

IMPLEMENTASI ALGORITMA K-MEANS UNTUK MENGENALI POLA JEMAAT DALAM KEGIATAN PELAYANAN GEREJA

Emerensye S. Y. Pandie¹, Tiwuk Widiastuti², Sebastianus A. S. Mola³, Bertha S. Djahi⁴

Program Studi Ilmu Komputer, Universitas Nusa Cendana, Jl. Adisucipto Kupang
Email: emerensyepandie@staf.undana.ac.id¹

INTISARI

Gereja adalah sebuah institusi yang tidak hanya mengelola kerohanian jemaatnya, namun juga merupakan wadah organisasi yang mengayomi dan melindungi jemaat serta menjamin keberlangsungan kehidupan sosial masyarakat jemaatnya. Data statistik juga menunjukkan jumlah pemeluk agama Kristen Protestan di kota Kupang mencapai angka 209.438 orang atau sebanyak 61,08% dari jumlah total penduduk di kota Kupang. Jumlah gereja Kristen Protestan sebanyak 150 gereja sehingga diasumsikan setiap gereja memiliki jemaat rata-rata 1.396 jemaat per gereja. *Data mining* merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan sehingga didapat pola tertentu. Penelitian ini akan mengimplementasikan algoritma data mining K-Means Clustering untuk mengenali pola jemaat yang menjadi salah satu target kegiatan pelayanan gereja yaitu jemaat yang belum dibaptis padahal sudah melewati usia baptis pada umumnya sebanyak. Luaran utama yaitu berupa rumusan/pola jemaat dan hasil rumusan itu akan dipakai oleh pihak gereja untuk melakukan pendekatan kepada jemaat dengan ciri/pola tersebut.

Kata kunci: *Data Mining, k-Means, Clustering, Pelayanan Gereja*

ABSTRACT

The church is an institution that not only manages the spirituality of the congregation, but also is an organization that protects and protects the congregation and guarantees the social life of the community of the congregation. Statistics also show that the number of Protestant Christians in Kupang city reached 209,438 people or 61.08% of the total population in the city of Kupang. There are 150 Protestant Christian churches, so it is assumed that each church has an average of 1,396 churches per church. Data mining is a process that employs one or more computer learning techniques (*machine learning*) to analyze and extract knowledge so that certain patterns are obtained. This research will implement the K-Means Clustering data mining algorithm to recognize the pattern of the congregation which is one of the targets of church service activities, namely unbaptized congregations even though they have passed the baptism age in general. The main output is the formulation/ pattern of the congregation and the results of the formulation will be used by the church to approach the congregation with these characteristics / patterns.

Keywords: *Data Mining, k-Means, Clustering, Church Service*

1. PENDAHULUAN

Dari data statistik jumlah pemeluk agama Kristen Protestan di kota Kupang mencapai angka 209.438 orang atau sebanyak 61,08% dari jumlah total penduduk di kota Kupang. Jumlah gereja Kristen Protestan sebanyak 150 gereja sehingga diasumsikan setiap gereja memiliki jemaat rata-rata 1.396 jemaat per gereja. Dari seluruh jemaat tersebut, pihak gereja melakukan pengelolaan data jemaat secara manual ataupun dengan bantuan *spreadsheet* komputer seperti *Microsoft Excell*. Data jemaat dikelola dengan baik oleh pihak gereja sebagai sarana untuk memudahkan pelayanan kepada jemaat antara lain diakonia, bantuan beasiswa, bantuan kematian, dan kegiatan pelayanan lainnya. Kegiatan ini dapat dilaksanakan secara optimal jika pemantauan kondisi sosial ekonomi jemaat gereja dapat dipantau dan dikelola secara baik. Salah satu kegiatan diakonia gereja adalah pelayanan terhadap jemaat yang belum dibaptis padahal sudah melewati batas usia baptis pada umumnya. Data yang digunakan sebanyak 3772 record. Jika gereja bisa mengetahui pola jemaat yang melakukan hal tersebut maka hal itu bisa dihindari atau diminimalisir. Pihak gereja juga bisa melakukan langkah antisipasi dengan melakukan pendekatan kepada jemaat untuk menghindari terjadinya kasus serupa.

