

Categorical Data Clustering with K-Modes Method on Fire Cases in DKI Jakarta Province

Widia Handa Riska, Dony Permana*, Atus Amadi Putra, dan Zilrahmi

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: donypermana@fmipa.unp.ac.id

Submitted : 06 Oktober 2023

Revised : 25 Januari 2024

Accepted : 20 Februari 2024

ABSTRACT

In DKI Jakarta province, the number of fire cases fluctuates every year. In order not to cause a lot of property losses and casualties, efforts are needed to prevent and reduce the risk of the fluctuating number of fire cases. BPBD DKI Jakarta has responsibility for this. But in order for these efforts to be effective, information is needed about the patterns of fire cases that often occur. Fire patterns can be seen using K-Modes categorical clustering analysis. The data used is fire data in DKI Jakarta in 2018 which totalled 690 fire cases. In the clustering process, each fire case that occurred was clustered. Thus, the optimal number of clusters is obtained as many as 6 clusters based on the Davies Bouldin Index value with the smallest DBI value of 6.22. Of the six clusters, cluster 3 is the cluster with the highest number of fire cases. Cluster 3 has a centroid which is fires that occur on Fridays, in November, in Cakung Sub-district, due to electrical short circuit, burning houses and rarely causing minor injuries, serious injuries, or death.

Keywords: *Categorical Data Clustering, Davies Bouldin Index, Fire Cases, K-Modes*

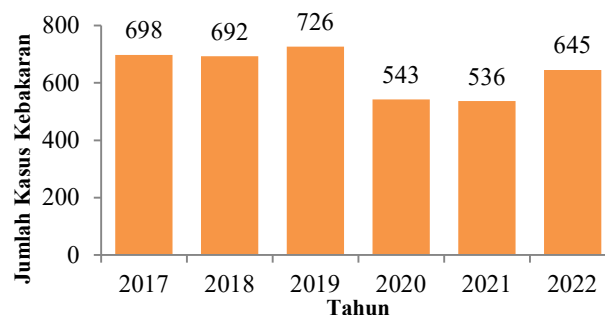


This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Kebakaran merupakan salah satu bencana yang terjadi secara tidak terduga yang dapat menimbulkan kerugian, baik harta benda, maupun korban jiwa. Kebakaran menurut Badan Nasional Penanggulangan Bencana (BNPB) adalah keadaan dimana bangunan pada suatu tempat seperti rumah/pemukiman, pasar, pabrik, gedung dan lain-lain dilahap api, sehingga menimbulkan korban dan/atau kerugian. Adapun kebakaran disebabkan oleh banyak hal, diantaranya, yaitu korsleting listrik, instalasi listrik yang tidak sesuai standar, tabung gas meledak, pembakaran sampah, membuang puntung rokok di sembarang tempat, dan pembakaran ladang Lathifah (2021: 15–16).

Daerah padat penduduk, sangat rentan terhadap bahaya kebakaran. Padatnya pemukiman penduduk mengakibatkan api menjalar secara meluas saat terjadinya kebakaran. Di samping itu, masyarakat perkotaan tidak terlepas dari pemakaian energi, seperti energi listrik maupun bahan bakar yang beraneka bentuk. Semakin banyaknya penggunaan energi yang dapat menimbulkan api, maka semakin tinggi pula potensi terjadinya kebakaran (Manalu dan Yotenka, 2013: 12–13). Menurut data dari Badan Pusat Statistik (BPS), provinsi dengan kepadatan penduduk tertinggi di Indonesia dipegang oleh Provinsi DKI Jakarta.



Gambar 1. Grafik Jumlah Kasus Kebakaran di DKI Jakarta Tahun 2017-2022

Dari Gambar 1 dapat dilihat bahwa jumlah kasus kebakaran di Jakarta mengalami fluktuasi dari tahun 2017 sampai 2022. Jika diperhatikan kembali, jumlah kasus kebakaran di tahun 2021 merupakan yang terendah semenjak

tahun 2017, lalu mengalami kenaikan yang cukup tinggi pada tahun 2022. Kenaikan jumlah kasus kebakaran ini tidak bisa dibiarkan. Diperlukan tidak lanjut dari pihak yang berwenang dalam pencegahan dan penurunan risiko bencana, yaitu Badan Penanggulangan Bencana Daerah (BPBD).

