

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

by Joins 5935

Submission date: 31-Mar-2022 11:28AM (UTC+0700)

Submission ID: 1797589218

File name: 5935-17989-1-BR.docx (1.29M)

Word count: 3748

Character count: 24887

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

Diterima: xx Juli 20yy; Direvisi: xx September 20yy; Disetujui: xx Nopember 20yy

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* paling optimal 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*, Indeks Silhouette Litbang Industri, Data Mining

Abstract

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 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 companies based on R&D statistics using the *K-Medoids* clustering method so that optimal grouping results are obtained. 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 the strong category. In addition, the results of this grouping can be used as an initial mapping in making R&D policies in the Indonesian industrial sector.

Keywords: *K-Means*, *K-Medoids*, Dunn's Index, Industrial R&D, Data Mining

1. PENDAHULUAN

Aktivitas penelitian dan pengembangan/*research and development*/litbang berperan penting untuk meningkatkan daya saing dan perekonomian suatu bangsa. Referensi [1] menyampaikan bahwa pengeluaran negara terhadap kegiatan R&D memiliki dampak positif pada pertumbuhan ekonomi 33 negara OECD. Adapun naskah [2] mengonfirmasi bahwa pengeluaran R&D antara tahun 1996-2010 pada 52 negara yang menjadi fokus penelitiannya memiliki pengaruh positif terhadap pertumbuhan ekonomi, khususnya dalam jangka panjang. Bahkan melalui kegiatan litbang, negara memiliki peran untuk mencapai *Sustainable Development Goals*/SDGs ke-9 terkait pengembangan industri, inovasi dan infrastruktur [3]. Berangkat dari pentingnya R&D pada suatu negara maka aktivitas tersebut perlu terus memantau perkembangan kapasitas litbang dari setiap sektor.

Aktivitas litbang pada sektor industri menjadi salah satu sektor yang memiliki peran penting dalam mendukung ekosistem litbang bahkan meningkatkan perekonomian [4]. Kegiatan litbang pada sektor industri dapat berkontribusi bagi ekspansi pengetahuan suatu industri, peningkatan efisiensi produksi serta peningkatan *output* [4]. Sektor tersebut memiliki peran untuk mendorong penciptaan suatu produk hasil penelitian yang memiliki nilai jual dan dapat diproduksi massal. Dalam industri tertentu, kegiatan litbang memiliki pengaruh yang signifikan terhadap peningkatan margin keuntungan perusahaan [5]. Oleh karena itu, penting untuk memantau sejauh mana kapasitas penelitian dan pengembangan yang dilakukan di sektor industri [6].

Salah satu pedoman universal yang digunakan untuk mengukur kapasitas litbang, termasuk litbang pada sektor industri, adalah *Research and Development Statistics*/RDS yang dikembangkan oleh OECD. Indikator ini mengukur indikator pengeluaran litbang/*R&D Expenditure* dan personil litbang/*R&D Personnel* [7]. Adapun statistik tersebut akan ditambah dengan indikator jumlah pekerja di perusahaan untuk memberikan pemahaman terhadap klasifikasi jenis perusahaan di negara tertentu [8]. Dalam hal ini, sektor industri menjadi salah satu kategori pengukuran statistik litbang nasional. Berbagai negara menggunakan statistik ini untuk menyediakan instrumen untuk manajemen, perencanaan, pembuatan keputusan dan kontrol dalam rangka kebijakan ilmiah nasional dan menyediakan badan statistik dengan informasi sesuai dengan peraturan internasional sehingga memungkinkan dilakukan perbandingan dengan negara-negara lainnya [9].

Pengelompokan kapasitas kegiatan litbang pada sektor industri secara umum memiliki bermacam manfaat. Meskipun pada dasarnya industri pelaku litbang terbesar adalah industri pada bidang manufaktur tetapi tidak menutup kemungkinan pada bidang industri lain memiliki aktivitas litbang yang cukup signifikan dalam memengaruhi eksistensi industrinya sehingga diperlukan pengelompokan terhadap kondisi litbang sektor industri. Hasil pengelompokan yang baik dapat melacak performa litbang dari tiap-tiap perusahaan terhadap pendapatannya [10]. Hasil pengelompokan dapat memberikan sebuah gambaran terkait kompleksitas kegiatan litbang di bidang-bidang tertentu yang butuh penanganan tersendiri [10]. Pada akhirnya pengelompokan industri berdasarkan kapasitas litbangnya memiliki peran untuk peningkatan ekonomi.

