



RANCANGAN SISTEM PENGELOMPOKKAN BUAH BERDASARKAN KADAR VITAMIN MENGGUNAKAN METODE K-MEDOIDS

Agung Purnomo Sidik^{1*}, Hermansyah², Dina Aulia³
Program Studi Sistem Komputer, Universitas Pembangunan Panca Budi

*Email: agung@dosen.pancabudi.ac.id

ABSTRACT

Lau Gumba Village is a village located in Brastagi District, Karo Regency, North Sumatra Province which is quite well known because it has quite interesting tourist objects to visit. A village that presents the location of tourist attractions in the village displays a natural charm that is quite amazing for local and foreign tourists. Lau Gumba Village has many plantation locations for various types of fruit grown by village farmers, this is what makes the tourist attraction quite good and healthy for tourists. The village of Lau Gumba has fruit farming with a variety of good fruits, with the importance of paying attention to the quality and type of vitamins, because when people sell or consume fruit based on vitamin levels, it is very important to pay attention to the shape, freshness, color of the fruit consumed. However, this requires technology that can provide information on fruit grouping based on vitamin levels. As is the case with information that provides an explanation in the information in grouping fruit based on vitamin levels, it must include the correct data. The method used in this research is using UML (Unified Modeling Language), namely using the k-medoids method, web programming, mysql as data storage. With the application of information technology, it will have an impact on fruit grouping based on vitamin levels, so that farmers can easily classify harvested fruit based on the levels of vitamins produced.

Keywords: Data Mining, Fruit Grouping, k-medoids)

PENDAHULUAN

Desa Lau Gumba merupakan Desa yang terletak di Kecamatan Brastagi Kabupaten Karo Provinsi Sumatera Utara yang cukup di kenal karena memiliki objek wisata yang cukup menarik untuk dikunjungi. Sebuah desa yang menyajikan lokasi objek wisata di Desa tersebut menampilkan pesona alam yang cukup menakjubkan untuk para wisatawan local maupun mancanegara. Desa Lau Gumba memiliki banyak tempat lokasi perkebunan berbagai jenis buah-buahan yang ditanam petani desa tersebut, hal ini yang menjadikan daya tarik objek wisata yang cukup baik dan menyehatkan untuk wisatawan. Desa Lau Gumba memiliki banyak tempat lokasi perkebunan berbagai jenis buah-buahan yang ditanam petani desa tersebut, hal ini yang menjadikan daya tarik objek wisata yang cukup baik dan menyehatkan untuk wisatawan. Desa Lau Gumba memiliki hasil pertanian buah buahan dengan berbagai macam buah yang baik, dengan penting nya memperhatikan kualitas dan jenis vitamin, karena Ketika masyarakat menjual atau pun konsumsi buah berdasarkan kadar vitamin merupakan hal yang sangat penting memperhatikan bentuk kesegaran warna buah yang dikonsumsi. Namun Hal ini dibutuhkan teknologi yang dapat menyajikan informasi pengelompokkan buah berdasarkan kadar vitamin. Seperti hal dengan informasi yang memberi penjelasan dalam informasi dalam pengelompokkan buah berdasarkan kadar vitamin harus meliputi data data yang benar. Metode yang di pakai dalam penelitian ini dengan menggunakan UML (Unifed Modeling Language) yaitu menggunakan metode k-medoids, pemograman web, mysql sebagai penyimpanan data. Dengan penerapan informasi teknologi akan berdampak pada pengelompokkan buah berdasarkan kadar vitamin, sehingga petani dapat dengan mudah mengelompokkan buah buah hasil panen berdasarkan kadar vitamin yang dihasilkan. Citra atau gambar dapat didefinisikan sebagai fungsi dua dimensi yang disimbolkan dengan $f(x, y)$ di mana x dan y adalah koordinat spasial (bidang), dan amplitudo f pada setiap pasangan koordinat (x, y) disebut intensitas atau tingkat keabu-abuan gambar. Nilai intensitas x, y , dan jumlah diskrit disebut sebagai gambar digital (digital



image). Bidang pemrosesan gambar (image processing) mengacu pada pemrosesan dengan menggunakan komputer digital, yang selanjutnya citra dapat diklasifikasi untuk mendapatkan suatu informasi atau pengetahuan baru.

