

USE CASE DATA SCIENCE PADA RUMAH SAKIT YANG MELAYANI PENGGUNA BPJS KESEHATAN

Pendahuluan

Data science ialah ilmu yang mengkombinasikan suatu keahlian di bidang ilmu tertentu dengan kemampuan pemrograman, matematika, serta statistik. Tujuannya ialah guna mengekstrak suatu pengetahuan ataupun data dari informasi. Penerapan *data science* juga populer di industri kesehatan, salah satunya BPJS Kesehatan di Indonesia.

Sebutan *data mining* mempunyai sebagian padanan, semacam *knowledge discovery* maupun *pattern recognition*. Sebutan *knowledge discovery* ataupun temuan pengetahuan sebab tujuan utama dari *data mining* memanglah buat memperoleh pengetahuan yang masih tersembunyi di dalam bongkahan informasi. Sebutan *pattern recognition* ataupun pengenalan pola juga tepat guna digunakan sebab pengetahuan yang hendak digali. Aktivitas inilah yang jadi garapan ataupun atensi utama dari disiplin ilmu *data mining*.

Secara sistematis, proses data mining dibagi menjadi tiga langkah utama:

1. Eksplorasi pra-pemrosesan data.
2. Membangun model dan memvalidasi model.
3. Menerapkan Implementasi berarti menerapkan model ke data baru untuk menghasilkan perkiraan/prediksi untuk pertanyaan yang sedang diselidiki.

Salah satu pemanfaatan data yang dapat digunakan untuk kepentingan rumah sakit adalah untuk menentukan anggaran tindakan medis yang ditanggung oleh BPJS Kesehatan. Rumah sakit seringkali menderita dengan menggunakan tarif rumah sakit yang lebih tinggi dari tarif standar yang ditawarkan oleh BPJS Kesehatan. Namun, dengan banyaknya data, bisa jadi sulit bagi rumah sakit untuk menentukan tingkat penyakit mana yang cenderung merugikan rumah sakit. Oleh karena itu diperlukan analisis data dalam hal ini yang dapat dilakukan dengan mengelompokkan tarif rawat inap. *Clustering* adalah suatu metode yang digunakan dalam *data mining* yang bekerja dengan cara mencari dan mengelompokkan data yang memiliki kesamaan karakteristik antara satu data dengan data lain yang telah diperoleh.

Teknik *data mining* ini dicirikan oleh sifatnya yang *unsupervised*, artinya penerapan teknik ini tidak memerlukan data latih, tidak ada pengajar, dan tidak ada target output. Metode *k-means* merupakan metode *clustering* yang efisien dan cepat yang dapat digunakan untuk membuat cluster objek berdasarkan atribut yang menjadi k partisi. Berdasarkan pertanyaan di atas, kita akan melihat bagaimana suatu aplikasi data mining dapat menggunakan algoritma *k-means* untuk menentukan apakah anggaran rawat inap sudah sinkron dengan bea pada INACBG sebagaimana mestinya. Temuan yang ingin melihat informasi baru mengenai anggaran tindakan medis di rumah sakit sejalan dengan INACBG (Indonesia Case Base Group) menggunakan algoritma *k-means*. Hasil dari penelitian ini bisa digunakan demi sistem pendukung keputusan rumah sakit dalam menetapkan bea rumah sakit.

Clustering menggunakan k-means

1. Memuat data

Berasal data yang telah ditransformasi, data tersebut dimasukkan ke dalam file baru. Dimana file inilah yang akan digunakan dalam proses *clustering* menggunakan *k means*.

NO	DESKRIPSI_INACBG	JMLH_DESK	SELISIH_TARIF
1	ABORSI MENGAMCAM	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 UTIN (RINGAN)	9	-1118889
10
210	TUMOR MYELOPROLIFERATIF LAIN-LAIN (RINGAN)	2	-2720200
211	TUMOR SISTEM SARAF & GANGGUAN DEGENERATIF (RINGAN)	2	1805000

2. Proses *training*

a. Perulangan 1

Tentukan K jumlah sentra cluster secara acak (random). Di percobaan pertama ini dipengaruhi tiga *random data* menjadi titik *centroid* awal buat perhitungan jeda asal semua grup *cluster* yang akan dibuat.

Jumlah *cluster* = 3

Jumlah data = 211

Jumlah atribut = 2

b. Hitung jeda tiap data menggunakan masing-masing *cluster* pusat dengan memakai persamaan Euclidean.

