

Pengelompokan Perusahaan Berdasarkan Statistik Penelitian dan Pengembangan Melalui Metode *K-Medoids*

Shiddiq Sugiono*¹, Nursinta Adi Wahanani², Laila Juwita Hendriani³

Badan Riset dan Inovasi Nasional

e-mail: ¹shid002@brin.go.id, ²nurs004@brin.go.id, ³lail003@brin.go.id

*Penulis Korespondensi

Diterima: 8 Maret 2022; Direvisi: 28 April 2022; Disetujui: 18 Mei 2022

Abstrak

Kapasitas penelitian dan pengembangan sektor industri memiliki peran dalam peningkatan ekonomi negara sehingga perlu terus dipantau untuk meningkatkan daya saing negara. Statistik litbang merupakan salah satu indikator yang dapat digunakan untuk mengukur sejauh apa kapasitas suatu perusahaan dalam melakukan litbang melalui pengukuran pengeluaran dan personil litbang. Diperlukan suatu pengelompokan perusahaan untuk melacak performa litbang dari sektor industri sehingga dapat menjadi referensi terhadap pemantauan kapasitas litbang. Informasi mengenai jumlah pekerja dalam suatu perusahaan turut menjadi indikator untuk memberikan gambaran mengenai ukuran perusahaan dalam hasil pengelompokan. Penelitian ini bertujuan untuk mengelompokkan perusahaan berdasarkan statistik litbang menggunakan metode clustering *K-Medoids* sehingga didapatkan hasil pengelompokan yang optimal. Adapun penggunaan metode tersebut dikarenakan data yang diperoleh memiliki pencilan. Evaluasi hasil pengelompokan akan ditinjau melalui indeks silhouette. Hasil penelitian ini menyebutkan bahwa pengelompokan data litbang industri dengan menggunakan metode *K-Medoids* membentuk empat kelompok. Adapun berdasarkan hasil evaluasi indeks silhouette, struktur dari hasil pengelompokan masuk dalam kategori kuat. Profil dari masing-masing kelompok turut disampaikan pada bagian interpretasi. Selain itu hasil pengelompokan ini dapat menjadi pemetaan awal dalam membuat kebijakan litbang dalam sektor industri Indonesia.

Kata kunci: *K-Medoids*, Analisis Cluster, Personil Litbang, Pengeluaran Litbang, Data Mining

Abstract

The research and development capacity of the industrial sector has a role in improving the country's economy so it needs to be continuously monitored to increase states competitiveness. R&D statistics are an indicator that can be used to measure the extent to which a company's capacity to carry out R&D through the measurement of R&D expenditures and personnel. A clustering of companies is needed to track the R&D performance of the industrial sector. Information on the number of workers in a company also becomes an indicator to provide an overview of the size of the company in the grouping results. This study aims to group the companies based on R&D statistics using the *K-Medoids* clustering method. The evaluation will be reviewed through the silhouette index. The results of this study indicate that the most optimal grouping of industrial R&D data using the *K-Medoids* method is to form four groups. Meanwhile, based on the results of the evaluation of the silhouette index, the structure of the grouping results is in a strong category. In addition, the results of this clustering can be used as an initial mapping in making R&D policies in the Indonesian industrial sector.

Keywords: *K-Medoids*, Cluster Analysis, R&D Personnel, R&D Expenditures, Data Mining

1. PENDAHULUAN

Kegiatan litbang atau *R&D* di suatu negara memiliki peran penting dalam mencapai *Sustainable Development Goals* (SDGs) dalam bidang industri, inovasi dan infrastruktur sehingga perlu direncanakan secara matang pengembangan jangka panjangnya [1]. Pada 33 negara OECD, pengeluaran negara terhadap *R&D* memiliki dampak positif pada pertumbuhan negara [2], begitu juga pengeluaran pada 52 negara pada rentang tahun 1966-2010 memiliki pengaruh positif terhadap pertumbuhan ekonomi dalam jangka panjang [3]. Berangkat dari pentingnya *R&D* pada suatu negara maka suatu negara perlu terus memantau perkembangan kapasitas litbang dari setiap sektor.

Litbang sektor industri berperan penting untuk mendukung ekosistem litbang diantaranya dengan berkontribusi pada ekspansi pengetahuan dalam peningkatan efisiensi produk, mendorong penciptaan produk yang memiliki nilai jual dan berpengaruh signifikan terhadap peningkatan margin keuntungan perusahaan [4,5]. Sektor litbang turut memiliki peran untuk mendorong penciptaan suatu produk hasil penelitian yang memiliki nilai jual dan dapat diproduksi massal. Oleh karena itu, penting untuk memantau sejauh mana kapasitas penelitian dan pengembangan yang dilakukan di sektor industri [6].

Pengukuran kapasitas litbang nasional merupakan instrumen untuk perencanaan, pengelolaan serta pendukung pengambilan keputusan dan kontrol untuk kebijakan ilmiah nasional [7]. Litbang industri termasuk dalam sektor yang berkontribusi terhadap kapasitas litbang nasional. Pengukuran litbang sektor industri diantaranya dikembangkan oleh OECD yang indikatornya antara lain pengeluaran litbang dan personil litbang [6]. Penambahan indikator jumlah pekerja untuk memberikan informasi terhadap klasifikasi jenis perusahaan di suatu negara [8]. Pengelompokan industri berdasarkan karakteristiknya, dalam hal ini statistik litbang, diperlukan untuk mengetahui performa litbangnya serta kompleksitas kegiatan litbang di bidang bidang industri tertentu sehingga diharapkan dapat memberikan dampak kepada sektor ekonomi secara keseluruhan [9].

