

# IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA ANALISIS PENYEBARAN PENYAKIT PASIEN PENGGUNA BADAN PENYELENGGARA JAMINAN SOSIAL KESEHATAN (BPJS) (STUDI KASUS : RUMAH SAKIT UMUM PUSAT HAJI ADAM MALIK MEDAN)

Parasian D P Silitonga, S.Kom., M.Cs<sup>(1)</sup>, Romanus Damanik S.Kom., M.Kom<sup>(2)</sup>

<sup>1</sup>Teknik Informatika Unika Santo Thomas  
Email : parasianirene@gmail.com

<sup>2</sup>Sistem Informasi Unika Santo Thomas  
Email : romanusdamanik@yahoo.com

## Abstract

*Clustering technique is a technique of grouping of records in a database based on certain criteria. The clustering results are given to the end user to provide a snapshot of what is happening in the database. One method that can be used in clustering technique is a method of K-Means Clustering. K-Means Clustering method divides the data into multiple groups and can accept input in the form of data without a label.*

*This study used the method K-Means Clustering data disease patients in an effort to find the tendency of the disease in a group of people who went to Haji Adam Malik Hospital in Medan. The results of this study are expected to be used as a reference for health education programs at once may be the anticipation of a priority service for patients especially BPJS users.*

**Keywords : BPJS, Clustering, K-Means Clustering**

## 1. Pendahuluan

Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan merupakan lembaga asuransi milik pemerintah yang dibentuk berdasarkan Peraturan Presiden Nomor 12 Tahun 2013. BPJS Kesehatan mencakup seluruh warga negara Indonesia dan resmi beroperasi sejak 1 Januari 2014 [3].

Adapun tujuan dari penyelenggaraan program ini adalah untuk memenuhi kebutuhan kesehatan yang layak yang diberikan kepada setiap anggota masyarakat yang telah membayar iuran atau iurannya dibayar oleh pemerintah maupun lembaga swasta. Dengan adanya program BPJS Kesehatan maka

seluruh lapisan masyarakat memperoleh hak layanan kesehatan dari rumah sakit maupun puskesmas yang telah ditunjuk oleh pemerintah sebagai rumah sakit ataupun puskesmas pemberi layanan BPJS Kesehatan.

Salah satu rumah sakit yang menyediakan fasilitas layanan kesehatan bagi masyarakat pengguna BPJS Kesehatan adalah Rumah Sakit Umum Pusat (RSUP) Haji Adam Malik Medan. RSUP Haji Adam Malik Medan merupakan rumah sakit pemerintah yang berada di kota Medan dan memiliki kategori kelas A.

Dengan diberlakukannya program BPJS Kesehatan dapat dipastikan bahwa jumlah pasien RSUP Haji Adam Malik Medan semakin meningkat. Seiring dengan meningkatnya jumlah pasien rumah sakit, diharapkan pihak rumah sakit dapat mengetahui klasifikasi penyakit yang ada di masyarakat serta penyebab penyakit tersebut. Dengan demikian pihak rumah sakit khususnya RSUP Haji Adam Malik Medan dapat memberikan masukan bagi pemerintah propinsi Sumatera Utara berkaitan dengan usaha pencegahan penyakit dan penyuluhan kesehatan ke daerah-daerah.

*Clustering* merupakan teknik pengelompokan *record* pada basis data berdasarkan kriteria tertentu. Hasil *clustering* diberikan kepada pengguna akhir untuk memberikan gambaran tentang apa yang terjadi pada basis data [5]. *Clustering* melakukan pengelompokan data tanpa berdasarkan kelas data tertentu. Bahkan clustering dapat dipakai untuk memberikan label pada kelas data yang belum diketahui itu. Karena itu clustering sering digolongkan sebagai metode *unsupervised learning*[4].

Salah satu metode yang dapat dilakukan untuk mengelompokkan data pada basis data adalah metode *K-Means Clustering*. Metode *K-Means Clustering*

membagi data menjadi beberapa kelompok serta dapat menerima masukan berupa data tanpa label kelas [2]. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

Hasil dari penelitian ini diharapkan dapat memberikan masukan bagi pemerintah propinsi Sumatera Utara berkaitan dengan usaha pencegahan penyakit dan penyuluhan kesehatan ke daerah-daerah sedangkan bagi pihak RSUP Haji Adam Maslik Medan dapat digunakan sebagai masukan bagi proses antisipasi prioritas pelayanan jika diketahui pola penyakit dengan kecenderungan tertinggi.