Pada penelitian ini data yang digunakan ada beberapa record untuk clustering dan beberapa record yang lain untuk prediksi. Data belum memiliki label, maka metode yang dapat digunakan untuk melakukan pengelompokan adalah K-Medoids. Metode tersebut merupakan metode partisional clustering dimana bertujuan untuk menemukan satu set k-cluster di antara data yang paling mencirikan objek dalam kumpulan suatu data. Kelebihan dari metode ini mampu mengatasi kelemahan dari metode k-means yang sensitive terhadap outlier (Sundari et.al., 2019).

Penelitian ini pernah dilakukan oleh penelitian sebelumnya yang telah dilakukan oleh Sindi et.al. (2020) yang berjudul “Analisis Algoritma K-medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 di Indonesia”. Perbedaan penelitian ini dengan sebelumnya yaitu data dan variasi pengujian clustering serta dilakukan prediksi menggunakan hasil pemodelan medoids dari hasil clustering sebelumnya. Hasil prediksi digunakan untuk memprediksi tingkat penyebaran pandemi di tanggal atau hari berikutnya.

Pendahuluan ditulis guna membawa pembaca untuk memahami permasalahan yang akan dibahas pada artikel ilmiah secara urut, jelas, dan terperinci. Bagian pendahuluan atau latar belakang menguraikan tentang pentingnya penelitian/ kajian atau alasan pemilihan judul penelitian. Penulis dapat mengaitkannya dengan isu-isu terkini serta melengkapi dengan data pendukung. Pendahuluan harus berisi (secara berurutan) latar belakang umum, kajian literatur terdahulu (state of the art) sebagai dasar pernyataan kebaruan ilmiah dari artikel, pernyataan kebaruan ilmiah, dan permasalahan penelitian atau hipotesis. Di bagian akhir pendahuluan harus dituliskan tujuan kajian artikel tersebut. Di dalam format artikel ilmiah tidak diperkenankan adanya tinjauan pustaka sebagaimana di laporan penelitian, tetapi diwujudkan dalam bentuk kajian literatur terdahulu (*state of the art*) untuk menunjukkan kebaruan ilmiah artikel tersebut.

Dalam penelitian ini, penulis menggunakan algoritma K-Medoids yang telah digunakan oleh Marlina (2018) yang berjudul “Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak”. Penelitian lain pernah dilakukan oleh Sangga (2018) yang berjudul “Perbandingan Algoritma K-Means dan Algoritma K-Medoids Dalam Pengelompokan Komoditas Peternakan di Provinsi Jawa Tengah Tahun 2015”, Preeti et.al. (2016) yang berjudul “Analysis of K-Means and K-Medoids Algorithm For Big Data” dan Nurhayati et.al.(2018) yang berjudul “Analysis of K-Means and K-Medoids’s Performance Using Big Data Technology”. Keempat penelitian tersebut menyimpulkan bahwa algoritma K-Medoids lebih baik dibandingkan dengan algoritma K-Means. Selain itu, dalam penelitian sebelumnya yang dilakukan oleh Asmiatun et.al.(2020) yang berjudul “Penerapan Metode K-Medoids Untuk Pengelompokan Kondisi Jalan di Kota Semarang” menghasilkan rata-rata nilai Silhouette Coefficient sebesar 0,57432 untuk 4 cluster dan mendekati 1 untuk 2 cluster. Dengan adanya masalah tersebut serta ada solusi untuk mengatasi keadaan seperti itu, maka penulis tertarik untuk membuat si penulis melakukan sebuah klasifikasi data yang berjudul : “Aplikasi Data Mining Pengelompokan Tanama Buah Berdasarkan Kadar Vitamin Metode Clustering Algoritma KMedoids Di Desa Lau Gumba .”

TINJAUAN PUSTAKA

Machine learning memungkinkan manusia untuk memprogram komputer sehingga mesin dapat mengenali pola atau belajar dari apa yang dimasukkan ke dalamnya. Konsep pembelajaran adalah proses mengubah informasi menjadi pengetahuan. Algoritma machine



learning adalah data pelatihan, yang mewakili pengalaman, dan hasilnya adalah beberapa keahlian, yang biasanya berupa program komputer lain yang dapat melakukan beberapa tugas (Ion C. Freeman, Ashley J. et al, 2018).

