

Implementasi Algoritma K-Means Clustering dalam Penentuan Klasifikasi Tingkat Pembangunan Perekonomian di Provinsi Bali

I Wayan Wijaya Kusuma Sandi^{a1}, Ida Bagus Gede Dwidasmar^{a2}

Informatics Department, Faculty of Math and Science, Udayana University South Kuta, Badung, Bali, Indonesia

¹wijayakusumasandi@gmail.com

²dwidasmar@unud.ac.id

Abstract

Inequality in economic development is one of the problems faced by regencies/cities in Bali Province. Even though Bali is one of the provinces with a fairly large economic contribution in Indonesia, most of its economic resources are still centered in one area. This study aims to form a cluster of districts/cities in Bali Province based on the performance of regional human and economic development using the K-Means Clustering method in order to support an even distribution of economic development in Bali by making regional-based policies that are in adaptability to the level of the economy. The study result showed that three clusters were formed with the first cluster consisting of Klungkung, Karangasem, Bangli, and Jembrana regencies, the second cluster consisting of Badung Regency and Denpasar City, and the third cluster consisting of Gianyar, Buleleng and Tabanan regencies.

Keywords: Pembangunan Ekonomi, K-Means Clustering, Bali.

1. Pendahuluan

Perekonomian merupakan salah satu prioritas utama dalam pelaksanaan pembangunan nasional, salah satu ciri negara maju adalah sistem perekonomian yang kuat dan berkeadilan. Pembangunan ekonomi adalah serangkaian upaya yang dilakukan dalam perekonomian untuk mengembangkan kegiatan ekonomi guna meningkatkan taraf hidup masyarakat. Pembangunan ekonomi nasional tidak dapat dipisahkan dari pembangunan ekonomi daerah [1], oleh karena itu pembangunan ekonomi daerah harus menjadi salah satu prioritas negara untuk menjadi negara yang lebih maju.

Keberhasilan pembangunan ekonomi daerah dapat diukur dengan beberapa indikator. Indikator utama yang digunakan adalah produk domestik bruto per kapita (PDB) daerah yang dapat digunakan sebagai indikator kegiatan ekonomi secara umum sebagai ukuran kemajuan daerah. Indikator kesejahteraan manusia juga harus dipadukan dengan Indeks Pembangunan Manusia (IPM). IPM dapat memberikan wawasan tentang harapan hidup, pendidikan dan standar hidup di wilayah tersebut. Indikator lain yang dapat digunakan untuk mengukur efisiensi pembangunan ekonomi suatu daerah adalah rasio gini dan pengeluaran per kapita.

Provinsi Bali merupakan salah satu daerah di kawasan Indonesia yang mengalami pertumbuhan ekonomi yang melambat di saat pandemi, perubahan tingkat perekonomian yang terjadi juga berpengaruh pada kesenjangan-kesenjangan perekonomian antar kabupaten/kota di Provinsi Bali [2]. Kesenjangan tersebut akan lebih mudah diatasi apabila setiap daerah memiliki kebijakan - kebijakan yang sesuai dengan tingkat perekonomian daerah tersebut, sehingga perekonomian Bali tidak hanya ditopang oleh kota-kota besar seperti kabupaten Badung dan kota Denpasar.

Berangkat dari hal tersebut, kajian ini bertujuan untuk membentuk kluster wilayah/kota berdasarkan aktivitas pada tingkat ekonomi wilayah. Melalui kluster yang akan dibentuk, pemerintah pusat dan daerah dapat memantau dan mengevaluasi kinerja perekonomian masing-masing kabupaten/kota. Dengan demikian, perbedaan pembangunan antar kabupaten/kota di Provinsi Bali tidak bertambah. Metode K-Means clustering digunakan dalam penelitian ini untuk membentuk cluster wilayah/kota karena metode ini dapat memperkenalkan model yang lebih

fleksibel dan memudahkan penyelesaian perhitungan masalah yang dirumuskan.

1.1. Data Mining

Data mining adalah proses pencarian informasi dalam kumpulan data yang memiliki kapasitas yang besar [3]. Data mining bertujuan untuk mencari relasi dan informasi yang terdapat dalam data sehingga dapat dijadikan ilmu pengetahuan yang berguna. Secara garis besar metode dalam data mining dapat dibagi kedalam lima bagian yaitu klasifikasi, prediksi, asosiasi, estimasi dan klustering.

