

Aplikasi Dynamic Cluster pada K-Means Berbasis Web untuk Klasifikasi Data Industri Rumahan

Web-based Application of Dynamic Cluster on K-Means for Classification of Home Industry Data

Hadi Santoso¹, Hilyah Magdalena², Helna Wardhana³

^{1,2}Institut Sains dan Bisnis Atma Luhur, Indonesia

³Universitas Bumigora, Indonesia

Informasi Artikel

Genesis Artikel:

Diterima, 25 Januari 2022

Direvisi, 24 Februari 2022

Disetujui, 21 April 2022

Kata Kunci:

Data Mining
Dynamic Cluster K-means
Industri rumahan
Klasifikasi
SEMMA

Keywords:

Data Mining
Dynamic Cluster K-means
Classification
Home Industry
SEMMA

ABSTRAK

Masalah utama yang dihadapi Pemerintah Daerah Provinsi Kepulauan Bangka Belitung saat ini adalah sulitnya mengklasifikasikan data industri rumahan berdasarkan Peraturan Menteri PPPA No 2 Tahun 2016 yaitu pemula, berkembang dan maju. Berdasarkan permasalahan tersebut diusulkan pengembangan algoritma *K-means* yaitu algoritma *Dynamic cluster* pada *K-means* dengan tujuan agar dapat menghasilkan kluster yang optimal dalam pengelompokan data industri rumahan dengan membangun aplikasi cerdas berbasis web. Penelitian ini menggunakan metode analisis data *mining* SEMMA, yang meliputi tahapan-tahapan seperti data sampel, deskripsi data, transformasi data, pemodelan data, dan evaluasi data. 3.466 industri rumah tangga digunakan sebagai sampel data. Kinerja algoritma dievaluasi menggunakan pengukuran validitas kluster *Davies Bouldin Index* (DBI). Hasil eksperimen menunjukkan bahwa algoritma *Dynamic cluster* pada *K-means* memberikan nilai yang optimal pada iterasi ke lima, dengan perolehan sebagai berikut: kluster pemula (C1) diperoleh sebanyak 3214, kemudian kluster berkembang (C2) diperoleh sebanyak 167 dan kluster maju (C3) diperoleh sebanyak 85. Hasil evaluasi validitas kluster menunjukkan bahwa algoritma *Dynamic cluster* pada *K-means* memperoleh nilai DBI lebih kecil dibandingkan dengan algoritma *K-means* dengan nilai DBI sebesar 0.184. Implementasi algoritma *dynamic cluster* pada *K-means* untuk pengelompokan data industri rumahan pada Dinas P3ACSKB di Provinsi Kepulauan Bangka Belitung terbukti menghasilkan kualitas *cluster* yang lebih optimal.

ABSTRACT

The main issue currently confronting the Regional Government of Bangka Belitung Islands Province is the difficulty in categorizing home industry data in accordance with Ministry Regulation PPPA No. 2 In 2016, namely beginners, develop, and advanced. Based on these issues, it is proposed to develop a *K-means* algorithm, namely the *Dynamic cluster* algorithm on *K-means*, with the goal of producing optimal clusters in grouping home industry data through the development of intelligent web-based applications. The SEMMA data mining analysis method is used in this study, which includes stages such as sample data, data description, data transformation, data modeling, and data evaluation. As data samples, 3,466 home industries were used. The performance of the algorithm was evaluated using the *Davies Bouldin Index* (DBI) cluster validity measurement. The experimental results show that the *dynamic cluster* algorithm on *K-means* gives an optimal value in the fifth iteration, with the following gains: The beginner cluster (C1) obtained 3214, the developing cluster (C2) obtained 167, and the advanced cluster (C3) obtained 85. The results of the cluster validity evaluation show that the *Dynamic cluster* algorithm on *K-means* has a lower DBI value than the *K-means* algorithm, which has a DBI value of 0.184. The implementation of the *dynamic cluster* algorithm on *K-means* for home industry data clustering at the P3ACSKB Office in the Province of the Bangka Belitung Islands has proven to produce a more optimal cluster quality.

This is an open access article under the [CC BY-SA](#) license.



Penulis Korespondensi:

Hadi Santoso,
Fakultas Teknologi Informasi,
Institut Sains dan Bisnis Atma Luhur,
Email: hadisantoso@atmaluhur.ac.id

1. PENDAHULUAN

Pada konsep Usaha Mikro, Kecil, dan Menengah (UMKM), industri rumahan termasuk kedalam kategori Usaha Mikro [1]. Industri rumahan di beberapa negara di dunia dikelompokkan kedalam sektor informal. Industri rumahan sebagian besar tidak memiliki legalitas yang resmi sebagai badan usaha, bahkan seringkali tidak terdaftar dalam mekanisme perpajakan usaha. Industri rumahan termasuk dalam Usaha Mikro yang memiliki potensi besar dalam menyerap tenaga kerja dengan berbagai keterbatasan dan kebanyakan belum dapat diatasi secara optimal. Pendampingan pemerintah daerah sangat dibutuhkan untuk mendorong pertumbuhan usaha mikro melalui akses produksi, teknologi, dan pemasaran yang didukung dengan penguatan sumber daya manusia (SDM). Selama ini pemerintah daerah masih kesulitan dalam melakukan pengelompokan data industri rumahan di Provinsi Kepulauan Bangka Belitung yang prosesnya dikerjakan secara manual, sehingga pekerjaan mengelompokkan data industri rumahan menjadi kurang efektif dan efisien terutama dalam menentukan klaster pemula, berkembang dan maju sesuai Permen PPPA No 2 Tahun 2016 dengan berbagai macam kriteria tertentu yang harus dipenuhi. Atas dasar permasalahan tersebut, maka penelitian ini mengusulkan membangun sebuah sistem cerdas dalam mengelompokkan data industri rumahan menggunakan beberapa kriteria yaitu tenaga kerja, teknologi produksi, sumber modal usaha, jumlah modal, lama usaha dan pola produksi yang akan menjadi atribut pada aplikasi cerdas berbasis web ini. Pengelompokan data industri rumahan dapat dibangun tentunya diperlukan sebuah algoritma pada sistem yang mampu menghasilkan pengelompokan data industri rumahan kedalam klasifikasi pemula, berkembang dan maju. Algoritma *K-means* dipilih oleh kebanyakan peneliti untuk menyelesaikan masalah dalam pengelompokan data [2–6]. Namun, algoritma *K-means* ternyata belum memberikan solusi untuk menghasilkan kualitas klaster yang optimal [7]. Berdasarkan permasalahan tersebut penelitian ini pada intinya merealisasikan pengelompokan data industri rumahan dengan membangun aplikasi berbasis web menggunakan pengembangan algoritma *K-means* yaitu algoritma *Dynamic cluster* pada *K-means* agar dapat menghasilkan klaster yang optimal pada kasus dataset industri rumahan di Provinsi Kepulauan Bangka Belitung sesuai Permen PPPA No. 2 Tahun 2016 tentang yaitu **pemula (C1), berkembang (C2) dan maju (C3)** [8].