Machine learning memungkinkan dalam klasifikasi data, aplikasi ini mengenali pola dalam data baik dengan pelatihan atau tanpa pelatihan. Di klasifikasi data disebut clustering dalam machine learning. Beberapa contoh algoritma pengelompokan termasuk K-Means, Farthest-First Maksimalisasi-Ekspektasi (EM), dan lainnya (Ion C. Freeman, Ashley J. et al, 2018).

Machine Learning merupakan cabang ilmu artificial intelligence yang memiliki konsep bahwa komputer sebagai mesin memiliki kemampuan untuk melakukan adaptasi terhadap lingkungan yang baru dan mampu mendeteksi pola dari fakta yang ada. Definisi pembelajaran yang dilakukan mesin adalah ketika mesin dari pengalaman E terhadap tugas T dan mengukur peningkatan kinerja P, jika kinerja tugas T diukur oleh kinerja P, meningkatkan pengalaman E.

Klasifikasi data adalah proses pengurutan dan pengelompokan data ke dalam berbagai jenis, bentuk atau kelas berbeda lainnya. Klasifikasi data memungkinkan pemisahan dan klasifikasi data sesuai dengan persyaratan kumpulan data untuk berbagai tujuan bisnis atau pribadi. Ini terutama merupakan proses manajemen data.

Analisis kelompok sebagai suatu metode untuk melakukan klasifikasi data menjadi beberapa kelompok dengan menggunakan metode pengukuran ukuran asosiasi, sehingga data yang sama berada dalam satu kelompok dan data yang memiliki perbedaan yang besar diletakkan dalam kelompok data lainnya. Masukan untuk sistem analisis kelompok adalah sebuah data set dan kesamaan ukuran antara kedua data tersebut. Sedangkan hasil dari analisis kelompok adalah sejumlah kelompok yang membentuk sebuah partisi atau struktur partisi dari kumpulan data dan deskripsi umum dari setiap kelompok, dimana hal ini sangat penting untuk analisis yang lebih dalam pada karakteristik yang terdapat pada data tersebut.

Pengelompokan data harus menggunakan sebuah pendekatan untuk mencari kesamaan dalam data sehingga mampu menempatkan data ke dalam kelompok-kelompok yang tepat. Pengelompokan data akan membagi kumpulan data ke dalam beberapa kelompok dimana kesamaan dalam sebuah kelompok lebih besar jika dibandingkan dengan kelompok-kelompok lainnya (Gorunescu, 2011).

Terdapat dua metode pembelajaran yang tersedia pada model klasifikasi yaitu :

- a. Eager learning adalah proses pembelajaran pada data latih secara intensif sehingga model dapat melakukan prediksi ke label kelas output dengan benar. Beberapa metode bersifat eager learning antara lain, Neural Network, Bayesian, decision tree, Support Vector Machine.
- b. Lazy learning adalah proses pembelajaran tanpa melakukan pelatihan dan hanya menyimpan nilai data latih untuk digunakan pada proses prediksi. Beberapa metode bersifat lazy learning antara lain: K-Nearest Neighbor, Regresi Linear, Fuzzy K-Nearest Neighbor.

Clustering atau klasterisasi adalah metode pengelompokan data. Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan cluster yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma clustering. Oleh karena itu, clustering sangat berguna dan bisa menemukan group atau



kelompok yang tidak dikenal dalam data. Clustering banyak digunakan dalam berbagai aplikasi seperti misalnya pada business intelligence, pengenalan pola citra, web search, bidang ilmu biologi, dan untuk keamanan (security). Di dalam business intelligence, clustering bisa mengatur banyak customer ke dalam banyaknya kelompok. Contohnya mengelompokkan customer ke dalam beberapa cluster dengan kesamaan karakteristik yang kuat.

Clustering juga dikenal sebagai data segmentasi karena clustering mempartisi banyak data set ke dalam banyak group berdasarkan kesamaannya. Selain itu clustering juga bisa sebagai outlier detection. Analisis berbasis cluster merupakan analisis pengelompokan kelompok atau cluster data objek berdasarkan informasi yang hanya ditemukan dalam data yang menggambarkan objek dan hubungan mereka. Tujuannya adalah agar objek-objek dalam suatu kelompok serupa (atau berhubungan) dengan satu sama lain dan berbeda dari k (atau tidak berhubungan dengan) objek-objek dalam kelompok-kelompok lain. Semakin besar kesamaan (homogenitas) antar objek dalam suatu kelompok dan semakin besar perbedaan antara cluster, maka clustering akan semakin baik (Tan et.al., 2005).