1.2. Clustering

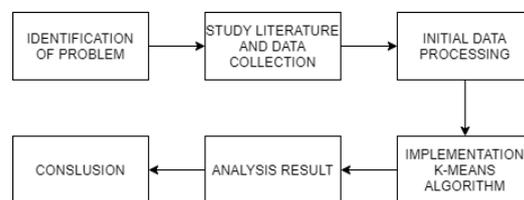
Clustering adalah metode membagi dataset menjadi beberapa kelompok sesuai dengan karakteristik data [4]. Dalam proses clustering, objek-objek dibagi menjadi dua atau lebih cluster sehingga objek-objek dalam satu cluster sangat mirip satu sama lain. Dengan clustering, lebih mudah untuk menemukan pola distribusi dalam data dan juga lebih mudah untuk menentukan hubungan antar atribut data. Beberapa persyaratan pengelompokan penambangan data meliputi skalabilitas, kemampuan untuk menangani berbagai jenis atribut, kemampuan untuk menangani dimensi tinggi, menangani data yang berisik dan mudah diterjemahkan.

1.3. K-Means Clustering

K-Means Clustering merupakan metode clustering yang paling umum digunakan, K-Means merupakan salah satu metode pembelajaran unsupervised data mining, yang membagi data menjadi beberapa kelompok, sehingga data dalam sebuah cluster memiliki sifat yang sama [5]. Algoritma K-Means dikenal karena kemampuannya untuk mempartisi sejumlah besar data dan outlier dengan cepat dan mudah. K-Means bekerja dengan memecah data menjadi beberapa cluster dan mengelompokkannya berdasarkan seberapa mirip data tersebut dengan data centroid. Algoritma K-Means adalah metode non-hierarkis yang terlebih dahulu mengambil beberapa komponen populasi untuk bertindak sebagai pusat cluster awal. Pada langkah awal algoritma K-Means, centroid dipilih secara acak dari kumpulan data, setelah itu data diuji kesamaannya dengan setiap centroid yang membentuk cluster. Posisi pusat dihitung kembali hingga semua komponen data pada setiap cluster terklasifikasi dan akhirnya terbentuk pusat cluster baru.

2. Metode Penelitian

Penelitian ini menggunakan data sekunder dari Badan Pusat Statistik (BPS). Unit analisis mencakup 9 kabupaten/kota di Provinsi Bali dengan acuan waktu tahun 2020. Variabel yang digunakan dalam penelitian ini adalah produk domestik bruto per kapita (PDRB/kapita), pertumbuhan pengeluaran per kapita, rasio gini, angka harapan hidup, rata-rata tahun sekolah dan IPM. Tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Flowchart Alur Penelitian

2.1. Identifikasi Permasalahan

Pada penelitian ini identifikasi permasalahan dilakukan dengan tujuan untuk mengklasifikasikan wilayah di provinsi Bali berdasarkan tingkat perekonomian. Hal ini bertujuan agar peneliti dan pembaca mengetahui tingkat perekonomian setiap daerah sehingga masyarakat dan pemerintah dapat mengambil kebijakan yang tepat untuk mengatasi kesenjangan perekonomian setiap daerah di Provinsi Bali.

2.2. Pengumpulan Data dan Studi Literatur

Pada tahapan ini dilakukan studi literatur untuk memperdalam pengetahuan dasar dalam memahami metode yang digunakan, yang mana data dalam penelitian ini merupakan data kondisi perekonomian Bali. Data tersebut bersumber dari Badan Pusat Statistik Provinsi Bali pada tahun 2020, rincian dari data yang digunakan dapat dilihat dalam Tabel 1.

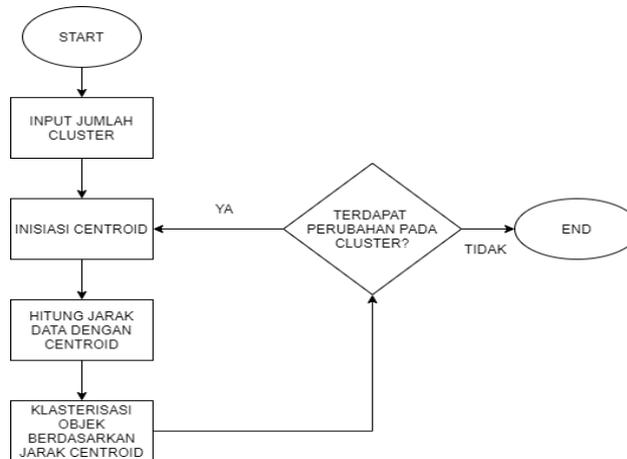
Table 1. Data Pertumbuhan Perekonomian Provinsi Bali