BPBD DKI Jakarta memiliki tugas melakukan upaya yang dapat mengurangi angka kebakaran dan menurunkan risiko kebakaran, agar tidak menimbulkan lebih banyak kerugian harta benda maupun korban jiwa. Namun, upaya tersebut memerlukan informasi mengenai pola kasus kebakaran yang sering terjadi dengan melakukan pengelompokan menggunakan analisis *clustering*. Pola kasus kebakaran dapat dikelompokkan berdasarkan waktu, lokasi, penyebab, ada atau tidaknya korban (luka ringan, luka berat, dan meninggal), dan sarana yang rusak oleh kebakaran. Semua variabel tersebut tersaji dalam bentuk data kategorik. Oleh karena itu, diperlukan suatu metode *clustering* yang dapat mengolah data kategorik.

Metode *clustering* yang sering digunakan adalah *K-Means*. Namun, metode ini hanya dapat mengolah data numerik. Pada tahun 1998, Huang mengembangkan metode *clustering* data kategorik yang diberi nama *K-Modes*. Huang menyatakan bahwa *K-Modes* merupakan pengembangan dari algoritma *k-means* untuk menangani data kategorik dimana means diganti oleh modus, sehingga *K-Modes* lebih efisien digunakan pada data kategorik (Prastya dkk, 2021). Oleh karena itu, metode *K-Modes* cocok untuk digunakan pada penelitian ini. Tujuan dari penelitian ini adalah untuk melihat kasus kebakaran seperti apa ataupun pola kebakaran yang sering terjadi di DKI Jakarta. *Cluster* dengan banyak kasus kebakaran tertinggi, akan menjadi prioritas bagi BPBD dalam upaya pencegahan dan penurunan. Sehingga, upaya tersebut dapat berjalan efektif.

Pada penelitian terdahulu, terdapat penelitian dengan metode *K-Modes Clustering* yang dilakukan oleh Jumeilah dan Pratama (2018). Penelitian tersebut mengelompokkan penduduk usia kerja di Provinsi Sumatera Selatan ke dalam 8 *cluster*. *Cluster* 1 mayoritas berisikan laki-laki yang tinggal di desa dan masih bekerja. *Cluster* 2 mayoritas berisikan penduduk yang berumur 25 tahun yang masih sekolah. *Cluster* 3 mayoritas berisikan perempuan tinggal di desa dan tidak bekerja, dan seterusnya. Pada penelitian terdahulu lainnya, yaitu yang dilakukan oleh Badruttamam dkk (2020). Tujuan penelitian tersebut adalah untuk mempermudah pengiklan untuk memilih kanal Youtube yang tepat untuk beriklan sesuai kriteria yang diinginkan. Dengan metode *clustering K-Modes* dan teknik validasi *Davies Bouldin Index*, diperoleh enam kelompok kanal Youtube dengan nilai DBI terkecil yaitu sebesar 1,08.

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data pada penelitian ini merupakan data sekunder, yang diambil dari situs data.go.id atau satu data Indonesia. Data tersebut berisi data kasus kebakaran di DKI Jakarta pada tahun 2018. Data ini merupakan data kebakaran yang terakhir kali dipublikasikan oleh BPBD DKI Jakarta. Data ini berisikan 690 kasus kebakaran. Adapun variabel pada penelitian ini ada delapan variabel kategorik, yaitu variabel Hari (X_1), Bulan (X_2), Kecamatan (X_3), Penyebab (X_4), Sarana rusak (X_5), Luka Ringan (X_6), Luka Berat (X_7) dan Meninggal (X_8). Berikut ini penjelasan dari kedelapan variabel tersebut.

Tabel 1. Variabel Penelitian

Variabel	Nama Variabel	Keterangan
X_1	Hari	Hari terjadinya kebakaran.
X_2	Bulan	Bulan terjadinya kebakaran
X_3	Kecamatan	Kecamatan dimana terjadinya kebakaran.
X_4	Penyebab	Penyebab kebakaran.
X_5	Sarana Rusak	Sarana atau bangunan yang rusak akibat kebakaran.
X_6	Luka Ringan	Ada atau tidaknya korban luka ringan.
X_7	Luka Berat	Ada atau tidaknya korban luka berat.
X_8	Meninggal	Ada atau tidaknya korban meninggal dunia.

B. Teknik Analisis Data

Teknik analisis data pada penelitian ini menggunakan analisis *clustering* data kategorik *K-Modes*. Analisis data dilakukan dengan bantuan *software Python* dan *Microsoft Excel*. Di bawah ini dijelaskan tahapan-tahapan analisis data.

