

Implementasi Metode K-Means Clustering Dalam Pengelompokan Bibit Tanaman Kopi Arabika

Implementation of the K-Means Clustering Method in the Grouping of Arabica Coffee Plant Seeds

Benny Ginting*¹, Fristi Riandari²

^{1,2}STMIK Pelita Nusantara

e-mail: bennyginting23@yahoo.com¹, Fristy.Rianda@gmail.com²

Abstrak

Munculnya bermacam-macam informasi bibit kopi yang baik untuk ditanam mendorong Dinas Pertanian dan Perkebunan melakukan pengelompokan bibit untuk direkomendasikan di sentra-sentra penanaman kopi di wilayah kerja Desa Sarimunthe Kec. Munte Kabupaten Karo. Data mining digunakan untuk mengekstrak informasi berharga dari sebuah dataset kemudian memaparkannya dalam format yang mudah dimengerti oleh manusia dengan tujuan untuk mengambil sebuah keputusan. Dalam penelitian ini dilakukan pengolahan data bibit Kopi Arabika terdiri dari 30 item, bidang Pertanian Kabupaten Karo dalam mempersiapkan bibit yang akan dibagikan kepada masyarakat, penilaian dibagi menjadi 3 fase yakni bibit kopi yang tidak menghasilkan (Fase 0-1 Tahun), belum menghasilkan (Fase 1-2 Tahun) dan menghasilkan (Fase 2 Tahun keatas). Hasil dari perhitungan algoritma K-Means berakhir pada Kombinasi-4 dengan Centroid awal $C1 = [6,8 \ 9,08 \ 9,21]$ $C2 = [23,05 \ 12,85 \ 38,65]$ $C3 = [29,37 \ 35,23 \ 9,34]$. Hasil akhir pengelompokan bibit kopi arabika ada 10 item direkomendasikan layak ditanam.

Kata Kunci: Bibit Kopi, Data Mining, Clustering, K-Means

Abstract

The emergence of various information on good coffee seeds to be planted has prompted the Agriculture and Plantation Service to group the seeds to be recommended in coffee planting centers in the working area of Sarimunthe Village, Kec. Munte Karo District. Data mining is used to extract valuable information from a dataset and then present it in a format that is easily understood by humans with the aim of making a decision. In this study, data processing for Arabica Coffee seedlings consisted of 30 items, in the Karo Regency Agriculture sector, in preparing the seeds to be distributed to the public, the assessment was divided into 3 phases, namely coffee seeds that did not produce (Phase 0-1 Year), immature (Phase 1-2 years) and produce (Phase 2 years and above). The results of the calculation of the K-Means algorithm end in Combination-4 with the initial Centroid $C1 = [6,8 \ 9,08 \ 9,21]$ $C2 = [23,05 \ 12,85 \ 38,65]$ $C3 = [29,37 \ 35,23 \ 9,34]$. The final result of the grouping of Arabica coffee seedlings is that there are 10 recommended items suitable for planting.

Keywords: Coffee Seeds, Clustering, Data Mining, Grouping, K-Means

Pendahuluan

Peraturan Menteri Pertanian Nomor 49/Permentan/OT.140/4/2014, pemilihan bahan tanam unggul merupakan langkah penting dalam praktek budidaya kopi yang baik. Munculnya keberagaman ciri-ciri bibit kopi yang baik mendorong perlunya diterapkan suatu metode pengelompokan bibit tanaman kopi berbasis teknologi sehingga membantu Dinas Pertanian dan Perkebunan dalam mengelompokkan bibit rekomendasi yang akan disalurkan dan ditanam petani di sentra-sentra penanaman kopi. Pada tanaman kopi bahan tanam dapat berupa varietas (diperbanyak secara generatif) dan berupa klon (diperbanyak secara vegetatif). Benih unggul pada tanaman kopi dapat diperoleh dengan cara-cara semaian biji, setek, Somatic Embryogenesis (SE), dan sambungan klon unggul.