Kabupaten Kota	PDRB (Milyar)	Umur Harapan Hidup	Pertumbuhan Pengeluaran Perkapita (%)	Gini Ratio	IPM	Rata-Rata Lama Sekolah
Jembrana	13.465	72,35	14,99	0,353	72,36	8,23
Tabanan	22.332	73,65	8,08	0,324	76,17	8,88
Badung	49.215	75,1	4,04	0,317	81,6	10,39
Gianyar	25.915	73,68	4,54	0,317	77,36	9,04
Klungkung	8.469	71,25	6,32	0,358	71,73	8,13
Bangli	6.722	70,52	12,65	0,283	69,36	7,17
Karangasem	16.453	70,47	14,97	0,327	67,35	6,32
Buleleng	33.359	71,83	4,3	0,285	72,55	7,24
Denpasar	49.583	74,82	12,37	0,33	83,93	11,47

2.3. Pemrosesan Awal Data

Pada tahapan ini, data yang diperoleh akan diproses menggunakan metode K-Means Clustering. Sebelumnya data tersebut akan diubah ke dalam bentuk integer dan akan dinormalisasi untuk mempermudah pemrosesan. Dalam penelitian ini software yang digunakan untuk membantu proses pengolahan data adalah aplikasi Rapid Miner, aplikasi tersebut digunakan untuk menentukan pengelompokan data sehingga dapat menampilkan hasil akhir pengelompokan clustering.

2.4. Implementasi Algoritma K-Means



Gambar 2. Flowchart Algoritma K-Means

Algoritma K-Means pada penelitian ini digunakan dalam proses klusterisasi data, langkah-langkah dalam implementasinya dapat dilihat sebagai berikut:

1. Tentukan jumlah kluster yang diinginkan
2. Pada iterasi awal, tentukan centroid secara acak
3. Hitung jarak setiap titik ke centroid menggunakan rumus Euclidian Distance
4. Hitung ulang centroid menggunakan nilai rata-rata setiap variabel dari masing-masing cluster
5. Ulangi langkah ketiga dan keempat hingga nilai centroid tidak berubah, jika anggota dari masing-masing cluster tidak terdapat perubahan maka proses klusterisasi sudah selesai

3. Hasil dan Pembahasan

3.1 Seleksi Data

Pada tahap pemilihan data, data yang dipakai adalah data tinjauan perekonomian provinsi Bali pada tahun 2020 yang terdiri dari 7 variabel yaitu nama kabupaten/kota, PDRB, umur harapan hidup, pertumbuhan pengeluaran perkapita, gini ratio, IPM dan rata-rata lama sekolah. Data yang didapatkan bisa langsung diolah tanpa harus melawati *preprocessing* terlebih dahulu.

3.2 Centroid Data

Centroid merupakan titik pusat data yang akan digunakan sebagai perhitungan dalam menentukan cluster, pada iterasi pertama nilai centroid ditentukan secara acak, berikut merupakan centroid awal dalam penelitian ini:

Table 2. Centroid Awal Data

Kabupaten Kota	PDRB (Milyar)	Umur Harapan Hidup	Pertumbuhan Pengeluaran Perkapita (%)	Gini Ratio	IPM	Rata-Rata Lama Sekolah	
Jembrana	13.465	72,35	14,99	0,353	72,36	8,23	C0
Gianyar	25.915	73,68	4,54	0,317	77,36	9,04	C1
Denpasar	49.583	74,82	12,37	0,33	83,93	11,47	C2

3.3 Menghitung Jarak Data ke Titik Pusat Cluster

Langkah selanjutnya adalah menghitung jarak setiap datum dari centroid asal dengan menggunakan persamaan jarak Euclidean. Pada titik ini, jarak terpendek antara data dengan cluster menentukan apakah data berada di cluster 0, cluster 1, atau cluster 2. Berikut adalah hasil perhitungan jarak data pada iterasi pertama:

Table 3. Hasil Iterasi pertama

Kabupaten/Kota	C0	C1	C2	Output
Jembrana	0	12449,96549	36118,56218	C0
Tabanan	8867,22363	3582,74195	27251,34159	C1
Badung	35750,35304	23300,39047	368,3132727	C2
Gianyar	12449,96549	0	23668,60236	C1
Klungkung	4995,937685	17445,89119	41114,49255	C0
Bangli	6742,931405	19192,89373	42861,49291	C0
Karangasem	2987,935402	9462,041979	33130,63494	C0
Buleleng	19894,1129	7444,152005	16224,45683	C1
Denpasar	36118,56218	23668,60236	0	C2

