Klastering K-Medoid Untuk Entrepreneur Sorgum

Nia Saurina[1], Endang Noerhartati [2], Marina Revitriani [3], Lestari Retnawati [4]
[1]Jurusan Informatika, Fakultas Teknik, Universitas Wijaya Kusuma Surabaya
[2]Jurusan Teknologi Industri Pertanian, Fakultas Teknik, Universitas Wijaya Kusuma Surabaya
[3]Jurusan Teknologi Industri Pertanian, Fakultas Teknik, Universitas Wijaya Kusuma Surabaya
[4]Jurusan Informatika, Fakultas Teknik, Universitas Wijaya Kusuma Surabaya
Email: [1]niasaurina@gmail.com, [2]endang_noer@uwks.ac.id, [3]marina_revi@uwks.ac.id, [4]lestari.047@gmail.com

Abstract. *Since 2009 Universitas Wijaya Kusuma Surabaya has established the Unit Entrepreneurship Sorghum (UES) and collaborated with sorghum producers in several areas on Java. UES wants to create a system that can help clearly classify the activities of each sorghum entrepreneur, so that UES can provide business assistance to sorghum entrepreneurs in a targeted manner. The number of groups receiving assistance to sorghum entrepreneurs is divided into 5 groups, namely: Sorghum entrepreneur who has the smallest business capital, Sorghum entrepreneur who has a large number of regular customers, Sorghum entrepreneur which has a small number of employees, Sorghum entrepreneur promoting on social media, Sorghum entrepreneur who has the least income. The researcher uses the k-medoids algorithm from the clustering technique in classifying sorghum entrepreneurs. The results show that there are 6 sorghum entrepreneurs in the 1st cluster, 9 sorghum entrepreneurs in the 2nd cluster, 4 sorghum entrepreneurs in the 3rd cluster, 6 sorghum entrepreneurs in the 4th cluster and 10 sorghum entrepreneurs in the 5th cluster. In addition, the results of the evaluation using the K-Medoids method obtained a Silhouette Index of 0.5787 and included in the Reasonable Structure has been founded criteria. Besides that, the results of testing the effect of the k-fold value are the best percentage found in the k-fold value = 6 with an accuracy value of 34.597%..*
Keywords: *entrepreneur sorgum, clustering, k-medoids, Unit Entrepreneurship Sorghum.*

Abstrak. *Sejak tahun 2009 Universitas Wijaya Kusuma Surabaya telah membentuk Unit Entrepreneurship Sorgum (UES) dan bekerjasama dengan produsen sorgum di beberapa daerah di Pulau Jawa. UES ingin membuat sebuah sistem yang dapat membantu mengklasifikasikan secara jelas kegiatan masing-masing entrepreneur sorgum, sehingga UES dapat memberikan pendampingan usaha kepada entrepreneur sorgum secara tepat sasaran. Jumlah kelompok yang menerima pendampingan pengusaha sorgum dibagi menjadi 5 kelompok, yaitu: Entrepreneur sorgum yang memiliki modal usaha paling kecil, entrepreneur sorgum yang memiliki pelanggan tetap yang banyak, Entrepreneur sorgum yang memiliki karyawan sedikit, Entrepreneur sorgum yang mempromosikan di media sosial, Entrepreneur sorgum yang berpenghasilan paling sedikit. Peneliti menggunakan algoritma k-medoids dari teknik clustering dalam mengklasifikasikan pengusaha sorgum. Hasil penelitian menunjukkan bahwa terdapat 6 pengusaha sorgum di klaster 1, 9 Entrepreneur sorgum di klaster 2, 4 Entrepreneur sorgum di klaster 3, 6 Entrepreneur sorgum di klaster 4 dan 10 Entrepreneur sorgum di klaster 5. Selain itu, hasil evaluasi menggunakan metode K-Medoids diperoleh Silhouette Index sebesar 0,5787 dan termasuk dalam kriteria "struktur yang masuk akal telah ditemukan". Selain itu hasil pengujian pengaruh nilai k-fold adalah persentase terbaik terdapat pada nilai k-fold=6 dengan nilai akurasi 34,597%.*
Kata Kunci*: entrepreneur sorgum, clustering, k-medoids, Unit Entrepreneurship Sorghum.*

1. Pendahuluan

Sejak tahun 2009 Universitas Wijaya Kusuma Surabaya telah membentuk Unit Entrepreneurship Sorgum (UES) dan bekerjasama dengan produsen sorgum di beberapa daerah di Jawa. ada beberapa unit yang sudah bergabung dan dapat dilihat pada tabel 1.

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

Tabel 1. Jumlah UES dari Tahun 2020 – 2022

No	*Produk Sorgum di UES*	*2020*	*2021*	*2022*
1	Biji Sorgum Merah	1	3	4
2	Tepung Sorgum	16	18	19
3	Biji Sorgum Putih	-	2	6
4	Sorgum untuk souvenir	2	4	4
5	Mie Sorgum	1	1	1
6	Kecap Sorgum	-	1	1

Melalui UES banyak kegiatan yang dilakukan seperti seminar, talkshow, short course, workshop, workshop, praktek bisnis, kerjasama bisnis, Entrepreneurship Expo, dll. UES juga membantu Entrepreneur muda untuk tumbuh dan berkembang dengan melibatkan UES dalam kegiatan penelitian dan pengabdian masyarakat. Semakin banyaknya UES yang bergabung dan semakin beragamnya produk yang dapat dihasilkan dari sorgum. Selain itu, tidak lengkapnya pelaporan kegiatan yang dilakukan oleh pengusaha sorgum seringkali menyebabkan UES kesulitan dalam memberikan pendampingan usaha. UES ingin membuat sebuah sistem yang dapat membantu mengklasifikasikan secara jelas kegiatan masing-masing pengusaha sorgum, sehingga UES dapat memberikan pendampingan usaha kepada Entrepreneur sorgum secara tepat sasaran.