1. Menyiapkan data dengan *data preprocessing*

Data preprocessing merupakan tahapan penting sebelum melakukan analisis data. Banyak kesalahan-kesalahan pada data yang dapat mengganggu proses analisis data itu sendiri, seperti adanya *missing value*, pencilan, ataupun format data yang tidak sesuai dengan sistem. Data berkualitas rendah jika dianalisis dengan suatu metode akan menghasilkan hasil analisis yang berkualitas rendah pula. Oleh karena itu perlu dilakukan pemrosesan data awal yang bertujuan untuk meningkatkan efisiensi dan kemudahan dalam proses analisis data (Marisa, 2021: 18). Menurut Marisa dkk (2021: 19), ada empat teknik dalam data preprocessing, yaitu *data cleaning*, *data transformation*, *data reduction*, dan *data integration*. Namun, tidak semua teknik tersebut harus dilakukan pada data. Empat teknik tersebut dapat dilakukan jika dibutuhkan.

Pada penelitian ini dilakukan tahapan *preprocessing* berupa *data cleaning* dan *data transformation*. Proses data cleaning yang dilakukan berupa memperbaiki ejaan atau tulisan yang salah (kesalahan struktural) dan menghilangkan *missing value* (data hilang). Lalu, dilakukan proses *data transformation* berupa pemilihan atribut (variabel) yang dibutuhkan dan memberi kode atau label pada tiap kategori variabel penelitian.

2. Analisis statistik deskriptif

Analisis statistik deskriptif bertujuan untuk melihat karakteristik data secara umum. Pada penelitian ini hanya terdapat variabel kategorik. Sehingga analisis statistik deskriptif menampilkan modus dan jumlah kategori tiap variabel.

3. Mengelompokkan data menggunakan *K-Modes clustering*

Kekurangan utama dari metode *clustering K-Means* adalah tidak bisa digunakan pada variabel kategorik (Aggarwal dan Reddy, 2014: 94). Lalu, pada tahun 1988, Huang mengemukakan metode *K-Modes clustering* yang merupakan pengembangan dari metode *K-Means Clustering* yang dapat digunakan untuk mengelompokkan data kategorik. Metode *K-Modes* mampu menghasilkan *cluster* yang lebih optimal dan komputasi lebih cepat daripada metode *K-Means* (Maulana dan Yotenka, 2023). *K-Modes* menggunakan suatu ukuran jarak (*dissimilarity*) berupa kecocokan nilai variabel tiap titik objek terhadap titik pusat/*centroid*. Pada metode *K-Modes*, *centroid* diperbarui dengan modus untuk meminimalkan jarak dari tiap objek ke masing-masing pusat *cluster* (Yulianton dkk, 2021). Objek yang di-*cluster* pada penelitian ini adalah kasus-kasus kebakaran yang terjadi.

Jarak antar objek pengamatan pada metode *K-Modes* dihitung dengan ukuran ketidaksamaan. Misalkan X dan Y adalah dua objek pengamatan dengan m variabel kategorik. Ukuran ketidaksamaan antara X dan Y dapat didefinisikan sebagai total ketidakcocokan dari variabel kategorik yang sesuai dari kedua objek tersebut. Semakin kecil nilai ketidakcocokan, semakin mirip kedua objek tersebut. Ukuran ini sering disebut dengan *simple matching dissimilarity* atau ukuran ketidaksamaan sederhana (Huang, 1998). Secara matematis dirumuskan dengan persamaan di bawah ini.

$$d(X, Y) = \sum_{j=1}^m \delta(x_j, y_j) \quad (1)$$

Dimana

$$\delta(x_j, y_j) = \begin{cases} 0, & x_j = y_j \\ 1, & x_j \neq y_j \end{cases} \quad (2)$$

Keterangan:

$d(X, Y)$: jarak objek (kasus kebakaran) X ke objek (kasus kebakaran) Y

m : banyak variabel

x_j : nilai objek X pada variabel ke-j

y_j : nilai objek Y pada variabel ke-j

$\delta(x_j, y_j)$: nilai ketidakcocokan antara objek X dan Y pada variabel ke-j

Menurut Hidayah dkk (2022), berikut langkah-langkah *clustering* menggunakan metode *K-Modes*.