3.4 Menghitung Nilai Centroid Baru

Langkah berikutnya hitung nilai centroid baru menggunakan nilai rata-rata setiap variabel dari masing-masing cluster. Jika nilai centroid berubah, ulangi kembali proses iterasi sehingga nilai centroid dan cluster tidak lagi ada perubahan

Table 4. Centroid Data Baru

C0	11277,1775	71,1475	12,2325	0,33025	70,2	7,4625
C1	27202,00667	73,05333333	5,64	0,308666667	75,36	8,386666667
C2	49399,365	74,96	8,205	0,3235	82,765	10,93

3.5 Hasil Akhir Iterasi

Setelah tidak terdapat perubahan dari hasil cluster maupun centroid maka hasil akhir dari perhitungan sudah bisa didapatkan. Berikut merupakan hasil akhir yang didapatkan dari perhitungan:

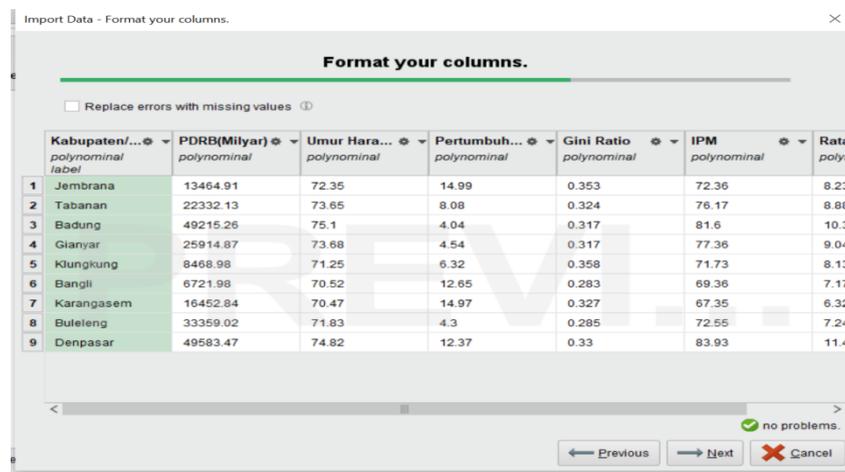
Table 4. Hasil Akhir Cluster

Kabupaten/Kota	C0	C1	C2	Output
Jembrana	2187,735769	13737,10023	35934,4572	C0
Tabanan	11054,95527	4869,87744	27067,2358	C1
Badung	37938,08542	22013,25443	184,1541884	C2
Gianyar	14637,69658	1287,139042	23484,49591	C1
Klungkung	2808,204222	18733,02715	40930,38667	C0
Bangli	4555,197649	20480,02897	42677,38758	C0
Karangasem	5175,664179	10749,17424	32946,5297	C0
Buleleng	22081,84406	6157,014315	16040,34915	C1
Denpasar	38306,29535	22381,46623	184,160388	C2

3.6 Implementasi Algoritma K-Means Menggunakan RapidMiner

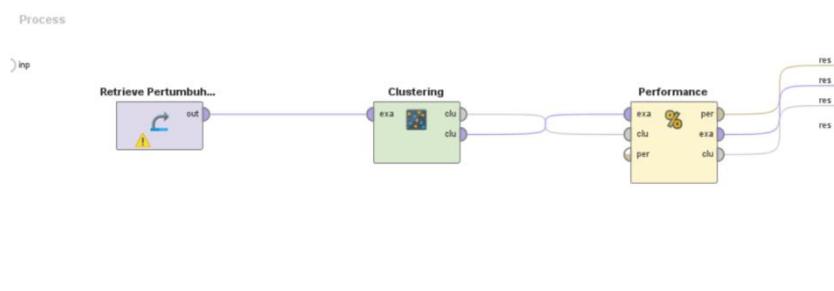
Untuk menerapkan metode K-Means pada data yang telah disiapkan, dataset harus diimport terlebih dahulu ke dalam aplikasi RapidMiner. Kemudian beralih ke desain proses, yang mencakup implementasi dan pengujian menggunakan metode operator RapidMiner yang ada. Penjelasan lebih rinci tentang langkah-langkah ini adalah sebagai berikut.