Bantuan usaha yang diberikan memiliki syarat dan ketentuan yang berlaku dan penentuan penerima manfaat juga memiliki aturan yang telah ditentukan. Jumlah kelompok penerima bantuan kepada pengusaha sorgum dibagi menjadi 5 kelompok, yaitu:

1) Entrepreneur sorgum yang memiliki modal usaha paling kecil
2) Entrepreneur sorgum yang memiliki banyak pelanggan tetap
3) Entrepreneur sorgum yang karyawannya sedikit
4) Promosi Entrepreneur sorgum di media sosial
5) Entrepreneur sorgum yang berpenghasilan paling kecil

Untuk memudahkan UES, dalam memetakan pendampingan usaha kepada Entrepreneur sorgum atas hasil kinerja yang telah dilakukan, diperlukan metode pengelompokan data untuk mengklasifikasikan data penilaian kinerja pengusaha sorgum sesuai dengan penetapan persyaratan pendampingan usaha. Salah satu metode pengelompokan data adalah clustering, clustering adalah metode pengelompokan data, dengan cara mengelompokkan data yang memiliki kesamaan karakteristik antara satu data dengan data lainnya (Mehdi Allahyari, 2017).

Penelitian merupakan salah satu aspek penting dalam pengembangan bidang keilmuan, berbagai penelitian yang telah dilakukan tidak hanya terkonsentrasi pada suatu domain tertentu, namun bisa jadi merupakan kolaborasi dari beberapa domain (Vit Zuraida, 2018). Penelitian ini menggabungkan klastering mengenai entrepreneur, dimana entrepreneur selalu mencari perubahan, menanggapinya, dan memanfaatkannya sebagai peluang. Entrepreneur memiliki pola pikir yang melihat kemungkinan daripada masalah yang diciptakan oleh perubahan. pengusaha telah meningkatkan perhatian mereka terhadap model bisnis mereka sehingga perusahaan mereka dapat tetap kompetitif di pasar saat ini (Montiel, 2021). Entrepreneur muda di negara ini dapat didukung untuk memulai usaha bahan-bahan tersebut, di dalam daerah di mana para petani menanam sorgum dan pengolah telah didirikan. Dan dapat mendukung pengolahan sektor swasta dan peningkatan produksi dari petani (Kipchumba Judith, 2021). Sorgum (sorghum bicolor) (l.) (moench) dibudidayakan di semua benua. meskipun terutama di daerah semi-kering yang lebih hangat di dunia, tanaman ini dibudidayakan di daerah tropis, subtropis, dan beriklim sedang karena ketahanannya terhadap kekeringan, potensi produksi, input yang rendah, dan biaya produksi.

Di Indonesia, sorgum telah dibudidayakan di beberapa provinsi antara lain Lampung Sumatera, Jawa Barat, Jawa Tengah, Jawa Timur, Sulawesi Selatan, Sulawesi Tenggara, Nusa Tenggara Barat (NTB) dan Nusa Tenggara Timur (NTT). Sebagian besar petani telah membudidayakan sorgum untuk sumber pangan dengan sistem monokultur, tumpang sari dan tumpang tindih. Sebelum tahun 1970, budidaya sorgum dilakukan karena kerawanan pangan dan untuk memenuhi kebutuhan pangan masyarakat. Sesuai program swasembada pemerintah Indonesia, prioritas pangan pokok sebagian besar adalah beras. Hal ini menyebabkan berkurangnya areal

budidaya sorgum, kemudian berubah menjadi padi, jagung dan kedelai. Meningkatnya permintaan dari konsumen akan makanan ringan bergizi yang memberikan nutrisi yang sehat dan untuk meningkatkan pemanfaatan sorgum dalam makanan sehari-hari masyarakat, diinginkan untuk mengembangkan produk baru dan bernilai tambah dari sorgum (Sarita Verma, 2018).

Pengenalan pola bertujuan menentukan kelompok atau kategori pola berdasarkan ciri-ciri yang dimiliki oleh pola tersebut. Dengan kata lain, pengenalan pola membedakan suatu objek dengan objek lain (Karina Indah Sari Sitanggang, 2020). Pengenalan pola berupa pengelompokan data adalah tugas pembelajaran tanpa pengawasan di mana kumpulan kategori yang terbatas diidentifikasi sebagai kelompok berdasarkan kesamaan intra kelas yang ada dalam data. Setiap cluster terdapat maksimum intra class similarity dan minimum inter class similarity. Dengan demikian teknik clustering diterapkan untuk mengidentifikasi pengelompokan intrinsik antara sekumpulan data yang tidak diberi label atau dikelompokkan. Teknik ini dapat digunakan ketika kelas tidak diketahui dan tidak menganalisis instance berlabel kelas seperti yang digunakan dalam klasifikasi. Proses klasifikasi yang digunakan perlu mempertimbangkan kemungkinan faktor multi class (Septiyawan Rosetya Wardhana, 2019). Atribut yang memberikan kemiripan yang baik harus diidentifikasi untuk meningkatkan metrik kesamaan antar cluster (Nurhayati, 2018). Properti cluster dapat dianalisis untuk mengidentifikasi profil yang membedakan satu cluster dari yang lain. Kinerja teknik clustering yang baik diukur dari kemampuannya untuk mengidentifikasi pola-pola yang tersembunyi dan menghasilkan kesamaan intra kelas yang maksimal dan mengurangi kemiripan antar kelas antar objek lain antar cluster.

Dari beberapa algoritma clustering, peneliti dapat menggunakan algoritma k-medoids dari teknik clustering dalam mengklasifikasikan Entrepreneur sorgum yang berhak menerima bantuan. Beberapa algoritma yang digunakan dalam metode unsupervised learning adalah K-Means dan K-Medoids. Algoritma K-Means adalah metode partisi yang terkenal untuk clustering (Han, 2012). K-Means adalah metode pengelompokan data non-hierarki yang berusaha mempartisi data yang ada menjadi satu atau lebih cluster atau grup sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam cluster yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam grup yang lain (Zhiguo Liu, 2020). Sedangkan K-Medoids menurut adalah algoritma clustering yang mirip dengan K-Means. Perbedaan dari kedua algoritma ini adalah algoritma K-Medoids menggunakan objek sebagai perwakilan (medoid) sebagai pusat cluster untuk setiap cluster, sedangkan K-Means menggunakan nilai mean sebagai pusat cluster (Samudi, 2020).