- Tentukan jumlah *cluster* (k) dan pilih secara acak sebanyak k objek untuk dijadikan *centroid* awal, tiap *cluster* memiliki satu *centroid*.
 - Hitung jarak masing-masing objek terhadap semua *centroid* awal dengan Persamaan (1) dan (2). Lalu, masing-masing objek dialokasikan ke dalam *cluster* yang *centroid*-nya memiliki jarak terdekat dengan objek.
 - Perbarui *centroid* masing-masing *cluster* dengan nilai kategori yang sering muncul pada tiap *cluster* (modus).
 - Ulangi langkah b dan c hingga posisi *centroid* pada tiap *cluster* tidak berubah atau tidak ada objek yang berpindah *cluster*.
 - Proses *K-Modes clustering* selesai jika sudah memenuhi syarat pada langkah d.
4. Memilih jumlah *cluster* optimal menggunakan *Davies Bouldin Index* (DBI)

Sesuai namanya, *Davies Bouldin Index* pertama kali diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. DBI merupakan teknik validasi *cluster* yang digunakan untuk menentukan jumlah *cluster* optimal setelah tahapan *clustering* selesai. DBI bertujuan untuk memaksimalkan jarak antar *cluster* dan meminimalkan jarak antar objek dalam satu *cluster*. Semakin kecil nilai DBI dimana $DBI \geq 0$, maka semakin baik *cluster* yang diperoleh (Az-Zahra dkk, 2021).

Menurut Butsianto dan Saepudin (2020), nilai DBI diperoleh dengan langkah-langkah sebagai berikut.

- a. Menghitung *Sum of Square Within-Cluster* (SSW)

SSW merupakan rata-rata jarak setiap objek dalam satu *cluster* terhadap *centroid*-nya atau kohesi. SSW diperoleh dengan persamaan berikut ini.

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (3)$$

Dimana

$$d(x_j, c_i) = \begin{cases} 0, & x_j = c_i \\ 1, & x_j \neq c_i \end{cases} \quad (4)$$

Keterangan:

SSW_i : nilai kedekatan setiap objek dalam *cluster* ke- i dengan *centroid*-nya

m_i : banyak objek pada *cluster* ke- i

x_j : nilai objek (X) ke- j pada suatu *cluster*

c_i : *centroid cluster* ke- i ($i = 1, 2, \dots, k$)

$d(x_j, c_i)$: jarak objek (X) ke- j pada *cluster* ke- i dengan *centroid cluster* ke- i

- b. Menghitung *Sum of Square Between-Cluster* (SSB)

SSB bertujuan untuk mengetahui jarak antar *cluster* atau separasi. SSB dihitung dengan rumus sebagai berikut.

$$SSB_{i,j} = d(C_i, C_j) \quad (5)$$

Dimana

$$d(C_i, C_j) = \begin{cases} 0, & c_i = c_j \\ 1, & c_i \neq c_j \end{cases} \quad (6)$$

Keterangan:

$SSB_{i,j}$: jarak antara *cluster* i (c_i) dengan *cluster* j (c_j) ($i = 1, 2, \dots, k; j = 1, 2, \dots, k; i \neq j$)

$d(C_i, C_j)$: jarak antara *centroid* ke- i dengan *centroid* ke- j , $i \neq j$ (dengan menggunakan rumus *simple matching dissimilarity*)

- c. Rasio

Perhitungan rasio bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j .

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (7)$$

Keterangan:

$R_{i,j}$: rasio antara *cluster* i dan *cluster* j , $i \neq j$

SSW_i : nilai kedekatan setiap objek dalam *cluster* i dengan *centroid*-nya

SSW_j : nilai kedekatan setiap objek dalam *cluster* j dengan *centroid*-nya

- d. Menghitung nilai DBI

Nilai rasio yang telah diperoleh digunakan untuk mencari nilai DBI menggunakan rumus sebagai berikut.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad ; j = 1, 2, \dots, k \quad (8)$$

Keterangan:

DBI : nilai *Davies Bouldin Index*

k : jumlah *cluster*

5. Menginterpretasi hasil *cluster*.

III. HASIL DAN PEMBAHASAN

Langkah pertama yang dilakukan adalah *data cleaning*. Proses *data cleaning* ada dua, yaitu memperbaiki kesalahan struktural dan menangani data hilang (*missing value*). Pertama, dilakukan proses memperbaiki kesalahan struktural pada data. Kesalahan struktural tersebut berupa penamaan yang aneh, salah ketik, atau kapitalisasi yang salah. Contohnya pada data adalah penulisan “Korselting Ilstrik”, yang seharusnya “Korsleting Listrik”. Cara menanganinya adalah dengan memperbaiki semua kesalahan struktural secara manual. Langkah kedua, yaitu penanganan data hilang. Ternyata terdapat 6 pengamatan dengan data hilang. Angka ini sangat kecil dibandingkan dengan besarnya data. Oleh karena itu, data hilang dapat diabaikan atau dihapus dari gugus data. Sehingga tersisa sebanyak 684 kasus kebakaran.