Pekerjaan terdahulu yang telah dilakukan oleh peneliti lainnya berdasarkan dengan topik yang serupa dengan penelitian ini adalah sebagai berikut: Widiarina dan Romi Satria Wahono (2015) melakukan penelitian pemetaan nasabah potensial menggunakan algoritma klaster dinamik pada *K-means*. Hasil segmentasi yang terbentuk dievaluasi menggunakan *Davies-Bouldin Index* (DBI) dan *purity*. Hasil penelitian menunjukan bahwa pengukuran validasi klaster dengan menggunakan *Davies-Bouldin Index* dan *purity*, membuktikan bahwa *K-means* dengan algoritma klaster dinamik menghasilkan kualitas klaster yang lebih optimal [9]. Artikel penelitian terdahulu ini sama-sama menggunakan metode algoritma klaster dinamik pada *K-means*, namun obyek penelitian yang dikelompokkan berbeda dengan penelitian pada artikel ini. Selain itu, perbedaan yang lain pada penelitian terdahulu ini ada pada tidak membangun aplikasi cerdas sebagaimana artikel pada penelitian ini.

Z. Hossain, N. Akhtar, R. B. Ahmad, dan M. Rahman (2019) melakukan penelitian pengelompokan data iris secara dinamis menggunakan algoritma klaster dinamis pada *K-means*. Algoritma yang diusulkan menghitung nilai ambang sebagai *centroid* dari *K-Means* [10]. Artikel terdahulu sama-sama menggunakan metode algoritma klaster dinamis pada *K-means*, namun obyek penelitian yang dikelompokkan berbeda dengan penelitian pada artikel ini. Selain itu, penelitian terdahulu tidak membangun aplikasi cerdas seperti yang dilakukan pada penelitian ini.

F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, Fachrudin, dan T. M. A. Aziz (2019) telah mengusulkan model segmentasi pelanggan untuk usaha kecil dan menengah (UKM) menggunakan algoritma *K-means* dan Model LRFM (*Length, Recency, Frequency, Monetary*) [11]. Obyek penelitian yang digunakan pada artikel terdahulu ini hampir serupa dengan penelitian di artikel ini yaitu usaha mikro atau industri rumahan, namun berbeda pada metode yang digunakan. Selain itu, penelitian terdahulu tersebut juga tidak membangun aplikasi cerdas sebagaimana artikel pada penelitian ini.

Komang Ariasa, I Gede Aris Gunadi dan I Made Candiasa (2020) melakukan penelitian pengelompokan kinerja akademik mahasiswa menggunakan algoritma klaster dinamis pada *K-means* [12]. Artikel terdahulu sama-sama menggunakan algoritma klaster dinamis pada *K-means*, namun obyek penelitian yang di kelompokkan berbeda. Perbedaan lainnya, penelitian terdahulu ini tidak membangun aplikasi cerdas sebagaimana yang dilakukan pada penelitian di artikel ini.

Romadansyah Siagian, Pahala Sirait, dan Arwin Halima (2021) melakukan penelitian mengenai segmentasi pelanggan berdasarkan data riwayat transaksi perusahaan menggunakan algoritma *K-means* dan Model LRFM (*Length, Recency, Frequency, Monetary*). Metode Elbow, Silhouette dan Davis-Bouldin Index digunakan untuk mengevaluasi dalam menentukan klaster k optimal pada *K-means* [13]. Obyek penelitian yang dilakukan pada artikel terdahulu ini berbeda dengan penelitian di artikel ini. Algoritma yang digunakan juga berbeda, namun ada metode evaluasi yang sama yaitu Davis-Bouldin Index yang digunakan dalam artikel penelitian ini. Perbedaan lainnya, penelitian terdahulu ini tidak membangun aplikasi cerdas seperti yang dilakukan pada penelitian di artikel ini.

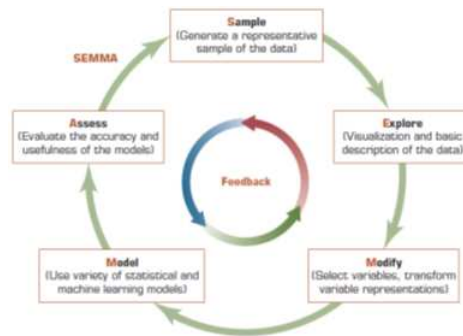
Yudhistira Arie Wijaya, Dedy Achmad Kurniady, Eddy Setyanto, Wahdan Sanur Tarihoran, Dadan Rusmana, dan Robbi Rahim (2021) melakukan penelitian mengenai pemetaan fasilitas sekolah. Metode Davis-Bouldin Index digunakan untuk mengoptimalkan pemetaan [14]. Artikel terdahulu ini menggunakan algoritma yang berbeda dengan penelitian pada artikel ini yaitu algoritma *K-means*. Obyek penelitian yang dilakukan pada artikel terdahulu ini berbeda dengan penelitian di dalam artikel ini, termasuk metode evaluasi yang digunakan dalam artikel penelitian ini. Perbedaan lainnya yaitu penelitian terdahulu ini tidak membangun aplikasi cerdas sebagaimana yang dilakukan pada penelitian di artikel ini.

Merujuk pada tinjauan pekerjaan penelitian terkait pada artikel penelitian terdahulu, algoritma *dynamic cluster* pada *K-means* yang digunakan pada penelitian ini bertujuan untuk mengelompokkan data industri rumahan agar dapat menghasilkan klaster dan proses iterasi yang optimal sesuai Permen PPA No 2 Tahun 2016 yaitu pemula, berkembang dan maju. Selain itu, penelitian ini juga memberikan kontribusi dalam membangun aplikasi cerdas berbasis web dengan menggunakan teknik *data mining* SEMMA (*Sample, Explore, Modify, Model, Assess*).

Adapun struktur organisasi penulisan dari artikel ini dibagi menjadi: pembahasan sub bagian ke 2 tentang metodologi penelitian. Sub bagian ke 3 membahas tentang hasil dan analisis. Sub bagian berikutnya membahas mengenai kesimpulan dari penelitian yang dilakukan yaitu tentang temuan dari penelitian ini, kebaharuan yang diberikan serta saran untuk penelitian lanjut.