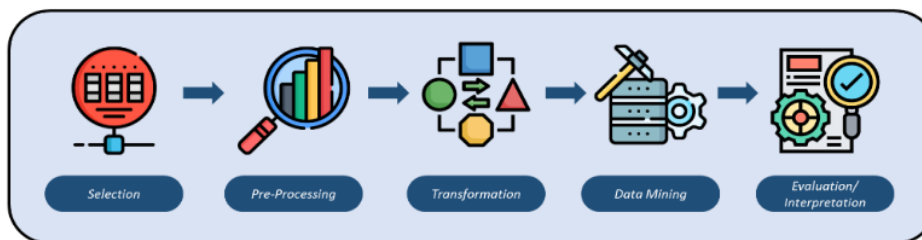
Melalui pentingnya pengelompokan litbang industri tersebut maka dibutuhkan suatu bentuk pengelompokan yang tepat melalui metode tertentu. Berbagai penelitian sebelumnya telah melakukan pemilihan metode pengelompokan/*clustering* yang paling optimal dari suatu data statistik dalam konteks litbang. Analisis *clustering* digunakan dalam pengelompokan negara berdasarkan negara yang terpengaruh akibat industri 4 berdasarkan *Global Innovative Index*, *Sustainable Development Goals Index*, *Logistic Performance Index* dan *Environmental Performance Index* melalui metode *clustering hierarchical* dan *non-hierarchical* [11]. Naskah [12] melakukan analisis *K-Means* terhadap negara melalui variabel persentase belanja litbang terhadap GDP, jumlah artikel jurnal, jumlah aplikasi paten, dan jumlah aplikasi merek dagang. Adapun referensi [13] melakukan pengelompokan negara bagian di Amerika Serikat berdasarkan intensitas litbangnya melalui metode *K-Means*. Berdasarkan beberapa penelitian sebelumnya

maka terdapat kesempatan untuk meneliti pengelompokan perusahaan dalam lingkup nasional berdasarkan jumlah belanja litbang, jumlah personil litbang dan jumlah pegawai.

Tujuan penelitian ini adalah untuk mengelompokkan perusahaan berdasarkan RDS melalui metode *k-medoids*. Alasan penggunaan *clustering* tersebut dikarenakan data yang akan diolah adalah data yang berjenis numerik. Adapun algoritma *k-medoids* lebih mampu menangani data yang bersifat pencilan/*outlier* dibandingkan metode lainnya [14]. 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*. Evaluasi pengelompokan akan ditinjau melalui indeks *silhouette*. Kebaruan yang disampaikan dalam penelitian ini adalah mengelompokkan perusahaan lingkup nasional berdasarkan indikator RDS. 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 [15]. *Flowchart* dari tiap-tiap fase KDD yang akan dilakukan pada penelitian ini disampaikan pada gambar 1.



Gambar 1. Fase dalam KDD. Sumber [15]

2.1. Selection

Fase seleksi ini merupakan tahap untuk mencari *field*/variabel 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 [16]. Adapun proses standarisasi data telah dipertimbangkan sebagai salah satu proses yang dapat membantu proses analisis *cluster* agar mendapatkan kelompok yang berkualitas [16]. Berikut rumus yang digunakan untuk melakukan standarisasi data:

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

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 distandarisasi 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*. Adapun berbagai penjelasan metode-metode tersebut telah disampaikan dalam penelitian-penelitian terdahulu seperti pada referensi [17].

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 [18]. Adapun kriteria evaluasi yang digunakan adalah statistik indeks *silhouette* dengan interpretasi sebagai berikut:

Tabel 1. Kriteria evaluasi indeks *silhouette*. Sumber [19]

Index Silhouette	Interpretasi
0.71 – 1.00	Bentuk Kelompok yang terbentuk memiliki struktur yang kuat
0.51 – 0.70	Bentuk Kelompok yang terbentuk memiliki struktur yang baik
0.26 – 0.50	Bentuk Kelompok yang terbentuk memiliki struktur yang lemah
≤ 0.25	Bentuk Kelompok yang terbentuk bersifat tidak terstruktur

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/*R&D expenditure*, jumlah

personil litbang/*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.

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

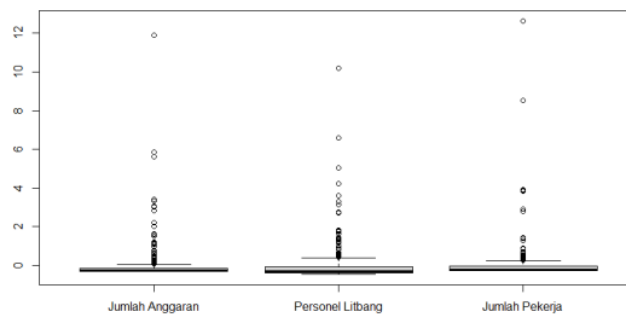
Kode Perusahaan	Anggaran_litbang	Personil_litbang	Jumlah_pekerja
1	0.200080845	-0.220080058	0.254322707
2	0.281668333	-0.368939958	0.001190576
3	-0.310632155	-0.443369908	-0.265575704
4	-0.294632620	-0.398711938	-0.225187752
5	-0.303191836	-0.398711938	-0.292329506
...	...	...	...
320	-0.12394691	-0.339167978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