Menurut (R. L. Novianto, 2019) melakukan penelitian menggunakan algoritma K-Means clustering untuk menganalisis bisnis perusahaan asuransi, dengan tidak menyertakan jumlah nasabah pada setiap produk asuransinya. Penelitian ini menghasilkan 3 jenis cluster yaitu nilai terendah, sedang dan tertinggi dalam pengelompokan premi, klaim dan nilai pertanggungan. Menurut (T. Taslim, 2016) dalam penelitiannya menggunakan algoritma K-Means untuk clustering data obat pada puskesmas Rumba, dengan menggunakan 2 atribut, yang menghasilkan 3 cluster untuk kategori obat yaitu klaster pertama merupakan obat yang tidak/kurang ada permintaan, klaster kedua terdiri dari obat yang ratarata permintaan nya > 300 buah per bulan dan klaster ketiga yang isinya adalah daftar obat yang sangat tinggi ratarata permintaannya >2000 buah per bulan. Peneliti lain (D. F. Pramesti, 2017) menggunakan metode K-Medoids untuk pengelompokan data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas yang menghasilkan 2 jenis cluster. Cluster 1 area hutan yang termasuk dalam potensi kebakaran yang tinggi dengan hasil rata-rata brightness sebesar 344.470K dan cluster 2 meliputi area dengan potensi kebakaran sedang dengan hasil rata-rata brightness sebesar 318.800K. Penelitian ini menghasilkan rata-rata confidence sebesar 58,73%. Selain itu penelitian (Zaenal Mustofa, 2018) dalam penelitiannya dengan algoritma K-Medoids, untuk mengelompokkan data EDGI E-government Surve 2014 kedalam 4 status EDGI. Model yang dihasilkan diuji untuk mendapatkan nilai Bouldin Index dari setiap algoritma sehingga didapat pengujian dengan menggunakan data dan setelah dilakukan pengujian dengan tools rapidminer didapat nilai Bouldin Index adalah 0.593. Penelitian yang dilakukan (Anggreini, 2019) menggunakan algoritma K-Medoids untuk metode clustering data yang dapat diterapkan untuk menghasilkan informasi-informasi yang dapat diusulkan kepada Direktur Politeknik TEDC Bandung agar dapat mendukung kegiatan promosi penerimaan mahasiswa baru dengan lebih efisien dan efektif.

Berdasarkan penelitian-penelitian sebelumnya didapatkan bahwa algoritma K-Medoid memiliki nilai sensitivitas yang lebih tinggi terhadap dataset yang berjumlah kecil (Zaenal Mustofa, 2018). Sehingga penelitian menggunakan algoritma K-medoid dikarenakan jumlah dataset pada penelitian ini berjumlah 35 entrepreneur sorgumm yang bertujuan untuk mengelompokkan penerima bantuan usaha oleh UES kepada Entrepreneur sorgum.

2. Tinjauan Pustaka

2.1 Entrepreneur

Kata "entrepreneur" berasal dari Perancis pada abad ke-17, ketika orang menganggap entrepreneur sebagai individu yang menjalankan bisnis komersial tertentu. Entrepreneur adalah orang yang memanfaatkan peluang potensial untuk menciptakan nilai secara kreatif dan inovatif dengan memperhitungkan sumber daya yang terbatas. Banyak pendapat tentang bisnis dan entrepreneurship telah terbentuk selama periode waktu yang begitu lama, dan banyak sudut pandang berbeda tentang konsep yang terakhir disediakan. Pemahaman entrepreneur saat ini dikaitkan dengan para ilmuwan seperti (Xing Li, 2015) (Bygrave, 2010), (Gintare Giriuniene, 2016). Beberapa sumber informasi, memberikan sinonim, menunjukkan bahwa bisnis, kewirausahaan dan inisiatif adalah sinonim.

2.2 Sorgum

Sorgum adalah rumput yang berasal dari Sorgum vulgare Pers, yang merupakan.tanaman liar asli Afrika yang tahan kekeringan dan merupakan anggota keluarga rumput tahan panas dan banyak varietas yang dibudidayakan dalam sejarah baru-baru ini berasal dari benua itu. Tanaman sorgum telah menjadi tanaman biji-bijian terbesar keempat di dunia, dan menempati posisi kelima dalam hal luas tanam, setelah gandum, padi, jagung, dan barley (Saddam Hossain, 2022). Bahkan tanaman ini merupakan makanan pokok di daerah tropis semi-kering. Di Indonesia, sorgum berpotensi untuk dikembangkan di lahan kering dan tadah hujan hingga 52,5 juta ha. Namun tanaman sorgum belum banyak dikenal di daerah Lampung, bahkan sulit untuk mendapatkan data resmi sejauh mana budidaya ini. Luas tanaman ini di Lampung masih sangat terbatas, dan hanya ada sekitar 25 ha di Lampung Selatan (Winarti, 2020).

2.3 Klastering K-Medoid

Teknik clustering sebagian besar unsupervised data mining proses ini banyak digunakan untuk mengatur data ke dalam kelompok - disebut cluster - berdasarkan kesamaan antara item data individu. Objek-objek dalam cluster cenderung serupa sedangkan objek-objek yang tergabung dalam cluster yang berbeda tidak serupa (Min Ren, 2016). Clustering adalah pengorganisasian data yang tidak berlabel ke dalam kelompok-kelompok kesamaan yang disebut cluster. Cluster adalah kumpulan item data yang "mirip" di antara mereka, dan "tidak mirip" dengan item data di cluster lain.

K- Medoids adalah algoritma pengelompokan yang berkaitan dengan algoritma K-Means dan algoritma medoids. Algoritma K-Medoids ini diusulkan pada tahun 1987 (Rika Elizabet Sihombing, 2019). Algoritma K- Medoids dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw. Algoritma K- Medoids menggunakan metode partisi clustering untuk mengelompokkan sekumpulan n objek menjadi sejumlah cluster. Algoritma ini menggunakan objek pada kumpulan objek untuk mewakili sebuah cluster. Objek yang terpilih untuk mewakili sebuah cluster disebut dengan medoid. Cluster dibangun dengan menghitung kedekatan yang dimiliki antara medoid dengan objek non-medoid.

2.3 K-Fold Cross Validation

K-Fold Cross Validation adalah salah satu dari jenis pengujian cross validation yang berfungsi untuk menilai kinerja proses sebuah metode algoritme dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K k-fold. Kemudian salah satu kelompok k-fold tersebut dapat dijadikan sebagai data uji sedangkan sisa kelompok yang lain akan dijadikan sebagai data latih (Rohani, 2017).

3. Metode Penelitian

K-medoids menggunakan k sebagai jumlah pusat cluster awal yang dibangkitkan secara acak pada awal proses clustering. Setiap objek yang lebih dekat ke pusat cluster dapat melihat dan

membentuk cluster baru. Algoritma kemudian secara acak menentukan pusat cluster baru dari setiap cluster yang terbentuk sebelumnya dan menghitung jarak antara objek dan pusat cluster baru yang dihasilkan. Jarak antara objek I dan j dihitung menggunakan fungsi pengukuran dissimilarity, salah satunya adalah Euclidean Distance Function. Langkah-langkah dari algoritma K-medoids adalah sebagai berikut (Han, 2012) (Dewi, 2017).