Setelah proses *data cleaning* selesai, dilakukanlah proses *data transformation*. *Data transformation* merupakan proses mengubah format data ke dalam bentuk yang sesuai untuk proses analisis data. Salah satu teknik

dalam *data transformation*, yaitu pemilihan atribut (variabel). Pada pemilihan atribut ini, akan dipilih atribut/variabel yang dibutuhkan dalam analisis data.

Setelah dilakukan *preprocessing data*, dilihat terlebih dahulu deskripsi dari masing-masing variabel penelitian dengan statistika deskriptif. Hal ini dilakukan untuk melihat karekeristik data secara umum. Deskripsi variabel penelitian disajikan pada Tabel 2.

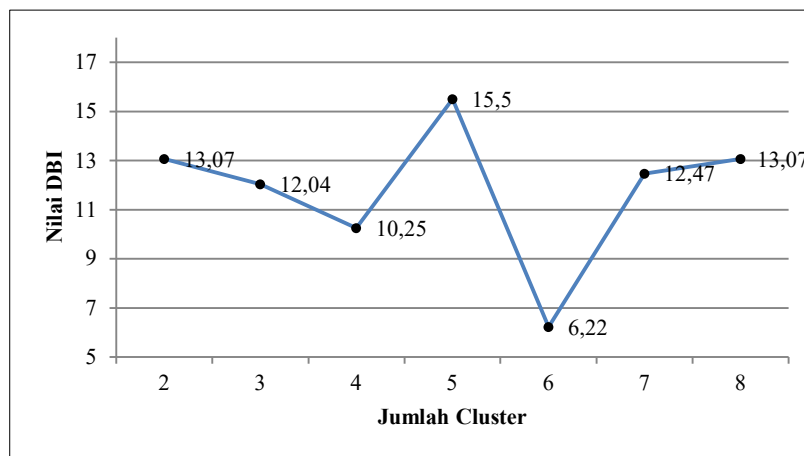
Tabel 2. Statistika Deskriptif

Nama Variabel	Banyak Kategori	Kategori dengan Frekuensi Terbanyak
Hari	7	Selasa
Bulan	12	Agustus
Kecamatan	61	Cakung
Penyebab	40	Korsleting Listrik
Sarana Rusak	90	Rumah Tinggal
Luka Ringan	2	Tidak Ada
Luka Berat	2	Tidak Ada
Meninggal	2	Tidak Ada

Dari Tabel 1 dapat dilihat bahwa variabel Sarana Rusak memiliki nilai atau kategori yang sangat beragam. Dimana kategori dengan frekuensi terbanyak dalam variabel Sarana Rusak adalah kategori “Rumah Tinggal”. Lalu, variabel Kecamatan juga sangat beragam dengan 61 kategori dengan modusnya, yaitu Kecamatan Cakung. Kemudian, diikuti oleh variabel Penyebab dengan 40 kategori dengan modusnya, yaitu Korsleting Listrik.

Sebelum dilakukan analisis *clustering K-Modes*, kategori-kategori pada masing-masing variabel penelitian harus diberi label atau kode. Jika tidak, hasil validasi *cluster* dengan menggunakan metode *Davies Bouldin Index* tidak akan muncul di *Python*. Pelabelan dilakukan dengan menggunakan *library* *LabelEncoder* yang di-*import* dari *sklearn.preprocessing*.

Langkah selanjutnya adalah analisis *clustering* dengan metode *K-Modes*. Tahapan ini dilakukan pada jumlah *cluster* yang berbeda-beda. Dari hasil *clustering* tersebut dipilih jumlah *cluster* optimal berdasarkan nilai *Davies Bouldin Index* (DBI). Hasil dari validasi *cluster* tersebut disajikan dalam Gambar 2.