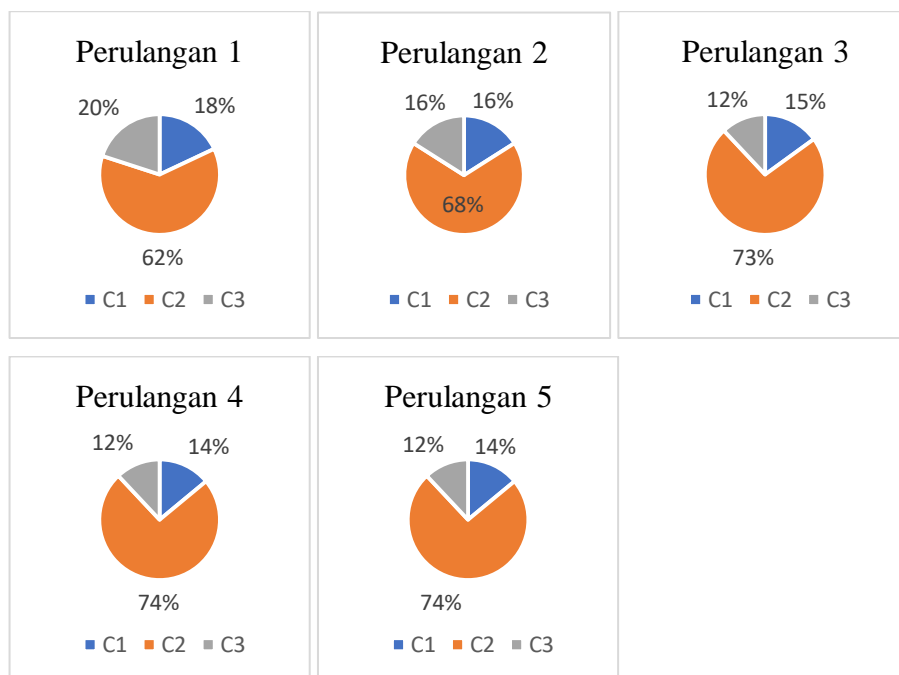
c. Perulangan 2

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

d. Ketika perulangan 2 tidak sama hasilnya terhadap perulangan 1 maka perlu dilakukan perulangan yang selanjutnya sampai perulangan dengan hasil yang sama.

3. Distribusi cluster

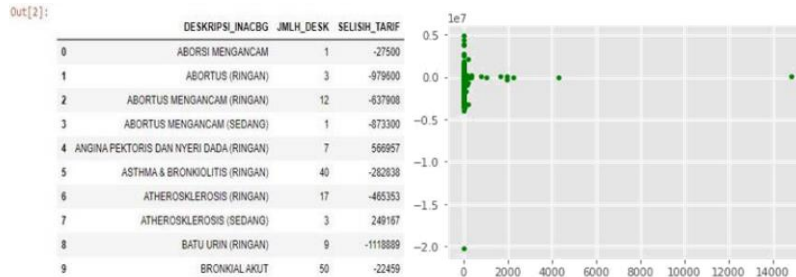
Distribusi cluster dipergunakan guna menunjukkan banyaknya data di setiap cluster pada proses perulangan. Distribusi cluster ditunjukan memakai diagram:



Implementasi menggunakan *Jupyter Notebook*

1. Memuat data

Melakukan *loading data* pada *data set* yang sudah ditransformasi untuk dimasukkan ke dalam program perangkat lunak *Jupyter Notebook*.



2. *Clustering* data

Cluster 1: [18, 22, 23, 24, 31, 41, 43, 55, 100, 110, 111, 113, 114, 117, 125, 126, 134, 140, 147, 148, 151, 152, 153, 156, 159, 160, 163, 175, 178, 180, 183, 184, 185, 187, 188, 189, 191, 201, 205] Jumlah data: 39

Cluster 2: [0, 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 13, 17, 20, 21, 25, 26, 28, 32, 33, 36, 39, 40, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 57, 58, 59, 61, 62, 63, 64, 65, 66, 67, 68, 70, 71, 74, 75, 76, 77, 82, 83, 84, 85, 86, 87, 88, 89, 91, 93, 94, 95, 96, 99, 102, 103, 106, 107, 109, 112, 115, 116, 118, 119, 120, 123, 124, 127, 128, 129, 131, 132, 133, 135, 136, 139, 142, 143, 144, 145, 146, 150, 155, 157, 158, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 176, 177, 179, 181, 186, 190, 192, 193, 194, 195, 196, 197, 203, 204, 205, 206, 207, 209] Jumlah data: 130

Cluster 3: [4, 11, 14, 15, 16, 19, 27, 29, 30, 34, 35, 37, 38, 56, 60, 69, 72, 73, 79, 79, 80, 81, 89, 92, 97, 98, 100, 104, 105, 121, 122, 130, 136, 137, 140, 154, 162, 166, 199, 200, 202, 210] Jumlah data: 42

3. Distribusi *cluster*

Distribusi *cluster* dipergunakan untuk membuktikan kontribusi relatif asal tiap *cluster* terhadap total keseluruhan data.