2. METODE PENELITIAN

Pengembangan sistem aplikasi cerdas pada penelitian ini menggunakan bahasa pemrograman PHP dan basis data Mysql. Data penelitian yang digunakan dalam penelitian ini menggunakan data industri rumahan yang diperoleh dari Dinas DP3ACSKB Provinsi Kepulauan Bangka Belitung, Indonesia bulan April tahun 2021. Sebanyak 3466 data industri rumahan yang menjadi sampel dalam penelitian ini. Faktor utama untuk menentukan klasifikasi industri rumahan berdasarkan tenaga kerja, teknologi produksi, sumber modal usaha, jumlah modal, lama usaha dan pola produksi. Teknik data *mining* yang digunakan dalam penelitian ini adalah teknik klastering dan menggunakan algoritma *Dynamic Cluster* pada *K-means*. Penelitian ini membangun sistem aplikasi cerdas data *mining* berbasis web menggunakan SEMMA (*Sample, Explore, Modify Model, Assess*) [15]. SEMMA mengacu pada standar proses data *mining* yang merupakan tahapan strategi dalam pemecahan masalah pada bisnis atau unit penelitian. Teknik data *mining* pada metode SEMMA dalam pengolahan data terbagi menjadi lima tahapan sebagai berikut (1) Tahap pengumpulan data (*Sample*), (2) Tahap deskripsi data (*Explore*), (3) Tahap transformasi data (*Modify*), (4) Tahap pemodelan data (*Model*), (5) Tahap evaluasi data (*Assess*) yang ditunjukkan pada Gambar 1 [16].



Gambar 1. Proses Data Mining SEMMA [16]

2.1. Tahap Pengumpulan Data Sampel

Pada tahap ini dilakukan pengumpulan data awal sampel sebagaimana ditunjukkan pada Tabel 1. Data yang dikumpulkan adalah data industri rumahan dari 7 Kabupaten/Kota Provinsi Kepulauan Bangka Belitung, data tersebut akan digunakan untuk proses penelitian. Data yang terkumpul dari hasil pengumpulan data industri rumahan sebanyak 3.466 data. Data industri rumahan sebanyak 3.466 tersebut kemudian dilakukan pemilahan data yang bertujuan untuk menentukan data mana yang digunakan (atau data yang di perlukan) untuk proses penelitian.

Tabel 1. Sampel Data Awal Industri Rumahan Sebelum di Normalisasi

No.	NIK	No. Kartu Keluarga	Nama Industri Rumahan	Alamat	Nama Produk	Tahun Berdiri	Perizinan	Jumlah Tenaga Kerja	Teknologi Produksi	Sumber Modal Usaha	Jumlah Modal	Lama Usaha	Pola Produksi
1	1901082703430000	1901082703430000	Toko Oleh-oleh Sahir	Jl.Pikas Rt.01 Rw.01	Makanan	2021	P-IRT	1-2 orang	Manual	Sendiri	< 5 jt	< 1 tahun	Tidak kontinyu
2	1901051501610000	1901051501610000	Toko Kue Yeyen	Jl.Pikas Rt.05 Rw.03	Makanan	2019	P-IRT	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt - < 25 jt	1 2 tahun	Semi kontinyu
3	1901085103710000	1901085103710000	Umah Crafter	Jl.Rustam Efendi Rt.06 Rw.02	Kerajinan	2019	P-IRT	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt - < 25 jt	1 2 tahun	Semi kontinyu
4	1901085504810000	1901085504810000	Meilani Cake	Jl.Terubuk Rt.05 Rw.02	Makanan	2021	P-IRT	1-2 orang	Manual	Sendiri	< 5 jt	< 1 tahun	Tidak kontinyu
5	1901084505920000	1901084505920000	Tenun Cual Ibu Maslina	Jl.Raya selindung Rt.03 Rw.01	Kerajinan	2020	P-IRT	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt - < 25 jt	1 2 tahun	Semi kontinyu
6	1901086304960000	1901086304960000	Pewter Tin HandiCraft	Jl.Yos Sudarso 3 Rt.01 Rw.01	Kerajinan	2019	P-IRT	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt - < 25 jt	1 2 tahun	Semi kontinyu
7	1901085104790000	1901085104790000	Asusi Rajutan	Jl.R hundani Rt.05 Rw.02	Kerajinan	2019	P-IRT	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt - < 25 jt	1 2 tahun	Semi kontinyu
8	1901086711730000	1901086711730000	Lukis Kaca	Jl.Selangat RT 06/RW 02 kel.selindung baru Rt.01/01	Kerajinan	2019	P-IRT	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt - < 25 jt	1 2 tahun	Semi kontinyu
9	1901085506670000	1901085506670000	Go-Ban-Art	Jl. Depati Hamzah RT. 08/02	Kerajinan	2019	P-IRT	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt - <	1 2 tahun	Semi kontinyu
10	1901084107580010	1901084107580010	Toko dani (zhafira rajut)	Jl. Air Mawar RT. 09 RW. 02 Kel. Air Mawar	Kerajinan	2021	P-IRT	1-2 orang	Manual	Sendiri	< 5 jt	< 1 tahun	Tidak kontinyu
.	.	.	.	.	.	.	.	.	.	.	.	.	.
3466	1901081406560000	1901081406560000	Sita Salon	Jl. Kalamaya RT. 003 RW. 001 Kel. bacang	Jasa	2021	NIB	1-2 orang	Semi Manual/Teknologi sederhana	Sendiri	< 5 jt	< 1 tahun	Semi kontinyu

2.2. Tahap Deskripsi Data (*Explore*)

Pada tahap ini dilakukan eksplorasi dari data yang terkumpul dengan cara menjelaskan data secara lengkap sesuai dengan atribut-atribut yang akan digunakan seperti nama industri rumahan, nama produk, sektor industri, alamat industri rumahan, kota/kabupaten, kecamatan, kelurahan/desa, tahun berdiri, *latitude* dan *longitude*, perizinan, tenaga kerja, teknologi produksi, sumber modal usaha, jumlah modal, lama usaha, dan pola produksi. Pada tahap ini juga dilakukan data *selection* dan data *pre-processing/data cleaning*. Adapun Teknik yang digunakan untuk menyeleksi data penelitian ini yaitu *feature selection* yang terdiri dari tahap *eliminate*, *engineer* dan *validate*.

1. *Eliminate*

Pada tahap ini dilakukan pemilihan atribut yang akan digunakan dalam penelitian seperti nama industri rumahan, tenaga kerja, teknologi produksi, sumber modal usaha, jumlah modal, lama usaha, dan pola produksi.

2. *Engineer*

Setelah proses *eliminate* selanjutnya data sumber tersebut dibentuk kembali menjadi data yang siap divalidasi. Hasil dari tahap *engineer* dapat dilihat pada Tabel 2.