1. Pilih secara acak titik 'k' dari data input ('k' adalah jumlah cluster yang dapat dibentuk). Ketepatan pilihan nilai k dapat dinilai dengan menggunakan metode seperti metode siluet.

2. Setiap titik data ditempatkan ke cluster tempat medoid terdekatnya berada.

3. Untuk setiap titik data cluster i, jaraknya dari semua titik data lainnya dihitung dan ditambahkan. Titik kluster ke-i yang jumlah jaraknya dari titik lain yang dihitung minimal ditetapkan sebagai medoid untuk kluster itu, dan peneliti menggunakan Jarak Euclidean

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

Keterangan:
d = jarak antara x dan y
x = pusat data cluster
y = data pada atribut
I = setiap data
n = jumlah data,
Xi = data di tengah cluster i
yi = data di setiap data ke i

4. Langkah (2) dan (3) diulangi sampai tercapai konvergensi yaitu medoid berhenti bergerak.

Dst

Gambar 1 menjelaskan algoritma K-Medoids yang diimplementasikan untuk klastering entrepreneur sorgum. Algoritma dimulai dengan memasukkan n sebagai banyaknya data dan k sebagai jumlah klaster yang pada penelitian ini, klaster dibagi menjadi 5 klaster. Kemudian algoritma dimulai dengan inisialisasi entrepreneur sorgum sebagai pusat klaster, lalu mulai menghitung jarak Euclidean menggunakan persamaan (1) dan menghitung total distance. Perhitungan total distance dapat berlanjut jika total distance baru dikurangi total distance lama nilainya kurang dari nol, maka hasil klastering untuk entrepreneur sorgum dapat divisualisasikan.

Gambar 1. Klastering K Medoid untuk Entrepreneur Sorgum

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

4. Hasil dan Pembahasan

4.1 Uji Coba Data

Data UES diambil dari dokumen pribadi UES kemudian data tersebut diekstraksi ke dalam data warehouse setelah data dihitung di data warehouse dan dikelompokkan menjadi 5 kelompok. Penelitian ini menggunakan bahasa pemrograman Python untuk membuat clustering, seperti yang telah dilakukan (Weksi Budiaji, 2019), (Alessio Martino, 2017), (Adya Hermawati, 2020). Gambar 1 menjelaskan tampilan dataset Entrepreneur Sorgum.

Gambar 2. Dataset Pengusaha Sorgum

Atribut yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2, dimana terdapat atribut, tipe data dan sampel data.

Tabel 2. Atribut K-Medoid Entrepreneur Sorgum

No	*Produk Sorgum di UES*	*Tipe Data*	*Contoh*
1	ID Entrepreneur	Integer	1
2	Jenis Sorgum	Character	Sorgum Numbu
3	Kota	Character	Sidoarjo
4	Hasil Olahan	Character	Tepung Sorgum
5	Penghasilan	Integer	Rp. 2.500.000
6	Modal awal	Integer	Rp. 2.500.000

Pengkodean atribut digunakan untuk memudahkan pengolahan data karena tipe datanya berbeda-beda. Misalnya data pendapatan, datanya berupa pendapatan Rp. 750.000, Rp. 1.200.000, Rp. 3.100.000.

Tabel 3. Kodefikasi Atribut

No	*Atribut*	*Kodefikasi*
1	Jenis sorgum	Sorghum Numbu, Kategori = 1 Sorghum Pahat, Kategori = 2 Sorghum Tortillero, Kategori = 3 Sorghum Milion, Kategori = 4
2	Pendapatan	Rp. 400.000 – Rp. 1.000.000 Kategori = 1 Rp. 1.100.000 – Rp. 2.000.000 Kategori = 2 Rp. 2.100.000 – Rp. 3.500.000 Kategori = 3 Rp. 3.600.000 – Rp. 5.000.000 Kategori = 4
3	Kota	Lamongan dan Sidoarjo, Kategori = 1 Malang dan Jombang, Kategori = 2 Kediri, Nganjuk dan Tuban, Kategori = 3

		Sumenep, Blitar dan Tulung Agung, Kategori = 4
4	Modal Awal	Rp. 1.000.000 – Rp. 1.500.000, Kategori = 1 Rp. 1.600.000 – Rp. 2.000.000, Kategori = 2 Rp. 2.100.000 – Rp. 3.000.000, Kategori = 3 Lebih dari Rp. 3.000.000 Kategori = 4
5	Produk Olahan	Biji Sorgum, Kategori = 1 Tepung Sorgum, Kategori = 2 Sorgum Souvenir, Kategori = 3 Mie dan Kecap Sorgum, Kategori = 4

Peneliti menentukan analisis pengelompokan pengusaha sorgum di UES menggunakan Algoritma K-Medoids. Berikut ini adalah langkah-langkah pengolahan data menggunakan Algoritma K-Medoids menggunakan data 2022 dengan data sesuai Tabel 2.

a. Menentukan data yang diproses
   Data yang digunakan adalah data UES tahun 2022
b. Menentukan jumlah cluster, atau k yang digunakan sebanyak 5 kelompok seperti yang dijelaskan sebelumnya
c. Penentuan Medoid Center dilakukan secara acak di setiap Cluster. Center Medoid yang digunakan dapat dilihat pada kotak merah pada Gambar 1 dan source code K Medoid dengan Python ditunjukkan pada Tabel 4.