## 2. Badan Penyelenggara Jaminan Sosial (BPJS)

Badan Penyelenggara Jaminan Sosial atau BPJS merupakan lembaga yang dibentuk untuk menyelenggarakan program jaminan sosial di Indonesia menurut Undang-undang Nomor 40 Tahun 2004 dan Undang-undang Nomor 24 Tahun 2011. Sesuai Undang-undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional, BPJS merupakan badan hukum nirlaba [3]

Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan merupakan lembaga asuransi milik pemerintah yang dibentuk berdasarkan Peraturan Presiden Nomor 12 Tahun 2013. BPJS Kesehatan mencakup seluruh warga negara Indonesia dan resmi beroperasi sejak 1 Januari 2014

Adapun tujuan dari penyelenggaraan program ini adalah untuk memenuhi kebutuhan kesehatan yang layak yang diberikan kepada setiap anggota masyarakat yang telah membayar iuran atau iurannya dibayar oleh pemerintah maupun lembaga swasta. Dengan adanya program BPJS Kesehatan maka seluruh lapisan masyarakat memperoleh hak layanan kesehatan dari rumah sakit maupun puskesmas yang telah ditunjuk oleh pemerintah sebagai rumah sakit ataupun puskesmas pemberi layanan BPJS Kesehatan.

## 3. Data Clustering

Seiring dengan perkembangan teknologi informasi yang pesat dan semakin meningkatnya kebutuhan manusia terhadap data dan informasi pada satu sisi mengakibatkan penumpukan data dalam jumlah besar namun miskin dalam informasi. Tidak

jarang tumpukan data tersebut dibiarkan seakan-akan menjadi kuburan data.

Salah satu fungsi *data mining* adalah *clustering*. *Clustering* merupakan teknik pengelompokkan *record* pada basis data berdasarkan kriteria tertentu. Hasil *clustering* diberikan kepada pengguna akhir untuk memberikan gambaran tentang apa yang terjadi pada basis data [5].

*Clustering* melakukan pengelompokkan data tanpa berdasarkan kelas data tertentu. Bahkan *clustering* dapat dipakai untuk memberikan label pada kelas data yang belum diketahui itu.

*Clustering* termasuk ke dalam *descriptive methods*, dan juga termasuk *unsupervised learning* [4]. Dimana tidak ada pendefinisian kelas objek sebelumnya. Sehingga *clustering* dapat digunakan untuk menentukan label kelas bagi data-data yang belum diketahui kelasnya.

Konsep dasar dari *clustering* adalah mengelompokkan sejumlah objek ke dalam *cluster* dimana *cluster* yang baik adalah *cluster* yang memiliki tingkat kesamaan yang tinggi antar objek di dalam suatu *cluster* dan tingkat ketidaksamaan yang tinggi dengan objek *cluster* yang lainnya.

Analisa *clustering* mengidentifikasi kumpulan objek yang memiliki kemiripan satu dengan yang lainnya. Metode *clustering* yang baik dapat menghasilkan *cluster* yang berkualitas untuk memastikan kesamaan data-data yang ada pada sebuah *cluster*.

Terdapat banyak algoritma *clustering* yang dalam penggunaannya tergantung pada tipe data yang akan dikelompokkan dan apa tujuan dari pembuatan aplikasinya. Algoritma tersebut dapat digunakan untuk mengelompokkan objek ke dalam *cluster-cluster*, kemudian dari hasil *clustering* akan dideteksi keberadaan outlier dalam data tersebut.

## 4. K-Means Clustering

*K-means* merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempartisi dataset kedalam beberapa klaster *k*. Algoritmanya cukup mudah untuk diimplementasi dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan.

Metode *K-Means* merupakan metode *clustering* yang paling populer dan banyak digunakan dalam dunia industri . Metode *K-Means Clustering* membagi data menjadi beberapa kelompok serta dapat menerima masukan berupa data tanpa label kelas [2]. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam

satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

Metode ini merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil [1].

Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster.

Secara umum metode K-Means Clustering dilakukan dengan langkah-langkah sebagai berikut [2] :

1. Pilih jumlah cluster k.
2. Inisialisasi pusat cluster k. Secara umum pusat-pusat cluster diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean sesuai dengan Persamaan :

$$D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

Dimana :

D(i,j) = Jarak data ke-i ke pusat cluster j

$x_{ki}$  = Data ke-i pada atribut data ke-k

$x_{kj}$  = Titik pusat ke-j pada atribut data ke-k

4. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu.
5. Periksa setiap objek pemakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah 3 sampai pusat cluster tidak berubah lagi.

## 5. Analisis Masalah

### 5.1.1. Data Primer

Data primer adalah data yang dikumpulkan sendiri oleh peneliti dari obyek penelitian. Data primer dalam penelitian ini diperoleh dan dikumpulkan langsung dari lokasi penelitian. Data primer yang diperoleh adalah berupa data pasien yang bersumber dari data rekam medik pasien di RSUP Haji Adam Malik Medan.