Tabel 2. Sampel Data Industri Rumahan Siap Divalidasi

No.	Nama Industri Rumahan	Tenaga Kerja	Teknologi Produksi	Sumber Modal Usaha	Jumlah Modal	Lama Usaha	Pola Produksi
1	Toko Oleh-oleh Rizki Sahir	1-2 orang	Manual	Sendiri	< 5 jt	< 1 tahun	Tidak kontinyu
2	Toko Kue Yeyen	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt-<25 jt	1-2 tahun	Semi kontinyu
3	Umah Crafter	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt-<25 jt	1-2 tahun	Semi kontinyu
4	Meilani Cake	1-2 orang	Manual	Sendiri	< 5 jt	< 1 tahun	Tidak kontinyu
5	Tenun Cual Ibu Maslina	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt-<25 jt	1-2 tahun	Semi kontinyu
6	Pewter Tin HandiCraft	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt-<25 jt	1-2 tahun	Semi kontinyu
7	Asusi Rajutan	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt-<25 jt	1-2 tahun	Semi kontinyu
8	Lukis Kaca	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt-<25 jt	1-2 tahun	Semi kontinyu
9	Go-Ban-Art	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt-<25 jt	1-2 tahun	Semi kontinyu
10	Toko dani (zhafira rajut)	1-2 orang	Manual	Sendiri	< 5 jt	< 1 tahun	Tidak kontinyu
...	...	...	...	...	...	...	...
3464	Sita Salon	1-2 orang	Manual	Sendiri	< 5 jt	< 1 tahun	Tidak kontinyu
3465	Toko Oleh-oleh Rizki Sahir	1-2 orang	Manual	Sendiri	< 5 jt	< 1 tahun	Tidak kontinyu
3466	Toko Kue Yeyen	1-2 orang	Manual	Sendiri	< 5 jt	< 1 tahun	Tidak kontinyu

2.3. Tahap Transformasi Data (*Modify*)

Pada tahap ini dilakukan proses transformasi data dengan tujuan untuk mengubah data menjadi format yang dapat diproses dengan mudah. Caranya yaitu dengan memodifikasi data hasil dari proses deskripsi data dengan memilih, dan mengubah variabel untuk fokus pada proses pemilihan model. Atribut yang semula bertipe *numeric* disesuaikan menjadi tipe kategori dengan tujuan agar proses *data mining* menjadi lebih mudah. Pada kasus penelitian ini atribut yang dikelompokkan dan diinisialisasi adalah atribut berdasarkan klasifikasi tenaga kerja, teknologi produksi, sumber modal usaha, jumlah modal, lama usaha, dan pola produksi. Hasil pengelompokan atribut yang dilakukan dapat dilihat pada Tabel 3.

Tabel 3. Transformasi Data Industri Rumahan

No.	Tenaga Kerja	Teknologi Produksi	Sumber Modal Usaha	Jumlah Modal	Lama Usaha	Pola Produksi	Inisialisasi Bobot
1	1-2 orang	Manual	Sendiri	<5 jt	<1 tahun	Tidak kontinyu	1
2	3-5 orang	Semi Manual/Teknologi sederhana	Sendiri+pinjaman dari LKM non formal	5 jt - < 25 jt	1 2 tahun	Semi kontinyu	2
3	6-10 orang	Teknologi tinggi	Sendiri+pinjaman dari LKM Formal	25 jt - < 50 jt	>2 tahun	Kontinyu	3

2.4. Tahap Pemodelan Data (Model)

Tahap selanjutnya yaitu mengerjakan desain rancangan antar muka (*interface*) yang dibutuhkan pada aplikasi cerdas yang dibangun menggunakan UML (*Unified Modeling Language*) serta membuat rancangan alur proses sistemnya, sebagaimana ditunjukkan pada Gambar 2.



Gambar 2. Diagram Use Case Rancangan Alur Sistem Industri Rumahan

Pada Gambar 3 menunjukkan bahwa seorang admin harus melakukan *login* terlebih dahulu untuk dapat mengakses sistem, setelah melakukan *login* admin dapat melihat tampilan *home/beranda/dashboard* (halaman admin). Setelah itu admin dapat menginputkan data industri rumahan di sistem, selain itu admin juga dapat mencari data industri yang telah diinputkan tadi di menu dan halaman cari industri. Selanjutnya, admin harus menginputkan dahulu kriteria untuk menentukan kelas industri rumahan berdasarkan tenaga kerja, teknologi produksi, sumber modal usaha, jumlah modal, lama usaha, dan pola produksi. Data kriteria yang sudah diinputkan tadi dilakukan transformasi terlebih dahulu. Setelah itu data kriteria yang sudah ditransformasi diproses menggunakan algoritma *Dynamic cluster* pada *K-means*. Setelah perhitungan dilakukan admin dapat melihat hasil pengelompokan industri rumahan menjadi 3 kelas usaha yaitu pemula, berkembang dan maju.

2.5. Tahap Evaluasi Data (Assess)

Pada tahap ini dilakukan proses pengukuran kualitas hasil klaster pada algoritma *Dynamic cluster* pada *K-means* dan algoritma *K-means* menggunakan metode evaluasi *Davies-Bouldin Index* (DBI). Persamaan yang digunakan Davies-Bouldin Index akan menghasilkan nilai paling optimum yakni nilai terkecil yang dihasilkan oleh persamaan (1) dan (2).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \{D_{i,j}\} \quad (1)$$

$$D_{i,j} = \frac{\bar{d}_i + \bar{d}_j}{d_{i,j}} \quad (2)$$

Berdasarkan persamaan (1) dan (2), DBI merupakan nilai Davies-Bouldin Index, k merupakan jumlah klaster, $D_{i,j}$ merupakan nilai dari jarak antara klaster i dan j . $\bar{d}_i$ dan $\bar{d}_j$ merupakan jarak rata-rata antara data yang ada di klaster i dan klaster j sementara. Semakin kecil nilai *Davies Bouldin Index* (DBI) yang diperoleh (non-negatif ≥ 0), maka semakin baik klaster yang diperoleh dari pengelompokan menggunakan algoritma *clustering*.

2.6. Algoritma Dynamic Cluster pada K-means

Pada bagian ini dijelaskan mengenai langkah-langkah melakukan klastering menggunakan algoritma *Dynamic Cluster* pada *K-means* dalam bentuk *pseudocode* yang digunakan dalam penelitian ini. Adapun langkah-langkah melakukan klastering data industri rumahan ditunjukkan pada Tabel 4.

