39198605517035)$$

Tabel 9. Contoh Matriks *Similarity*

0	-1,3919861	-10,272	-9,044	-16
-1,392	0	-4,528	-4,6969	-9,392
-10,272	-4,528	0	-8,3606	-10,272
-9,0436	-4,6968641	-8,3606	0	-1,044
-16	-9,3919861	-10,272	-1,044	0

Contoh hasil perhitungan data target yang telah di normalisasi menjadi matriks *similarity* dapat dilihat pada Tabel 9.

Algoritma *Affinity Propagation* sangat bergantung pada nilai *preferences*, *damping factor* dan iterasi untuk menghasilkan nilai *Silhouette Coefficient* yang yang baik. Nilai *Silhouette Coefficient* berada pada *range* -1 hingga 1. Nilai *Silhouette Coefficient* yang mendekati 1 maka semakin baik pula kualitas *cluster* yang dihasilkan. Oleh karena itu, sehingga pada penelitian ini dilakukan beberapa kali percobaan. Uji coba dilakukan dengan mengkombinasikan parameter *preference*, *damping* dan *iterasi*.

Pemilihan nilai *preference* akan mempengaruhi banyaknya eksemplar walau pun hal ini juga dipengaruhi oleh metode pertukaran pesan. Nilai *preference* yang bisa digunakan adalah nilai dari median matriks *similarity* untuk mendapatkan *cluster* berjumlah sedang, atau nilai dari minimum matriks *similarity* untuk mendapatkan *cluster*

berjumlah kecil [7]. Oleh karena itu, uji hasil untuk *Affinity Propagation* dengan penentuan nilai parameter dilakukan sebanyak dua kali untuk uji coba *preference* dengan nilai *preference* minimum, median.

Pemilihan nilai *damping* berfungsi untuk mempertahankan nilai bobot dan mencegah osilasi numerik saat setiap titik saling memperbarui pesan. Jangkauan nilai *damping* berada dari 0,5 hingga kurang dari 1 [7]. Oleh karena itu, dilakukan lima kali uji coba untuk *damping* dengan nilai *damping* 0.5, 0.6, 0.7, 0.8, 0.9.

Pemilihan banyak iterasi akan mempengaruhi perulangan pada prosedur pertukaran tiap titik serta kecepatan dalam proses *clustering*. Sehingga dilakukan empat kali untuk uji coba iterasi dengan nilai iterasi 100, 200, 300 dan 400. Jika dikombinasikan, percobaan untuk *Affinity Propagation* dengan penentuan nilai parameter dilakukan sebanyak empat puluh kali.

Tabel 9 memiliki nilai diagonal 0, nilai diagonal tersebut harus diganti dengan nilai

Tabel 10. Hasil Matriks Similarity dengan preference median

-4,6969	-1,3919861	-10,272	-9,044	-16
-1,392	-4,69686	-4,528	-4,6969	-9,392
-10,272	-4,528	-4,6969	-8,3606	-10,272
-9,0436	-4,6968641	-8,3606	-4,6969	-1,044
-16	-9,3919861	-10,272	-1,044	-4,6969

preference. Nilai preference yang dapat digunakan merupakan nilai minimum yang berasal dari kuartil 1 atau nilai median dari matriks similarity yang terbentuk. Nilai minimum preference dari matriks tersebut yaitu -9.39198 dan nilai preference median yaitu -4.69686. Sehingga nilai diagonal yang semula 0 perlu diganti dengan salah satu preference tersebut. Table 10 merupakan contoh nilai diagonal dari matriks similarity yang telah diganti menggunakan nilai median preference

Algoritma Affinity Propagation melakukan message-passing dimana clustering berdasarkan pesan yang dikirim dan di-

terima di antara setiap titik data. Message-passing berupa nilai responsibility dan availability dengan data lain dan dilakukan iterasi hingga konvergen. Pada awalnya availability diinisialisasi dengan 0. Sehingga perhitungan dapat dilanjutkan untuk mencari nilai responsibility. Perhitungan responsibility untuk tahap awal dapat menggunakan Persamaan (3) dan ketika memperbaharui pesan perhitungan responsibility diperkecil dengan sebuah faktor pengecil (damping factor) untuk menghindari hasil yang bergelombang atau tidak konvergen seperti pada Persamaan (4) [8].