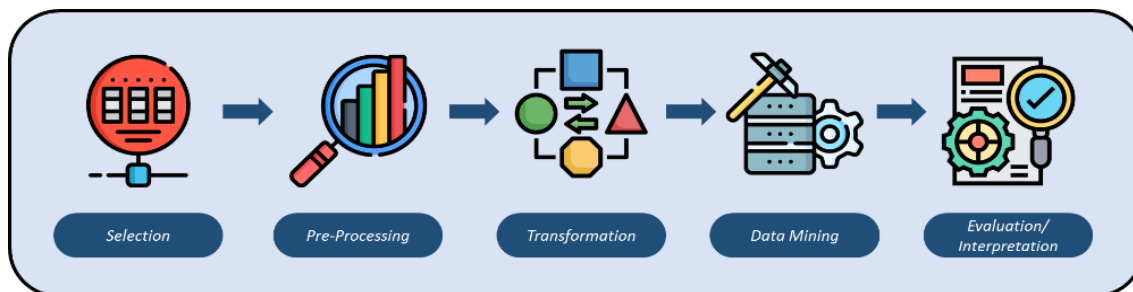
Metode data mining untuk mengelompokkan objek berdasarkan kesamaan karakteristik dalam data atau disebut *clustering* telah digunakan untuk memetakan potensi litbang. Penerapan analisis clustering hierarchial dan non-hierarchial antara lain pengelompokan berbagai negara berdasarkan negara yang terpengaruh akibat industri 4 berdasarkan *Global Innovative Index*, *Sustainable Development Goals Index*, *Logistic Performance Index* dan *Environmental Performance Index* [10]. *Clustering* dengan K-Means digunakan untuk mengelompokkan berbagai negara berdasarkan variabel persentase belanja litbang terhadap GDP, jumlah artikel jurnal, jumlah aplikasi paten, dan jumlah aplikasi merek dagang [11]. Terdapat kesempatan untuk meneliti pengelompokan perusahaan secara khusus dalam lingkup nasional berdasarkan statistik litbang industri dan dengan metode lainnya selain K-Means.

Tujuan penelitian ini adalah untuk mengelompokkan perusahaan berdasarkan RDS melalui metode K-Medoids. Dalam hal ini statistik litbang merupakan gambaran dari kegiatan litbang berbagai jenis perusahaan, baik skala besar maupun kecil, yang ada di Indonesia sehingga memungkinkan adanya data yang bersifat pencilan (*outlier*). Adapun algoritma *K-Medoids* lebih mampu menangani data numerik yang bersifat pencilan dibandingkan metode lainnya [12]. Evaluasi pengelompokan akan ditinjau melalui indeks *silhouette*. Kebaruan dari penelitian ini adalah pengelompokan perusahaan lingkup negara Indonesia dengan menggunakan data primer dan berdasarkan indikator Statistik Penelitian dan Pengembangan OECD. Adapun pengolahan data yang dilakukan pada penelitian ini akan menggunakan perangkat lunak *open source* Rstudio versi 2021.09.2-382.

2. METODE PENELITIAN

Penelitian ini akan mengimplementasikan kerangka kerja *Knowledge Discovery in Database* (KDD). Proses KDD adalah sekumpulan tahapan untuk menggali dan menemukan

pengetahuan-pengetahuan tersembunyi yang bermanfaat dari suatu data yang berasal dari basis data [13]. *Flowchart* dari tiap-tiap fase KDD yang akan dilakukan pada penelitian ini disampaikan pada gambar 1.



Gambar 1. Fase dalam KDD [13]

2.1. Selection

Fase seleksi ini merupakan tahap untuk mencari variabel (*field*) dalam database yang relevan terhadap tujuan yang akan dicapai. Adapun data yang diambil dari proses ini akan disimpan pada tempat yang terpisah dari database operasional. Data yang diambil pada bagian ini belum dapat diproses melalui data mining sehingga diperlukan tahap selanjutnya untuk menyiapkan data yang siap untuk diolah.

2.2. Pre-processing

Fase pre-processing bertujuan untuk membersihkan data yang diambil dari suatu database. Adapun proses ini mencakup berbagai macam jenis perlakuan seperti membuang data ganda/duplikasi data, memeriksa unit data yang tidak konsisten, memperbaiki kesalahan penulisan pada data (*typo*) dan pengamatan *missing value*. Selain itu dalam proses ini dimungkinkan juga menambah informasi eksternal lainnya yang diperlukan dalam proses data mining.