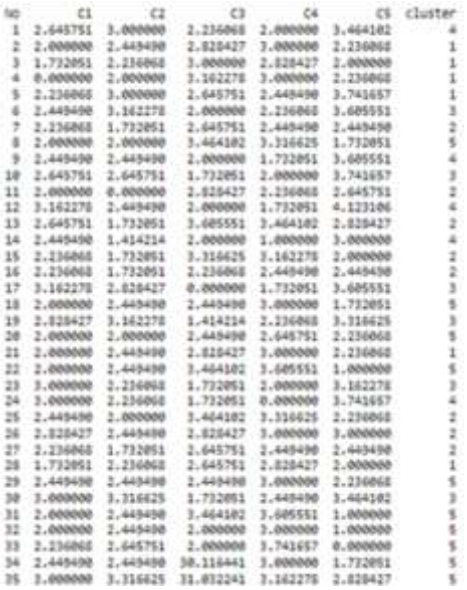
Kode 1. Source Code K-Medoid

```
def __start_algo(self):
        self.__initialize_medoids()
        self.clusters, self.cluster_distances =
self.__calculate_clusters(self.medoids)
        self.__update_clusters()

def __update_clusters(self):
        for i in range(self.max_iter):
            cluster_dist_with_new_medoids = self.__swap_and_recalculate_clusters()
            if self.__is_new_cluster_dist_small(cluster_dist_with_new_medoids) ==
True:
                self.clusters, self.cluster_distances =
self.__calculate_clusters(self.medoids)
            else:
                break
```

d. Menghitung Nilai Euclidian
   Untuk menghitung jarak antara titik Medoid dan titik setiap objek menggunakan Jarak Euclidian. Rumus untuk menghitung Nilai Euclidian dengan rumus pada (1).
   Sehingga didapatkan tabel jarak dari Medoid dan cari nilai minimum dari ketiga Medoid tersebut. Tabel Jarak dari Medoid yang dapat dilihat pada Gambar 2.

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

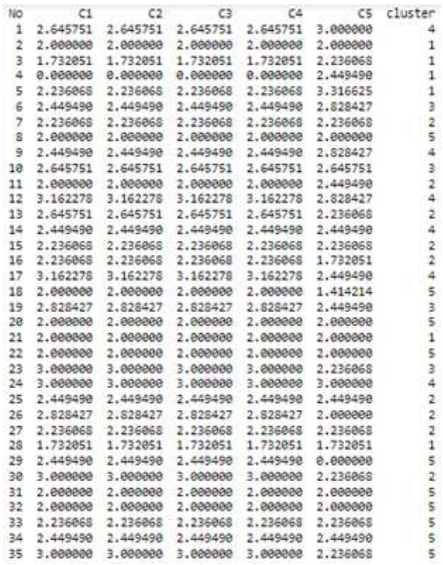


No	C1	C2	C3	C4	C5	cluster
1	2.645751	3.000000	2.236068	2.000000	3.464102	4
2	2.000000	2.449490	2.828427	3.000000	2.236068	1
3	1.732051	2.236068	3.000000	2.828427	2.000000	1
4	0.000000	2.000000	3.162278	3.000000	2.236068	1
5	2.236068	3.000000	2.645751	2.449490	3.741657	1
6	2.449490	3.162278	2.000000	2.236068	3.605551	3
7	2.236068	1.732051	2.645751	2.449490	2.449490	2
8	2.000000	2.000000	3.464102	3.316625	1.732051	5
9	2.449490	2.449490	2.000000	1.732051	3.605551	4
10	2.645751	2.645751	1.732051	2.000000	3.741657	3
11	2.000000	0.000000	2.828427	2.236068	2.645751	2
12	3.162278	2.449490	2.000000	1.732051	4.123106	4
13	2.645751	1.732051	3.605551	3.464102	2.828427	2
14	2.449490	1.414214	2.000000	1.000000	3.000000	4
15	2.236068	1.732051	3.316625	3.162278	2.000000	2
16	2.236068	1.732051	2.236068	2.449490	2.449490	2
17	3.162278	2.828427	0.000000	1.732051	3.605551	3
18	2.000000	2.449490	2.449490	3.000000	1.732051	5
19	2.828427	3.162278	1.414214	2.236068	3.316625	3
20	2.000000	2.000000	2.449490	2.645751	2.236068	5
21	2.000000	2.449490	2.828427	3.000000	2.236068	1
22	2.000000	2.449490	3.464102	3.605551	1.000000	5
23	3.000000	3.236068	1.732051	2.000000	3.162278	3
24	3.000000	2.236068	1.732051	0.000000	3.741657	4
25	2.449490	2.000000	3.464102	3.316625	2.236068	2
26	2.828427	2.449490	2.828427	3.000000	3.000000	2
27	2.236068	1.732051	2.645751	2.449490	2.449490	2
28	1.732051	2.236068	2.645751	2.828427	2.000000	1
29	2.449490	2.449490	2.449490	3.000000	2.236068	5
30	3.000000	3.316625	1.732051	2.449490	3.464102	3
31	2.000000	2.449490	3.464102	3.605551	1.000000	5
32	2.000000	2.449490	2.000000	3.000000	1.000000	5
33	2.236068	2.645751	2.000000	3.741657	0.000000	5
34	2.449490	2.449490	30.116441	3.000000	1.732051	5
35	3.000000	3.316625	31.032241	3.162278	2.828427	5

Gambar. 3. Hasil Clustering Result pada Iterasi Pertama

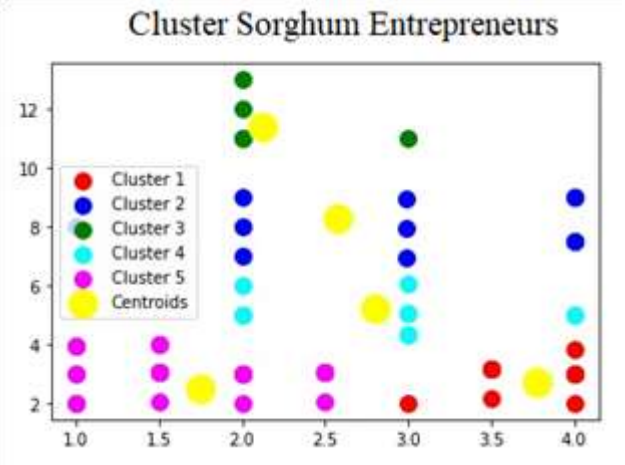
e. Menghitung Nilai Euclidian

Untuk menghitung jarak antara titik Medoid dan titik setiap objek menggunakan Jarak Euclidian. Rumus untuk menghitung Nilai Euclidian dengan rumus pada (1) sehingga didapatkan tabel jarak dari Medoid dan cari nilai minimum dari ketiga Medoid tersebut. Tabel Jarak dari Medoid yang dapat dilihat pada Gambar 2.



