Data Mining Pengelompokan Biaya Pengeluaran Rumah Sakit Oleh BPJS Kesehatan Menggunakan Algoritma K-Means

Fuaida nabyla[1], Mukrodin[2]
email : nabilafuaida@gmail.com[1], mukrodins@gmail.com[2]
universitas peradaban

Abstrak

Sejak dioperasionalkan BPJS Kesehatan sebagai pelaksana Jaminan Kesehatan Nasional (JKN), berbagai kalangan mengkhawatirkan tarif yang diberlakukan dengan mengacu kepada INACBG. Sering kali rumah sakit mengalami kerugian karena tidak sesuainya biaya yang dikeluarkan rumah sakit terhadap tarif INACBG. Pengelompokan dalam data mining dapat digunakan untuk menganalisa kelompok biaya tiap-tiap diagnosa penyakit yang mengalami kerugian atau biaya yang rumah sakit keluarkan tidak sesuai dengan INACBG. Salah satu metode yang dapat digunakan dalam teknik pengelompokan adalah algoritma KMeans. Algoritma K-Means yang merupakan metode data clustering non hirarki yang mempartisi data ke dalam cluster 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 lain. Tujuan dari penelitian ini mengelompokan biaya yang rumah sakit keluarkan terhadap tarif INACBG apakah telah sesui dengan semestinya. Hasil dari penelitian ini dapat digunakan sebagai sistem pendukung keputusan rumah sakit dalam menetapkan tarif rumah sakit.

Kata Kunci : *Algoritma K-Means, Clustering, INACB.*

1. Pendahuluan

Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan merupakan lembaga asuransi milik Pemerintah yang dibentuk berdasarkan Undang Undang Nomor 40 Tahun 2004 Tentang Sistem Jaminan Sosial Nasional, dan Undang-Undang Nomor 24 Tahun 2011 Tentang Badan Penyelenggara Jaminan Sosial. 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. Salah satu rumah sakit yang menyediakan fasilitas layanan kesehatan bagi masyarakat pengguna BPJS Kesehatan adalah Rumah Sakit Umum Muhammadiyah Siti Aminah Bumiayu. Salah satu pemanfaatan data yang bisa digunakan untuk kepentingan rumah sakit yaitu sebagai penentuan biaya tindakan medis yang ditanggung oleh BPJS Kesehatan. Dimana seringkali rumah sakit mengalami kerugian karena tarif rumah sakit yang digunakan lebih besar dari tarif standar yang diberikan oleh pihak BPJS Kesehatan. Namun dengan data yang sangat banyak, pihak rumah sakit sulit untuk mengenali penyakit-penyakit mana saja yang cenderung mendapati tarif yang merugikan rumah sakit. Maka perlunya analisis data dalam hal ini, salah satu yang bisa dilakukan yaitu dengan pengelompokan tarif rumah sakit. Clustering adalah metode yang digunakan dalam data mining yang cara kerjanya mencari dan mengelompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang telah diperoleh (Ong, 2013). Ciri khas dari teknik data mining ini adalah mempunyai sifat tanpa arahan (unsupervised), yang dimaksud adalah teknik ini diterapkan tanpa perlunya data training dan tanpa ada teacher serta tidak memerlukan target output. Metode clustering yang mempunyai sifat efesien dan cepat yang dapat digunakan salah satunya adalah metode k-means, metode ini bertujuan untuk membuat cluster objek menurut atribut yang menjadi k partisi. Berdasakan permasalahan diatas peneliti akan melihat bagaiaman penerapan data mining tersebut menggunakan algoritma k-means untuk menentukan apakah biaya tindakan di rumah sakit telah sesuai dengan tarif pada INACBG. Harapannya adalah untuk melihat penemuan informasi baru mengenai biaya tindakan medis di RSU Muhammadiyah Siti Aminah Bumiayu sudah sesuai dengan INACBG (Indonesia Case Base Groups) dengan menggunakan algoritma k-means.