2.3. Transformation

Fase ini merupakan suatu proses *coding* untuk mentransformasi data yang telah dipilih sehingga dapat dioperasionalkan secara baik pada tahap data mining. Adapun pada fase ini data yang akan diolah akan dilakukan standarisasi data sehingga setiap variabel yang memiliki unit pengukuran berbeda dapat disamakan [14]. Adapun proses standarisasi data telah dipertimbangkan sebagai salah satu proses yang dapat membantu proses analisis *cluster* agar mendapatkan kelompok yang berkualitas [14]. Berikut rumus yang digunakan untuk melakukan standarisasi data:

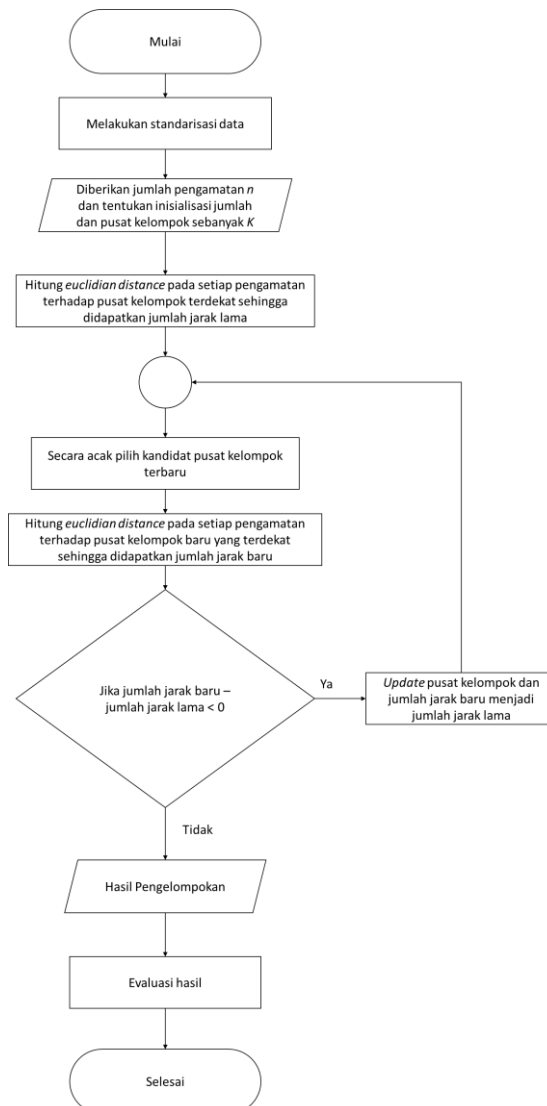
$$Z_i = \frac{X_i - \bar{X}}{\sigma} \tag{1}$$

Dimana Z_i adalah nilai hasil standarisasi *Z score* pengamatan ke- i , X_i adalah data pengamatan ke- i yang akan distandarisasi, \bar{X} adalah rata-rata dari dataset dan σ adalah standar deviasi dari dataset. Adapun data yang telah terstandarisasi akan memiliki rata-rata 0 dan standar deviasi 1.

2.4. Data Mining

Fase data mining merupakan suatu proses untuk mencari pola atau informasi yang tersembunyi dalam suatu data hasil seleksi dengan menggunakan suatu teknik maupun metode tertentu. Adapun pada tahap ini pula akan dilakukan pengujian asumsi seperti pengujian sampel

yang representatif dan pengujian multikolinearitas. Pada tahap ini, proses data mining akan dilakukan melalui metode K-Medoids clustering. Gambar 2 merupakan alur tahapan metode K-Medoids [15,16]:



Gambar 2. Alur metode K-Medoids [15]

Adapun penjelasan alur tersebut adalah sebagai berikut:

- Melakukan standarisasi data.
- Melakukan inisialisasi jumlah kelompok dan pusatnya sebanyak K .
- Memetakan setiap pengamatan atau objek ke pusat kelompok terdekat dengan menggunakan rumus jarak *Euclidian Distance* melalui persamaan berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

dimana n adalah jumlah pengamatan atau objek.

- Secara acak memilih objek di masing-masing kelompok sebagai kandidat pusat terbaru.
- Menghitung jarak dari setiap pengamatan yang berada pada masing-masing kelompok dengan kandidat pusat terbaru.
- Hitung total simpangan dengan menghitung jumlah jarak baru – jumlah jarak lama. Jika total simpangan < 0 maka tukar pusat kelompok dengan objek lainnya untuk membentuk sekumpulan K objek baru sebagai pusat.

- g. Ulangi langkah ke-3 sampai ke-5 hingga tidak terjadi kembali perubahan pusat kelompok, sehingga didapatkan kelompok dan anggotanya masing-masing.
- h. Melakukan perhitungan nilai evaluasi sebagai indikator untuk menilai konsistensi kelompok.

2.5. Evaluation and Interpretation

Fase ini merupakan proses untuk menafsirkan informasi maupun pola yang diperoleh dari metode data mining sehingga mudah dimengerti oleh pihak-pihak yang berkepentingan. Adapun fase ini turut menilai validitas dari metode yang telah digunakan sehingga diperoleh kelompok yang memiliki konsistensi tinggi dimana karakteristik anggota suatu kelompok berbeda dengan kelompok yang lainnya. Kriteria yang akan digunakan dalam mengevaluasi hasil pengelompokan ada indeks *silhouette* dimana indeks tersebut mengukur seberapa mirip suatu titik terhadap pusat kelompoknya jika dibandingkan dengan pusat kelompok lainnya [17]. Adapun kriteria evaluasi yang digunakan adalah statistik indeks *silhouette* dengan interpretasi pada tabel 1 sebagai berikut:

Tabel 1. Kriteria evaluasi indeks *silhouette* [18]

