Jurnal TAM (Technology Acceptance Model) Volume 8, No2, Desember 2017 *Hal. 85-89*
p-ISSN : 2339-1103
e-ISSN : 2579-4221

KLASTERISASI PENDUDUK LANJUT USIA SUMATERA SELATAN MENGGUNAKAN ALGORITMA K-MODES

Fithri Selva Jumeilah[1], Dicky Pratama[2]
[1,2]*Prodi Sistem Informasi STMIK Global Informatika Palembang*
[1,2]*Jl. Rajawali no. 14 Palembang*
E-mail: fithri.selva@mdp.ac.id, dqpratama@mdp.ac.id

ABSTRAK
Saat ini Indonesia termasuk kedalam negara berpenduduk struktur tua karena penduduk lanjut usianya lebih dari 7% dari total penduduk dan 2% berasal dari Sumatera selatan. Besarnya jumlah penduduk lanjut usia membutuhkan kebijakan khusus pemerintah untuk merumuskan kebijakan dan program khusus penduduk lanjut usia sehingga dapat meringankan beban bagi masyarakat. Untuk membantu pemerintah terutama pemerintah Sumatera Selatan menentukan kebijakan dan program tersebut maka dibutuhkan klasterisasi penduduk lanjut usia dengan menggunakan algoritma K-modes yang ada di R-Studio. Penelitian ini menggunakan data sensus penduduk Sumatera selatan tahun 2010 yang diperoleh dari Bapan Pusat Statistik dengan jumlah sampel 47.358 jiwa. Dari hasil penelitian ini diperoleh 4 *cluster* yaitu: K1 berjumlah 16244 jiwa, K2 6061 jiwa, K3 18681 jiwa dan K4 6372 jiwa. K1 merupakan kelompok lansia yang sebagian besar laki-laki yang tinggal di desa dan masih bekerja di bidang pertanian dan perkebunan. K2 adalah *cluster* wanita yang masih bekerja dan tinggal di desa. *Cluster* ketiga K3 yang merupakan kelompok lansia yang tidak bekerja yang sebagian besar tinggal di kota dan 25%nya tinggal sendiri. Terakhir K4 yang merupakan *cluster* perempuan yang tidak bekerja lagi, tinggal di desa dan 73% buta huruf. Dengan adanya *cluster* tersebut pemerintah dapat menentukan kebijakan apa yang paling tepat untuk masing-masing *cluster*.

Keywords: Clustering, K-modes, Sensus Penduduk, Lansia, Sumatera selatan.

ABSTRACT
Currently, Indonesia is included in a country with a population of old structures because of its advanced population of more than 7% of the total population and 2% comes from southern Sumatra. The large number of elderly citizens required a special government policy to formulate policies and special programs the population can use to alleviate the community. To help local government of South Sumatera government to determine the policy and program hence needed clustering elderly population by using K-mode algorithm existing in R-Studio. This study uses population census data of South Sumatera in 2010 obtained from Bapan Pusat Statistik with 47,358 data sample. From the results of this study made 4 clusters: K1 16244 people, K2 6061 people, K3 18681 people, and K4 6372 people. K1 is an elderly group of mostly men who live in the village and still work in agriculture and plantations. K2 is a cluster of women who still work and live in the village. The third K3 cluster is an elderly unemployed group that mostly lives in the city and 25% lives alone. The last K4 is a cluster of women who do not work anymore, live in the village and 73% illiterate. With the cluster the government can determine what is most appropriate for each cluster.

Keywords: Clustering, K-modes, Population Census, Elderly, South Sumatera.

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Menurut hasil sensus penduduk 2010 Indonesia memiliki jumlah penduduk sebanyak 237.641.326 jiwa dan 18,1 juta jiwa adalah penduduk lanjut usia (lansia) [1]. Hal ini menunjukan bahwa Indonesia termasuk negara yang memasuki era penduduk menua karena jumlah penduduk lansia melebihi 7% [2]. Sumatera selatan (Sumsel) adalah provinsi yang memiliki penduduk lansia 2% dari penduduk lansia indonesia. Meningkatnya jumlah lansia menunjukan peningkatan harapan hidup dan sekaligus menunjukan bahwa pemerintah berhasil melaksanakan pembangunan dibidang kesehatan.