$$r(i, k) \leftarrow s(i, k) - \max_{k' \text{ s.t. } k' \neq k} \{a(i, k') + s(i, k')\} \quad (3)$$

$$r(i, k)_t = \lambda r(i, k)_{t-1} + (1 - \lambda)r(i, k) \quad (4)$$

Keterangan :

s(i,k): Nilai similarity antara data i dengan data k

a(i,k'): Nilai availability antara data i dengan data k', di mana k'≠k

s(i,k'): Nilai similarity antara data i dengan data k', di mana k'≠k

Perhitungan responsibility dari Tabel 10 untuk elemen data (1,2) dapat dilihat sebagai berikut:

$$r(1,2) \leftarrow s(1,2) - \max_{k' \text{ s.t. } k' \neq k} \{a(1, 2') + s(1, 2')\}$$

$$r(1,2) = (-1,39198606) - (0 + (-4,69686))$$

$$r(1,2) = 3,30487394$$

Tabel 11. Contoh hasil perhitungan *Responsibility*

-3,30487	3,30487394	-3,30487	-3,30487	-3,30487
3,136014	-3,30487394	-3,13601	-3,30487	-8,00001
-5,74378	0,16886	-0,16886	-3,83263	-5,744
-7,99955	-3,65286411	-7,31663	-3,65286	3,65286
-14,956	-8,34798606	-9,228	3,65286	-3,65286

Contoh hasil dari perhitungan persamaan (5) untuk elemen data non-*responsibility* dapat dilihat pada Tabel 11.

Tahap berikutnya adalah tahap perhitungan *availability* baru menggunakan persamaan (6) untuk elemen data diagonal dan persamaan (7) untuk memperbaharui *availability* [8].

$$a(i, k) \leftarrow \min \left\{ 0, r(k, k) + \sum_{i' \text{ s.t. } i' \neq \{i, k\}} \max\{0, r(i', k)\} \right\} \quad (5)$$

$$a(k, k) \leftarrow \sum_{i' \text{ s.t. } i' \neq \{i, k\}} \max\{0, r(i', k)\} \quad (6)$$

$$a(i, k)_t = \lambda a(i, k)_{t-1} + (1 - \lambda)a(i, k) \quad (7)$$

Keterangan:

λ : Nilai *damping factor*, antara 0 sampai 1

$a(k, k)$: Nilai *availability* diagonal

$a(i, k)$: Nilai *availability* sekarang antara data i dengan data k

$a(i, k)_{t-1}$: Nilai *availability* iterasi sebelum antara data i dengan data k

$r(i, k)$: Nilai *responsibility* sekarang antara data i dengan data k

$r(i, k)_{t-1}$: Nilai *responsibility* iterasi sebelum antara data i dengan data k

Perhitungan *availability* baru dengan menggunakan *damping* dengan nilai 0.9 menggunakan nilai *responsibility* dengan elemen data (2,2) pada Tabel 11 dan nilai *responsibility* maksimal dari kolom 2 pada Tabel 11 dan menghasilkan matriks *availability* dengan elemen data (1,2) pada Tabel 12 dapat dilihat sebagai berikut:

$$a(1,2) \leftarrow \min \left\{ 0, r(2,2) + \sum_{i' \text{ s.t. } i' \neq \{1,2\}} \max\{0, r(i', 2)\} \right\}$$

$$a(1,2) \leftarrow -3,30487394 + (0,16886)$$

$$a(1,2) \leftarrow -3,13601394$$

$$a(1,2)_2 = \lambda a(1,2)_1 + (1 - \lambda)a(1,2)$$

$$a(1,2)_2 = 0,9 * 0 + (1 - 0,9)(-3,13601394)$$

$$a(1,2)_2 = -0,31360139$$

Tabel 12. Contoh hasil perhitungan *Availability*

0,313601	-0,31360139	-0,01689	0	0
-0,33049	0,347373394	-0,01689	0	0
-0,01689	0	0	0	0
-0,01689	0	-0,01689	0,365286	-0,36529
-0,01689	0	-0,01689	-0,36529	0,365286

Contoh hasil dari perhitungan *availability* baru dengan menggunakan damping dengan nilai 0.9 dapat dilihat pada Tabel 12.