Berdasarkan gambar 2 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()* 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 3. 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 lebih dari 0.5/ $KMO > 0.5$ [20]. 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. Berikut tabel hasil perhitungan nilai korelasi *pearson* tiap variabel dengan menggunakan aplikasi Rstudio dan melalui bantuan fungsi *cor()*.

```
> cor(Data_Litbang)
```

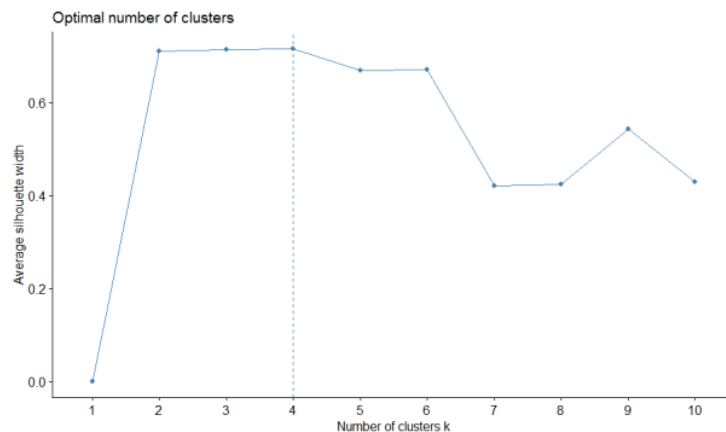
	Jumlah Anggaran	Personel Litbang	Jumlah Pekerja
Jumlah Anggaran	1.0000000	0.3378785	0.4482759
Personel Litbang	0.3378785	1.0000000	0.2165811
Jumlah Pekerja	0.4482759	0.2165811	1.0000000

Gambar 4. 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 [21] 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.

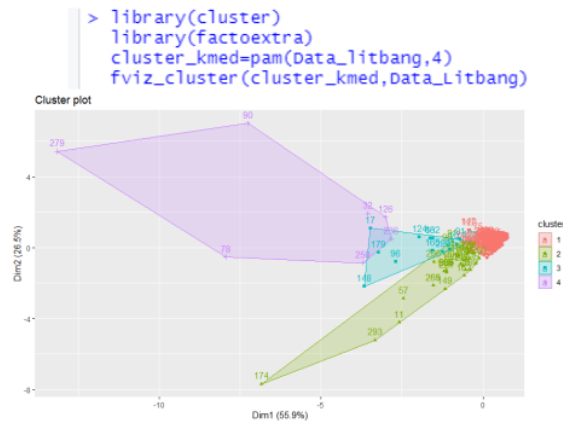
D. Metode K-Medoids

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 5 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 5. 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 4 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 4 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 6 menyajikan *script* dan visualisasi hasil pengelompokan data statistik litbang industri tahun 2020.

Gambar 6. Peta Persebaran Kelompok melalui metode *K-Medoids*

Berdasarkan peta kelompok pada gambar 6 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*. Tabel 6 menyajikan nilai Indeks *Silhouette* dari tiap-tiap metode.

Gambar 7. Nilai evaluasi metode.

```
> 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
```

Berdasarkan tabel 7 tersebut maka dapat disampaikan bahwa Indeks *Silhouette* pada hasil pengelompokan 4 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%)

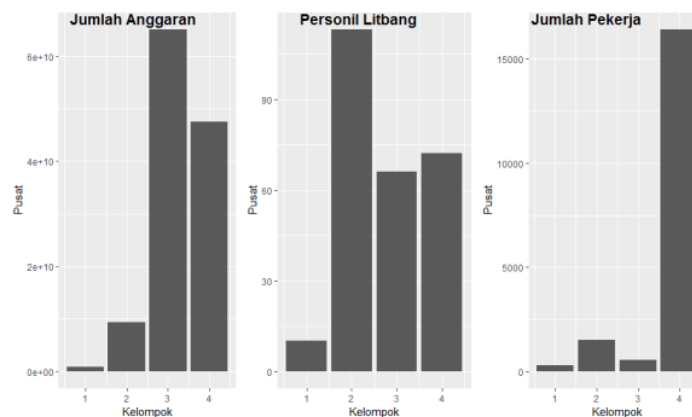
Nilai pada tabel 5 diperoleh dengan mengeluarkan aspek *medoids* dan ukuran/size dari hasil pengimplementasian *K-Medoids*. Nilai tersebut akan ditransformasi kembali menjadi nilai asli sehingga interpretasi dari karakteristik dari tiap kelompok dapat diinterpretasikan dengan lebih mudah. Tabel 6 menyajikan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

Tabel 6. Nilai Pusat Hasil Pengembalian dari Nilai Standarisasi

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 8 menyajikan nilai pusat dari tiap-tiap kelompok sehingga dapat diketahui karakteristik dari tiap-tiap kelompok. Pembuatan grafik pada gambar 8 dibantu dengan *package ggplot2* di Rstudio.