Data mining digunakan untuk mengekstrak informasi berharga dari sebuah dataset kemudian memaparkannya dalam format yang mudah dimengerti oleh manusia dengan tujuan untuk mengambil sebuah keputusan [1]. Ada tiga fase pada data mining input, proses dan output. Salah satu teknik pengelompokan dalam data mining adalah metode clustering, yaitu pengelompokan sejumlah data atau objek ke dalam cluster (group) sehingga setiap dalam cluster tersebut akan berisi data yang semirip mungkin [2]. Data Clustering salah satu metode Data Mining yang bersifat tanpa arahan

*) Penulis Korespondensi : bennyginting23@yahoo.com

(unsupervised). Algoritma K-Means merupakan algoritma pengelompokan iterative yang melakukan partisi set data ke dalam jumlah K cluster yang sudah ditetapkan diawal. Algoritma K-Means sederhana untuk diimplementasikan dan dijalankan, relative cepat, mudah beradaptasi, umum dalam penggunaannya dalam praktek [3].

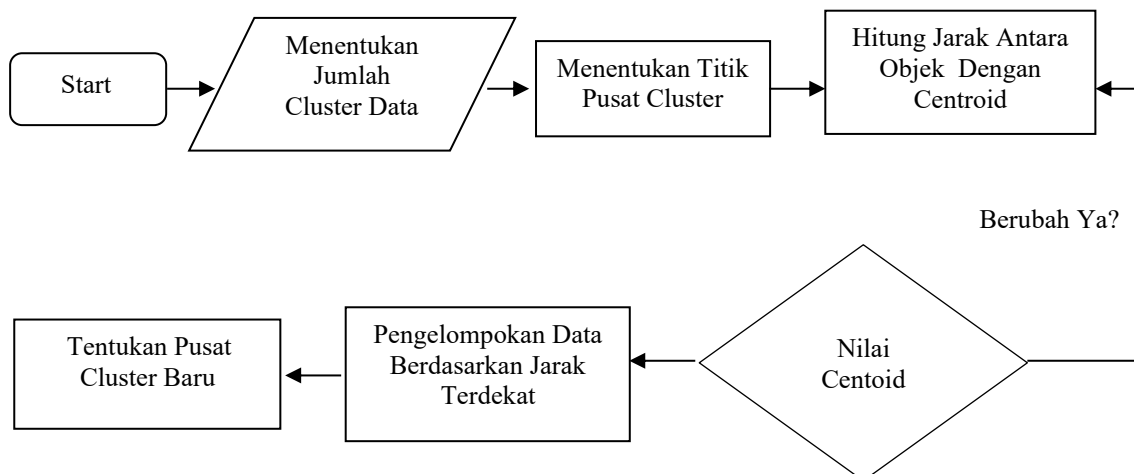
Penelitian Penerapan Algoritma C4.5 untuk mengukur tingkat kepuasan mahasiswa menjelaskan penerapan data mining dalam mengukur kepuasan mahasiswa [4]. Penelitian data mining penjualan suku cadang sepeda motor menggunakan algoritma K-Means menguraikan pengumpulan suku cadang yang tersedia akan dibagi menjadi beberapa kelompok untuk mendapatkan suku cadang yang paling banyak dibeli pelanggan untuk setiap transaksi [5]. K-Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode k-means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Penelitian menerapkan Datamining dengan metode K-Means Cluster pada tanaman kelapa sawit produktif di SIPEF Bukit Maradja Estate dapat mengatasi masalah dalam mengetahui pengelompokan jenis tanaman kelapa sawit produktif sehingga dapat meningkatkan produksi dan kesejahteraan karyawan PT. SIPEF Bukit Maradja Estate [6]

Dalam penelitian ini akan dilakukan pengolahan data bibit Kopi Arabika terdiri dari 30 item, akan dikelompokkan dalam 3 bagian yaitu direkomendasikan, tidak direkomendasikan dan tidak layak. Bidang Pertanian Kabupaten Karo dalam mempersiapkan bibit yang akan dibagikan kepada masyarakat, penilaian dibagi menjadi 3 fase yakni bibit kopi yang tidak menghasilkan (Fase 0-1 Tahun), belum menghasilkan (Fase 1-2 Tahun) dan menghasilkan (Fase 2 Tahun ke atas). K-Means digunakan untuk mempartisi data non hirarki ke dalam bentuk satu atau lebih cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [7].

Penelitian berjudul Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara melakukan tahapan data mining dengan menguji dari setiap komponen yang terdapat dalam populasi data dan menandai komponen ke dalam salah satu pusat cluster yang telah diklarifikasikan sebelumnya tergantung jarak minimum antara komponen dengan tiap-tiap pusat cluster. Penerapan metode k-means clustering untuk mengelompokkan dalam beberapa penelitian terdahulu cenderung untuk prediksi dan hanya pengelompokan jenis item objek yang diteliti, dalam penelitian ini dilakukan pengelompokan bibit kopi yang layak ditanam untuk direkomendasikan sebagai bibit kopi yang baik untuk ditanam [8].