Data rekam medik pasien pengguna BPJS Kesehatan yang diperoleh terdiri dari No. Rekam Medik, Nama Pasien, Tgl. Masuk, Tgl. Keluar, Tgl. Lahir, Umur, Kode INACBG (Kode Penyakit Pasien) dan Deskripsi Penyakit Pasien. Hasil pengumpulan data rekam medik pasien pengguna BPJS Kesehatan untuk tahun 2014.

### 5.1.2. Data Sekunder

Data sekunder dalam penelitian ini diperoleh dari dokumentasi resmi Peraturan Menteri Kesehatan Republik Indonesia Nomor 69 Tahun 2013 tentang Standar Tarif Pelayanan Kesehatan Pada Fasilitas Kesehatan. Data yang diperoleh adalah berupa data INA-CBG yang terdiri dari Kode INA-CBG, Deskripsi Kode INA-CBG, dan Tarif INA-CBG.

Data INA-CBG terdiri dari 789 Kode INA-CBG untuk tarif yang berlaku pada Rumah Sakit Kelas A.

INA-CBG merupakan sebuah singkatan dari Indonesia *Case Base Groups* yaitu sebuah aplikasi yang digunakan rumah sakit untuk mengajukan klaim pada pemerintah.

Sistem INA-CBG dikembangkan dari sistem *casemix* dari UNU-IIGH (*The United Nations University-International Institute for Global Health*) dan berpedoman pada *International Classification Of Diseases (ICD)*.

### 5.1.3 Transformasi Data

Tahapan transformasi data merupakan proses merubah data ke dalam bentuk yang sesuai untuk di mining. Perubahan awal yang dilakukan adalah dengan mengubah format data untuk mengetahui jumlah diagnosa penyakit untuk tiap bulan (januari sampai dengan desember). Hasil transformasi data yang dilakukan adalah sesuai dengan bentuk tabel seperti pada Tabel 1.

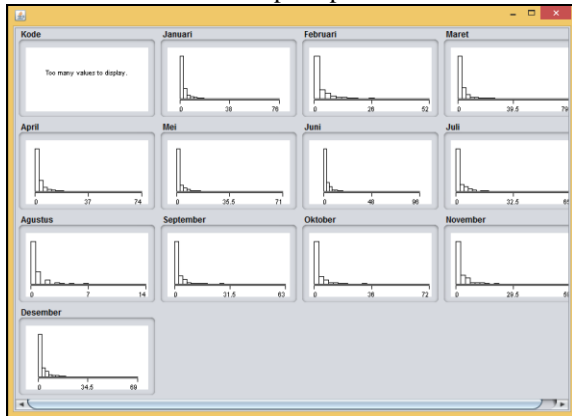
Tabel 1. Tabel Transformasi Data

Atribut	Keterangan
KodeInaCBG	Kode Penyakit
Bulan	Bulan Kalender
Jumlah	Jumlah diagnosa penyakit

## 6. Hasil

Proses transformasi dilakukan untuk memperoleh kecenderungan data penyakit pasien setiap bulan selama 2 tahun penelitian yaitu tahun 2014 dan 2015, sekaligus untuk menghilangkan data yang kosong.

Hasil pembacaan data pasien dilakukan dengan menggunakan aplikasi Weka disajikan dalam bentuk visualisasi data seperti pada Gambar 1.

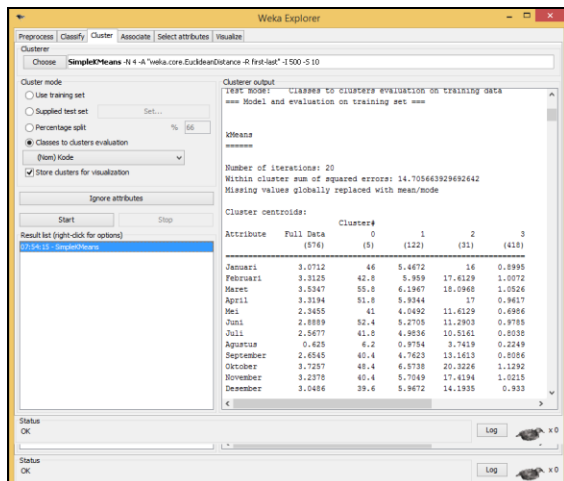


Gambar 1. Visualisasi Data Pasien

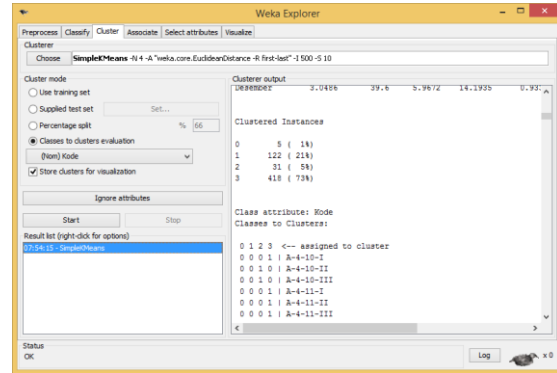
Proses *clusterisasi* data dengan menggunakan *K-Means Clustering* menghasilkan 4 *cluster* dengan masing-masing *cluster* memiliki persentase data sebagai berikut :