Hal ini bisa dilakukan dengan menganalisa data jemaat secara keseluruhan dengan ilmu data mining, untuk menemukan pola khusus dari jemaat yang belum dibaptis. Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban dkk. 2005). Terdapat beberapa istilah lain yang memiliki makna sama dengan data mining, yaitu *Knowledge discovery in databases* (KDD), ekstraksi pengetahuan (*knowledge extraction*), Analisa data/pola (*data/pattern analysis*), kecerdasan bisnis (*business intelligence*) dan *data archaeology* dan *data dredging* (Larose, 2005).

K-means adalah salah satu algoritma *clustering* dengan tujuan membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Pembelajaran dalam K-Means ini termasuk dalam *unsupervised learning*. Masukan yang diterima adalah data atau objek dan K buah kelompok (*cluster*) yang ditentukan. Algoritma ini akan mengelompokkan data atau objek ke dalam K buah kelompok tersebut. Pada setiap *cluster* terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut. K-means ditemukan oleh beberapa orang yaitu Lloyd (1957, 1982), Forgy (1965), Friedman and Rubin (1967), and McQueen (1967). Ide dari *clustering* pertama kali ditemukan oleh Lloyd pada tahun 1957, namun hal tersebut baru dipublikasi pada tahun 1982.

2. MATERI DAN METODE

Metode yang digunakan yaitu metode data mining sebagai berikut. (a) Tahap pengumpulan data, (b) Tahap pengolahan data, (c) Tahap Clustering dan (d) Tahap Analisis.

2.1 Data Mining

Data mining sering disebut juga *Knowledge Discovery in Databases* (KDD) adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Penemuan pengetahuan tersebut bisa berupa penjelasan tentang masa lalu ataupun prediksi masa depan. *Data mining* menggunakan teknik statistika, matematika, kecerdasan buatan, dan *machine learning* untuk mendapatkan informasi dan pengetahuan yang bermanfaat dalam database. Keluaran dari *data mining* bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan.

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor (Kusrini, 2009).

- Pertumbuhan yang cepat dalam kumpulan data.
- Penyimpanan data dalam *warehouse* sehingga seluruh perusahaan memiliki akses ke dalam database.
- Adanya peningkatan akses data melalui navigasi web dan intranet.
- Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
- Perkembangan teknologi perangkat lunak untuk *data mining*
- Perkembangan dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

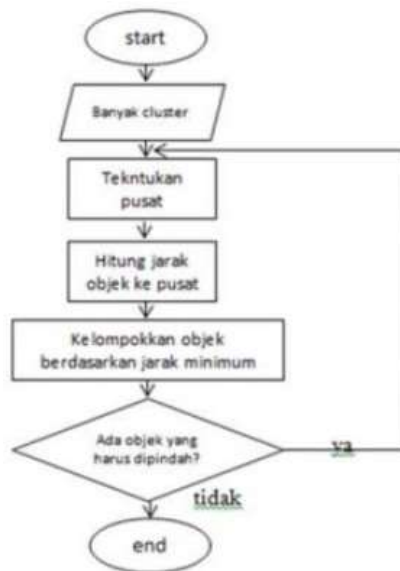
2.2 Clustering

Clustering adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode Data Mining, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu 'wilayah' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah' yang lain.

Terdapat beberapa pendekatan yang digunakan dalam mengembangkan metode clustering. Dua pendekatan utama adalah clustering dengan pendekatan partisi dan clustering dengan pendekatan hirarki. Clustering dengan pendekatan partisi mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam cluster-cluster yang ada. Clustering dengan pendekatan hirarki atau hierarchical clustering mengelompokkan data dengan membuat suatu hirarki dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan. Gambar 1 menunjukkan flowchart dalam menentukan cluster dengan K-Means (Windarto, 2017).