2. Metode Penelitian

Penelitian ini menggunakan algoritma K-Means untuk menentukan cluster yang terbaik. Cluster terbaik ini dipergunakan untuk pemilihan mahasiswa-mahasiswa terbaik yang dapat diikutkan lomba. Sehingga peluang untuk mendapatkan juara dalam lomba bisa semakin besar.

A. Data Mining

Istilah data mining memiliki beberapa padanan, seperti *knowledge discovery* ataupun *pattern recognition*. Istilah *knowledge discovery* atau penemuan pengetahuan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang masih tersembunyi di dalam bongkahan data. Istilah *pattern recognition* atau pengenalan pola pun tepat untuk digunakan karena pengetahuan yang hendak digali. Kegiatan inilah yang menjadi garapan atau perhatian utama dari disiplin ilmu data mining (Susanto, 2010: 2).

Gorunescu (dalam Prasetyo, 2014: 7) menjelaskan bahwa secara sistematis, ada tiga langkah utama dalam proses data mining:

1. Eksplorasi pemrosesan awal pada data.
2. Membangun model dan melakukan validasi terhadap model.
3. Penerapan Penerapan berarti menerapkan model pada data yang baru untuk menghasilkan perkiraan/prediksi masalah yang diinvestigasi.

B. Clustering

Proses pengelompokan sekumpulan obyek kedalam kelas-kelas obyek yang sama disebut clustering/pengelompokan. Set variabel Cluster adalah suatu set variabel yang merpresentasikan karakteristik yang dipakai objek- objek. Solusi Analisis. Solusi Cluster secara keseluruhan bergantung pada variabel-variabel yang digunakan sebagai dasar untuk menilai kesamaan. Penambahan atau pengurangan variabel-variabel yang relevan dapat mempengaruhi substansi hasil analisis Cluster (Ediyanto, Mara, & Satyahadewi, 2013).

C. Algorita K-Means

Analisis Cluster merupakan salah satu metode objek mining yang bersifat tanpa latihan (unsupervised analysis), sedangkan K-Means Cluster Analysis merupakan salah satu metode Cluster analisis non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih Cluster atau kelompok objek. Tujuan pengelompokan adalah untuk meminimalkan objective function yang di set dalam proses Clustering (Ediyanto et al., 2013).

Menurut Gustientiedina, dkk (2019), Secara umum metode K-Means menggunakan algoritma sebagai berikut :

1. Tentukan k sebagai jumlah Cluster yang di bentuk.
2. Bangkitkan k Centroid (titik pusat Cluster) awal secara random, dilakukan secara random/acak dari objekobjek yang tersedia sebanyak k Cluster.

Rumus sebagai berikut: $v = \sum xi\ n\ i{=}1\ n$; i = 1,2,3, … n dimana: v : centroid pada Cluster xi : objek ke-i n : banyaknya objek/jumlah objek yang menjadi anggota Cluster.

3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing Cluster.

Euclidean Distance.

$(x, y) = \| x - y \| = \sqrt{\sum}(xi - yi)\ 2$; i = 1,2,3, … $n\ n\ i{=}1$

dimana: xi : obyek x ke-i yi : daya y ke-i n : banyaknya objek.

4. Alokasikan masing-masing objek ke dalam centroid yang paling terdekat.

5. Lakukan Iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan 1.

6. Ulangi langkah 3 jika posisi centroid baru tidak sama.

D. INACBG (Indonesia *Case Base Groups*)

INACBG merupakan suatu sistem pengklasifikasian penyakit yang mengkombinasikan antara sekelompok penyakit dengan karakteristik klinis serupa dengan biaya perawatan disuatu rumah sakit. INACBG merupakan sistem pembayaran dengan sistem "paket", berdasarkan penyakit yang diderita pasien. Rumah Sakit akan mendapatkan pembayaran berdasarkan tarif INACBG yang merupakan rata-rata biaya yang dihabiskan oleh untuk suatu kelompok diagnosis (Info BPJS Kesehatan, 2014).