1. Data yang sudah disimpan dalam format *.xls diimport terlebih dahulu ke dalam RapidMiner sesuai dengan lokasi penyimpanan data. Selanjutnya tentukan atribut yang digunakan sebagai label indikator, dalam penelitian ini label indikator yang digunakan adalah kabupaten/kota



Gambar 3. Import Data ke Dalam RapidMiner

2. Setelah diimport, data diinput ke dalam desain proses dan dilanjutkan dengan memilih metode yang akan digunakan untuk pemrosesan data, yang mana dalam penelitian ini menggunakan metode K-Means



Gambar 4. Desain Proses Pada RapidMiner

3. Tahapan selanjutnya adalah menentukan jumlah cluster yang diinginkan dan maksimal iterasi yang dilakukan dalam proses. Untuk lebih jelasnya dapat dilihat dalam gambar berikut

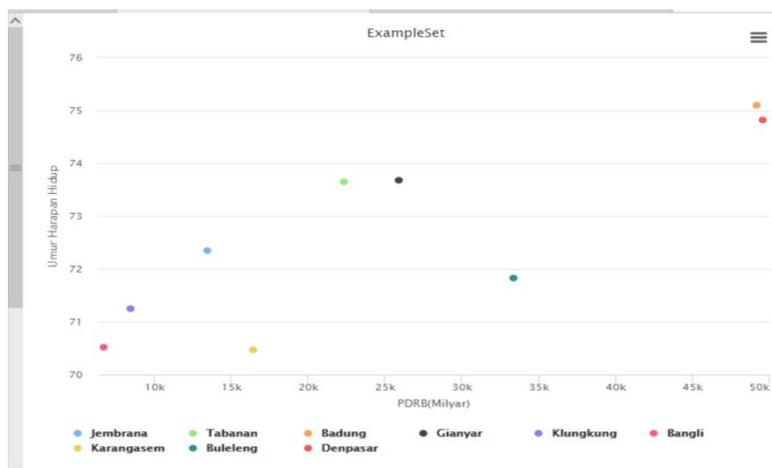
Gambar 5. Input Parameter K-Means

4. Setelah parameter klasterisasi ditentukan maka proses siap dijalankan. Hasil Clustering dari algoritma K-Means dapat dilihat pada gambar 6 dan 7

Open in Turbo Prep Auto Model Filter (9 / 9 examples): all

id	Kabupaten/...	cluster	PDRB(Milyar)	Umur Harap...	Pertumbuha...	Gini Ratio	IPM	Rata-Rata L...
1	Jembrana	cluster_0	13464.910	72.350	14.990	0.353	72.360	8.230
2	Tabanan	cluster_2	22332.130	73.650	8.080	0.324	76.170	8.880
3	Badung	cluster_1	49215.260	75.100	4.040	0.317	81.600	10.390
4	Gianyar	cluster_2	25914.870	73.680	4.540	0.317	77.360	9.040
5	Klungkung	cluster_0	8468.980	71.250	6.320	0.358	71.730	8.130
6	Bangli	cluster_0	6721.980	70.520	12.650	0.283	69.360	7.170
7	Karangasem	cluster_0	16452.840	70.470	14.970	0.327	67.350	6.320
8	Buleleng	cluster_2	33359.020	71.830	4.300	0.285	72.550	7.240
9	Denpasar	cluster_1	49583.470	74.820	12.370	0.330	83.930	11.470

Gambar 6. Hasil Clustering Dalam Bentuk Tabel



Gambar 7. Hasil Clusterisasi Dalam Bentuk Plot

Dari gambar 5 dapat dilihat hasil klusterisasi terbagi menjadi 3 kelompok, Cluster 0 terdiri dari kabupaten Jembrana, Klungkung, Bangli dan Karangasem, Cluster 1 terdiri dari kabupaten Badung dan kota Denpasar dan yang terakhir Cluster 2 terdiri kabupaten Tabanan, Gianyar dan Buleleng

3.7 Rata-Rata Variabel Pada Masing - Masing Klaster

Table 4. Rata-Rata Variabel Pada Klaster

Cluster	PDRB (Milyar)	Umur Harapan Hidup	Pertumbuhan Pengeluaran Perkapita (%)	Gini Ratio	IPM	Rata-Rata Lama Sekolah
C0	11.277	71,14	12,23	0,33025	70,2	7,46
C1	49.399	74,96	8,20	0,3235	82,76	10,93
C3	27.202	73,05	5,64	0,3086	75,36	8,38