No	C1	C2	C3	C4	C5	cluster
1	2.645751	2.645751	2.645751	2.645751	3.000000	4
2	2.000000	2.000000	2.000000	2.000000	2.000000	1
3	1.732051	1.732051	1.732051	1.732051	2.236068	1
4	0.000000	0.000000	0.000000	0.000000	2.449490	1
5	2.236068	2.236068	2.236068	2.236068	3.316625	1
6	2.449490	2.449490	2.449490	2.449490	2.828427	3
7	2.236068	2.236068	2.236068	2.236068	2.236068	2
8	2.000000	2.000000	2.000000	2.000000	2.000000	5
9	2.449490	2.449490	2.449490	2.449490	2.828427	4
10	2.645751	2.645751	2.645751	2.645751	2.645751	3
11	2.000000	2.000000	2.000000	2.000000	2.449490	2
12	3.162278	3.162278	3.162278	3.162278	2.828427	4
13	2.645751	2.645751	2.645751	2.645751	2.236068	2
14	2.449490	2.449490	2.449490	2.449490	2.449490	4
15	2.236068	2.236068	2.236068	2.236068	2.236068	2
16	2.236068	2.236068	2.236068	2.236068	1.732051	2
17	3.162278	3.162278	3.162278	3.162278	2.449490	4
18	2.000000	2.000000	2.000000	2.000000	1.414214	5
19	2.828427	2.828427	2.828427	2.828427	2.449490	3
20	2.000000	2.000000	2.000000	2.000000	2.000000	5
21	2.000000	2.000000	2.000000	2.000000	2.000000	1
22	2.000000	2.000000	2.000000	2.000000	2.000000	5
23	3.000000	3.000000	3.000000	3.000000	2.236068	3
24	3.000000	3.000000	3.000000	3.000000	3.000000	4
25	2.449490	2.449490	2.449490	2.449490	2.449490	2
26	2.828427	2.828427	2.828427	2.828427	2.000000	2
27	2.236068	2.236068	2.236068	2.236068	2.236068	2
28	1.732051	1.732051	1.732051	1.732051	1.732051	1
29	2.449490	2.449490	2.449490	2.449490	0.000000	5
30	3.000000	3.000000	3.000000	3.000000	2.236068	2
31	2.000000	2.000000	2.000000	2.000000	2.000000	5
32	2.000000	2.000000	2.000000	2.000000	2.000000	5
33	2.236068	2.236068	2.236068	2.236068	2.236068	5
34	2.449490	2.449490	2.449490	2.449490	2.449490	5
35	3.000000	3.000000	3.000000	3.000000	2.236068	5

Gambar. 4 Hasil Klastering pada Iterasi Kedua



Gambar. 5. Visualisasi Klaster Entrepreneur Sorgum

Gambar 4 menunjukkan bahwa terdapat 6 pengusaha sorgum di klaster 1, 9 pengusaha sorgum di klaster 2, 4 pengusaha sorgum di klaster 3, 6 pengusaha sorgum di klaster 4 dan 10 pengusaha sorgum di klaster 5.

Algoritma PAM (*Partitioning Around Medoid*) merupakan algoritma yang diwakili oleh sebuah cluster yaitu sebuah medoid. Algoritma ini lebih populer dengan nama K-Medoids Algorithm. Algoritma K-Medoids lebih baik daripada K-Means, karena pada K-Medoids ditemukan k sebagai representasi objek untuk meminimalkan jumlah ketidaksetaraan objek data, sedangkan pada K-Means digunakan jumlah jarak Euclidean untuk objek data (Marlina, 2018). Jika dibandingkan dengan algoritma K-Means, perbedaannya terletak pada titik pusat cluster, K-Means menggunakan nilai rata-rata (mean) sebagai pusat cluster, sedangkan K-Medoids menggunakan objek sebagai pusat cluster. dan objek harus mewakili setiap cluster.

4.2 Silhouette Coefficient

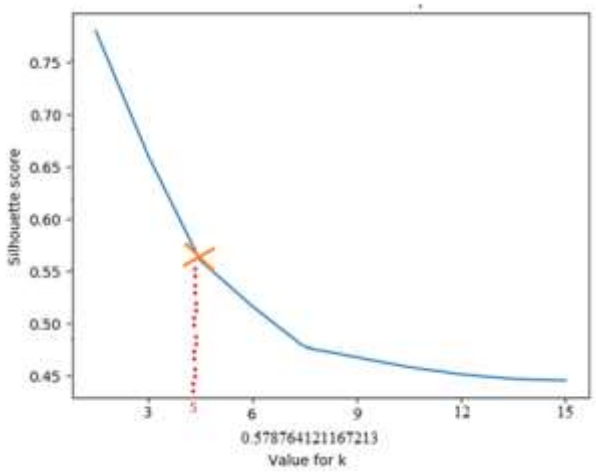
Silhouette Coefficient yang berfungsi untuk menguji kualitas cluster kemudian diambil keputusan apakah hasil data mining dapat digunakan atau tidak. Koefisien Silhouette terletak pada rentang antara nilai -1 hingga 1. Rentang nilai ini digunakan untuk menunjukkan kemiripan yang dekat dari data yang dikelompokkan dalam suatu cluster. Jika nilai rata-rata Silhouette Coefficient mendekati 1, maka cluster tersebut semakin baik. Sebaliknya jika nilai rata-rata Silhouette Coefficient mendekati -1, maka cluster tidak baik. Berikut kriteria pengukuran nilai Silhouette Coefficient (Chatti, 2014) pada Tabel 4.

Tabel 4. Kriteria Penilaian Koefisien Silhouette

No	*Koefisien Silhouette*	*Kriteria Penilaian*
1	$0.7 < SC \leq 1.0$	struktur yang kuat telah ditemukan
2	$0.5 < SC \leq 0.7$	struktur yang masuk akal telah ditemukan
3	$0.25 < SC \leq 0.5$	struktur yang lemah telah ditemukan
4	$SC \leq 0.25$	tidak ditemukan struktur yang substansial

Pada bagian ini dapat dilakukan uji coba untuk menentukan jumlah cluster yang tepat berdasarkan nilai SSE (*Sum of Square Error*) yang mengalami penurunan drastis. Semakin tinggi nilai SSE maka kualitas cluster semakin rendah, begitu juga sebaliknya. Semakin kecil nilai SSE maka semakin baik kualitas cluster (Chatti, 2014). Pada Gambar 6 dapat dilihat bahwa ketika jumlah cluster k=1 menunjukkan nilai SSE tertinggi, maka ketika jumlah cluster k=2 nilai SSE menurun secara signifikan. Saat jumlah cluster k = 3 maka nilai SSE kembali menurun, begitu seterusnya hingga jumlah cluster k = 5 juga berkurang. Berdasarkan Gambar 5, hasil evaluasi dengan menggunakan metode K-Medoids diperoleh Silhouette Index sebesar 0,5787 dan termasuk dalam kriteria "struktur yang masuk akal telah ditemukan".