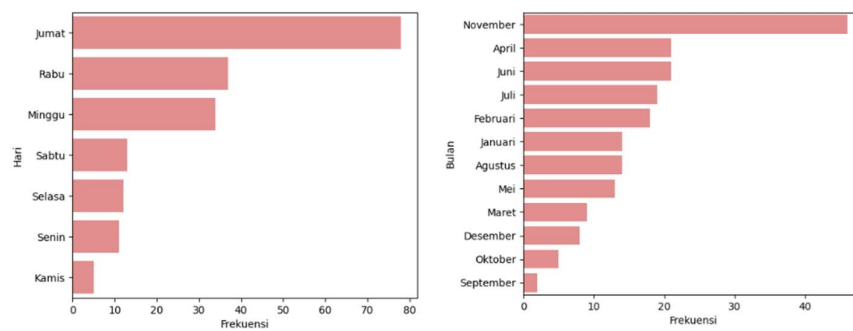
Gambar 2. Nilai DBI Pada Tiap Jumlah Cluster

Gambar 2 menunjukkan nilai indeks Davies Bouldin terkecil berada pada *cluster* 6, yaitu sebesar 6,22. Artinya jumlah *cluster* optimal adalah sebanyak 6 *cluster*. Berikut ini disajikan *centroid* dan jumlah dari keenam *cluster* tersebut pada Tabel 3.

Tabel 3. *Centroid* dan Jumlah Anggota *Cluster* yang Terbentuk

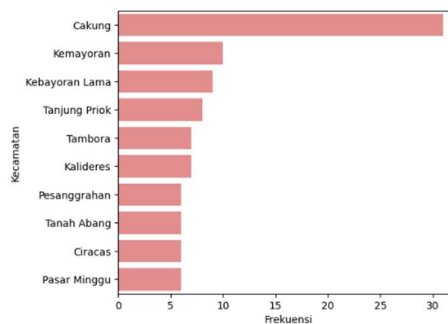
<i>Cluster</i>	<i>Centroid</i>	Jumlah Kasus Kebakaran
1	Selasa; Oktober; Duren Sawit; Korsleting Listrik; Rumah Tinggal; Tidak Ada; Tidak Ada; Tidak Ada	162
2	Senin; September; Penjaringan; Korsleting Listrik; Rumah Tinggal; Tidak Ada; Tidak Ada; Tidak Ada	105
3	Jumat; November; Cakung; Korsleting Listrik; Rumah Tinggal; Tidak Ada; Tidak Ada; Tidak Ada	190
4	Kamis; Agustus; Cilincing; Korsleting Listrik; Rumah Tinggal; Tidak Ada; Tidak Ada; Tidak Ada	121
5	Sabtu; Desember; Keramat Jati; Korsleting Listrik; Rumah Tinggal; Tidak Ada; Tidak Ada; Tidak Ada	88
6	Kamis; November; Grogol Petamburan; Korsleting Listrik; Rumah Tinggal; Tidak Ada; Tidak Ada; Tidak Ada	18

Pada Tabel 3 dapat dilihat bahwa *cluster* 3 memiliki kasus kebakaran tertinggi. Itu artinya *cluster* 3 memiliki pola kebakaran yang sering terjadi di DKI Jakarta pada Tahun 2018. Pola *cluster* yang memiliki jumlah kasus terbanyak dapat menjadi prioritas bagi BPBD DKI Jakarta dalam pencegahan dan pengurangan risiko kebakaran. Dari Tabel 3 juga dapat dilihat bahwa nilai *centroid* antar *cluster* tidak memiliki perbedaan yang besar. *Centroid* untuk setiap *cluster* bernilai sama pada 5 dari 8 variabel penelitian. Sehingga, *centroid* tidak mengandung informasi yang berguna. Untuk itu perlu ditelusuri lebih dalam pola pada tiap variabelnya. Langkah selanjutnya dikhususkan untuk melihat pola pada *cluster* 3 yang merupakan *cluster* dengan jumlah kasus kebakaran tertinggi. Pertama, ditinjau dari variabel yang mengandung waktu terjadinya kebakaran, yaitu variabel Hari dan Bulan.