Tahap selanjutnya adalah menghitung matriks *criterion* menggunakan Persamaan

$$criterion(i,k) = r(i,k) + a(i,k) \quad (8)$$

Contoh perhitungan matriks *criterion* menggunakan elemen data (1,2) pada Tabel 11 dan elemen data (1,2) pada Tabel 12 untuk

$$\begin{aligned} criterion(1,2) &= r(1,2) + a(1,2) \\ criterion(1,2) &= 3,30487394 + (-0,31360139) \\ criterion(1,2) &= 2,991272546 \end{aligned}$$

Contoh hasil dari perhitungan matriks *criterion* dapat dilihat pada Tabel 13.

Jadi, pada Tabel 13 dapat diketahui bahwa baris 1,2,4 dan 5 adalah nilai *criterion* positif yang artinya pada iterasi pertama baris tersebut adalah *exemplar*. Serta nilai yang sama antara baris 4 dan 5 menandakan bahwa baris ini berada dalam cluster yang sama.

(8). Lalu penentuan exemplar dilakukan menggunakan nilai *criterion* tertinggi dari setiap baris ditetapkan sebagai exemplar. Baris yang memiliki exemplar yang sama berada di cluster yang sama [8].

menghasilkan *criterion* dengan elemen data (1,2) pada Tabel 13 dapat dilihat sebagai berikut:

Tahap berikutnya adalah validasi apakah data tersebut sudah konvergen atau tidak. Jika data belum konvergen, maka dilakukan kembali perhitungan *availability*, penjumlahan antara matriks *availability* dengan *similarity*, perhitungan *responsibility*, perhitungan *availability* baru dan program akan melakukan validasi kembali hingga data konvergen atau sesuai dengan iterasi.

Tabel 13. Contoh hasil perhitungan *Criterion*

-2,99127	2,991272546	-3,32176	-3,30487	-3,30487
2,805527	-2,95750055	-3,1529	-3,30487	-8,00001
-5,76066	-0,16886	-0,16886	-3,83263	-5,744
-8,01644	-3,65286411	-7,33352	-3,28757	3,287574
-14,9729	-8,34798606	-9,24489	3,287574	-3,28757

HASIL DAN PEMBAHASAN

Uji coba nilai *preferences*, *damping factor* dan iterasi untuk menghasilkan nilai *Silhouette Coefficient* yang tinggi dilakukan sebanyak empat puluh kali. Berdasarkan uji coba tersebut, dihasilkan 4 uji coba yang tidak berhasil mendapat *cluster*. Hal ini disebabkan karena matriks belum konvergen sedangkan batas iterasi sudah mencapai

maksimal. Hasil uji coba juga menunjukkan bahwa implementasi Algoritma *Affinity Propagation* paling baik menggunakan percobaan ke-37 dengan *preference median*, *damping factor* 0,9 dan iterasi sebanyak 100 kali, sehingga menghasilkan sebanyak 2 *cluster* dengan nilai *silhouette coefficient* sebesar 0,426 dan *running time* 0,56 detik. Detail dari uji coba dapat dilihat pada Tabel 14.