4. Pembaruan *centroid*

Pembaruan *centroid* dilakukan guna menetapkan *centroid* yang terbaik.

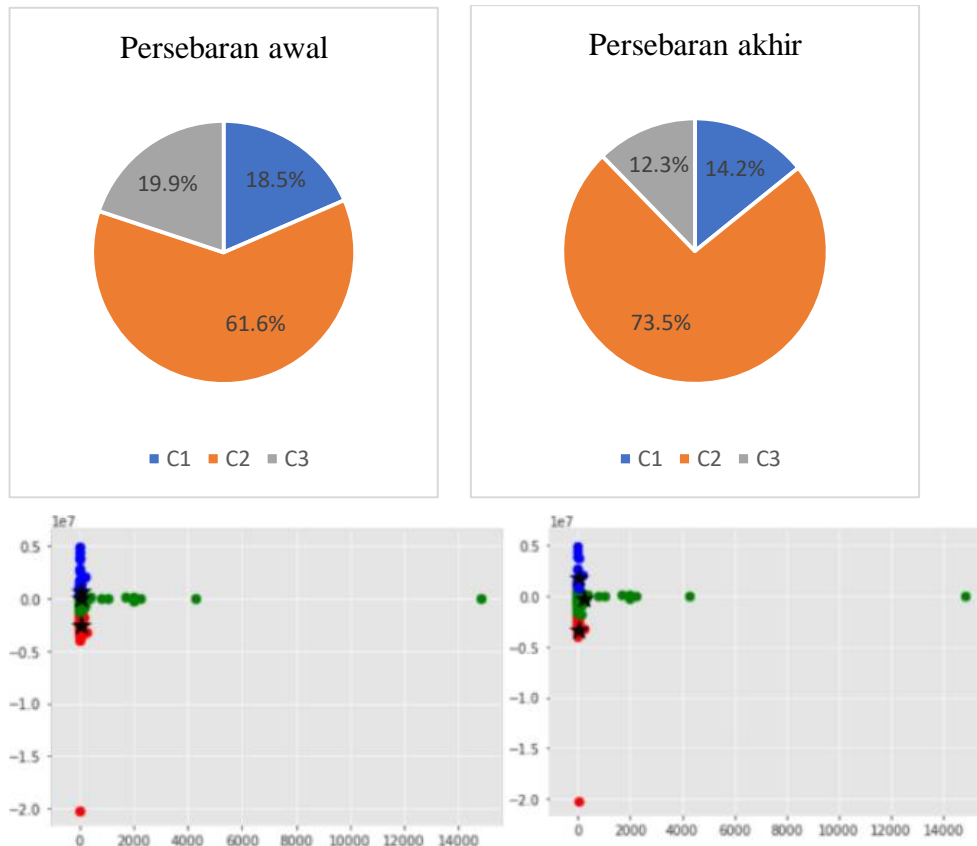
5. Persebaran *cluster*

Persebaran *cluster* terbanyak yaitu di C2 dimana ada distribusi *cluster* sebesar 73,5%, sedangkan didapati pada C1 dan C3 distribusi *cluster* yang lebih kecil yakni 14,2% serta 12,3%.

6. Visualisasi kembali

Melihat data yang telah dilakukan persebaran serta data bisa dicermati pada persebaran dan visualisasi *cluster*.

7. Persebaran dan visualisasi *cluster*



Kesimpulan

Klusterisasi anggaran yang rumah sakit keluarkan bisa ditarik kesimpulan bahwa gerombolan anggaran yang termasuk menguntungkan bagi rumah sakit yaitu memiliki rentang anggaran yakni Rp798,945 hingga Rp4,860,600. Anggaran yang termasuk menguntungkan serta tidak menguntungkan bagi rumah sakit mempunyai rentang anggaran yakni Rp-1,747,330 hingga Rp633,565. Sedangkan anggaran yang termasuk tidak menguntungkan bagi rumah sakit mempunyai rentang anggaran dari Rp20,223,379 hingga Rp-1,814,000. Hal ini tentunya menjadi perhatian spesifik bagi pihak rumah sakit dalam menentukan anggaran rumah sakit terhadap anggaran tindakan medis pengguna BPJS Kesehatan, agar lebih sinkron terhadap INACBG yang digunakan dalam hal klaim yang ditagihkan rumah sakit.