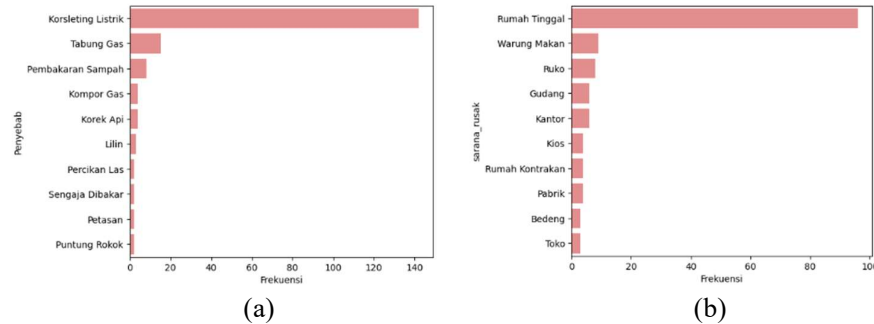
Gambar 3. (a) Frekuensi Kategori pada variabel Hari *cluster* 3, (b) Frekuensi Kategori pada variabel Bulan *cluster* 3

Dari Gambar 3(a) dapat dilihat bahwa pada hari Jumat paling sering terjadi kebakaran. Lalu pada Gambar 3(b) menunjukkan bahwa kebakaran paling sering terjadi di bulan November. Namun, tidak diketahui alasan pasti mengapa kebakaran sering terjadi di hari dan bulan tersebut. Selanjutnya ditinjau dari segi lokasi kebakaran, yaitu variabel Kecamatan.



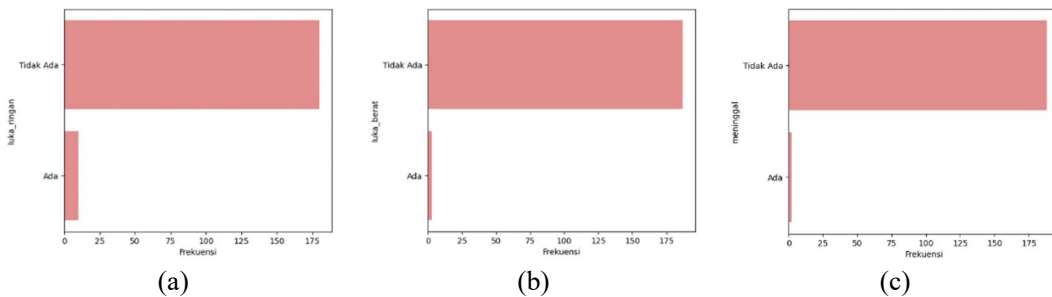
Gambar 4. Frekuensi 10 kategori teratas pada variabel Kecamatan *cluster* 3

Dari Gambar 4 dapat dilihat bahwa kecamatan Cakung merupakan kecamatan di DKI Jakarta yang paling sering mengalami kebakaran. Kecamatan Cakung merupakan kecamatan terpadat di Kota Jakarta Timur. Kemudian, kecamatan kedua yang paling sering terjadi kebakaran adalah Kecamatan Kemayoran. Kecamatan Kemayoran merupakan salah satu kecamatan terpadat di wilayah Kota Jakarta Pusat. Itu artinya, kecamatan dengan kepadatan penduduk yang tinggi, memiliki potensi yang lebih besar dalam terjadinya kebakaran. Kemudian, ditinjau dari segi penyebab dan sarana rusak.



Gambar 5. (a) Frekuensi 10 teratas pada variabel Penyebab cluster 3, (b) Frekuensi 10 kategori teratas pada variabel Sarana Rusak cluster 3

Dari Gambar 5 dapat dilihat korsleting listrik paling sering menjadi penyebab terjadinya kebakaran, terutama kebakaran di rumah tinggal. Hal ini disebabkan karena di rumah tinggal listrik digunakan di setiap hari dan sepanjang hari untuk memenuhi kebutuhan penghuni rumah. Lalu, tabung gas cukup sering menjadi penyebab kebakaran. Warung makan juga sering mengalami kebakaran, diakibatkan oleh pemakaian tabung gas untuk memasak makanan. Kemudian, ditinjau dari ada atau tidaknya korban akibat kebakaran.



Gambar 6. (a) Frekuensi kategori pada variabel Luka Ringan cluster 3, (b) Frekuensi kategori pada variabel Luka Berat cluster 3, dan (c) Frekuensi kategori pada variabel Meninggal cluster 3

Dari Gambar 6 dapat dilihat bahwa jumlah kasus kebakaran dengan tidak adanya korban, baik luka ringan, luka berat maupun meninggal sangatlah tinggi. Namun jika dilihat dari adanya korban, jumlah kasus dengan korban luka ringan, luka berat ataupun meninggal sangatlah kecil. Jumlah adanya korban luka ringan lebih besar dari jumlah adanya korban luka berat dan meninggal. Jumlah adanya korban luka berat dan korban meninggal hampir sama.