Tabel 4. Pseudocode Algoritma *Dynamic Cluster* pada K-Means

Algoritma <i>Dynamic Cluster</i> pada K-Means
Input: Data set $D = \{x_1, x_2, \dots, x_n\}$; Jumlah cluster K.
Output: Cluster K dibagi ke dalam cluster $c = \{c_1, c_2, \dots, c_n\}$
1. Mulai
2. Masukkan dataset industri rumahan
3. <i>Preprocessing</i> data
4. Masukan jumlah klaster K, nilai K=3.
5. Inisialisasi titik k dengan memilih secara acak titik k untuk dijadikan pusat <i>cluster</i> atau <i>centroid</i>
6. Kelompokkan semua data atau objek yang memiliki kemiripan dengan cara menghitung jarak ke titik <i>centroid</i> terdekat dengan jarak <i>euclidean distance</i> menggunakan persamaan (1)
7. Hitung kembali titik <i>centroid</i> baru berdasarkan keanggotaan yang terbentuk. Titik <i>centroid</i> baru adalah rata-rata dari semua data atau objek dalam klaster tertentu
8. Jika anggota claster sudah mencapai konvergensi maka lakukan pada proses hitung intra <i>cluster</i> dan juga inter klaster
9. Jika intra baru < intra lama dan inter baru > inter lama, maka jumlah <i>cluster</i> akan ditambahkan dengan $k = k + 1$ dan kembali pada proses kedua menentukan titik <i>centroid</i> data dan mengalokasikan data. Jika tidak maka proses klaster akan berhenti pada k=k
10. Hasil klaster C1, C2 dan C3
11. Selesai

Pengelompokan data atau objek yang memiliki kemiripan agar dapat ditentukan suatu data atau objek tersebut masuk ke dalam suatu klaster tertentu dihitung menggunakan jarak *Euclidean* menggunakan persamaan ((3)).

$$d_{i,j} = \sqrt{(X_{1i}, X_{1j}^2) + (X_{2i}, X_{2j}^2) + \dots + (X_{ni}, X_{nj})^2} \tag{3}$$

Persamaan (3) menjelaskan bahwa $d_{i,j}$ adalah jarak data ke i ke pusat klaster j , kemudian x_{ni} merupakan data ke i pada atribut data ke n dan x_{nj} adalah Titik pusat atau *centroid* ke j pada atribut ke n . Didalam algoritma *Dynamic cluster* pada K-means terdapat istilah *inter cluster* yang merupakan jarak minimum antar titik *centroid* atau pusat klaster dengan pusat klaster lainnya. Inter ini biasanya digunakan untuk mengukur tingkat kemiripan minimum pada klaster lain atau pemisahan antar klaster yang ditunjukkan pada persamaan (4) [[17]].

$$inter = \min(\sum_{i=0, j=0}^{N-1} \text{dist}(X_k, X_j)) \forall k = 1, 2, \dots, k - 1 \text{ and } k = k + 1, \dots, k_n \tag{4}$$

Persamaan (4) menjelaskan bahwa x_k adalah jarak ke pusat klaster sebelumnya, sedangkan x_j adalah jarak ke pusat klaster berikutnya, dan k adalah jarak pusat klaster. Berikutnya didalam algoritma *Dynamic cluster* pada K-means juga terdapat istilah *intra* yang merupakan jarak antara data dengan titik *centroid* dalam suatu klaster. Intra digunakan untuk mengukur tingkat kemiripan maksimum suatu klaster yang ditunjukkan oleh persamaan (5) [18].

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i, X_m)^2} \tag{5}$$

Persamaan (5) menjelaskan bahwa n merupakan jumlah data, x_i merupakan data, dan x_m merupakan *centroid*.

3. HASIL DAN ANALISIS

Sistem aplikasi cerdas yang dikembangkan ini menggunakan bahasa pemrograman PHP. Proses rancang bangun pada sistem aplikasi cerdas menghasilkan sejumlah *sub-window* sebagai berikut:

3.1. Sub-Window dari Form Login

Form pada Gambar 3 berfungsi untuk memberikan hak akses kepada seorang admin (administrator) untuk menggunakan sistem. Form login admin ini meliputi *textbox email*, *textbox password*, tombol login, tombol register dan kembali. Tombol login berfungsi untuk memproses email dan password masuk ke dalam sistem apakah data yang diinputkan pada *textbox* sudah sesuai atau tidak.



Gambar 3. Halaman Form Login

3.2. Sub-Window Halaman Admin

Halaman utama admin setelah dari *form login* akan menampilkan menu-menu untuk menggunakan sistem sebagaimana ditunjukkan pada Gambar 4. Pada halaman ini terdiri dari 7 menu, yaitu: *Dashboard*, *Data Industri*, *Cari Industri*, *Kriteria*, *Assessment*, *Dynamic K-Means* dan *Cluster Results*.



Gambar 4. Halaman Admin

3.3. Sub-Window Data Industri

Menu data industri yang ditunjukkan pada Gambar 5 menampilkan data yang sudah ditambahkan (di input kan) melalui tombol tambah sesuai dengan atribut yang disediakan dan terdapat tombol di sebelah kanan tabel yaitu tombol edit dan hapus. Pada halaman ini juga terdapat beberapa tombol *export* data kedalam format csv dan excel.

Gambar 5. Halaman Data Industri

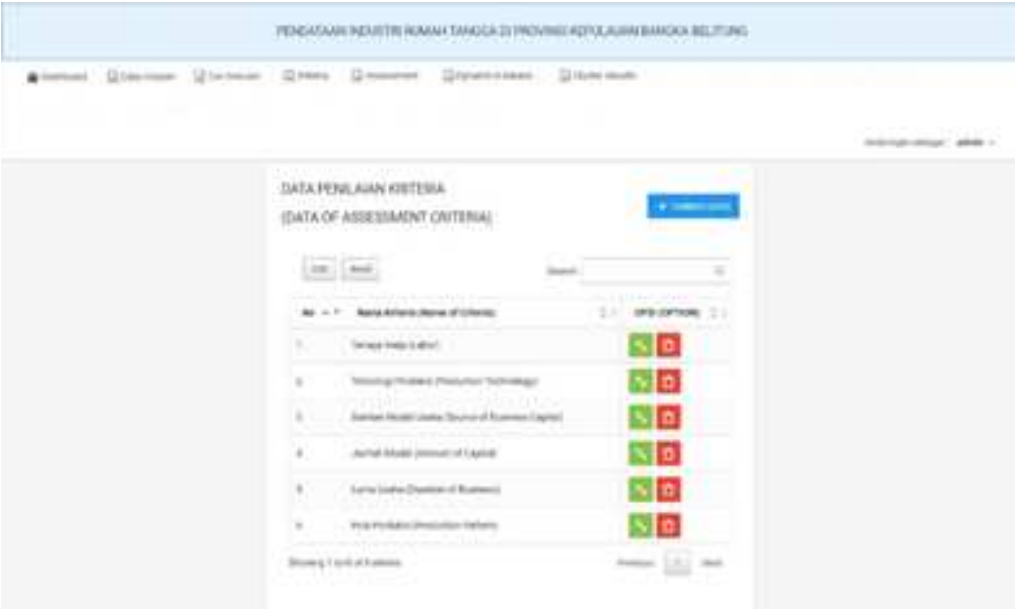
3.4. Sub-Window Cari Industri

Gambar 6. Halaman Cari Industri

Gambar 6 diatas adalah tampilan menu cari data industri. Menu ini menampilkan data yang sudah ditambahkan (di input kan) melalui menu data industri, kemudian pada halaman ini dapat mencari data industri rumahan berdasarkan Kabupaten/Kota dan kelas usaha industri rumahan. Dibagian kanan atas terdapat tombol cetak. Pada halaman ini juga terdapat beberapa tombol *export* data kedalam format csv dan excel.