Metode clustering juga harus dapat mengukur kemampuannya sendiri dalam usaha untuk menemukan suatu pola tersembunyi pada data yang sedang diteliti. Terdapat berbagai metode yang dapat digunakan untuk mengukur nilai kesamaan antar objek-objek yang dibandingkan. Salah satunya ialah dengan weighted Euclidean Distance. Euclidean distance menghitung jarak dua buah point dengan mengetahui nilai dari masing-masing atribut pada kedua poin tersebut. Berikut formula yang digunakan untuk menghitung jarak dengan Euclidean distance:

$$Distance(p, q) = \left(\sum_k^n \mu_k |p_k - q_k|^r \right)^{1/r}$$

Keterangan:

N = Jumlah record data

K = Urutan field data

r = 2

μ_k = Bobot field yang diberikan user

Jarak adalah pendekatan yang umum dipakai untuk menentukan kesamaan atau ketidaksamaan dua vektor fitur yang dinyatakan dengan ranking. Apabila nilai ranking yang dihasilkan semakin kecil nilainya maka semakin dekat/tinggi kesamaan antara kedua vektor tersebut. Teknik pengukuran jarak dengan metode Euclidean menjadi salah satu metode yang paling umum digunakan. Pengukuran jarak dengan metode euclidean dapat dituliskan dengan persamaan berikut:

$$j(v_1, v_2) = \sqrt{\sum_{k=1}^N (v_1(k) - v_2(k))^2}$$

dimana v_1 dan v_2 adalah dua vektor yang jaraknya akan dihitung dan N menyatakan panjang vektor.

K-Medoid merupakan kelompok metode partitional Clustering yang meminimalkan jarak antara titik berlabel berada dalam cluster dan titik yang ditunjuk sebagai pusat klaster itu. Berbeda dengan algoritma K-Means, K-Medoid memilih data points sebagai



pusat (medoids). Perbedaan dari kedua algoritma ini yaitu algoritma K-Medoids atau PAM menggunakan objek sebagai perwakilan (medoid) sebagai pusat cluster untuk setiap cluster, sedangkan K-Means menggunakan nilai rata-rata (mean) sebagai pusat cluster (Kaur, dkk., 2014). Algoritma K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada algoritma K-Means yang sensitif terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Kelebihan lainnya yaitu hasil proses Clustering tidak bergantung pada urutan masuk dataset (Furqon, dkk., 2015).

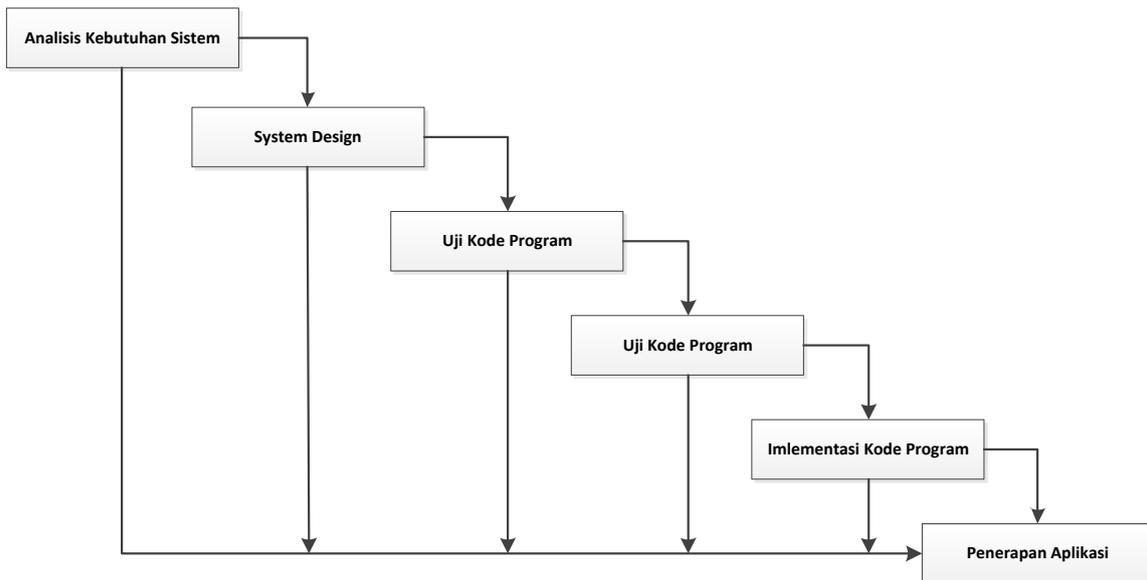
K-Medoid adalah teknik partisi klasik Clustering yang mengelompokkan data set dari n objek ke dalam kelompok k yang dikenal apriori. Dibandingkan dengan K-Means, K-Medoid lebih kuat untuk mengatasi kebisingan (noise) dan pencilan (outlier) karena meminimalkan sejumlah dissimilarities berpasangan, bukan jumlah kuadrat jarak Euclidean. Sebuah medoid dapat didefinisikan sebagai objek cluster yang rata-rata perbedaan untuk semua objek dalam cluster minimal yaitu titik paling berlokasi di cluster.