Metode Penelitian

Tahapan penelitian sebagai berikut :



Gambar 1. Tahapan Penelitian

Algoritma K-Means dimulai dengan pemilihan secara acak K, K merupakan banyaknya cluster yang ingin dibentuk. Pengelompokan (clustering) pada data mining mengelompokkan data, yang tidak diketahui label kelasnya, ke dalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya [9]. Pada

tahapan ini dilakukan proses utama yaitu segmentasi atau pengelompokan data bibit kopi yang diakses dari database, yaitu sebuah metode clustering algoritma K-Means [10]. Proses perhitungan implementasi K-Means Clustering pada sistem pemilihan bibit kopi Arabika, dengan asumsi bahwa parameter input adalah jumlah data set sebanyak n data dan jumlah inisialisasi centroid $K = 3$.

Tahapan penerapan algoritma K-Means [11] :

- 1) Tentukan jumlah cluster.
- 2) Menentukan titik pusat.
- 3) Hitung centroid/rata-rata dari data yang ada di masing-masing cluster.
- 4) Alokasikan masing-masing data ke centroid/rata-rata terdekat.
- 5) Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid.
- 6) Pengelompokan data.
- 7) Menentukan pusat cluster baru.

Proses KDD (*Knowledge Discovery in Database*) diantaranya secara berurutan selection, preprocessing, transformation, data mining, dan Interpretation/evaluation. Relokasi sebuah data dalam *cluster* yang diikuti dapat dinyatakan dengan nilai keanggotaan a yang bernilai 0 atau 1. Nilai 0 jika tidak menjadi anggota sebuah *cluster* dan 1 jika menjadi anggota dari *cluster*, maka nilai a sebuah data pada semua *cluster*, hanya satu yang bernilai 1, sedangkan yang lainnya 0 seperti dinyatakan oleh persamaan berikut:

Hasil dan Pembahasan

Rekomendasi clustering bibit kopi Arabika, digunakan metode *K-Means Clustering* untuk menentukan kelompok yang sesuai dengan kriteria bibit kopi yang baik untuk ditanam. Metode ini memiliki ketelitian cukup tinggi terhadap ukuran objek, sehingga lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Selain itu metode *K-Means* tidak terpengaruh terhadap urutan objek. Kualitas bibit kopi dapat ditinjau dari pengamatan dan penilaian pertumbuhan bibit kopi yang dilakukan oleh pihak Dinas Pertanian, Perekebunan dan Peternakan Kabupaten Karo dalam mempersiapkan bibit yang akan dibagikan kepada masyarakat, penilaian dibagi menjadi 3 fase yaitu bibit kopi yang tidak menghasilkan (Fase 0-1 Tahun), belum menghasilkan (Fase 1-2 Tahun) dan menghasilkan (Fase 2 Tahun keatas). Sumber daftar bibit kopi yang akan dikelompokkan, ada 30 jenis bibit, nama bibit kopi diberi kode, dapat dilihat pada Tabel 1.

Pada tahapan ini dilakukan proses utama yaitu segmentasi atau pengelompokan data bibit kopi yang diakses dari database, yaitu sebuah metode clustering algoritma K-Means. Proses perhitungan implementasi K-Means Clustering pada sistem pemilihan bibit kopi Arabika, dengan asumsi bahwa parameter input adalah jumlah data set sebanyak n data dan jumlah inisialisasi centroid $K = 3$.

Ketentuan untuk implementasi K-Means Clustering pada sistem pengelompokan bibit kopi Arabika:

- 1) Menentukan k sebagai jumlah *cluster* yang akan dibentuk.

Ditentukan nilai $k = 3$, terdapat 3 kelompok sebagai target clustering, yaitu :

C1: Rekomendasi (Nilai Awal di Record yang nilai tinggi).

C2: Tidak Rekomendasi (Nilai Awal di Record menengah).

C3: Tidak Layak (Nilai Awal di Record yang nilai rendah).

Data bibit Kopi Arabika terdiri dari 30 item, memiliki 3 atribut, yaitu Menghasilkan (M), Tidak Menghasilkan (TM), Belum Menghasilkan (BM).