- a. Cluster 0 : 1%
- b. Cluster 1 : 21%
- c. Cluster 2 : 5%
- d. Cluster 3 : 73%

Hasil perincian *clustering* tersebut disajikan seperti pada Gambar 2 dan Gambar 3.

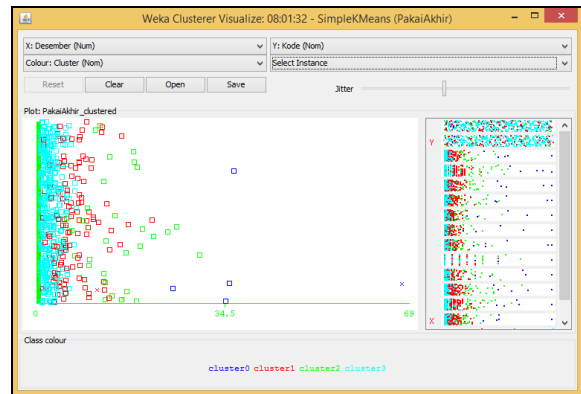


Gambar 2. Clustering Data Kode Penyakit InaCBG



Gambar 3. Persentase Clustering Data Kode Penyakit InaCBG

Sebaran data pada setiap *cluster* disajikan dalam bentuk plot grafik seperti pada Gambar 4.



Gambar 4 Grafik Sebaran Data InaCBG

## 7. Kesimpulan

Berdasarkan hasil *clustering* yang dilakukan terhadap data pasien penggunaan BPJS Kesehatan di RSUP Haji Adam Malik Medan dihasilkan kesimpulan antara lain :

- a. Bahwa proses *clustering* dengan menggunakan algoritma *K-Means Clustering* menghasilkan 4 *cluster* yang menyimpan data-data penyakit pasien dengan karakteristik yang berdekatan.
- b. Kode penyakit berdasarkan kode INACBG yang paling banyak terjadi adalah kode A-4-10-I yaitu kode penyakit Septikemia Ringan.

## 8 Daftar Pustaka

Pada Fasilitas Kesehatan Tingkat Pertama Milik Pemerintah Daerah.

- [1] Agusta, Y, 2007, *K-means - Penerapan, Permasalahan dan Metode Terkait*, Jurnal Sistem dan Informatika Vol. 3 (Februari 2007) : 47-60.
- [2] Berkhin Pavel, 2002, *Survey of Clustering Data Mining Techniques*. Accrue Software, Inc.
- [3] [http://id.wikipedia.org/wiki/BPJS\\_Kesehatan..](http://id.wikipedia.org/wiki/BPJS_Kesehatan..), 27 Maret 2015., 09.45 wib.
- [4] Ian H and Eibe Frank, 2005, *Data Mining Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, San Francisco.
- [5] Jiawei Han and Micheline Kember, 2006, *Data Mining: Concepts and Techniques Second Edition* , Morgan Kaufmann Publishers, San Francisco.
- [6] Larose, Daniel, *Discovery Knowledge in Data*, A Jhon Wiley & Sons, Inc Publication. Canada: 2005
- [8] Manning, Christopher D., Prabhakar Raghavan, Hinrich Schutze, 2009, *An Introduction to Information Retrieval*, Cambridge: Cambridge University Presss.
- [9] Mardiana T, Rudy D, 2015, *Kluster Bag-of-Word Menggunakan Weka*, Jurnal Edukasi dan Penelitian Informatika (JEPIN) Vol. 1, No. 1, (Juni 2015) ISSN 2460-7041.
- [10] N. Sharma, A. Bajpai, and R. Litoriya, *Comparison the Various Clustering Algorithms of WEKA Tools*, Int. J. Emerg. Technol. Adv. Eng., vol. 2, No. 5, May 2012.
- [11] Peraturan Presiden Republik Indonesia Nomor 12 Tahun 2013 Tentang Jaminan Kesehatan.
- [12] Peraturan Presiden Republik Indonesia Nomor 32 Tahun 2014 Tentang Pengelolaan dan Pemanfaatan Dana Kapitasi Jaminan Kesehatan Nasional