Tarif yang dimaksudkan berbentuk paket yang mencakup seluruh komponen biaya RS. Berbasis pada data costing dan coding penyakit mengacu *International Classification of Diseases* (ICD) yang disusun WHO. Menggunakan ICD 10 untuk mendiagnosis 14.500 kode dan ICD 9 *Clinical Modifications* yang mencakup 7.500 kode. Sedangkan tarif INACBG terdiri dari 1.077 kode CBG yang terdiri dari 789 rawat inap dan 288 rawat jalan dengan tiga tingkat keparahan.

3. Hasil dan Pembahasan

3.1 *Clustering* Menggunakan *K-Means*

1). Memuat data

Dari data yang sudah di transformasi, data tersebut dimasukan ke dalam file baru. Dimana file inilah yang akan digunakan dalam proses *cluster*ing menggunakan *k-means*.

Tabel 1. Data Set Baru

NO	DESKRIPSI_INACBG	JMLH_DESK	SELISIH_TARIF
1	ABORSI MENGANCAM	1	-27500
2	ABORTUS (RINGAN)	3	-979600
3	ABORTUS MENGANCAM (RINGAN)	12	-637908
4	ABORTUS MENGANCAM (SEDANG)	1	-873300
5	ANGINA PEKTORIS DAN NYERI DADA (RINGAN)	7	566957
6	ASTHMA & BRONKIOLITIS (RINGAN)	40	-282838
7	ATHEROSKLEROSIS (RINGAN)	17	-465353
8	ATHEROSKLEROSIS (SEDANG)	3	249167
9	BATU URIN (RINGAN)	9	-1118889
10	...	...	...
210	TUMOR MYELOPROLIFERATIF LAIN-LAIN (RINGAN)	2	-2720200
211	TUMOR SISTEM SARAF & GANGGUAN DEGENERATIF (RINGAN)	2	1805000

2). Proses Traning

a. Iterasi 1

Tentukan K jumlah pusat cluster secara acak (random). Pada percobaan pertama ini ditentukan 3 data secara acak sebagai titik centroid awal untuk perhitungan jarak dari seluruh kelompok cluster yang akan dibentuk.

Jumlah cluster = 3
Jumlah data = 211
Jumlah atribut = 2

b. Hitung jarak tiap data dengan masing-masing cluster pusat dengan menggunakan persamaan Euclidean.

Anggota dipilih dari terkecil diantara 3 cluster jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 39 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 130 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 42 data.

c. Lakukan iterasi ke 2

Tentukan posisi centroid baru dengan cara menghitung rata-rata dari data-data yang ada pada centroid yang sama atau anggota yang sama.

Centroid x:
Rerata = Total / Jumlah Data
Anggota C1 » Rerata = 724/39 = 18,5641
Anggota C2 » Rerata = 33843/130 = 260,3308
Anggota C3 » Rerata = 473/42 = 11,2619
Centroid y:
Rerata = Total / Jumlah Data
Anggota C1 » Rerata = -113037333/39 = -2899675,205
Anggota C2 » Rerata = -37439823/130 = -287998,6385
Aggota C3 » Rerata = 54824663/42 = 1305349,119

d. Karena hasil iterasi ke-2 tidak sama dengan iterasi ke-1 sehingga perlu dilakukan kembali perhitungan ke iterasi ke-3 dan seterusnya sampai mendapatkan hasil yang sama. Lakukan iterasi ke 3.

Centroid x:
Rerata = Total / Jumlah Data
Anggota C1 » Rerata = 710/34 = 20,88235
Anggota C2 » Rerata = 33954/144 = 235,7917 Aggota C3 » Rerata = 376/33 = 11,39394 Centroid y:
Rerata = Total / Jumlah Data
Anggota C1 » Rerata = -105738633/34 = -3109959,794 Anggota C2 » Rerata = -41245712/144 = -286428,5556 Anggota C3 » Rerata = 51281852/33 = 1553995,515