Dengan tingginya jumlah penduduk lansia di Sumsel, secara langsung menuntut pemerintah untuk membuat kebijakan dan program khusus untuk meringankan beban dari keluarga dan masyarakat. Pada Peraturan Pemerintah Nomor 43 Tahun 2004 tentang upaya peningkatan kesejahteraan lanjut usia dan Undang-Undang Nomor 36 Tahun 2009 pada pasal 138 telah ditentukan aspek-aspek apa saja yang harus disediakan pemerintah untuk penduduk lansia. Untuk mempermudah pemerintah menentukan kebijakan dan program khusus tersebut pemerintah membutuhkan gambaran kelompok *(cluster)* penduduk lansia.

Sudah banyak penelitian tentang *clustering* diantaranya: klasterisasi penjualan produk dengan

p-ISSN : 2339-1103
e-ISSN : 2579-4221

metode K-means yang menghasilkan 2 klaster yaitu laris dan tidak laris [3]; klasterisasi mahasiswa berdasarkan nilai akademik dengan metode K-Means dengan hasil 4 buah klaster [4]; pencarian pola pelanggan yang memberikan keuntungan tinggi, bernilai tinggi, dan beresiko rendah [5]. Algoritma *clustering* yang digunakan adalah algoritma *clustering* yang ada pada IBM I-Miner. Dari hasil penelitian, satu *cluster* memiliki pola pelanggan yang nilai tinggi dan beresiko rendah sebanyak 20% dari total pelanggan. Tenyata 80% dari pendapatkan diperoleh dari *cluster* tersebut.

Selanjutnya adalah penelitian [6] *clustering* dengan algoritma AGRID+ untuk mengektrak data penangkapan ikan di sekitar Samudra Hindia dengan data dari tahun 2000 sampai dengan tahun 2004 yang diambil dari PT. Perikanan Nusantara Indonesia. Suatu daerah dinyatakan potensial jika jumlah tangkapan sama atau lebih 5 di tempat yang sama. Jumlah tersebut diambil dari nilai minimum tangkapan. Dari hasil penelitian ini diperoleh 22 tempat potensial untuk harian, 31 zona untuk mingguan dan 21 zona untuk penangkapan bulanan.

Data yang digunakan dalam *clustering* bisa jadi memiliki berbagai jenis data seperti numerik, nominal dan lainnya. Untuk mengelompokkan data numerik, pengukuran jarak berdasarkan konsep geometris seperti *Euclidean distance* atau *Manhattan distance*. Sedangkan untuk data nominal tidak mungkin menggunakan jarak geometris maka dapat diganti dengan *simple matching* atau pengukuran *mismatching* seperti yang ada pada algoritma K-modes [7]. Data yang digunakan pada penelitian ini adalah data sensus penduduk lansia Sumatera selatan pada Tahun 2010 yang memiliki berbagai jenis data. Dengan demikian penelitian ini akan menggunakan algoritma K-modes yang ada di R-Studio.

1.2 Rumusan Masalah

Research question pada penelitian ini adalah:

1. Belum dilakukan klasterisasi penduduk lansia di Sumsel.
2. Belum diketahui hal baru yang ditemui dari pengelompokan dan analisis klasterisasi data penduduk lansia Sumatera Selatan.

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah sebagai berikut: mengklaster data penduduk lansia Sumatera selatan dan menganalisis kemiripannya yang dapat digunakan sebagai landasan pengambilan kebijakan dan program untuk lansia bagi pemerintah untuk meningkatkan kesejahteraan penduduk Sumatera selatan.

II. Tinjauan Pustaka

2.1. *Data Mining*

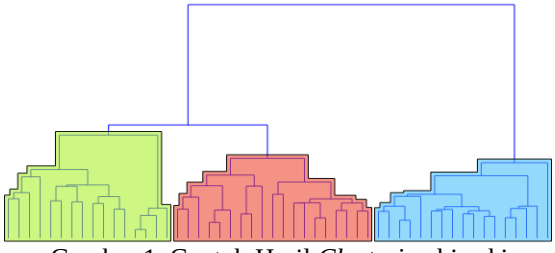
Data mining sering juga dikenal dengan *Knowledge Discovery from Databases (KDD)*. *Data mining* adalah proses ektraksi informasi yang tersembunyi, yang sebelumnya tidak diketahui dari data yang berskala besar. Ada banyak metode dari *data mining*, antara lain adalah klasifikasi, regresi, *clustering*, dan asosiasi. Langkah-langkah *data mining* dalam proses penemuan pengetahuan adalah sebagai berikut [8]:

1. *Data Cleaning*: penghapusan data yang tidak lengkap dan data yang tidak konsisten.
2. *Data Integration*: mengintegrasikan data yang berasal dari beberapa sumber.
3. *Data Selection*: hanya data yang relevan yang akan digunakan untuk analisis.
4. *Data Transformation*: transformasi data ke dalam bentuk format yang dibutuhkan.
5. *Data mining*: menggunakan metode data mining untuk melakukan ekstrak pola dari data.
6. *Pattern evaluation*: mengidentifikasi pola yang menarik.
7. *Knowledge presentation*: menyajikan pola yang telah diperoleh dengan visualisasi dan representasi yang menarik.

2.2. *Clustering*

Clustering adalah proses pengorganisasian objek ke dalam *class* / kelompok dengan cara mencari kemiripan setiap objek [7]. Sebuah *cluster* adalah kumpulan objek data yang mirip satu sama lain dalam *cluster* yang sama dan memiliki perbedaan dengan *cluster* yang lain. *Clustering* merupakan pembelajaran yang tidak terawasi dimana tidak memerlukan target *output* atau yang sering disebut *unsupervised learning*.

Clustering dibagi menjadi 2 metode yaitu: *hierarchical* dan *non-hierarchical*. *Hierarchical* adalah metode pengelompokan data yang diawali dengan mengelompokkan beberapa objek yang memiliki kesamaan paling dekat. Kemudian diteruskan ke objek lain yang memiliki kedekatan lainnya. Sehingga, *cluster* akan membentuk pohon/hirarki dari yang paling mirip sampai yang tidak mirip (Gambar 1).

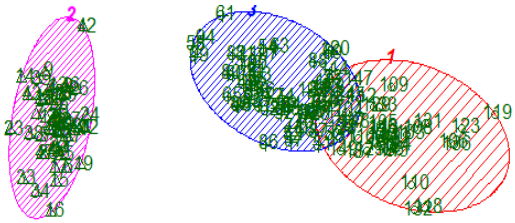


Gambar 1. Contoh Hasil *Clustering* hirarki

Sedangkah metode non-hirarki dimulai dengan menentukan terlebih dahulu jumlah *cluster* baru dilakukan proses *clustering*. Contoh hasil visual non-hirarki dapat dilihat pada Gambar 2. Hirarki digunakan apabila belum diketahui jumlah *cluster*, sedangkan

p-ISSN : 2339-1103
e-ISSN : 2579-4221

metode non-hirarki bertujuan mengelompokan objek ke kedalam k buah *cluster* [8].



Gambar 2. Contoh Hasil *Clustering* non-hirarki

2.3. K-Modes

Ada banyak metode clustering salah satunya adalah K-modes. K-modes adalah hasil modifikasi dari algoritma k-means. K-means merupakan algoritma yang sangat handal dalam mengelompokan data besar tetapi k-means tidak bisa diterapkan pada data selain numerik [9]. Pendekatan K-mode memodifikasi proses k-means standar untuk mengelompokan data dengan menganti fungsi jarak *Euclidean* dengan jarak *simple matching* dan menggunakan *mode* untuk mewakili pusat *cluster* [7]. Selain itu k-modes menggunakan metode frekuensi untuk memperbaharui mode [8]. Fungsi utama dari algoritma k-modes adalah persamaan 1.

$$F = \sum_{k=1}^{k} \sum_{x_i \in C_k} d(X_i, M_k) \qquad (1)$$

Dimana M_k adalah *mode* dari *cluster* C_k. Nilai atribut yang paling sering muncul akan dipilih sebagai *mode*. Setiap objek akan dibandingkan dengan *mode* menggunakan jarak *simple matching* hasil dari persamaan 2 dan setiap objek akan dialokasikan ke *cluster* terdekat.

$$d(X, Y) = \sum_{j=1}^{D} d(x_j, y_j), \qquad (2)$$

dimana:

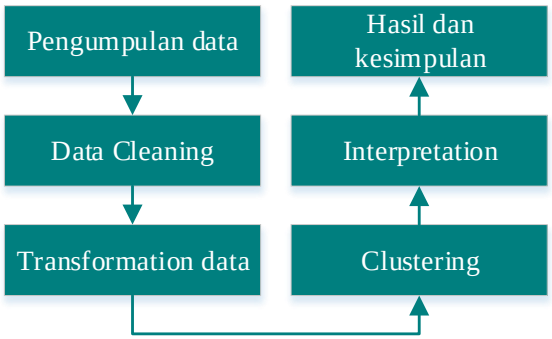
$$d(x_j, y_j) = \begin{cases} 0; \; if \; x_j = y_j \\ 1; otherwise \end{cases}$$

Ukuran ketidaksamaan antara X dan Y dapat didefinisikan oleh total ketidakcocokan dari X dan Y [10,11]. Semakin kecil jumlah ketidakcocokan, semakin mirip dua objek tersebut. Adapun cara kerja dari metode k-modes adalah sebagai berikut [10]:

1. Pilih k objek yang paling sering muncul sebagai pusat awal (*mode*) untuk setiap *cluster*.
2. Hitung jarak masing-masing objek ke *mode cluster* dengan menggunakan *simple matching* dengan persamaan 2. Tetapkan *cluster* setiap objek dengan melihat jarak pusat *cluster* terdekat.
3. Ulangi langkah 2 sampai semua objek dialokasikan ke *cluster* masing-masing.
4. Uji ulang objek dengan *mode* baru dan bandingkan dengan *mode* sebelumnya. Jika ada *cluster* yang berubah, kembali ke langkah 2; jika tidak, stop.

III. Metode Penelitian

Adapun tahap-tahap penelitian ini dapat dilihat pada Gambar 3.



Gambar 3. Langkah-langkah Penelitian

1. Pengumpulan data
   Pengumpulan data sensus peduduk dilakukan melalui wawancara tatap muka secara langsung untuk setiap penduduk Indonesia. Berdasarkan Undang-undang Statistik No. 16 tahun 1997 Pasal 8 ayat 1 ada tiga macam sensus di Indonesia. Salah satunya adalah sensus penduduk yang dilaksanakan pada tahun yang berakhiran 0. Data yang digunakan dalam penelitian ini adalah data sensus penduduk Sumatera selatan yang usianya > 60 tahun dengan total sampel 47.358 jiwa.
2. *Data Cleaning*
   Dari data yang diperoleh hanya dibutuhkan pemilihan kolom mana saja yang akan digunakan seperti kolom provinsi karena semua nilainya sama yaitu 16 yang merupakan kode provinsi sumatera selatan. Selain itu, juga dilakukan penghapusan atribut bulan dan tanggal lahir. Selain pengurangan kolom juga dilakukan penghapusan beberapa data karena data tersebut tidak relevan.
3. *Transformation data*
   Untuk melakukan *clustering* data harus ditransformasikan terlebih dahulu. Dari data sensus kolom yang dilakukan transformasi adalah kolom tahun lahir menjadi kolom umur.
4. *Clustering*
   Setelah data ditransformasikan maka sudah dapat digunakan untuk proses *clustering*. Algoritma yang digunakan pada penelitiian ini adalah K-modes dan *tools* yang digunakan adalah R-Studio.
5. *Interpretation*

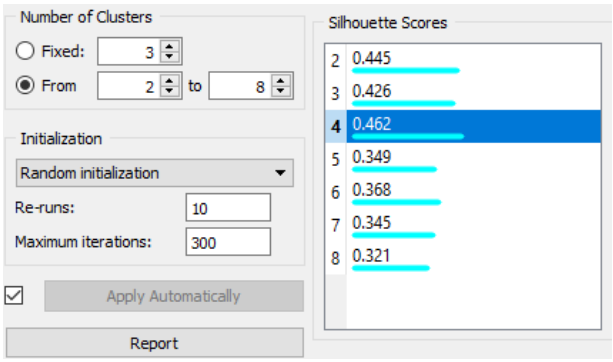
p-ISSN : 2339-1103
e-ISSN : 2579-4221

Dari hasil klasterisasi yang diperoleh akan dilakukan Analisa hasil pola yang ditemukan setiap *cluster*.