<i>Index Silhouette</i>	Interpretasi
0.71 – 1.00	Sebuah struktur kelompok yang kuat telah terbentuk. Pada tingkatan ini, suatu pusat kelompok dapat secara jelas memisahkan tiap-tiap pengamatan menjadi anggota kelompok tertentu berdasarkan karakteristiknya.
0.51 – 0.70	Sebuah struktur kelompok yang layak telah terbentuk. Pada tingkat ini, pusat kelompok dapat memisahkan tiap-tiap pengamatan menjadi anggota kelompok tertentu tetapi terdapat pengamatan yang tidak terpisah secara jelas.
0.26 – 0.50	Struktur yang terbentuk lemah. Pada tingkatan ini perlu dicoba pengelompokan dengan metode lainnya.
≤ 0.25	Tidak ada struktur kelompok yang terbentuk sehingga kelompok tidak dapat dijelaskan secara substansial.

3. HASIL DAN PEMBAHASAN

Pengelompokan perusahaan berdasarkan statistik litbang di Negara Indonesia pada tahun 2020 akan dilakukan melalui kerangka kerja KDD sebagai berikut:

3.1. Selection

Data diambil dari hasil kegiatan Survei Litbang pada sektor industri yang telah dilakukan oleh Pusat Data dan Informasi Badan Riset dan Inovasi Nasional pada tahun 2020. Data ini diambil dari database yang dimiliki oleh Pusat Data dan Informasi BRIN. Variabel yang akan digunakan adalah nama perusahaan, jumlah pengeluaran litbang atau *R&D expenditure*, jumlah personil litbang atau *R&D personnel* dan jumlah pekerja tetap dan tidak tetap dari tiap-tiap perusahaan yang melakukan kegiatan litbang pada tahun berjalan. Adapun perusahaan yang dipilih adalah perusahaan yang melakukan kegiatan penelitian dan pengembangan di tahun 2020.

3.2. Pre-processing

Pengecekan terhadap kesalahan penulisan, duplikasi data, *missing value*, dan konsistensi data telah dilakukan untuk menghasilkan data yang siap diolah. Adapun dalam hal ini *field* yang digunakan adalah nama perusahaan, anggaran litbang, personil litbang dan jumlah pekerja dengan jumlah pengamatan sebesar 320 orang.

3.3. Transformation

Pada tahap ini nama perusahaan diubah menjadi kode perusahaan sebagai pertimbangan dalam menjamin kerahasiaan data. Transformasi lainnya yang digunakan adalah proses standarisasi data. Proses tersebut perlu dilakukan jika terdapat unit pengukuran yang berbeda. Adapun proses tersebut menggunakan fungsi *scale()* pada Rstudio. Tabel 2 menyajikan hasil standarisasi data.

Tabel 2. Tabel hasil standarisasi

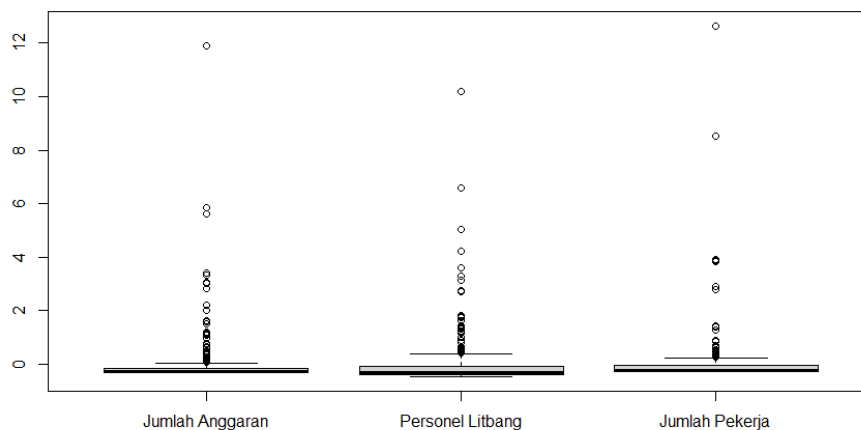
Kode Perusahaan	Anggaran_litbang	Personil_litbang	Jumlah_pekerja
1	0.200080845	-0.220080058	0.254322707
2	0.281668333	-0.368939958	0.001190576
3	-0.310632155	-0.443369908	-0.265575704
4	-0.294632620	-0.398711938	-0.225187752
5	-0.303191836	-0.398711938	-0.292329506
...
320	-0.12394691	-0.339167978	0.014824726

3.4. Data Mining

Fase ini merupakan proses untuk mencari informasi yang tersembunyi dalam data dengan suatu metode tertentu. Fase ini akan turut menguji beberapa asumsi yang diperlukan dalam analisis *cluster*.

A. Analisis Deskriptif Distribusi Data

Analisis deskriptif ditujukan untuk melihat dan memahami konteks maupun kondisi secara umum dari data yang akan diolah. Grafik yang digunakan pada analisis deskriptif ini adalah *boxplot* dengan tujuan untuk mengetahui keberadaan data *outlier* dan struktur dari suatu data.



Gambar 3. Boxplot distribusi data

Berdasarkan gambar 3 maka dapat dikatakan bahwa data aktivitas litbang yang akan diolah memiliki *outlier* di setiap variabelnya. Kondisi ini disebabkan karena data yang diperoleh merupakan hasil survei dari berbagai industri di seluruh Indonesia baik yang berskala besar maupun kecil. Kondisi data seperti ini seharusnya dapat ditangani oleh penggunaan metode *K-Medoids* yang dinilai secara baik dapat mengatasi data *outlier*. Adapun Tabel 3 menyajikan ringkasan data untuk melihat pemusatan terhadap data statistik litbang tahun 2020.