K-Medoids clustering, juga dikenal sebagai Partitioning Around Medoids (PAM), adalah varian dari metode K-Means. Hal ini didasarkan pada penggunaan medoids bukan dari pengamatan mean yang dimiliki oleh setiap cluster, tujuannya untuk mengurangi sensitivitas dari partisi sehubungan dengan nilai ekstrim yang ada di dalam dataset (Vercellis, 2009).

Kemudian menurut Kaufman dan Rousseeuw (1990), algoritma yang digunakan dalam K-Medoids clustering (PAM) didasarkan pada pencarian objek yang mewakili k di antara objek dari data yang ditetapkan. Objek yang mewakili sebuah cluster atau k disebut dengan medoids. Setelah menemukan satu set objek yang mewakili k (medoids), kemudian k cluster dibangun dengan menugasi setiap objek dari data (non medoids) yang ditetapkan ke medoids terdekat. K-Medoids clustering hadir untuk mengatasi kelemahan K-Means clustering yang sensitif terhadap outlier karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data. Tetapi algoritma K-Medoids seperti halnya dengan K-Means dalam memilih pusat cluster pada awalnya dipilih secara acak. Kemudian proses berikutnya pada algoritma K-Medoids, proses iterasi dalam mengganti objek-objek yang menjadi pusat cluster dengan objek-objek lainnya berlanjut terus hingga kualitas clustering yang dihasilkan tidak bisa lagi ditingkatkan dengan pergantian atau dengan kata lain tidak stabil. Kualitas ini diukur dengan fungsi rerata ketidakmiripan (jarak simpangan) di antara suatu objek dengan objek yang menjadi pusat clusternya (Han et.al., 2011).

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif, karena analisis data dipaparkan secara verbal, untuk mendapatkan informasi secara menyeluruh. Komponen dalam metode penelitian kualitatif meliputi antara lain; alasan menggunakan metode kualitatif, tempat atau lokasi penelitian, instrumen penelitian, informan dan sumber data penelitian, teknik pengumpulan data, teknik analisis data rencana pengujian keabsahan data (Sugiyono, 2010, hlm. 145). Metode penelitian kualitatif yang peneliti gunakan dalam penelitian ini dengan alasan karena permasalahan yang diteliti kompleks, dinamis dan penuh makna sehingga tidak mungkin data pada situasi sosial yang kompleks, dinamis dan belum jelas problemnya tersebut dijaring dengan metode penelitian kuantitatif.



Gambar 1. Metodologi Penelitian Dengan Diagram Waterfall

Berdasarkan gambar 1. penulis dapat menjelaskan beberapa kerangka kerja yang akan dilakukan dalam penelitian ini, yaitu sebagai berikut:

1. Mempelajari Literatur

Pada penelitian ini dipelajari literatur yang berhubungan dengan permasalahan. Kemudian literatur yang dipelajari diseleksi untuk dapat ditentukan literatur mana yang akan digunakan dalam penelitian. Sumber literatur didapatkan dari perpustakaan, jurnal, artikel dan konsep-konsep lain yang mendukung dalam menyelesaikan sistem yang akan dibangun termasuk referensi.