3.5. Sub-Window Kriteria

Gambar 7 dibawah ini adalah tampilan menu kriteria. Menu ini menampilkan data kriteria yang sudah ditambahkan (di input kan) melalui tombol tambah. Pada halaman ini juga terdapat beberapa tombol *export* data kedalam format csv dan excel.



Gambar 7. Halaman Data Kriteria

3.6. Sub-Window Assessment

Gambar 8 dibawah ini merupakan tampilan menu *assessment*. Menu ini menampilkan data yang sudah ditambahkan (di input kan) melalui menu data industri dan kriteria. Penilaian dapat diubah melalui tombol edit.



Gambar 8. Halaman Assessment

3.7. Sub-Window Dynamic K-means

Halaman ini merupakan halaman yang digunakan untuk memulai atau membuka halaman penghitungan algoritma *Dynamic cluster* pada *K-means* (sebagaimana disajikan pada Gambar 9).



Gambar 9. Halaman *Dynamic K-Means*

3.8. Sub-Window Cluster Results

Gambar 10 dibawah ini merupakan tampilan menu *cluster result* atau hasil klaster. Menu ini menampilkan data hasil klaster yang sudah ditambahkan (di input kan) melalui menu data industri dan kriteria kemudian dihitung oleh algoritma *Dynamic cluster* pada *K-means* di menu *Dynamic k-means*.



Gambar 10. Halaman *Cluster Results*

3.9. Penghitungan Dynamic Cluster pada K-means

Penghitungan algoritma *Dynamic cluster* pada *k-means* dilakukan untuk mendapatkan hasil pengelompokan industri rumahan yang optimal sesuai permen PPA No. 2 tahun 2016, yaitu pemula (C1), berkembang (C2) dan maju (C3).

3.10. Proses Pembobotan

Pembobotan pada tahap ini diperlukan untuk memberikan kemudahan dalam mendapatkan hasil pengelompokan.

Tabel 5. Tabel Hasil dari Proses Pembobotan

No.	Nama Industri Rumahan	Tenaga Kerja	Teknologi Produksi	Sumber Modal Usaha	Jumlah Modal	Lama Usaha	Pola Produksi
1	Toko Oleh-oleh Rizki Sahir	2	0	1	1	1	1
2	Toko Kue Yeyen	2	0	1	1	1	1
3	Umah Crafter	2	0	1	1	1	1
4	Meilani Cake	2	0	1	1	1	1
5	Tenun Cual Ibu Maslina	2	0	1	1	1	1
6	Pewter Tin HandiCraft	2	0	1	1	1	1
7	Asusi Rajutan	3	0	1	1	1	1
8	Lukis Kaca	2	0	1	1	1	1
9	Go-Ban-Art	1	0	1	1	1	1
10	Toko dani (zhafira rajut)	2	0	1	1	1	1
3464	Sita Salon	1	0	1	1	1	1
3465	Usaha Penjualan Kue Basah	2	0	1	1	1	1
3466	Kerupuk Kemplang	2	0	1	1	1	1

3.11. Proses Menentukan Pusat Kluster

Tabel 6. Hasil Pusat Kluster

Centroid	Tenaga Kerja	Teknologi Produksi	Sumber Modal Usaha	Jumlah Modal	Lama Usaha	Pola Produksi
C1	2	0	1	1	1	1
C2	2	0	1	1	1	1
C3	2	0	1	1	1	1

3.12. Pencarian Jarak Terdekat Pada Iterasi Pertama

Pencarian jarak terdekat dalam algoritma *Dynamic cluster* pada *K-means* diperlukan untuk mengetahui kelompok kelas data industri rumahan terdekat berdasarkan C1, C2, C3, sehingga data industri rumahan dapat dikategorikan termasuk kedalam kluster salah satu dari C1, C2, dan C3.

Tabel 7. Hasil Kluster Iterasi Pertama

No.	Nama Industri Rumahan	C1	C2	C3	Jarak	Kuadrat	Kelompok Data
1	Toko Oleh-oleh Rizki Sahir	0	12	14,69693	0	0	C1
2	Toko Kue Yeyen	8,48528	12	12	8,48528	70	C1
3	Umah Crafter	12	8,48528	8,48528	8,48528	70	C2
4	Meilani Cake	13,4164	10,3923	6	6	36	C3
5	Tenun Cual Ibu Maslina	12	8,48528	8,48528	8,48528	70	C2
6	Pewter Tin HandiCraft	13,4164	10,3923	6	6	36	C3
7	Asusi Rajutan	8,48528	8,48528	12	8,48528	70	C1
8	Lukis Kaca	14,69693	8,48528	0	0	0	C3
9	Go-Ban-Art	8,48528	8,48528	12	8,48528	70	C1
10	Toko dani (zhafira rajut)	8,48528	12	12	8,48528	70	C1
3464	Sita Salon	6	10,3923	13,4164	6	36	C1
3465	Usaha Penjualan Kue Basah	14,69693	8,48528	0	0	0	C3
3466	Kerupuk Kemplang	13,4164	6	6	6	36	C2

Tabel 7 menunjukkan hasil dari proses iterasi pertama. Hasil iterasi pertama didapatkan C1 (pemula) sebanyak 3231, C2 (berkembang) sebanyak 153 dan C3 (maju) sebanyak 82.

3.13. Pencarian Jarak Terdekat Pada Iterasi Kedua

Hasil perhitungan pada iterasi pertama ternyata masih menunjukkan perubahan pada *centroid*, sehingga proses masih berlanjut pada iterasi ke 2.

Tabel 8. Hasil Pusat Kluster

Centroid	Tenaga Kerja	Teknologi Produksi	Sumber Modal Usaha	Jumlah Modal	Lama Usaha	Pola Produksi
C1	0,21774	0,05529	0,29377	0,28686	0,18317	0,15552
C2	5,45454	0,872727	2,50909	5,23636	5,56363	4,03636
C3	5,17647	6	6	4	4,47058	5,29411

Tabel 8 menunjukkan hasil pusat kluster pada iterasi kedua dengan nilai terbaru. Proses pencarian nilai pusat kluster baru dilakukan dikarenakan masih ada perubahan pada *centroid*. Setelah itu pada masing-masing kluster dilakukan pencarian jarak terdekat menggunakan jarak Euclidean seperti yang ditunjukkan pada Tabel 9.