2.3 Algoritma K-Means

Algoritma K-means pada dasarnya melakukan 2 proses yakni proses pendeteksian lokasi pusat cluster dan proses pencarian anggota dari tiap-tiap cluster. Proses clustering dimulai dengan mengidentifikasi data yang akan dikluster, X_{ij} ($i=1, \dots, n$; $j=1, \dots, m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap cluster dite-tapkan secara bebas (sembarang), C_k ($k=1, \dots, k$; $j=1, \dots, m$). Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster (Soni dan Ganatra, 2012).



Gambar 1. Tahapan K-Means Clustering

Proses dasar algoritma k-means dapat dilihat di bawah ini:

1. Tentukan jumlah kluster yang ingin dibentuk dan tetapkan pusat cluster K.
2. Menggunakan jarak euclidean kemudian hitung setiap data ke pusat cluster

$$d(i, k) = \sqrt{\sum_i^m C_{ij} - C_{kj}^2} \dots\dots\dots (1)$$

3. Kelompokkan data ke dalam cluster dengan jarak yang paling pendek dengan persamaan

$$\min \sum_k^i - a_{ik} = \sqrt{\sum_i^m C_{ij} - C_{kj}^2} \dots\dots\dots (2)$$

4. Hitung pusat cluster yang baru menggunakan persamaan

$$C_{kj} = \frac{\sum_k^i x_{ij}}{p} \dots\dots\dots (3)$$

dengan: X_{ij} s Kluster ke k p = banyaknya anggota kluster ke - k

5. Ulangi langkah dua sampai dengan empat sehingga sudah tidak ada lagi data yang berpindah ke kluster yang lain (Sadewo dkk., 2017)

Berikut tahapan metode data mining:

A. Tahap pengumpulan data

Data jemaat sebanyak 3772 record dipilah dengan mengambil 3 variabel yaitu status baptis, umur dan jenis kelamin (gambar 1).

B. Tahap Pengolahan Data

Selanjutnya data melewati tahapan cleaning dengan merubah status baptis dan jenis kelamin (tabel 1). Selanjutnya data disimpan dengan format CSV atau *comma delimited* untuk diproses lebih lanjut (gambar 2)

Tabel 1. Konversi data

Variabel	Nilai	Konversi
Jenis Kelamin	Perempuan	0
	Laki-laki	1
Status Baptis	.F. (Belum)	0
	.T. (Sudah)	1

Halaman : 1 / 141

No.	NAMA JEMAAT	STATUS BAPTIS	UMUR JEMAAT	JENIS KELAMIN
1	EKTRINATA GILI GA	F.	26	Perempuan
2	GRACELIN ATALUPAH	F.	6	Perempuan
3	MELKIAS LEOBISA	T.	38	Laki-Laki
4	A.H. Y. KOLLOH	F.	68	Laki-Laki
5	A.MARGRET S. M. RUNDUNGAN	T.	31	Perempuan
6	A.MUTIARA LARASATI ISUKO	T.	10	Perempuan
7	ABIA DEREK FUDIKOA	T.	41	Laki-Laki
8	ABIA M. FUDIKOA	T.	27	Laki-Laki
9	ABIA P. ADAM	T.	13	Laki-Laki
10	ABIATER NABUNOME	F.	26	Laki-Laki
11	ABIMELEK BOLANG	F.	30	Laki-Laki
12	ABIMELEK NEONINE	F.	25	Laki-Laki
13	ABNER M. Y. JEMI	F.	27	Laki-Laki
14	ABNETA LIKA HANJA	T.	43	Perempuan
15	ABRAHAM CODEY	T.	30	Laki-Laki
16	ABRAHAM KOLLOH	T.	63	Laki-Laki
17	ABRAHAM LOLOMSAIT	F.	34	Laki-Laki
18	ABRAHAM OLIN	T.	53	Laki-Laki
19	ABRAHAM V. D. KOLLOH	T.	25	Laki-Laki
20	ABRAM A. ERYAH	F.	26	Laki-Laki
21	ABRISEN BERNABAS ADU	T.	31	Laki-Laki
22	ABROSIUS NAKAMNANU	T.	30	Laki-Laki
23	ABYO Y. FUDIKOA	T.	21	Laki-Laki
24	ACCO SAEFATU	T.	23	Laki-Laki
25	ADAM ELLIEK	T.	55	Laki-Laki
26	ADAM M. KOLLOH	T.	10	Laki-Laki
27	ADBEL A. ADD	F.	27	Laki-Laki