Tabel 3. Ringkasan Pemusatan Data

Kode Perusahaan	Anggaran_litbang	Personil_litbang	Jumlah_pekerja
Nilai Minimal	-0.3126	-0.47314	-0.29336
Kuantil Pertama	-0.3031	-0.39871	-0.26660
Median	-0.2812	-0.32428	-0.21991
Rata-Rata	0.0000	0.0000	0.0000
Kuantil Ketiga	-0.1588	-0.08238	-0.06158
Nilai Maksimal	11.8852	10.20011	12.64571

Berdasarkan tabel 3 tersebut dapat disampaikan bahwa nilai median lebih rendah dari nilai rata-rata. Nilai median perlu menjadi perhatian dalam data ini karena nilai pemusatan rata-rata menjadi sangat sensitif ketika terjadi pencilan dalam kumpulan data. Adapun jarak antara nilai paling rendah dan nilai paling tinggi dapat dikatakan relatif tinggi sehingga secara deskriptif dapat dikatakan bahwa data statistik litbang industri tahun 2020 mengalami penyebaran yang cukup luas.

B. Sampel Representatif

Salah satu asumsi untuk melakukan analisis *cluster* yang baik adalah terpenuhinya kondisi sampel yang representatif. Adapun untuk memenuhi asumsi tersebut maka dilakukan pengujian *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* melalui R dengan fungsi *KMO()* seperti yang dapat dilihat di Gambar 4 sebagai berikut:

```
> KMO(Data_Litbang)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = Data_Litbang)
Overall MSA = 0.59
MSA for each item =
  Jumlah Anggaran Personel Litbang Jumlah Pekerja
0.56 0.66 0.59
```

Gambar 4. Hasil *Kaiser-Meyer-Olkin Measure of Sampling Adequacy*

Berdasarkan hasil *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* diperoleh statistik $KMO = 0.59$. Adapun suatu data dikatakan memenuhi kondisi sampel yang representatif jika memiliki nilai statistik $KMO > 0.5$ [19]. Melalui statistik $KMO = 0.59$ maka dapat dikatakan bahwa data statistik litbang industri tahun 2020 memenuhi asumsi sampel yang representatif.

C. Uji Multikolinearitas

Salah satu asumsi dari analisis *cluster* adalah tidak ditemukannya kondisi multikolinearitas antara variabel pada data yang akan diolah. Adapun indikasi multikolinearitas dapat dilihat dari nilai korelasi *pearson* antara variabel. Gambar 5 menunjukkan hasil perhitungan nilai korelasi *pearson* tiap variabel dengan menggunakan aplikasi Rstudio dan melalui bantuan fungsi *cor()*.

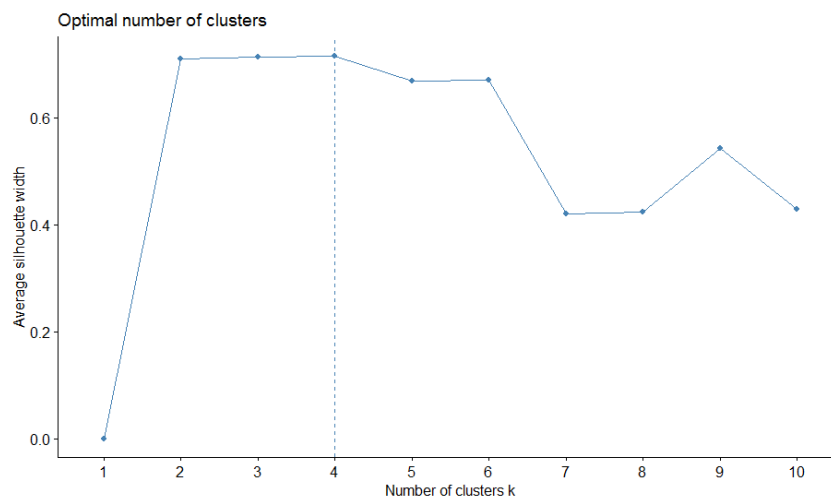
```
> cor(Data_Litbang)
      Jumlah Anggaran Personel Litbang Jumlah Pekerja
Jumlah Anggaran 1.0000000 0.3378785 0.4482759
Personel Litbang 0.3378785 1.0000000 0.2165811
Jumlah Pekerja 0.4482759 0.2165811 1.0000000
```

Gambar 5. Nilai korelasi tiap-tiap variabel

Berdasarkan tabel nilai korelasi antar variabel tersebut maka dapat disampaikan bahwa nilai korelasi antar variabel jumlah anggaran dengan personel litbang adalah 0.3378785, jumlah anggaran dengan jumlah pekerja adalah 0.4482759, dan personel litbang dengan jumlah pekerja adalah 0.2165811. Referensi [20] menyampaikan bahwa multikolinearitas terjadi ketika terdapat nilai korelasi antar variabel yang lebih besar dari 0.8. Berdasarkan kondisi tersebut maka dapat disampaikan bahwa tidak terjadi multikolinearitas pada data statistik litbang industri tahun 2020 yang akan diolah.