Tabel 9. Hasil Kluster Iterasi Kedua

No.	Nama Industri Rumahan	C1	C2	C3	Jarak	Kuadrat	Kelompok Data
1	Toko Oleh-oleh Rizki Sahir	0,52612	10,55817	12,75968	0,52612	0,25	C1
2	Toko Kue Yeyen	8,17833	10,24351	9,96394	8,17833	65,61	C1
3	Umah Crafter	11,66646	8,18327	8,34465	8,18327	65,61	C2
4	Meilani Cake	12,95509	7,47252	5,91988	5,91988	34,81	C3
5	Tenun Cual Ibu Maslina	11,66646	8,18327	8,34465	8,18327	65,61	C2
6	Pewter Tin HandiCraft	13,01577	8,38085	4,41764	4,41764	19,36	C3
7	Asusi Rajutan	8,1834	8,49719	11,10316	8,1834	65,61	C1
8	Lukis Kaca	14,21154	6,58804	2,74145	2,74145	7,29	C3
9	Go-Ban-Art	8,21375	7,15938	10,91077	7,15938	50,41	C2
10	Toko dani (zhafira rajut)	8,13256	9,3761	10,03454	8,13256	65,61	C1
3464	Sita Salon	5,73013	9,19992	12,28045	5,73013	32,49	C1
3465	Usaha Penjualan Kue Basah	10,0136	8,68763	8,55351	8,55351	72,25	C3
3466	Kerupuk Kemplang	8,25907	8,26287	10,84588	8,25907	67,24	C1

Tabel 9 menunjukkan hasil dari proses pengelompokan dari masing-masing kluster pada iterasi ke 2 berdasarkan jarak terdekat, dimana diperoleh kluster C1 (pemula) sebanyak 3233, C2 (berkembang) sebanyak 152 dan C3 (maju) sebanyak 82. Tabel 10 merupakan hasil perhitungan inter dan intra kluster pada iterasi kedua yang menunjukkan nilai inter lama lebih kecil dari inter baru dan intra lama lebih besar intra baru, maka proses iterasi dilanjutkan ke proses iterasi selanjutnya. Begitu pula untuk iterasi ketiga dan keempat dilakukan proses dan perhitungan dengan cara yang sama seperti yang dilakukan pada proses iterasi pertama dan kedua.

Tabel 10. Hasil Perhitungan Inter dan Intra Kluster Iterasi Kedua

Parameter	Lama	Baru
Inter C1 dan C2	0	5,48832
Inter C1 dan C3	0	7,26706
Inter C2 dan C3	0	1,778732
Intra C1	14,21154	1
Intra C2	11,95379	1
Intra C3	12,75968	1

Hasil dari perhitungan proses iterasi berhenti pada iterasi ke lima, dimana pada iterasi ke lima ini pusat kluster dan kelompok data tidak ada yang berubah. Hasil dari perhitungan proses iterasi kelima ditampilkan pada Tabel 11.

Tabel 11. Hasil Kluster Iterasi Kelima

No.	Nama Industri Rumahan	C1	C2	C3	Jarak	Kuadrat	Kelompok Data
1	Toko Oleh-oleh Rizki Sahir	0,14843	8,67227	11,43124	0,14843	0,01	C1
2	Toko Kue Yeyen	8,35816	9,22488	9,09249	8,35816	68,89	C1
3	Umah Crafter	11,90304	9,11703	7,43269	7,43269	54,76	C3
4	Meilani Cake	13,31318	7,41838	6,76877	6,76877	44,89	C3
5	Tenun Cual Ibu Maslina	11,90304	9,11703	7,43269	7,43269	54,76	C3
6	Pewter Tin HandiCraft	13,28995	9,66468	4,08331	4,08331	16,64	C3
7	Asusi Rajutan	8,392426	7,73173	9,99081	7,73173	59,29	C2
8	Lukis Kaca	14,56493	8,15506	3,97697	3,97697	15,21	C3
9	Go-Ban-Art	8,423927	6,97874	9,68588	6,97874	47,61	C2
10	Toko dani (zhafira rajut)	8,39505	7,6546	9,55221	7,6546	57,76	C2
3464	Sita Salon	5,96124	6,95033	11,39369	5,96124	34,81	C1
3465	Usaha Penjualan Kue Basah	14,56493	8,15506	3,97697	3,97697	15,21	C3
3466	Kerupuk Kemplang	8,38716	8,67227	9,23281	8,38716	68,89	C1

Tabel 11 menunjukkan hasil dari proses pengelompokan dari masing-masing kluster pada iterasi ke 5 berdasarkan jarak terdekat, dimana diperoleh kluster C1 (pemula) sebanyak 3214, C2 (berkembang) sebanyak 167 dan C3 (maju) sebanyak 85.

Tabel 12. Hasil Perhitungan Inter dan Intra Kluster Iterasi Kelima

Parameter	Lama	Baru
Inter C1 dan C2	2,76052	2,76052
Inter C1 dan C3	764.811	764.811
Inter C2 dan C3	488.758	488.758
Intra C1	1.456.493	1.456.493
Intra C2	1.102.243	1.102.243
Intra C3	1.143.124	1.143.124

Tabel 12 merupakan hasil dari perhitungan inter dan intra kluster iterasi kelima yang menunjukkan nilai inter lama sama dengan inter baru dan intra lama sama dengan intra baru, maka proses berhenti pada iterasi kelima.

3.14. Perbandingan Hasil Dinas P3ACSKB, Hasil *Dynamic Cluster* pada *K-means* dan Hasil *K-means*

Tabel 13. Perbandingan Hasil Pengelompokan Dinas P3ACSKB, Hasil *Dynamic Cluster* pada *K-means* dan Hasil *K-means*.

Tabel 13. Perbandingan Hasil Pengelompokan Dinas P3ACSKB, Hasil *Dynamic Cluster* pada *K-means* dan Hasil *K-means*

No.	Proses	Jumlah Data Industri Rumahan	Pemula (C1)	Berkembang (C2)	Maju (C3)
1	Pengelompokan oleh Dinas P3ACSKB Provinsi Kepulauan Bangka Belitung	3466	3224	160	82
2	Pengelompokan menggunakan Dynamic Cluster pada K-Means	3466	3214	167	85
3	Pengelompokan menggunakan K-Means	3466	3234	154	78

Tabel 13 menunjukkan hasil perbandingan dari pengelompokan data industri rumahan berdasarkan pengelompokan oleh Dinas P3ACSKB Provinsi Kepulauan Bangka Belitung dengan pengelompokan *Dynamic cluster* pada *K-means* pada penelitian ini. Hasil yang diperoleh dari proses pengelompokan dengan *Dynamic cluster* pada *K-means* tidak berbeda jauh dengan hasil proses pada pengelompokan manual yang dilakukan oleh Dinas P3ACSKB Provinsi Kepulauan Bangka Belitung. Berdasarkan hasil tersebut dapat disimpulkan bahwa aplikasi berbasis web yang dibangun menggunakan algoritma klusterisasi *Dynamic cluster* pada *K-means* mampu membantu Dinas P3ACSKB Provinsi Kepulauan Bangka Belitung dalam proses pengelompokan data industri rumahan di Kabupaten/Kota Provinsi Kepulauan Bangka Belitung.