Berdasarkan Tabel 7 dapat diketahui karakteristik setiap cluster berdasarkan enam variabel yang digunakan dengan rata-rata dari setiap variabel dari setiap cluster yang terbentuk. Karakteristik masing-masing cluster dijelaskan di bawah ini:

1. Klaster 0 dicirikan dengan rata-rata tingkat pembangunan perekonomian paling rendah tinggi, dimana rata-rata PDRB pada klaster ini hanya menyumbang 12% dari rata-rata keseluruhan PDRB Provinsi Bali. Rata-rata pertumbuhan pengeluaran perkapita Klaster ini merupakan yang terbesar pada tahun 2020 dikarenakan klaster ini tidak mengalami penurunan penghasilan yang signifikan karena pandemi covid-19. Dari sisi pendidikan penduduknya, kabupaten/kota dalam Klaster 0 memiliki rata – rata RLS sebesar 7,46 tahun. Gini Ratio pada klaster ini merupakan yang paling tinggi, hal tersebut mengindikasikan bahwa distribusi pendapatan kabupaten/kota Klaster 0 memiliki ketimpangan yang paling tinggi dibandingkan klaster-klaster lainnya.
2. Klaster 1 merupakan klaster yang terdiri dari dua wilayah, yaitu Kabupaten Badung dan Kota Denpasar. Rata-rata PDRB cluster 1 merupakan yang tertinggi diantara klaster lainnya, namun secara pertumbuhan pendapatan perkapita masih lebih kecil jika dibandingkan dengan klaster 0. umur harapan hidup klaster 1 relatif tinggi dibandingkan klaster yang lain, dan kualitas pendidikan penduduk juga berada pada posisi paling tinggi dengan rata-rata bernilai sebesar 10,93 tahun.
3. Klaster 2 dicirikan dengan gini ratio yang paling rendah, yang berarti bahwa ketimpangan pendapatan pada klaster ini lebih rendah dari dua klaster lainnya. Rata-rata pertumbuhan

pendapatan kabupaten/kota klaster ini pada tahun 2020 adalah sebesar 5,64%. PDRB per kapita pada klaster ini berada posisi kedua setelah klaster 1 j, yaitu bernilai 27,202 milyar per tahun. Sama seperti PDRB-nya, secara pendidikan klaster ini berada pada posisi kedua dengan RLS 8,38 tahun.

4. Kesimpulan

Berdasarkan penelitian yang sudah dilakukan sebelumnya, klasterisasi indeks tingkat pembangunan perekonomian di Provinsi Bali dapat dibagi menjadi 3 klaster yaitu klaster C0 yang terdiri dari Kabupaten Klungkung, Jembrana, Karangasem dan Bangli, klaster C1 yang terdiri dari Kabupaten Badung dan Kota Denpasar, dan yang terakhir klaster C2 yang terdiri dari Kabupaten Tabanan, Gianyar dan Buleleng. Penelitian ini juga membuktikan metode Clustering dengan algoritma KMeans dapat digunakan untuk mengelompokkan tingkat pembangunan perekonomian kabupaten/kota di Provinsi Bali. Adanya penelitian ini diharapkan dapat meningkatkan kesadaran masyarakat dan pemerintah akan ketimpangan sosial yang terjadi sehingga kebijakan yang tepat dapat diambil sesuai dengan kebutuhan dan tingkat pembangunan perekonomian masing-masing daerah.

Daftar Pustaka

- [1] Fatmawati, K., & Windarto, A. P., Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi, CESS (Journal of Computer Engineering System and Science), Vol. 3, No. 2, 2018
- [2] Adnyani, N. K. S., & Agustini, D. A. E. Digitalisasi Sebagai Pemulihan Perekonomian Di Sektor Kerajinan Dalam Mendukung Kebangkitan Umkm Di Provinsi Bali. Jurnal Pengabdian Kepada Masyarakat Media Ganesha FHIS, 1(2), 87-96. 2020
- [3] Jollyta, D., Ramdhan, W., & Zarlis, M. Konsep Data Mining Dan Penerapan. Deepublish. 2020
- [4] Dharmayanti, D., Bachtiar, A. M., & Prasetyo, A. C. Penerapan Metode Clustering Untuk Membentuk Kelompok Belajar Menggunakan Di Smpn 19 Bandung. Komputa: Jurnal Ilmiah Komputer dan Informatika, 6(2), 49-56. 2017
- [5] Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. Jurnal Teknologi Dan Sistem Informasi, 2(2), 100-108. 2021

halaman ini sengaja dibiarkan kosong