Jumlah cluster ada 3, yaitu :

Cluster 1 (C1) : Rekomendasi,

Cluster 2 (C2) : Tidak Rekomendasi,

Cluster 3 (C3): Tidak Layak

- 2) Menentukan titik pusat cluster

Tetapkan C pusat cluster awal secara random dengan menentukan nilai awal titik pusat atau centroid cluster. Inisialisasi centroid dapat ditentukan secara manual ataupun random. Langkah awal sebelum proses clustering, terlebih dahulu memilih nilai data jenis bibit kopi Arabika sebanyak 3 angka secara random yang akan digunakan sebagai nilai centroid. Dari data set diperoleh : $C1 = [72, 90, 67]$, $C2 = [75, 71, 71]$, $C3 = [63, 66, 56]$.

- 3) Hitung jarak setiap data dengan centroid dari masing-masing *cluster* menggunakan rumus *Euclidean Distance*.

- 4) Pengelompokan data berdasarkan jarak terdekat. Mengklasifikasi setiap data berdasarkan kedekatan dengan titik pusat data. Keanggotaan iterasi 1.

5) Tentukan kembali titik pusat cluster yang baru berdasarkan rata-rata. Setelah data dipartisi, maka selanjutnya nilai centroid harus dihitung ulang untuk menentukan jarak minimum yang baru, berikut perhitungan centroid baru. Untuk memperoleh titik centroid baru dilakukan dengan nilai rata-rata data yang ada pada cluster dengan persamaan $ck = \frac{1}{n_k} \sum d_i$. n_k = Jumlah data dalam cluster; d_i = Jumlah nilai jarak yang masuk dalam setiap cluster.

Nilai centroid masih berubah, maka perhitungan jarak diulangi pada setiap data dengan centroid. Update nilai centroid hingga nilai centroid tidak berubah. Pada Kombinasi-1, dihitung rasio antara besaran BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*). WCV yaitu dengan memilih jarak terkecil antara data dengan centroid pada masing-masing cluster. Jarak perhitungan centroid iterasi ke-2. Dengan mengklasifikasi setiap data berdasarkan kedekatan dengan titik pusat data. Iterasi ke-2, dengan centroid diperoleh C1 = [75, 80,3, 63,3]; C2 = [80,3, 83,6, 62,5]; C3 = [63,3, 70,2, 54,6].

Tabel 1. Hasil Akhir Pengelompokan Bibit Kopi Arabika

Kode Bibit	Jarak Terdekat	Cluster	Kelompok	Kode Bibit	Jarak Terdekat	Cluster	Kelompok
V1	68,3	C2	Tidak Rekomendasi	BL2	63,47	C2	Tidak Rekomendasi
VB1	34,4	C1	Rekomendasi	PB2	65,35	C3	Tidak Layak
VLM1	67,76	C1	Rekomendasi	PP2	80,1	C1	Rekomendasi
VK1	26,44	C3	Tidak Layak	BI2	51,64	C1	Rekomendasi
K1	68,69	C2	Tidak Rekomendasi	SB2	48,37	C1	Rekomendasi
BL1	21,74	C2	Tidak Rekomendasi	V3	59,91	C1	Rekomendasi
PB1	36,95	C2	Tidak Rekomendasi	VB3	61,58	C3	Tidak Layak
PP1	66,6	C3	Tidak Layak	VLM3	58,25	C1	Rekomendasi
BI1	101,78	C3	Tidak Layak	VK3	9,34	C3	Tidak Layak
SB1	24,2	C2	Tidak Rekomendasi	K3	57,09	C3	Tidak Layak
V2	39,53	C3	Tidak Layak	BL3	55,45	C3	Tidak Layak
VB2	39,53	C3	Tidak Layak	PB3	44,19	C3	Tidak Layak
VLM2	144,43	C1	Rekomendasi	PP3	41,06	C3	Tidak Layak
VK2	87,1	C2	Tidak Rekomendasi	BI3	52,19	C3	Tidak Layak
K2	88,59	C1	Rekomendasi	SB3	61,58	C1	Rekomendasi

Kombinasi-4 : Centroid awal C1 = [6,8 9,08 9,21] C2 = [23,05 12,85 38,65] C3 = [29,37 35,23 9,34]

hasil kombinasi-4. BCV = 105,85; WCV = 119566,8. Besar Rasio = BCV/WCV = 0,000885

Bila dibandingkan maka rasio sekarang (0,000885) sudah tidak lagi lebih besar dari rasio sebelumnya (0,00391) oleh karena itu algoritma akan dihentikan. Pada perhitungan centroid iterasi ke-4 dan kombinasi ke-4, Nilai centroid akhir adalah :

Centroid-1 = C1 = [6,8 9,08 9,21] ;

Centroid-2 = C2 = [23,05 12,85 38,65];

Centroid-3 = C3 = [29,37 35,23 9,34].