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)



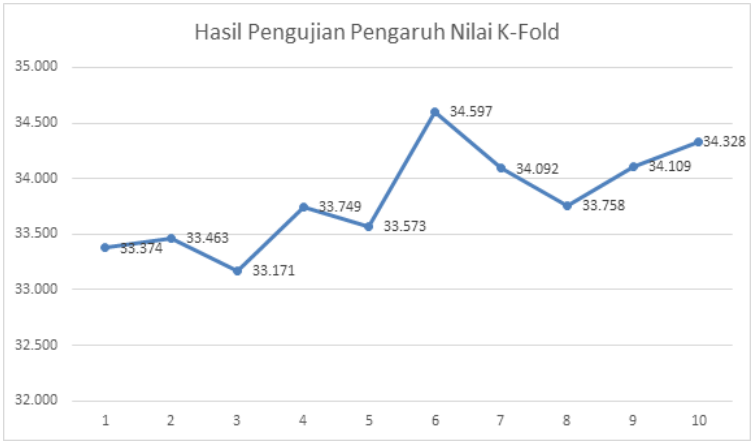
Gambar. 6. Hasil Silhouette untuk Entrepreneur Sorgum

Kode 2. Source Code Koefisien Silhouette

```
# Calculate Silhoutte Score
score = silhouette_score(X, km.labels_, metric='euclidean')
# Print the score
print('Silhouetter Score: %.3f' % score)
```

4.3 Hasil pengujian pengaruh nilai k-fold

Pengujian pengaruh nilai k-fold dilakukan mengunakan nilai k-fold=2 hingga k-fold=10. Pada tiap k-fold dilakukan inputan nilai K=1 hingga K=10. Hasil pengujian pengaruh nilai kfold ini menghasilkan nilai akurasi yang berbeda sehingga dilakukan perhitungan nilai rata-rata untuk mengetahui akurasi terbaik. Kesimpulan dari analisis hasil pengujian pengaruh nilai k-fold adalah persentase terbaik terdapat pada nilai k-fold=6 dengan nilai akurasi 34,597%. Pada pengujian tersebut mempunyai hasil nilai k-fold terbaik tidak berada di satu nilai tertentu karena dari hasil grafik menunjukkan persentase akurasi terdapat nilai yang naik turun pada masing-masing kfold. Hasil rata-rata pengujian pengaruh nilai kfold dapat dilihat pada Gambar 7. Berdasarkan pengujian pengaruh nilai K dan pengaruh nilai k-fold tidak terlihat perbedaan yang signifikan dikarenakan data yang berupa bilangan biner. Selain itu, gejala pada data latih tidak menunjukkan persamaan gejala yang dominan meskipun data latih terdapat pada kelas yang sama. Sebaliknya, data latih tersebut terdapat banyak kemiripan gejala sementara data latih tersebut ada pada kelas yang berbeda.



Gambar. 7. Grafik pengujian pengaruh nilai k-fold

5. Kesimpulan

Terdapat lima kelompok data pengusaha sorgum yang terdiri dari Pengusaha Sorgum yang memiliki modal usaha paling kecil, Pengusaha Sorgum yang memiliki banyak pelanggan tetap, Pengusaha Sorgum yang memiliki jumlah karyawan paling sedikit, Pengusaha Sorgum yang berpromosi di media sosial, Pengusaha Sorgum yang melakukan promosi di media sosial. memiliki pendapatan tertinggi. sedikit. Dengan menggunakan klaster k medoid didapatkan hasil bahwa pada klaster 1 terdapat 6 pengusaha sorgum, 9 pengusaha sorgum pada cluster 2, 4 pengusaha sorgum pada cluster 3, 6 pengusaha sorgum pada cluster 4 dan 10 pengusaha sorgum pada cluster 5 gugus. Selain itu, hasil evaluasi menggunakan metode K-Medoids diperoleh Silhouette Index sebesar 0,5787 dan termasuk dalam kriteria "struktur yang masuk akal telah ditemukan". Dari analisis hasil pengujian pengaruh nilai k-fold adalah persentase terbaik terdapat pada nilai k-fold=6 dengan nilai akurasi 34,597%. Pada pengujian tersebut mempunyai hasil nilai k-fold terbaik tidak berada di satu nilai tertentu karena dari hasil grafik menunjukkan persentase akurasi terdapat nilai yang naik turun pada masing-masing kfold.

Sebagai pekerjaan masa depan, penelitian dapat menerapkan metode pengelompokan lain untuk menghasilkan Indeks Silhouette kurang dari sama dengan 1,0. Selain itu dapat penelitian ini dapat dikembangkan dengan menambahkan dataset entrepreneur sorgum di Indonesia.

Referensi

Adya Hermawati, Sri Jumini, Mardiah Astuti, Fajri Ismail, Robbi Rahim. 2020. *Unsupervised Data Mining with K-Medoids Method in Mapping Areas of Student and Teacher Ratio in Indonesia*. TEM Journal. Volume 9, Issue 4, Pages 1614- 1618, ISSN 2217- 8309. https://doi.org/10.18421/TEM94- 37

Alessio Martino, Antonello Rizzi and Fabio Massimo Frattale Mascioli. 2017. *Efficient Approaches for Solving the Large-Scale k-medoids Problem*. Proceedings of the 9th International Joint Conference on Computational Intelligence (IJCCI 2017), pages 338-347. 2017. https://doi.org/10.5220/0006515003380347.

Bygrave, D.W., Zacharakis. 2010. *A. Entrepreneurship*. New York: Wiley. Page 619.

C. Dewi, B. Y. Gautama, and P. A. Mertasana, 2017. *Analysis of Clustering for Grouping of Productive Industry by K-Medoid Method*. Int. J. Eng. Emerg. Technol., vol. 2, no. 1, p. 26. https://doi.org/10.24843/ijeet.2017.v02.i01.p06.