Gambar 1. Data Jemaat

A	B	C	D	E
No.	NAMA JEMAAT	status baptis	UMUR JEMAAT	JK
1	EKTRINATA GILI GA	0	26	0
2	GRACELIN ATALUPAH	0	6	0
3	MELKIAS LEOBISA	1	38	0
4	A.H. Y. KOLLOH	0	68	1
5	A.MARGRET S. M. RUNDUNGAN	0	31	1
6	A.MUTIARA LARASATI ISUKO	1	10	0
7	ABIA DEREK FUDIKOA	1	41	0
8	ABIA M. FUDIKOA	1	27	1
9	ABIA P. ADAM	1	13	1
10	ABIATER NABUNOME	1	26	1
11	ABIMELEK BOLANG	0	30	1
12	ABIMELEK NEONINE	0	25	1
13	ABNER M. Y. JEMI	0	27	1
14	ABNETA LIKA HANJA	0	43	1
15	ABRAHAM CODEY	1	30	0
16	ABRAHAM KOLLOH	1	63	1
17	ABRAHAM LOLOMSAIT	1	34	1
18	ABRAHAM OLIN	0	53	1
19	ABRAHAM V. D. KOLLOH	1	25	1
20	ABRAM A. ERYAH	1	26	1
21	ABRISEN BERNABAS ADU	0	31	1
22	ABROSIUS NAKAMNANU	1	30	1
23	ABYO Y. FUDIKOA	1	21	1
24	ACCO SAEFATU	1	23	1
25	ADAM ELLIEK	1	55	1
26	ADAM M. KOLLOH	1	10	1
27	ADBEL A. ADD	1	27	1
28	ADRIAN F.P.P. SUNBANU	0	25	1
29	ADE NUSITA DETHAN	1	26	1
30	ADRIAN F.P.P. SUNBANU	0	25	1
31	ADE NUSITA DETHAN	1	26	1

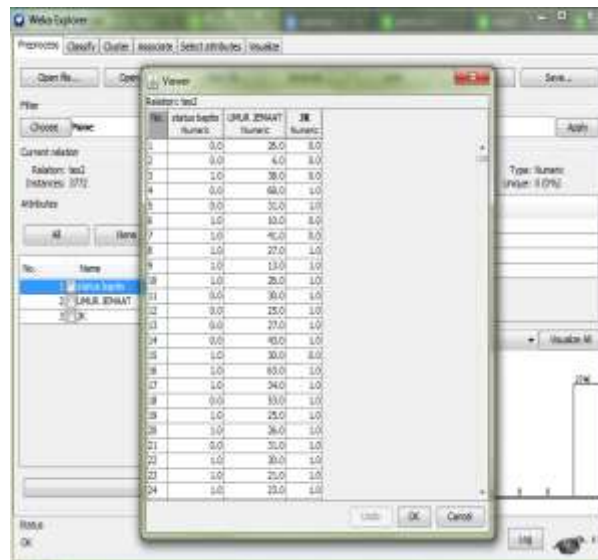
Tabel 2. Data hasil konversi

C. Tahap Clustering

Data selanjutnya diproses dengan bantuan aplikasi Weka versi 3.6.11 dengan jumlah cluster yang akan dibentuk ditentukan sejumlah 4 cluster (gambar 2).