3.15. Hasil Evaluasi Kluster menggunakan *Davies Bouldin Index* (DBI)

Hasil evaluasi kluster menggunakan *Davies Bouldin Index* pada Tabel 14 secara keseluruhan menunjukkan bahwa algoritma *Dynamic cluster* pada *K-means* memperoleh nilai lebih kecil dibandingkan dengan nilai algoritma *K-means* dengan nilai DBI sebesar 0,184. Nilai DBI yang lebih kecil menunjukkan bahwa skema kluster pada penelitian ini semakin baik keakuratan hasil klasternya dan lebih optimal.

Tabel 14. Perbandingan Hasil Pengujian Menggunakan Evaluasi *Davies Bouldin Index* (DBI)

Algoritma	Cluster	DBI
<i>K-means</i>	4	0,525
<i>Dynamic Cluster</i> pada <i>K-means</i>	5	0,184

4. KESIMPULAN

Implementasi algoritma *Dynamic cluster* pada *K-means* untuk pengelompokan data industri rumahan pada Dinas P3ACSKB di Provinsi Kepulauan Bangka Belitung memberikan kesimpulan bahwa algoritma mampu membantu kinerja administrasi Dinas P3ACSKB secara otomatis berdasarkan PERMEN PPPA No.2 Tahun 2016. Hasil akhir pengelompokan industri rumahan menggunakan algoritma *Dynamic cluster* pada *K-means* memberikan nilai yang optimal pada iterasi ke lima dengan perolehan kluster pemula (C1) diperoleh sebanyak 3214, kemudian kluster berkembang (C2) diperoleh sebanyak 167 dan kluster maju (C3) diperoleh sebanyak 85. Hasil evaluasi validitas kluster menunjukkan bahwa algoritma *Dynamic cluster* pada *K-means* memperoleh nilai DBI lebih kecil dibandingkan dengan algoritma *K-means* dengan nilai DBI sebesar 0,184. Penelitian ini memiliki keunggulan yaitu membangun sistem cerdas berbasis *website* dan dapat mengelompokkan industri rumahan berdasarkan PERMEN PPPA No.2 Tahun 2016 tentang industri rumahan. Penelitian ini juga masih perlu dikembangkan dengan melakukan penelitian dengan jumlah data besar lainnya.

UCAPAN TERIMA KASIH

Penelitian ini telah mendapatkan dukungan *finansial* dari Kemdikbudristek berdasarkan surat keputusan No.1326/SP2H/LT/LL2/2021 hibah penelitian terapan tahun 2019 yang dilaksanakan pada tahun 2021 dan LPPM ISB Atma Luhur yang telah memberikan dukungan hingga terlaksananya penelitian ini. Berdasarkan hal itu, kami mengucapkan banyak terima kasih atas dukungan tersebut.

REFERENSI

- [1] D. Kurniasih and F. Akbar, "E-Commerce Pandemic Covid-19 Home Industries and SMEs," *International Journal of Research and Applied Technology (INJURATECH)*, vol. 1, no. 1, pp. 70–75, jun 2021.
- [2] S. N. Sinha and R. L. Yadav, "Analysis of Data Using K-Means Clustering Algorithm with Min Max Function," *International Journal of Computer Trends and Technology*, vol. 58, no. 2, pp. 82–84, 2018.
- [3] N. Buslim, R. P. Iswara, and F. Agustian, "The Modeling of "Mustahiq" Data Using K-Means Clustering Algorithm and Big Data Analysis (Case Study: Laz)," *Jurnal Teknik Informatika*, vol. 13, no. 2, pp. 213 – 230, feb 2020.
- [4] I. H. Rifa, H. Pratiwi, and R. Respatiwan, "Clustering of Earthquake Risk in Indonesia Using K-Medoids and K-Means Algorithms," *Media Statistika*, vol. 13, no. 2, pp. 194–205, 2020.

- [5] C. Kamila, M. Adiyatma, G. R. Namang, and R. R. F. Syah, "Systematic Literature Review: Penggunaan Algoritma K-Means untuk Clustering di Indonesia dalam Bidang Pendidikan," *Informatika dan Teknologi (Intech)*, vol. 2, no. 1, pp. 19–24, 2021.
- [6] C. Satria and A. Anggrawan, "Aplikasi K-Means Berbasis Web untuk Klasifikasi Kelas Unggulan," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 1, pp. 111 – 124, nov 2021.
- [7] X. Ran, X. Zhou, M. Lei, W. Tepsan, and W. Deng, "A Novel K-Means Clustering Algorithm with A Noise Algorithm for Capturing Urban Hotspots," *Applied Sciences*, vol. 11, no. 23, pp. 1 – 21, 2021.
- [8] M. P. P. d. P. A. R. Indonesia, "PERMEN PPPA No 2 Tahun 2016 Tentang Industri Rumahang," pp. 1–41, 2016.
- [9] W. Widiarina and R. S. Wahono, "Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam Pemetaan Nasabah Potensial," *Journal of Intelligent Systems*, vol. 1, no. 1, pp. 33–36, 2015.
- [10] M. Z. Hossain, M. N. Akhtar, R. B. Ahmad, and M. Rahman, "A Dynamic K-Means Clustering for Data Mining," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 2, pp. 521–526, 2019.
- [11] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, and T. M. A. Aziz, "Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) Using K-Means Clustering and LRFM Model," *The International Journal of Integrated Engineering*, vol. 11, no. 3, pp. 169–180, 2019.
- [12] K. Ariasa, I. G. A. Gunadi, and I. M. Candiasa, "Optimasi Algoritma Klaster Dinamis pada K-Means dalam Pengelompokkan Kinerja Akademik Mahasiswa (Studi Kasus: Universitas Pendidikan Ganesha)," *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, vol. 9, no. 2, pp. 181–193, 2020.
- [13] R. Siagian, P. Sirait, and A. Halima, "E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model," *JITE (Journal of Informatics and Telecommunication Engineering)*, vol. 5, no. 1, pp. 152–159, 2021.
- [14] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM Journal*, vol. 10, no. 3, pp. 1099–1103, 2021.
- [15] Dean Jared, *Big Data, Data Mining and Machine Learning: Value Creation for Business Leaders and Practitioners*. Hoboken: John Wiley & Sons, 2014.
- [16] R. Sharda, D. Delen, and E. Turban, *Business Intelligence, Analytics and Data Science : A Managerial Perspective*, 4th ed. Pearson Education, 2018.
- [17] C. Sreedhar, N. Kasiviswanath, and P. Chenna Reddy, "Clustering Large Datasets Using K-Means Modified Inter and Intra Clustering (KM-I2C) in Hadoop," *Journal of Big Data*, vol. 4, no. 27, pp. 1 – 19, 2017.
- [18] R. M. Alguliyev, R. M. Aliguliyev, and F. J. Abdullayeva, "PSO+K-Means Algorithm for Anomaly Detection in Big Data," *Statistics, Optimization and Information Computing*, vol. 7, no. 2, pp. 348–359, 2019.