C Winarti, A B Arif, A Budiyanto and N Richana. 2020. *Sorghum development for staple food and industrial raw materials in East Nusa Tenggara, Indonesia: A review*. IOP Conference Series Earth and Environmental Science. 2020. https://doi.org/10.1088/1755-1315/443/1/012055

Chatti Subbalakshmi, G Rama Krishna, S Krishna Mohan Rao, P Venketeswa Rao. 2014. *A Method to Find Optimum Number of Clusters Based on Fuzzy Silhouette on Dynamic Data Set*. International Conference on Information and Communication Technologies (ICICT). https://doi.org/10.1016/j.procs.2015.02.030

D. F. Pramesti, M. T. Furqon, and C. Dewi, 2017. *Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan / Lahan Berdasarkan Persebaran Titik Panas (Hotspot),* JPTIIK, vol. 1, no. 9, pp. 723–732,

D. Marlina, N. Lina, A. Fernando, and A. Ramadhan. 2018. *Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak.* J. CoreIT. vol. 4, no. 2, p. 64, 2018, https://doi.org/10.24014/coreit.v4i2.4498.

Gintare Giriuniene, Lukas Giriunas, Gintaras Cernius. 2016. *Identification Research of the Concept of Entrepreneurship: The Theoretical Aspect*. International Journal of Economics and Financial Issues. ISSN: 2146-4138.

Han, J, Kamber, M, & Pei, J. 2012. *Data Mining: Concept and Techniques, Third Edition*. Waltham: Morgan Kaufmann Publishers.

Héctor Montiel-Campos. 2021. *Entrepreneurial Alertness, Innovation Modes, And Business Models in Small- And Medium-Sized Enterprises: An Exploratory Quantitative Study*. J. Technol. Manag. Innov. Volume 16, Issue 1. https://doi.org/10.4067/S0718-27242021000100023

Karina Indah Sari Sitanggang, Dian Ahkam Sani, Mochammad Zoqi Sarwani. 2020. *Aplikasi Alarm Memanfaatkan Speech Recognition dengan Menggunakan Metode KNN*. INTEGER: Journal of Information Technology, Vol 5, No 2.

Kipchumba Judith, Edith Gathungu, Oscar Ingasia Ayuya, Paul Kimurto. 2021. *Effects of Production and Market Innovations on the Level of Competitiveness of Sorghum Small Scale Agrienterprises*. Modern Economy.https://doi.org/10.4236/me.2021.127060

Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assef, et. al. 2017. *A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques*. KDD Bigdas, August 2017, Halifax, Canada. https://doi.org/10.48550/arXiv.1707.02919

Md. Saddam Hossain, Md. Nahidul Islam Md, Mamunur Rahman, Mohammad Golam Mostofa, Md. Arifur Rahman Khan. 2022. *Sorghum: A prospective crop for climatic vulnerability, food and nutritional security*. Journal of Agriculture and Food Research. Volume 8. https://doi.org/10.1016/j.jafr.2022.100300

Min Ren, Peiyu Liu, Zhihao Wang, and Jing Yi. 2016. *A Self-Adaptive Fuzzy C-Means Algorithm for Determining the Optimal Number of Clusters*. Computational Intelligence and Neuroscience. https://doi.org/10.1155/2016/2647389

N. L. Anggraeni. 2019. *Teknik Clustering Dengan Algoritma K-Medoids Untuk Menangani Strategi Promosi Di Politeknik TEDC Bandung*, Jurnal Teknologi Informasi dan Pendidikan, vol. 12 no. 2 pp. 1-7.

Nurhayati, Nadika Sigit Sinatrya, Luh Kesuma Wardhani, Busman. 2018. *Analysis of K-Means and K-Medoids's Performance Using Big Data Technology*. The 6th International Conference on Cyber and IT Service Management (CITSM 2018). https://doi.org/10.1109/CITSM.2018.8674251

R. L. Novianto, and Goeirmanto, 2019. *Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi*, JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 6, no. 1, pp. 85–95, doi: 10.35957/jatisi.v6i1.150.

Rika Elizabet Sihombing, Dewi Rachmatin, Jarnawi Afgani Dahlan. 2019. Program Aplikasi Bahasa R Untuk Pengelompokan Objek Menggunakan Metode K-Medoids Clustering. Jurnal EurekaMatika.

Rohani, Abbas., et al., 2017. *A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I).* Renewable Energy, 115, 411-422.

Samudi Samudi, S. Widodo, Herlambang Brawijaya. 2020. *The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic*. Sinkron : Jurnal dan Penelitian Teknik Informatika, Volume 5, Number 1. https://doi.org/10.33395/sinkron.v5i1.10649

Sarita Verma, Neelam Khetrapaul. and Vandana Verma. 2018. *Development and Standardization of Protein Rich Sorghum Based Cereal Bars*. Int. J. Curr. Microbiol. App. Sci. Volume 7(5): 2842-2849. https://doi.org/10.20546/ijcmas.2018.705.330

Septiyawan Rosetya Wardhana, Diana Purwitasari. 2019. *Klasifikasi Multi Class Pada Analisis Sentimen Opini Pengguna Aplikasi Mobile Untuk Evaluasi Faktor Usability*. INTEGER: Journal of Information Technology, Vol 4, No 1.

T. Taslim and F. Fajrizal, 2016. *Penerapan algorithma k-mean untuk clustering data obat pada puskesmas rumbai*, Digit. Zo. J. Teknol. Inf. dan Komun., vol. 7, no. 2, pp. 108–114, doi:10.31849/digitalzone.v7i2.602.

Weksi Budiaji and Friedrich Leisch. 2019. *Simple K-Medoids Partitioning Algorithm for Mixed Variable Data*. MDPI Journal Algorithm. https://doi.org/10.3390/a12090177

Vit Zuraida, Diana Purwitasari, Chastine Fatichah. 2018. *Cross-Domain Topic Learning Berbasis Frase untuk Pemodelan Topik pada Rekomendasi Kolaborasi Penelitian*. INTEGER: Journal of Information Technology, Vol 3, No 2.

Xing Li, Ying Jia. 2015. *Characteristics influence for Entrepreneurship behavior ability.* International Conference on Education, Management, Commerce and Society (EMCS 2015).

Zaenal Mustofa, Iman Saufik Suasana. 2018. *Algoritma Clustering K-Medoids Pada E-Government Bidang Information and Communication Technology Dalam Penentuan Status Edgi*. Jurnal Teknologi Informasi dan Komunikasi, ISSN:2087-0868, Volume 9 Nomor 1.

Zhiguo Liu, Changqing Ren and Wenzhu Cai. 2020. *Overview of clustering analysis algorithms in unknown protocol recognition*. MATEC Web of Conferences 309. MATEC Web of Conferences 309, 03008 (2020) https://doi.org/10.1051/matecconf/202030903008