D. Tahap Analisa

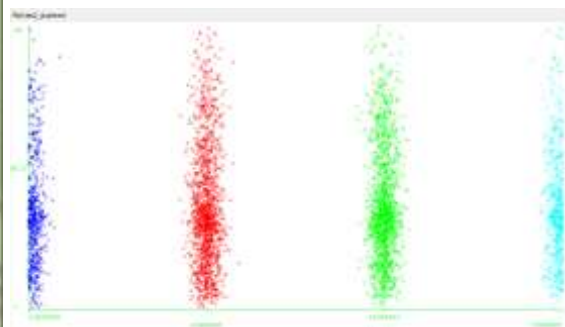
Hasil perhitungan jarak dengan menggunakan aplikasi Weka memberikan hasil seperti pada gambar 3 dan gambar 4):



Gambar 2. Data Jemaat format CSV

No.	status bapts Numeric	UMUR_JEMAAT Numeric	JK Numeric	Cluster Nominal
1	0.0	26.0	0.0	cluster3
2	0.0	6.0	0.0	cluster3
3	1.0	38.0	0.0	cluster1
4	0.0	68.0	1.0	cluster0
5	0.0	31.0	1.0	cluster0
6	1.0	10.0	0.0	cluster1
7	1.0	41.0	0.0	cluster1
8	1.0	27.0	1.0	cluster2
9	1.0	13.0	1.0	cluster2
10	1.0	26.0	1.0	cluster2
11	0.0	30.0	1.0	cluster0
12	0.0	25.0	1.0	cluster0
13	0.0	27.0	1.0	cluster0
14	0.0	43.0	1.0	cluster0
15	1.0	30.0	0.0	cluster1
16	1.0	63.0	1.0	cluster2
17	1.0	34.0	1.0	cluster2
18	0.0	53.0	1.0	cluster0
19	1.0	25.0	1.0	cluster2
20	1.0	26.0	1.0	cluster2
21	0.0	31.0	1.0	cluster0
22	1.0	30.0	1.0	cluster2
23	1.0	21.0	1.0	cluster2
24	1.0	23.0	1.0	cluster2

Gambar 3. Hasil Clustering data jemaat



Gambar 4. Sebaran Clustering berdasarkan umur

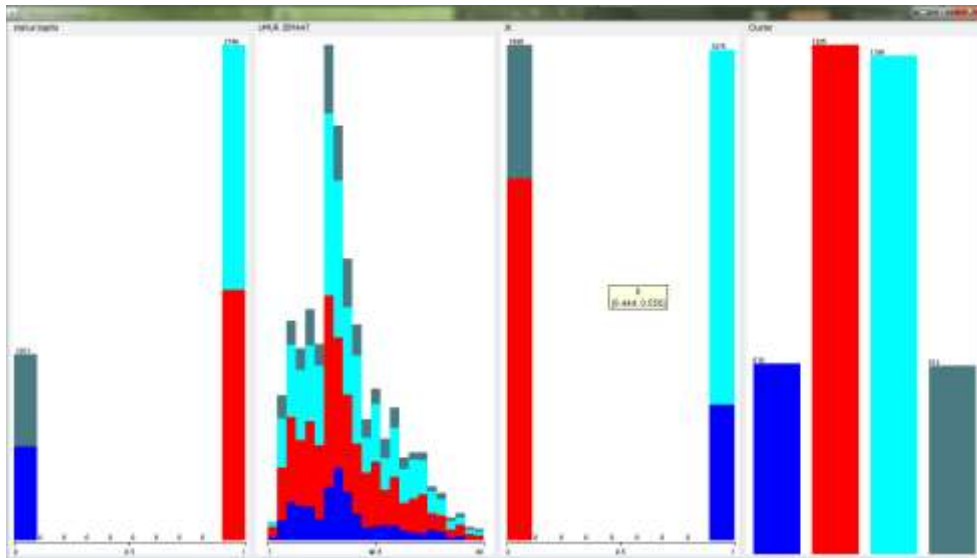
3. HASIL DAN PEMBAHASAN

4.

Data jemaat berjumlah 3772 record diproses dengan menerapkan algoritma Simple K-Means pada aplikasi Weka. Jumlah cluster inputan adalah 4 cluster. Hasil clustering dengan iterasi sebanyak 8 kali menghasilkan cluster 0 sejumlah 519 record (14%), cluster 1 sejumlah 1385 record (37%), cluster 2 sejumlah 1356 record (36%) dan cluster 3 sejumlah 512 record (14%).