2. Pengumpulan Data

Dalam melakukan penelitian ini, pengumpulan data dan informasi pada tahap ini dilakukan untuk mengetahui mengenai sistem yang diteliti. Dari data dan informasi yang dikumpulkan akan didapat data untuk pendukung penelitian serta pengumpulan data dilakukan untuk mengetahui kebutuhan dari pengguna. Metode yang digunakan penulis untuk pengumpulan data adalah sebagai berikut :

a. Observasi

Observasi berguna untuk melakukan pengumpulan data dan observasi dengan langsung terjun kelapangan pada pihak-pihak yang terkait dalam menyelesaikan penelitian ini dimana informasi dan materi akan diperoleh sebagai bahan dari rancang bangun sistem.

b. Wawancara

Melakukan wawancara pada pihak yang berkaitan dengan alur permasalahan. Wawancara ini dilakukan untuk mendapatkan bahan penulisan dan penjelasan pengamatan yang dilakukan.

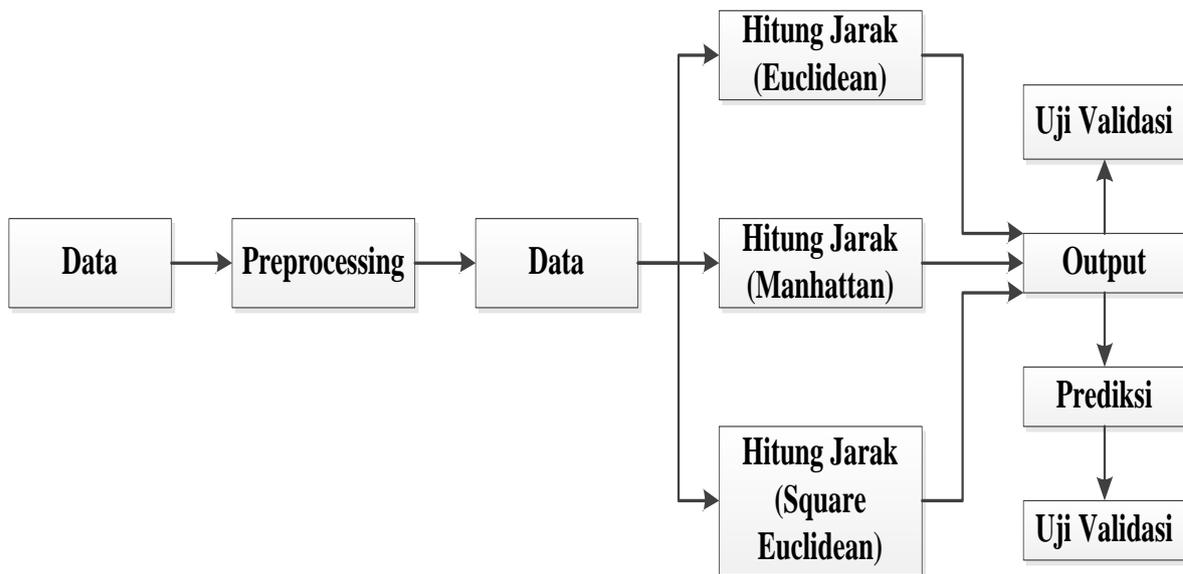
3. Analisa Kebutuhan

Analisis dapat didefinisikan sebagai penguraian dari suatu sistem informasi yang utuh kedalam bagian-bagian komponennya. Analisa kebutuhan ini bertujuan untuk mengetahui apa saja yang dibutuhkan dalam perancangan sistem yang di bangun dan mengetahui kebutuhan-kebutuhan pendukung dari perancangan sistem.

4. Desain dan Perancangan Sistem

Kegiatan desain sistem dilakukan untuk sebagai awal dari perancangan sistem yang akan dibangun sesuai kebutuhan. Dan pada tahap ini dilakukan perancangan antar

- muka terhadap sistem yang akan dibuat.
5. Implementasi Sistem
Implementasi sistem dilakukan sesuai desain dan rancangan antar muka aplikasi yang akan dibangun. Pada tahap ini melakukan pengkodean atau pembuatan program sehingga sistem yang dirancang dapat digunakan oleh pengguna.
 6. Pengujian Sistem
Pengujian sistem dilakukan untuk mengetahui uji kelayakan sistem yang telah dibangun sesuai yang diharapkan dan dengan dilakukannya pengujian dapat mengetahui kelemahan serta kelebihan dari sistem yang dirancang sehingga dapat dilakukan perbaikan pada tahap selanjutnya.
 7. Maintenance
Pemeliharaan Sistem.