Gambar 8. Grafik Besaran Pusat Variabel



Berdasarkan nilai pusat kelompok tersebut maka dapat disampaikan bahwa 4 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 diantara kelompok lainnya. Dapat dikatakan perusahaan ini merupakan skala perusahaan berskala kecil yang memiliki kapasitas litbang dibanding kelompok lainnya.

- b. 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 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. 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. Hal ini menunjukkan bahwa secara umum kapasitas litbang dari berbagai perusahaan di Indonesia masih perlu ditingkatkan kembali sehingga pada akhirnya dapat memunculkan inovasi pada setiap aspek. Pengelompokan ini diharapkan dapat memberikan gambaran kondisi kapasitas litbang industri sehingga menjadi dasar pembentukan kebijakan publik dalam meningkatkan iklim litbang di Industri.

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 *outlier*/pencilan.

DAFTAR PUSTAKA

- [1] A. A. Kadir, W. Afriana, and 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.
- [2] E. Gumus and F. Celikay, "R&D Expenditure and Economic Growth: New Empirical Evidence," *Margin*, vol. 9, no. 3, pp. 205–217, 2015, doi: 10.1177/0973801015579753.
- [3] 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.
- [4] OECD, "Working Party of National Experts on Science and Technology Indicators," 2015.
- [5] J. Xu and 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] Metrics, "R&D Expenditures and Personnel," 2022, [Online]. Available:

- <https://metrics.ekt.gr/en/research-development>.
- [8] OECD, *Frascati Manual 2015 The Measurement of Scientific, Technological and Innovation Activities*. 2015.
- [9] 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.
- [10] A. Elizondo-Noriega, N. Tiruvengadam, D. Güemes-Castorena, V. G. Tercero-Gómez, and M. G. Beruvides, "Classification of R&D activities in industrial environments," *IISE Annu. Conf. Expo 2019*, no. May, 2019.
- [11] M. D. Anuslu and S. U. Firat, "Clustering Analysis application on Indutstry 4 driver global indexes," in *Procedia Computer Science*, 2019, pp. 145–152.
- [12] K. R.S and 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.
- [13] J. Sanchez, W. Lin, and V. N. Vu, "Role of Science and Engineering Education and R & D in U . S . State Growth," 2013.
- [14] F. Alfiah, A. Almadayani, D. Al Farizi, and 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.
- [15] 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.
- [16] I. Bin Mohamad and 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.
- [17] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, and 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.
- [18] 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.
- [19] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," *Techno.Com*, vol. 20, no. 2, pp. 186–197, 2021, doi: 10.33633/tc.v20i2.4556.
- [20] D. Ls, Y. A. Lesnussa, M. W. Talakua, and M. Y. Matdoan, "Analisis Klaster untuk Pengelompokan 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.
- [21] H. Juliansyah and 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.

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

ORIGINALITY REPORT

6%

SIMILARITY INDEX

6%

INTERNET SOURCES

3%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1

id.123dok.com

Internet Source

1%

2

text-id.123dok.com

Internet Source

1%

3

conference.binadarma.ac.id

Internet Source

<1%

4

eprints.uny.ac.id

Internet Source

<1%

5

www.slideshare.net

Internet Source

<1%

6

sigaa.ufrn.br

Internet Source

<1%

7

elar.naiau.kiev.ua

Internet Source

<1%

8

jatim.suara.com

Internet Source

<1%

9

www.qjmst.ir

Internet Source

<1%

10	bagawanabiyasa.wordpress.com Internet Source	<1 %
11	core.ac.uk Internet Source	<1 %
12	ejurnal.seminar-id.com Internet Source	<1 %
13	journal.fkpt.org Internet Source	<1 %
14	Oktaviani Manurung, Penda Sudarto Hasugian. "Analisa Algoritma Apriori Untuk Peminjaman Buku Pada Perpustakaan SMA 1 Silima Pungga-Pungga Parongil", remik, 2019 Publication	<1 %
15	id.scribd.com Internet Source	<1 %
16	journal.lppmunindra.ac.id Internet Source	<1 %
17	press.umsida.ac.id Internet Source	<1 %
18	www.scribd.com Internet Source	<1 %
19	brainstormink.wordpress.com Internet Source	<1 %
20	Samudi Samudi, Slamet Widodo, Herlambang Brawijaya. "The K-Medoids Clustering Method	<1 %

for Learning Applications during the COVID-19 Pandemic", SinkrOn, 2020

Publication

Exclude quotes Off

Exclude matches Off

Exclude bibliography On