D. Metode K-Medoids

Proses ini merupakan tahap untuk mencari pola yang tersembunyi di dalam data melalui analisis *cluster*. Proses yang pertama kali dilakukan adalah menentukan jumlah kelompok optimal yang akan dibentuk. Penelitian ini menggunakan indeks *silhouette* sebagai indikator untuk menentukan jumlah kelompok yang optimal. Gambar 6 menyajikan jumlah kelompok yang optimal pada metode K-Medoids. Proses penghitungan indeks *silhouette* pada aplikasi Rstudio dibantu dengan *package factoextra* yang berisikan fungsi untuk melakukan analisis statistik multivariat.



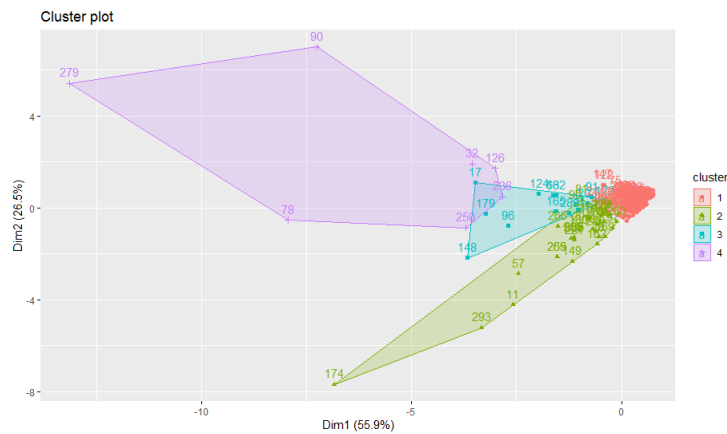
Gambar 6. Penentuan jumlah kelompok optimal pada metode K-Medoids

Berdasarkan indeks *silhouette* maka dapat disampaikan bahwa jumlah kelompok yang optimal dari data litbang industri tahun 2020 adalah sebanyak empat kelompok. Hal ini disebabkan nilai indeks *silhouette* kelompok ke-4 memiliki nilai tertinggi dibandingkan kelompok-kelompok lainnya. Sesuai dengan penjelasan indeks *silhouette*, bahwa semakin tinggi nilai indeks tersebut maka semakin kuat struktur kelompok yang terbentuk. Berdasarkan hasil ini maka perusahaan pada data litbang industri tahun 2020 akan dikelompokkan menjadi empat kelompok.

Tahap Selanjutnya adalah melakukan pengelompokan melalui metode K-Medoids melalui aplikasi Rstudio. Adapun metode tersebut dapat ditemukan pada *package cluster* dengan fungsi *pam()*. Selain itu visualisasi terhadap hasil pengelompokan dapat menggunakan fungsi *fviz_cluster()* pada *package factoextra*. Gambar 7 menyajikan *script* dan visualisasi hasil pengelompokan data statistik litbang industri tahun 2020 dapat dilihat di Gambar 8.

```
> library(cluster)
library(factoextra)
cluster_kmed=pam(Data_litbang,4)
fviz_cluster(cluster_kmed,Data_Litbang)
```

Gambar 7. Script pengelompokan data statistik



Gambar 8. Peta Persebaran Kelompok melalui metode *K-Medoids*

Berdasarkan peta kelompok pada gambar 8 maka dapat ditentukan keterangan mengenai keanggotaan kelompok dari masing-masing pengamatan. Tabel 4 menyajikan data yang telah ditambahkan label terkait keanggotaan kelompok suatu perusahaan yang dibentuk melalui metode *K-Medoids*.

Tabel 4. Tabel daftar keanggotaan kelompok berdasarkan *K-Medoids*

Kode Perusahaan	Anggaran_litbang	Personil_litbang	Jumlah_pekerja	Anggota Cluster
1	0.200080845	-0.220080058	0.254322707	1
2	0.281668333	-0.368939958	0.001190576	1
3	-0.310632155	-0.443369908	-0.265575704	1
...
320	-0.12394691	-0.339167978	0.014824726	1

3.5. Evaluation and Interpretation

Fase ini merupakan proses untuk mengevaluasi hasil pengelompokan yang telah dilakukan melalui suatu kriteria tertentu. Evaluasi dilakukan untuk menilai seberapa kuat struktur dari kelompok yang sudah dibangun dari suatu metode analisis *cluster* tertentu. Dalam penelitian ini, kriteria yang digunakan adalah indeks *silhouette*.

```
> index_kmed$data
  clusters  y
1         1 0.0000000
2         2 0.7096494
3         3 0.7141414
4         4 0.7147568
5         5 0.6677271
6         6 0.6704929
7         7 0.4196925
8         8 0.4237041
9         9 0.5425449
10        10 0.4292249
```

Gambar 9. Nilai evaluasi metode

Berdasarkan gambar 9 tersebut maka dapat disampaikan bahwa indeks *silhouette* pada hasil pengelompokan empat kelompok berdasarkan metode analisis *cluster K-Medoids* bernilai 0.7147568. Nilai tersebut menunjukkan bahwa kelompok yang dibentuk memiliki struktur paling kuat dibandingkan nilai indeks *silhouette* dari kelompok yang lainnya. Adapun nilai tersebut pula menunjukkan bahwa kelompok yang dibentuk memiliki struktur yang kuat dimana anggota antar kelompok memiliki karakteristik yang berbeda dan sesama anggota kelompok memiliki karakteristik yang homogen. Berikut adalah interpretasi profil dari karakteristik masing-masing kelompok yang terbentuk melalui metode *K-Medoids*.