Tabel 14. Tabel Uji Coba Implementasi Algoritma *Affinity Propagation*

Uji Coba Ke-	Parameter			Hasil		
	Preference	Damping	Iterasi	Banyak Cluster	Silhouette Coefficient	Running Time
1	Minimum	0,5	100	26	0,295	2,464
2	Minimum	0,5	200	26	0,295	2,465
3	Minimum	0,5	300	26	0,295	2,467
4	Minimum	0,5	400	26	0,295	2,470
5	Minimum	0,6	100	26	0,303	1,766
6	Minimum	0,6	200	26	0,304	1,936
7	Minimum	0,6	300	26	0,303	2,286
8	Minimum	0,6	400	26	0,303	2,354
9	Minimum	0,7	100	27	0,319	1,836
10	Minimum	0,7	200	27	0,319	1,857
11	Minimum	0,7	300	27	0,319	1,860
12	Minimum	0,7	400	27	0,319	1,985
13	Minimum	0,8	100	27	0,316	3,309
14	Minimum	0,8	200	27	0,316	3,387
15	Minimum	0,8	300	27	0,316	3,390
16	Minimum	0,8	400	27	0,316	3,811
17	Minimum	0,9	100	27	0,306	4,500
18	Minimum	0,9	200	27	0,305	4,645
19	Minimum	0,9	300	27	0,309	4,799
20	Minimum	0,9	400	27	0,309	4,809
21	Median	0,5	100	-	-	-
22	Median	0,5	200	-	-	-
23	Median	0,5	300	-	-	-
24	Median	0,5	400	-	-	-
25	Median	0,6	100	58	0,346	3,501
26	Median	0,6	200	59	0,346	3,576
27	Median	0,6	300	59	0,349	3,587

Uji Coba Ke-	Parameter			Hasil		
	Preference	Damping	Iterasi	Banyak Cluster	Silhouette Coefficient	Running Time
28	Median	0,6	400	60	0,346	3,737
29	Median	0,7	100	58	0,351	2,625
30	Median	0,7	200	58	0,351	2,699
31	Median	0,7	300	58	0,351	2,733
32	Median	0,7	400	58	0,351	2,911
33	Median	0,8	100	57	0,356	4,022
34	Median	0,8	200	57	0,355	4,051
35	Median	0,8	300	57	0,355	4,162
36	Median	0,8	400	59	0,355	4,293
37	Median	0,9	100	2	0,426	0,566
38	Median	0,9	200	2	0,426	0,567
39	Median	0,9	300	2	0,426	0,573
40	Median	0,9	400	2	0,426	0,597

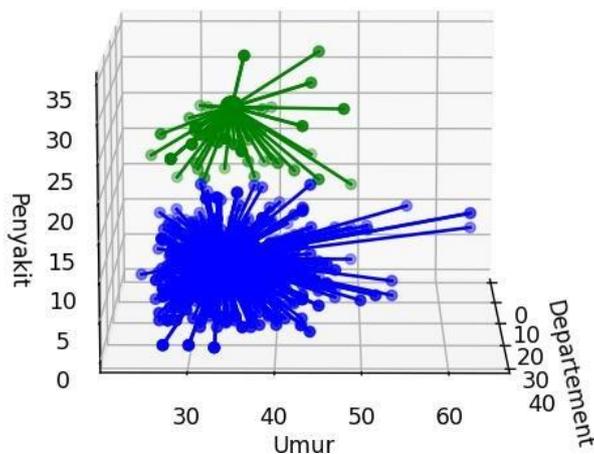
Clustering menggunakan nilai preference pertama menghasilkan 651 item set dan maksimum, damping factor bernilai 0,9 dan cluster kedua menghasilkan 56 item set. Hasil iterasi sebanyak 100 menghasilkan cluster clustering dapat dilihat pada Tabel 10 dan yang optimal sebanyak dua cluster. Cluster visualisasi nya dapat dilihat pada Gambar 1.

Tabel 15. Hasil Clustering

Cluster Ke-	Banyak Item	Rincian		
		Department (Total)	Range Usia	Penyakit (Total)
1	632	Cellular Operation (90) sebagai centeroid Permit Relation (54) IT (34) Network Management (33) Accounting (33) Project Rollout Cellular (32) Planning & Engineering (30) Business Support (27) Operation Support (26) SCM (23) Human Resources (22)	23-64	Gastritis/Tukak Lambung (200) sebagai centeroid Infeksi saluran pernafasan bagian atas termasuk influenza, Tonsilitis, Pharingitis, Laringitis, Sinusitis, Rhinitis dll. (140) Diare, Dysentri (Peradangan usus) (68) Fever (Demam) (52) Typus Abdominalis, paratyphus (45) Hipertensi (darah tinggi) (35) Dyspepsie (rasa tidak nyaman pada perut) (33) Cephalgia (sakit kepala) (22) Operasi/internis (13) Conjungtivitis, Keratitis, Skleritis. Katarak (6) Saraf kejepit (5) Stomatitis Caries (5) Kontrol kehamilan (3) Dermatitis Kontak (2) Vertigo (2) Dental (Gigi) (1)