IV. KESIMPULAN

Analisis *clustering* pada penelitian dilakukan dengan mengclusterkan kasus-kasus kebakaran di DKI Jakarta tahun 2018. Adapun metode *clustering* yang digunakan adalah metode *K-Modes*. Dari analisis *clustering*, diperoleh jumlah *cluster* optimal sebanyak 6 *cluster* berdasarkan nilai *Davies Bouldin Index* yang terkecil, yaitu sebesar 6,22. Dari 6 *cluster* tersebut, *cluster* 3 memiliki jumlah kasus kebakaran terbanyak. Adapun pola kebakaran yang sering terjadi di *cluster* 3 (*centroid*) adalah terjadi di hari jumat, bulan November, di Kecamatan Cakung, oleh karena korsleting listrik, membakar rumah tinggal dan jarang sekali ada korban luka ringan, luka berat maupun meninggal.

Untuk peneliti selanjutnya disarankan untuk menggunakan data kasus kebakaran di DKI Jakarta yang terbaru. Dari data terbaru tersebut dapat diperoleh hasil penelitian relevan dengan keadaan terkini. Selain itu, peneliti selanjutnya juga dapat melakukan penelitian dengan data kasus kebakaran di daerah lain yang sering terjadi kebakaran.

DAFTAR PUSTAKA

- Aggarwal, C.C., & Reddy, C.K. (2014). *Data Clustering Algorithms and Applications*, New York: CRC Press.
- Az-Zahra, A. A., Marsaoly, A. F., Lestyani, I. P., Salsabila, R., & Madjida, W. O. Z. (2021), "Penerapan Algoritma K-Modes *Clustering* dengan Validasi Davies Bouldin Index pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta", *Jurnal Matematika Dan Statistika Serta Aplikasinya*, Vol.9, No. 1, hal. 24–36.
- BNPB. *Definisi Bencana*. Diakses pada 20 Maret 2023, dari <https://www.bnpb.go.id/definisi-bencana>
- Butsianto, S., & Saepudin, N. (2020), "Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means", *Jurnal Nasional Komputasi Dan Teknologi Informasi*, Vol. 3, No.1, hal. 51–59.
- Hidayah, I. N., Suhery, C., & Hidayati, R. (2022), "Implementasi Metode K-Modes Dalam Pengelompokan Penerima Bantuan Langsung Tunai", *JURIKOM (Jurnal Riset Komputer)*, Vol. 9, No. 6, hal. 2102–2110.
- Huang, Z. (1998), "Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values", *Data Mining and Knowledge Discovery*, Vol.2, hal. 283–304.
- Jumeilah, F. S., & Pratama, D. (2018), "Identifikasi Cluster Penduduk Usia Kerja Pada Provinsi Sumatera Selatan Menggunakan K-Modes", *Jurnal Komputer Terapan*, Vol. 4, No. 1, hal. 1–9.
- Lathifah. (2021). *Panduan Keselamatan saat Kebakaran*, Yogyakarta: Diva Press.
- Manalu, M. P., & Sandur, K. (2013). *Kebakaran di Jakarta Ancaman dan Solusinya*, Jakarta: Lestari Kiranatama.
- Marisa, F., Maukar, A.L., & Akhriza, T.M. (2021). *Data Mining Konsep dan Penerapannya*, Yogyakarta: Deepublish.
- Maulana, A.A, & Yotenka, R. (2023), "Analisis Clustering K-Modes Terhadap Kepuasan Sistem Rekrutmen Online Mitra Kerja BPS Kabupaten Bima Tahun 2022", *Emerging Statistics and Data Science Journal*, Vol. 1, No. 1, hal. 81-91.
- Prastya, S.E., Nurhaeni, & Zulfadhilah, M. (2021), "Penentuan Pola Kecelakaan Lalu Lintas Menggunakan K-Modes Clustering", *Jurnal Edik Informatika*, Vol. 8, No. 1, hal. 27–39.
- Yulianton, H., Sutanto, F., & Mulyani, S. (2021), "Pengelompokan Mahasiswa Berbasis Categorical Variables Menggunakan Metode K-Modes Clustering", *Proceeding SENDI_U*, hal. 424-429.