6. Hasil dan kesimpulan
   Dari pola setiap *cluster* maka dapat ditarik kesimpulan dari penelitian ini.

IV. Hasil dan Pembahasan

Sebelum dilakukan klasterisasi menggunakan k-modes perlu ditentukan jumlah *cluster* (k) terlebih dahulu. Untuk menentukan nilai k dibutuhkan perhitungan *silhouette scores*. *Silhouette scores* yang paling mendekati 1 menentukan k yang paling baik. Data sesus penduduk lansia 2010 sumatera selatan diperoleh k=4 dimana *silhouette scores* dapat dilihat pada Gambar 4.

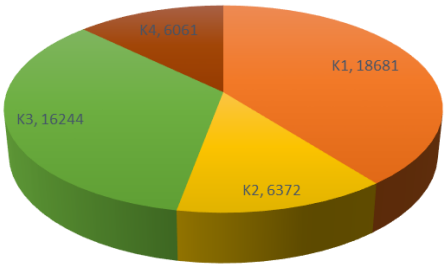


Gambar 4. Perhitungan *Silhouette Scores* untuk Setiap k

Setelah diperoleh jumlah *cluster*, maka data penduduk lansia dapat kelompokan dengan metode k-modes. Untuk melakukan *clustering* dengan k-modes harus menggunakan *code* program di bawah ini:

```
cluster.results <-kmodes(data.to.
cluster [,1:35], 4, iter.max = 500,
weighted = FALSE)
```

Dimana nilai 1:35 menunjukkan jumlah atribut atau kolom dari *dataset*. Sedangkan nilai 4 adalah jumlah *cluster* (k). Hasil *clustering* dapat diperoleh jumlah objek untuk setiap *cluster* yang dapat dilihat pada Gambar 5. K1 berjumlah 16244 orang, K2 6061 orang, K3 18681 orang dan K4 6372 orang.

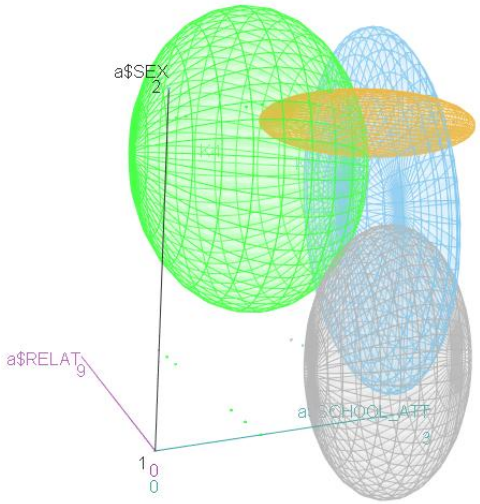


Gambar 5. Diagram Pai Jumlah Objek Setiap Cluster

Dari hasil Analisa diperoleh similariti dari setiap cluster, dimana similaritinya sebagai berikut:

1. K1: 95,8% bekerja, 94% kepala keluarga, 87% laki-laki, 80% status kawin, 77% tinggal di desa, 70 % bekerja di bidang pertanian dan perkebunan, 40% dibantu buruh tidak tetap dan 12 % tidak pernah sekolah.
2. K2: 99% wanita, 94% bekerja, 85% tinggal di desa, 74% bekerja dibidang pertanian dan perkebunan milik keluarga, 74% status kawin.
3. K3: 99,8% tidak bekerja dan tidak bersedia lagi bekerja, 58% tinggal di kota dimana 40% tinggal di kota Palembang, 68% perempuan, 55% status kawin dan 41% cerai mati dan 25% tinggal sendiri. Pada cluster ini juga banyak lansia mengalami kesulitan tinggat parah pada: penglihatan (3,2%), berjalan dan naik tangga (3,4), mendengar (2,3%), mengurus diri sendiri (2,5%) dan mengingat (2%).
4. K4: 97% tidak bekerja dan tidak bersedia bekerja lagi, 90% tinggal di desa, 83% perempuan, 79% buta huruf, 73% tidak pernah sekolah, 25% sekolah tidak tamat SD dan 73% cerai mati. Pada cluster ini juga tinggi jumlah lansia yang memiliki kesulitan pada penglihatan (5,8%), pendengaran (5,87%), berjalan dan naik tangga (6,4%), menginggat (5%) dan mengurus diri sendiri (5%). Dari 6061 orang pada cluster ini terdapat 90 orang yang mengalami kesulitan untuk semuanya.