Pengelompokan data set :

Cluster 1 (C1) : Rekomendasi

Cluster 2 (C2) : Tidak Rekomendasi

Cluster 3 (C3): Tidak Layak.

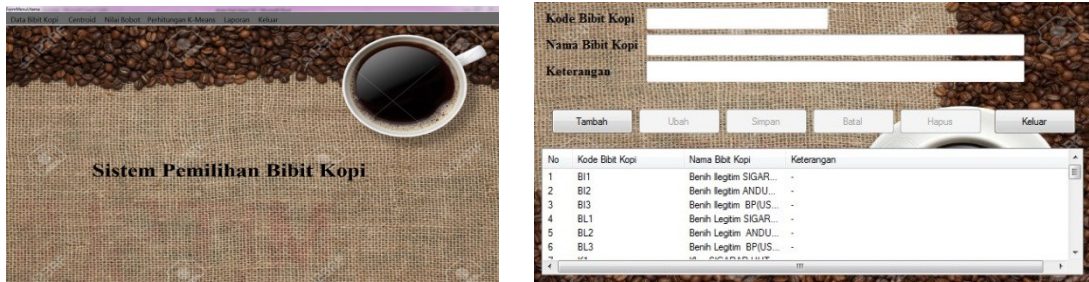
Hasil dari perhitungan algoritma K-Means yang telah di kelompokkan ke dalam cluster, dapat disimpulkan bahwa bibit kopi yang direkomendasikan (C1) terdiri dari 10 item, bibit kopi yang tidak direkomendasikan (C2) terdiri dari 7 jenis bibit kopi dan bibit kopi yang tidak layak (C3) terdiri dari 13 jenis bibit kopi.

Implementasi sistem pengelompokan bibit tanaman kopi arabika menghasilkan tampilan:

a. Menu utama, tampil setelah penginputan *username* dan *password*, sukses.

Pada Menu utama tersedia Menu Data Bibit Kopi, Menu Centroid, Menu Nilai Bobot, Menu Perhitungan K-Mean dan Menu Laporan. Menu Data Bibit Kopi, tersedia untuk menampilkan data

bibit kopi. Admin, dapat melakukan penambahan data bibit kopi (klik tombol Tambah), mengubah (Tombol Ubah), Menyimpan (Tombol Simpan), untuk membatalkan klik Tombol Batal, untuk keluar klik Tombol Keluar. Pada Gambar 2, tampilan menu utama untuk memulai menjalankan program, terdapat menu-menu utama aplikasi. Gambarr 2 menu halaman data bibit kopi, operator dapat melihat daftar bibit kopi yang sudah terinput sebelumnya.



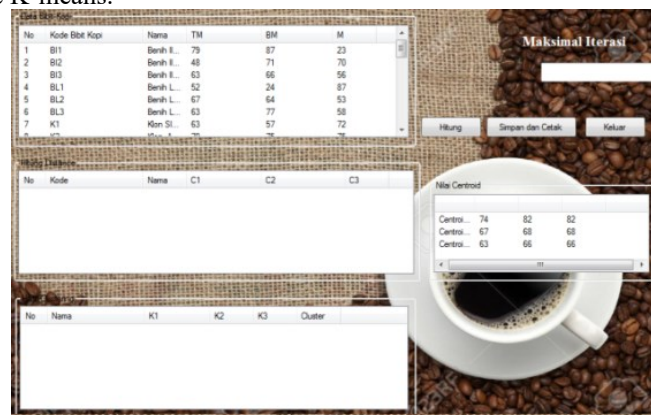
Gambar 2. Tampilan Menu Utama dan Menu Data Kopi

- b. Halaman Menu Centroid tersedia untuk menginput nilai Centroid yang telah ditentukan sebelumnya. Menu Data Bobot berfungsi untuk melakukan penginputan nilai atribut kriteria dari setiap variabel. Admin menginput nilai atribut berdasarkan data-data yang diperlukan pada sistem selanjutnya sistem akan menyimpan nilai atribut ke dalam database bibit kopi. Pada Gamabr 3 dapat digunakan untuk menentukan centroid dan menginput bobot data.