Cluster Ke-	Banyak Item	Rincian		
		Department (Total)	Range Usia	Penyakit (Total)
		<i>Data Center & Analysis (21)</i>		
		<i>Non Cellular Operation (21)</i>		
		<i>Finance (19)</i>		
		<i>Research & Development (18)</i>		
		<i>Legal (16)</i>		
		<i>Corporate & Whosales (15)</i>		
		<i>Technical Software & Engineering (14)</i>		
		<i>PMO (14)</i>		
		<i>Internal Audit (14)</i>		
		<i>Project Rollout Non Cellular (13)</i>		
		<i>NOC (10)</i>		
		<i>Marketing Non Bali (7)</i>		
		<i>General Affair (7)</i>		
		<i>Ftx Project (5)</i>		
		<i>Marketing Bali (4)</i>		
		<i>Ftx Corporate & Whosales (4)</i>		
		<i>Network Planning (4)</i>		
		<i>Sales & Marketing (4)</i>		
		<i>Tower & Network Operation (3)</i>		
		<i>Government Sales (3)</i>		
		<i>Retail Sales (2)</i>		
		<i>Corporate Secretary (2)</i>		
		<i>Mcp Project Rollout (2)</i>		
		<i>Mcp Planning (2)</i>		
		<i>Ftx Data Center & Analysis (1)</i>		
		<i>Project & Marketing Cellular (1)</i>		
		<i>Carrier Network (1)</i>		
		<i>Business Development (1)</i>		

Cluster Ke-	Banyak Item	Rincian		
		Department (Total)	Range Usia	Penyakit (Total)
2	56	NOC (9) sebagai centeroid Cellular Operation (7) Human Resources (7) Permit Relation (4) Technical Software & Engineering (4) Finance (4) Network Management (3) IT (3) Internal Audit (3) Marketing Bali (2) Marketing Non Bali (2) Project Rollout Cellular (2) Accounting (2) Planning & Engineering (2) Business Support (1) Operation Support (1) General Affair (1) Corporate & Whosales (1)	24-49	Infeksi saluran pernafasan lain termasuk Bronchitis, Pleuritis, Pneumonia, Asma dan lain-lain. (23) sebagai centeroid Dermatitis Alergi (5) Pulpitis Gingivitis (4) Dental (Gigi) (4) Kelainan darah lainnya sebutkan (4) Dehidrasi (4) Vertigo (3) Myalgia, athralgia (3) Malaria (3) Dermatitis Kontak (2) Stomatitis Caries (1)



Gambar 1. Visualisasi Clustering

Hasil menunjukkan bahwa *cluster* pertama memiliki *item set* sebanyak 632 dan *centroid* departemen pada *cluster* ini adalah *Cellular Operation* dengan jumlah data sebanyak 90 data. *Centroid* dalam *cluster* ini untuk penyakit adalah penyakit Gastritis/Tekak Lambung sebanyak 200 data. *Range* usia dalam *cluster* berkisar 23 hingga 64 tahun.

Cluster kedua memiliki *item set* sebanyak 56 dan *centroid* departemen pada *cluster* ini adalah departemen NOC dengan jumlah data sebanyak 9 data. *Centroid* dalam *cluster* ini untuk penyakit adalah penyakit Infeksi Saluran Pernafasan lain termasuk Bronchitis, Pleuritis, Pneumonia, Asma dan lain-lain sebanyak 23 data. *Range* usia dalam *cluster* berkisar 24 hingga 49 tahun.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis yang telah dilakukan dengan metode *clustering* menggunakan Algoritma *Affinity Propagation*, maka diperoleh kesimpulan bahwa hasil *clustering* dari Algoritma *Affinity Propagation* dipengaruhi oleh parameter seperti, *preferences*, *damping factor* dan iterasi. Parameter terbaik pada penelitian ini menggunakan *preference median* (kuartil 2), *damping factor* 0,5 dan iterasi sebanyak 100 kali, sehingga menghasilkan sebanyak 2 *cluster* dengan nilai *silhouette coefficient* sebesar 0,426 dan *running time* 0,56 detik.