Hasil clustering disimpan dalam tabel dan dianalisa menghasilkan kesimpulan sebagai berikut:

- 1) Cluster 0: jumlah data sebanyak: **519 record**, status baptis:**Belum**, jenis kelamin : **Laki-Laki**, Usia: **1 tahun-77 tahun**
- 2) Cluster 1: jumlah data sebanyak: 1385 record, status baptis:Sudah, jenis kelamin : Perempuan, Usia: 4 tahun-78 tahun
- 3) Cluster 2: jumlah data sebanyak: 1356 record, status baptis:Sudah, jenis kelamin : Laki-Laki, Usia: 4 tahun-80 tahun
- 4) Cluster 3: jumlah data sebanyak: **512 record**, status baptis:**Belum**, jenis kelamin : **Perempuan**, Usia: **4 tahun-80 tahun**



Gambar 5. Visualisasi hasil cluster

5. KESIMPULAN DAN SARAN

Berdasarkan hasil clustering maka dapat menjadi catatan bagi pihak gereja yaitu jemaat yang belum dibaptis berjenis kelamin laki dan perempuan dengan kisaran usia menyebar di antara 1 tahun sampai dengan 80 tahun dan jumlah jemaat yang belum dibaptis mencapai 1031 orang.

Untuk kedepannya variabel pendukung bisa ditambahkan untuk dapat menganalisa faktor-faktor yang menyebabkan jemaat belum dibaptis, misalnya pendidikan, pekerjaan, jumlah anggota keluarga dan faktor sosial lainnya.

DAFTAR PUSTAKA

- Afrisawati. Implementasi data mining pemilihan pelanggan potensial menggunakan algoritma k-means. *Jurnal Pelita Informatika Budi Karma*, 5(3), 2013.
- Nielza Atthina and Lizda Iswari. Klasterisasi data kesehatan penduduk untuk menentukan rentang derajat kesehatan daerah dengan metode k-means. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, volume 1, 2014.
- Emha Taufiq Luthfi Kusri. Algoritma data mining. *Yogyakarta: Andi Offset*, 2009.
- Ahlihi Masruro, Kusri Kusri, and Emha Taufiq Luthfi. Sistem penunjang keputusan penentuan lokasi wisata menggunakan k-means clustering dan topsis. *Data Manajemen dan Teknologi Informasi*, 15(4), 2014.
- Johan Oscar Ong. Implementasi algoritma k-means clustering untuk menentukan strategi marketing president university. *Jurnal Ilmiah Teknik Industri*, 12(1):10–20, 2013.
- Eko Prasetyo. Data mining konsep dan aplikasi menggunakan matlab. *Yogyakarta: Andi*, 2012.
- Mhd Gading Sadewo, Agus Perdana Windarto, and Dedy Hartama. Penerapan datamining pada populasi daging ayam ras pedaging di indonesia berdasarkan provinsi menggunakan k-means clustering. *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, 2(1): 60–67, 2017.
- Neha Soni and Amit Ganatra. Categorization of several clustering algorithms from different perspective: a review. *International Journal of*, 2012.
- Suprihatin. Klastering k-means untuk penentuan nilai ujian. *JUSI*, 1(1), 2011.
- Bambang Tri Wahyu and Angga Angriawan. Sistem rekomendasi paket wisata se-malang raya menggunakan metode hybrid content based and collaborative. *Jurnal Ilmiah Teknologi Informasi Asia*, 9(1):6–13, 2015.
- Anindya Khrisna Wardhani. K-means algorithm implementation for clustering of patients disease in kafen clinic of pekalongan. *Jurnal Transformatika*, 14(1):30–37, 2016.
- Agus Perdana Windarto. Implementation of data mining on rice imports by major country of origin using algorithm using k-means clustering method. *International Journal of Artificial Intelligence Research*, 1(2), 2017.