Tabel 5. Pusat kelompok berdasarkan metode *K-Medoids*

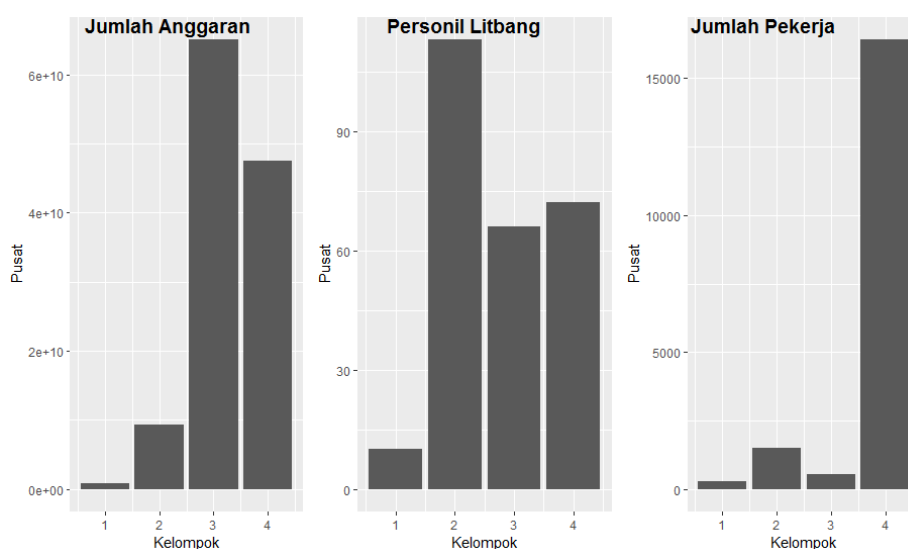
Kelompok	Anggaran_litbang	Personil_litbang	Jumlah_pekerja	Jumlah perusahaan
1	-0.27816489	-0.3391680	-0.27816489	267 (83%)
2	0.04432551	1.1940890	0.09688687	33 (10%)
3	2.19180241	0.4944475	-0.15573076	13 (4%)
4	1.51740084	0.5837634	3.92936907	7 (2%)

Nilai pada Tabel 5 diperoleh dengan mengeluarkan aspek *medoids* dan ukuran (*size*) dari hasil pengimplementasian *K-Medoids*. Nilai tersebut akan ditransformasi kembali menjadi nilai asli sehingga interpretasi dari karakteristik dari tiap kelompok dapat diinterpretasikan dengan lebih mudah. Tabel 6 menyajikan hasil transformasi pusat kelompok.

Tabel 6. Nilai pusat hasil pengembalian dari nilai standardisasi

Kelompok	Anggaran_litbang	Personil_litbang	Jumlah_pekerja	Jumlah perusahaan
1	900,000,000	10	300	267 (83%)
2	9,275,721,885	113	1519	33 (10%)
3	65,050,000,000	66	537	13 (4%)
4	47,534,440,000	72	16417	7 (2%)

Dalam memudahkan perbandingan antar kelompok, Gambar 9 menyajikan nilai pusat dari tiap-tiap kelompok sehingga dapat diketahui karakteristik dari tiap-tiap kelompok.



Gambar 9. Grafik Besaran Pusat Variabel

Berdasarkan nilai pusat kelompok tersebut maka dapat disampaikan bahwa empat kelompok yang terbentuk memiliki karakteristik sebagai berikut:

- Kelompok ke-1 berisikan 267 perusahaan dimana anggota-anggotanya memiliki nilai variabel anggaran litbang, personil litbang dan jumlah pekerja paling kecil di antara kelompok lainnya. Dapat dikatakan perusahaan ini merupakan skala perusahaan berskala kecil yang memiliki kapasitas litbang dibanding kelompok lainnya.
- Kelompok ke-2 berisikan 33 perusahaan dimana anggotanya memiliki karakteristik jumlah personil litbang yang relatif besar tetapi dengan jumlah anggaran dan jumlah anggaran yang relatif rendah. Perusahaan ini memiliki kapasitas litbang yang tinggi pada personilnya.

- c. Kelompok ke-3 berisikan 13 perusahaan dimana anggotanya memiliki jumlah pekerja yang relatif rendah dan jumlah personil litbang yang relatif tinggi tetapi memiliki anggaran litbang yang paling tinggi jika dibandingkan kelompok-kelompok lainnya.
- d. Kelompok ke-4 berisikan 7 perusahaan yang anggotanya memiliki jumlah pekerja yang paling tinggi tetapi karakteristik jumlah personil litbang tidak sebesar kelompok ke-2. Adapun perusahaan di kelompok ke-4 memiliki anggaran litbang yang relatif tinggi meskipun tidak sebesar kelompok ke-3. Dapat dikatakan bahwa kelompok ini merupakan perusahaan yang berskala besar dari sisi jumlah pekerjanya.

4. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa analisis *clustering* melalui metode K-Medoids berdasarkan data statistik litbang dan jumlah pekerja secara optimal membentuk empat kelompok perusahaan. Adapun berdasarkan evaluasi pengelompokan tersebut diperoleh struktur kelompok yang kuat dimana pusat kelompok dapat secara jelas memisahkan setiap pengamatan menjadi anggota kelompok tertentu berdasarkan karakteristiknya. Melalui nilai indeks *silhouette* yang menghasilkan struktur kelompok yang kuat, dapat dikatakan pula bahwa metode K-Medoids dapat secara baik membentuk kelompok pada data yang memiliki pencilan (*outlier*). Selain itu dapat disampaikan bahwa saat ini sebagian besar perusahaan di Indonesia lebih banyak masuk ke dalam kelompok dengan nilai statistik litbang yang berada pada median data.

5. SARAN

Penelitian selanjutnya dapat membandingkan berbagai macam jenis metode *clustering* baik itu metode *clustering* yang bersifat hierarki maupun non-hierarki. Adapun penelitian lainnya dapat melihat bagaimana validitas suatu metode *clustering* selain dalam penelitian ini ketika berhadapan dengan data yang bersifat pencilan (*outlier*).

DAFTAR PUSTAKA

- [1] OECD, "OECD Main Science and Technology Indicators. R&D Highlights in the February 2020 Publication," no. February, pp. 2018–2021, 2020, [Online]. Available: www.oecd.org/sti/msti2020.pdf.
- [2] A. A. Kadir, W. Afriana, dan H. A. Azis, "The Effects Of R&D Expenditures On Economic Growth In Oecd Countries," *Econder Int. Acad. J.*, vol. 1, no. 2, pp. 203–220, 2020, doi: 10.35342/econder.665074.
- [3] E. Gumus dan F. Celikay, "R&D Expenditure and Economic Growth: New Empirical Evidence," *Margin*, vol. 9, no. 3, pp. 205–217, 2015, doi: 10.1177/0973801015579753.
- [4] OECD, *Frascati Manual 2015 The Measurement of Scientific, Technological and Innovation Activities*. 2015.
- [5] J. Xu dan Z. Jin, "Research on the Impact of R&D Investment on Firm Performance in China's Internet of Things Industry," *J. Adv. Manag. Sci.*, no. March 2016, pp. 112–116, 2016, doi: 10.12720/joams.4.2.112-116.
- [6] RISTEKDIKTI, "Lanskap Ilmu Pengetahuan dan Teknologi di Indonesia," 2017.
- [7] Instituto Nacional de Estadística, "Statistics on R+D Activities. Latest data," 2020. https://www.ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736176754&menu=ultiDatos&idp=1254735576669.
- [8] Metrics, "R&D Expenditures and Personnel," 2022, [Online]. Available: <https://metrics.ekt.gr/en/research-development>.
- [9] A. Elizondo-Noriega, N. Tiruvengadam, D. Güemes-Castorena, V. G. Tercero-Gómez,

- dan M. G. Beruvides, "Classification of R&D activities in industrial environments," *IISE Annu. Conf. Expo 2019*, no. May, 2019.
- [10] M. D. Anuslu dan S. U. Firat, "Clustering Analysis application on Indutstry 4 driver global indexes," in *Procedia Computer Science*, 2019, pp. 145–152.
- [11] K. R.S dan K. R.K, "K-Means Clustering for Analyzing Productivity in Light of R & D Spillover," *Int. J. Inf. Technol. Model. Comput.*, vol. 4, no. 2, pp. 55–64, 2016, doi: 10.5121/ijitmc.2016.4204.
- [12] F. Alfiah, A. Almadayani, D. Al Farizi, dan E. Widodo, "Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020," *J. Ilm. Sains*, vol. 22, no. 1, p. 1, 2021, doi: 10.35799/jis.v22i1.35911.
- [13] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [14] I. Bin Mohamad dan D. Usman, "Standardization and its effects on K-means clustering algorithm," *Res. J. Appl. Sci. Eng. Technol.*, vol. 6, no. 17, pp. 3299–3303, 2013, doi: 10.19026/rjaset.6.3638.
- [15] D. Marlina, N. Lina, A. Fernando, dan A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [16] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, dan D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020, doi: 10.36294/jurti.v4i1.1296.
- [17] A. Al-Najafi, "Comparative performance of three methods to classify smokers data," *Austrian J. Stat.*, vol. 50, no. 3, pp. 32–40, 2021, doi: 10.17713/ajs.v50i3.1013.
- [18] A. Struyf, M. Hubert, dan P. J. Rousseeuw, "Clustering in an object-oriented environment," *J. Stat. Softw.*, vol. 1, pp. 1–30, 1996, doi: 10.18637/jss.v001.i04.
- [19] D. Ls, Y. A. Lesnussa, M. W. Talakua, dan M. Y. Matdoan, "Analisis Klaster untuk Pengelompokkan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Pendidikan dengan Menggunakan Metode Ward," *J. Stat. dan Apl.*, vol. 5, no. 1, pp. 51–60, 2021, doi: 10.21009/jsa.05105.
- [20] H. Juliansyah dan S. Ulfa, "Faktor-Faktor Yang Mempengaruhi Pemberdayaan Industri Kecil Masyarakat Di Kecamatan Baktiya Kabupaten Aceh Utara," *J. Ekon. Reg. Unimal*, vol. 01, no. April, pp. 30–37, 2018.
-