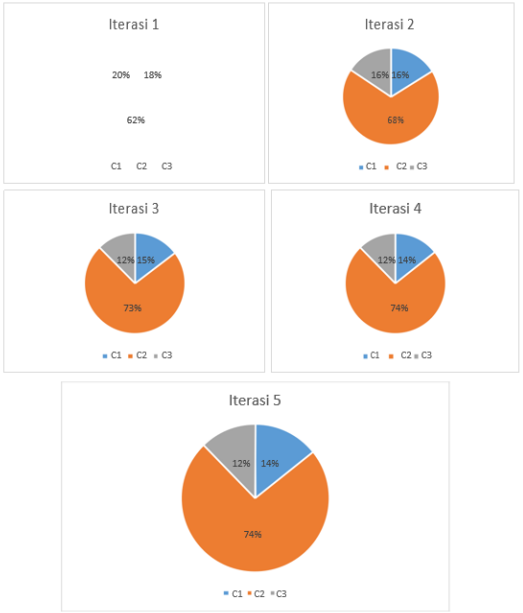
e. Karena hasil iterasi ke-3 tidak sama dengan iterasi ke-2 sehingga perlu dilakukan kembali perhitungan ke iterasi ke-4 dan seterusnya sampai mendapatkan hasil yang sama. Lakukan iterasi ke 4.

P-ISSN :2089-676X
E-ISSN : 2549-0796

Centroid x:
Rerata = Total / Jumlah Data
Aggota C1 » Rerata = 66/31 = 21,48387
Aggota C2 » Rerata = 34048/154 = 221,0909
Aggota C3 » Rerata = 326/26 = 12,53846
Centroid y:
Rerata = Total / Jumlah Data
Anggota C1 » Rerata = -100847285/31 = -3253138
Anggota C2 » Rerata = -42154499/144 = -273730,51
Anggota C3 » Rerata = 47299291/26 = 1819204
f. Karena hasil iterasi ke-4 tidak sama dengan iterasi ke-3 sehingga perlu dilakukan kembali perhitungan ke iterasi ke-4 dan seterusnya sampai mendapatkan hasil yang sama. Lakukan iterasi ke 5..
Centroid x:
Rerata = Total / Jumlah Data
Anggota C1 » Rerata = 538/30 = 17,93333
Anggota C2 » Rerata = 34176/155 = 220,4903 Anggota C3 » Rerata = 326/26 = 12,53846 Centroid y:
Rerata = Total / Jumlah Data
Anggota C1 » Rerata = -99099955/30 = -3303332 Anggota C2 » Rerata = -4390129/155 = -283238 Anggota C3 » Rerata = 47299291/26 = 1819204

3). Distribusi Cluster

Distribusi cluster digunakan untuk menunjukan banyaknya data disetiap cluster dalam proses iterasi. Distribusi cluster ditunjukan menggunakan diagram berikut:



Gambar 1. Distribusi Cluster Setiap Iterasi

3.2 Implementasi Menggunakan Jupyter Notebook

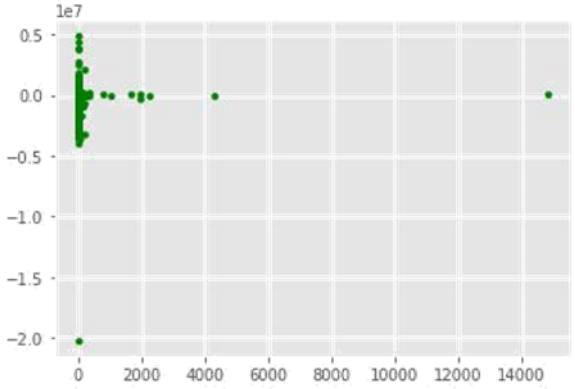
1). Memuat data

Melakukan loading data pada file data Set yang sudah di transformasi untuk dimasukan ke dalam *software* jupyter nootebok.

Out[2]:

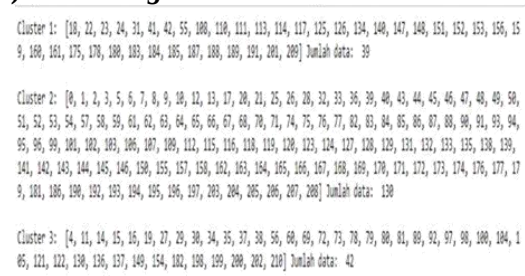
	DESKRIPSI_INACBG	JMLH_DESK	SELISIH_TARIF
0	ABORSI MENGANCAM	1	-27500
1	ABORTUS (RINGAN)	3	-979600
2	ABORTUS MENGANCAM (RINGAN)	12	-637908
3	ABORTUS MENGANCAM (SEDANG)	1	-873300
4	ANGINA PEKTORIS DAN NYERI DADA (RINGAN)	7	566957
5	ASTHMA & BRONKIOLITIS (RINGAN)	40	-282838
6	ATHEROSKLEROSIS (RINGAN)	17	-465353
7	ATHEROSKLEROSIS (SEDANG)	3	249167
8	BATU URIN (RINGAN)	9	-1118889
9	BRONKIAL AKUT	50	-22459

Gambar 2. Memuat Data



Gambar 3. Visualisasi Data Set

2). *Clustering* Data



Gambar 4. *Cluster* Awal

3). Distribusi Cluster

Distribusi cluster digunakan untuk menunjukan kontribusi relatif dari tiap cluster terhadap total keseluruhan data.

4). Visualisasi Data

Update centroid digunakan untuk menetapkan *centroid* terbaik.

5). Distribusi Akhir

Distribusi *cluster* terbanyak dimiliki pada C2 dimana terdapat distribusi *cluster* sebanyak 73,5%, sedangkan C1 dan C3

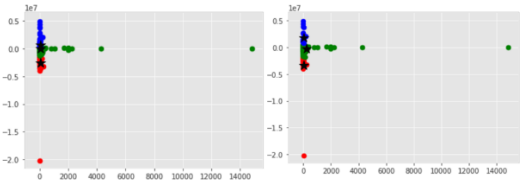
mendapati distribusi *cluster* yang lebih kecil yaitu 14.2% dan 12,3%.

6). Visualisasikan Kembali

Melihat data yang sudah di distribusi dan data dapat dilihat pada distribusi serta visualisai cluster.

7). Distribusi dan Visualisai Cluster

Gambar 5 Distribusi Cluster



Gambar 6 Visualisasi Cluster

4. Kesimpulan

Klusterisasi biaya yang Rumah Sakit keluarkan dapat ditarik kesimpulan bahwa kelompok biaya yang termasuk menguntungkan bagi rumah sakit yaitu memiliki rentang biaya dari Rp.798.945 sampai dengan Rp. 4.860.600, dan biaya yang termasuk menguntungkan dan merugikan bagi rumah sakit memiliki rentang biaya dari Rp. -1.747.330 sampai dengan Rp. 633.565, sedangkan biaya yang termasuk merugikan bagi rumah sakit memiliki rentang biaya dari Rp. -20.223.379 sampai dengan Rp. -1.814.000. Hal ini tentunya menjadi perhatian khusus bagi rumah sakit dalam menentukan biaya rumah sakit terhadap biaya tindakan medis pengguna BPJS Kesehatan, agar lebih sesuai dengan INACBG yang digunakan dalam hal klaim yang ditagihkan rumah sakit.

5. Daftar Pustaka

[1] Ediyanto, Mara, M. N., & Satyahadewi, N. (2013). Pengklasifikasian Karakteristik Dengan Metod K-Means Cluster Analysis. Buletin Ilmiah, 02(2), 133–136.

[2] Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. Jurnal Nasional Teknologi Dan Sistem Informasi, 5(1), 17–24. https://doi.org/10.25077/teknosi.v5i1.2019.17-24

[3] Info BPJS Kesehatan. (2014). Perubahan Tarif INA-CBGs. In BPJS Kesehatan (Vol. 8).

[4] Prasetyo, Eko. (2014). Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab. Bojonegoro: Universitas Negeri Malang.

[5] Project Jupyter . (2019, Desember). Jupyter. Retrieved from Jupyter: https://jupyter.org/

[6] Susanto, Sani. (2010). Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data. Yogyakarta : CV ANDI OFFSET.

[7] Ong, J. O. (2013). Implementasi Algotritma K-means clustering untuk menentukan strategi marketing president university. Jurnal Ilmiah Teknik Industri, vol.12, no(juni), 10–20.

P-ISSN :2089-676X
E-ISSN : 2549-0796