Dari hasil *clustering* dapat kita visualisasikan gambaran setiap *cluster* terhadap 3 atribut yang dominan seperti yang ada pada Gambar 6.



Gambar 6. Visualisasi 3D Hasil *Clustering*

V. Kesimpulan dan Saran

5.1. Kesimpulan

Adapun kesimpulan penelitian ini adalah:

p-ISSN : 2339-1103
e-ISSN : 2579-4221

1. Dari hasil penelitian ini diperoleh 4 *cluster* lansia dimana setiap *cluster* memiliki pola yang berbeda. *Cluster* K1 adalah laki-laki yang masih bekerja di bidang pertanian dan perkebunan yang sebagian besar tinggal di desa. K2 adalah *cluster* dimana sebagian besarnya adalah perempuan yang masih bekerja di bidang pertanian dan perkebunan di desa. *Cluster* ketiga (K3) adalah kelompok lansia yang tidak bekerja dan hampir sebagian tinggal di Palembang. *Cluster* terakhir K4 adalah kelompok sebagian besar janda yang tinggal di desa dan buta huruf.
2. Pola yang dihasilkan dari keempat cluster menunjukkan bahwa banyaknya lansia (laki-laki dan perempuan) yang tinggal di desa dan masih bekerja namun sebagian besar bekerja pada sektor pertanian dan perkebunan, selebihnya sudah tidak mau bekerja lagi. Sedangkan lansia perempuan yang tinggal dikota sebagian besar menyandang status janda (cerai mati) dan sudah tidak mau bekerja lagi karena mengalami kesulitan penglihatan, pendengaran dan lain sebagainya.

5.2. Saran

Dari keempat cluster yang dihasilkan penelitian ini, diharapkan pemerintah dapat menentukan kebijakan atau program yang paling tepat untuk setiap cluster. Sehingga meningkatkan kesejahteraan lansia dan meringankan beban masyarakat di provisni Sumatera selatan. Untuk peneliti selanjutnya diharapkan mampu melakukan analisis dengan metode lain yang lebih baik.

Daftar Pustaka

[1] Badan Pusat Statistik Suamtera Selatan, (2011). Statistik Penduduk Lanjut Usia Sumatera Selatan 2010. Palembang: Badan Pusat Statistik.
[2] Badan Pusat Statistik Suamtera Selatan, (2016). Statistik Penduduk Lanjut Usia Sumatera Selatan 2015. Palembang: Badan Pusat Statistik.
[3] Melpa, B., dan Latipa, H., (2015). Analisis Clustering Menggunakan Metode K-means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila. Jurnal Media Infotama, Vol.11, No.2, pp.110-118.
[4] Asroni, and Andrian, R., (2015). Penerapan Methode K-means untuk Clustering Mahasiswa Berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus Jurusan Teknik Informatika UMM Magelang. Jurnal Ilmiah Semesta Teknika, Vol. 18, No.1, pp.76-82.
[5] Rajagopal, Sankar. (2011). Customer Data Clustering Using Data Mining Technique. International Journal of Database Management Systems (IJDMS), Vol.3, No.4, pp. 1-11.
[6] Fitrianah, D. et al., (2016). A Data Mining based Approach for Determining the Potential Fishing Zones. International Journal of Information and Education Technology, Vol. 6, pp. 187-191.
[7] Aranganayagi, S., and Thangavel, K., (2009). Improved K-Modes for Categorical Clustering Using Weighted Dissimilarity Measure. International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol.3, No.3, pp. 729-735.
[8] Han, J., and Kamber, M., (2006). Data Mining: Concepts and Techniques 2nd. United States of America: Elsevier.
[9] Xiang, Z., and Zahidul, M., (2014). Hartigan's Method for K-modes Clustering and Its Advantages. Proceedings of the Australasian Data Mining Conference, Vol.158.
[10] Huang, Z., (1997) A Fast Clustering Algorithm to cluster Very Large Categorical Datasets in Data Mining", In Proc. SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery.
[11] Zhou, Zhang, and Liu, (2017). A Global-Relationship Dissimilarity Measure for the k-Modes Clustering Algorithm. Computational Intelligence and Neuroscience, Vol. 2017.