Gambar 3. Tampilan Menu Centroid dan Menu Bobot Data

- c. Tampilan Menu Perhitungan K-Means
 Pada halaman Menu Perhitungan K-Means, admin menginput jumlah cluster dan maksimal iterasi selanjutnya klik Tombol Hitung. Hasil perhitungan dapat disimpan dan dicetak. Setelah selesai penginputan bobot perhitungan, pada Gambar 4 dapat dijalankan perhitungan dan melihat hasil akhir penerapan metode K-means.



Gambar 4. Tampilan Menu Perhitungan K-means

Tahap awal, pada Menu Bibit Data Kopi, Admin dapat memilih bibit kopi yang tersedia pada database. Sebelum proses clustering, terlebih dahulu memilih nilai data jenis bibit kopi Arabika sebanyak 3 angka secara random yang akan digunakan sebagai nilai centroid. Alokasikan data ke dalam cluster secara random. Mengklasifikasi setiap data berdasarkan kedekatan dengan titik pusat data. Inisialisasi centroid dataset yaitu menghitung centroid/rata-rata dari data yang ada pada masing-masing cluster. pengulangan berikutnya (pengulangan ke-1 sampai selesai), centroid baru dihitung dengan menghitung nilai rata-rata data pada setiap cluster. Jika centroid baru berbeda dengan centroid sebelumnya, maka proses dilanjutkan ke langkah berikutnya. Jika centroid yang baru dihitung sama dengan centroid sebelumnya, maka proses clustering selesai.

Kesimpulan

Berdasarkan hasil analisis Informasi rekomendasi bibit kopi yang layak untuk ditanam dengan menghitung jarak terdekat dari satu data ke titik centroid, dari ujicoba 30 item diperoleh Kombinasi-4 : Centroid awal $C1 = [6,8 \ 9,08 \ 9,21]$ $C2 = [23,05 \ 12,85 \ 38,65]$ $C3 = [29,37 \ 35,23 \ 9,34]$ dan Bibit kopi yang direkomendasikan (C1) terdiri dari 10 item, bibit kopi yang tidak direkomendasikan (C2) terdiri dari 7 jenis bibit kopi dan bibit kopi yang tidak layak (C3) terdiri dari 13 jenis bibit kopi.

Daftar Pustaka

- [1] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.
- [2] W. Duhita, "Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita," *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 160–174, 2015.
- [3] Sanjani, H. Fahmi, and A. Sindar, "Implementasi Data Mining Penjualan Produk Pakaian Dengan Algoritma Apriori," vol 4, no 1, pp. 23-29, 2019.
- [4] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019.
- [5] S. M. Hutabarat and A. Sindar, "Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, p. 126, 2019.
- [6] F. Nur, M. Zarlis, and B. B. Nasution, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 2, pp. 100–105, 2017.
- [7] N. Sirait, "IMPLEMENTASI K-MEANS CLUSTERING PADA PENGELOMPOKAN MUTU BIJI SAWIT (Studi Kasus : PT . Multimas Nabati Asahan)," vol. 6, pp. 170–174, 2017.
- [8] R. A D. Saragih, S. Tambunan, "Penerapan Metode K-Means Clustering Untuk Mengelompokkan Kelapa Sawit Produktif", Seminar Nasional Matematika dan Terapan, Volume 1, Desember, pp: 362-370, 2019.
- [9] J. S. Parapat and A. S. Sinaga, "Data Mining Algoritma C4 . 5 Pada Klasifikasi Kredit Koperasi Simpan Pinjam," *J. Ilmu Tek. Elektro Komput. dan Inform.*, vol. 4, no. 2, pp. 144–154, 2018.
- [10] S. Syahidatul Helma *et al.*, "Clustering pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K-Means," *Puzzle Res. Data Technol. Fak. Sains dan Teknol.*, vol. 1, no. November, p. 4, 2019.
- [11] E. S. Susanto and H. Al Fatta, "Informatika Universitas Amikom Yogyakarta Menggunakan Metode K-Nearest Neighbor," *J. Teknol. Inf.*, vol. XIII, pp. 67–72, 2018.