Hasil *clustering* data cuti sakit menghasilkan dua *cluster*. *Cluster* pertama me-

iliki *item set* sebanyak 632 dan didominasi oleh departemen *Cellular Operation*, *Permit Relation* dan *IT*. Dominasi penyakit pada *cluster* ini adalah penyakit Gastritis/Tekak Lambung, Infeksi Saluran Pernafasan bagian atas termasuk Influenza, Tonsilitis, Pharyngitis, Laryngitis, Sinusitis, Rhinitis serta Diare dan Disentri (Peradangan usus) dengan *range* usia berkisar 23 hingga 64 tahun. *Cluster* kedua memiliki *item set* sebanyak 56 dan didominasi oleh departemen NOC, *Cellular Operation* dan *Human Resources* dengan dominasi penyakit Infeksi Saluran Pernafasan lain termasuk Bronchitis, Pleuritis, Pneumonia, Asma dan lain-lain, penyakit Dermatitis Alergi dan Pulpitis Gingivitis dengan *range* usia berkisar 24 hingga 49 tahun.

Clustering data cuti sakit pada penelitian ini perlu dikembangkan lagi karena dalam mengimplementasikan Algoritma *Affinity Propagation* standar ini memiliki keterbatasan yaitu sulitnya menentukan nilai parameter *preferences*, *damping factor* dan iterasi yang sesuai agar diperoleh hasil *clustering* yang baik. Pengembangan selanjutnya dapat menggunakan pendekatan Algoritma *Affinity Propagation* seperti, *Adaptive Affinity Propagation*, *Fuzzy Statistic Affinity Propagation*, *Partition Affinity Propagation* dan lain sebagainya sehingga dapat menghasilkan *cluster* dengan kualitas yang lebih baik. Pengembangan juga dapat dilakukan dengan melakukan kombinasi menggunakan algoritma lain. Sehingga tidak hanya terbatas pada analisa namun juga dapat memberikan rekomendasi dari permasalahan yang ada.

DAFTAR PUSTAKA

- [1] Davies dan P. Beynon, Database Systems Third Edition, New York: Palgrave Macmillan, 2004.
- [2] E. Irwansyah dan M. Faisal, Advanced Clustering: Teori dan Aplikasi, Yogyakarta: DeePublish, 2015.
- [3] T. Alfina, dkk, "Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem Kerja Praktek Teknik Industri ITS)," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, pp. 1-10, 2018.
- [4] F. Wiza, "Klasterisasi karakteristik kekerasan seksual terhadap anak," *Jurnal teknologi informasi dan Komunikasi*, pp. 44-53, 2019.
- [5] A. Sarica, M. G. Vaccaro, A. Quattrone dan A. Quattrone, "Pengelompokan Kognitif Parkinsonism melalui Affinity Propagation," *Algorithms*, vol. 14, pp. 2-22, 2021.
- [6] MAG Learning, "Skor Standar (Z-Score) Untuk Screening Data Dan Regresi," in MAG, [Daring], 2019. <https://maglearning.id/2019/02/27/skor-standar-z-score-untuk-screening-data-dan-regresi/> [Diakses: 24 Desember 2021].
- [7] R. Dermawan, F. A. Bachtiar dan P. P. Adikara, "Peringkasan Teks Untuk Deteksi Kejadian Pada Dokumen Twitter Berbahasa Indonesia Dengan Metode Affinity Propagation," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, pp. 2208-2214, 2018.
- [8] J. Kang, K. Lerman and Plangprasopchok, "Analisis Microblogs menggunakan Affinity Propagation," *Social Media Analytics*, pp. 67-70, 2010.