Gambar 2. Gambaran Umum Penelitian

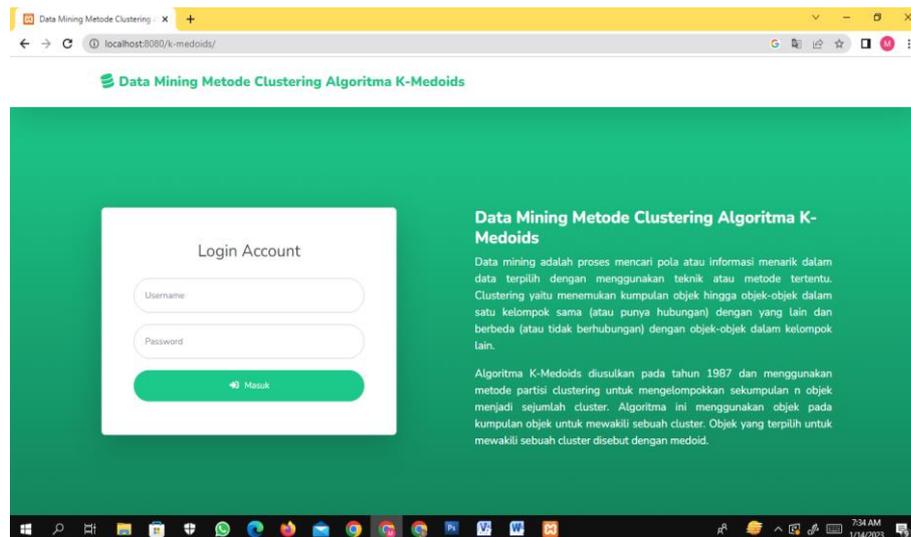
Pada Gambar 2 proses clustering dimulai dari input data mentah yang berupa data buah buahan dengan kadar vitamin berbeda yang diambil dari beberapa desa yang ada di Kabupaten Karo. Tahap kedua yaitu preprocessing yang terdiri dari beberapa tahap, antara lain: tahap data cleaning untuk membuang data yang tidak konsisten dan mengisi nilai atribut yang kosong (missing value), untuk mengatasi missing value dilakukan dengan mengisi nilai rata-rata dari setiap atribut. Tahap selanjutnya adalah tahap data selection untuk memilih atribut yang relevan dengan penelitian. Setelah itu, dilakukan tahap transformasi data, data yang berjenis alfabet seperti nama provinsi harus dilakukan inialisasi data terlebih dahulu ke dalam bentuk angka/numerik. Kemudian juga dilakukan normalisasi min-max, bertujuan supaya atribut dari data bisa jatuh dalam range tertentu. Proses preprocessing terakhir adalah perangkian atribut menggunakan PCA untuk memilih atribut yang menghasilkan hasil optimal. Tahap ketiga yaitu K-Medoids clustering, disini tahap K-Medoids clustering adalah mengelompokan data dengan menggunakan tiga distance measure yaitu Euclidean Distance, Square Euclidean Distance dan Manhattan Distance. Kemudian tahap terakhir, setelah mendapatkan hasil dari K-Medoids clustering dengan ketiga distance measure tersebut, maka dilakukan uji validasi dengan Silhoutte Index di setiap distance measure-nya, selain itu hasil model medoids dari clustering akan dilakukan prediksi dengan data baru kemudian akan dilakukan uji validasinya menggunakan Silhoutte Index. Beberapa data record akan digunakan untuk clustering dan beberapa lainnya record

untuk prediksi clustering.

Di dalam pengujian clustering dan prediksi menggunakan variasi pengujian (jumlah cluster, distance measure, normalisasi, jumlah atribut) yang sama di setiap pengujiannya. Pada penelitian ini penulis menggunakan PCA untuk mengetahui jumlah atribut yang menghasilkan uji validasi optimal, serta tanpa PCA untuk membandingkan hasilnya dengan PCA. Langkah pertama untuk melakukan pengujian yaitu dimulai dari jumlah cluster (3 dan 4), langkah kedua yaitu distance measure (euclidean, manhattan, square euclidean), ketiga yaitu normalisasi (min-max dan z-score), langkah terakhir yaitu jumlah atribut untuk PCA (3-9 atribut), sedangkan tanpa PCA (10 atribut).

HASIL PENELITIAN DAN DISKUSI

Hasil Penelitian



Gambar 3. Menu Utama

Pada tampilan menu utama pada gambar di atas merupakan menu utama untuk login ke dalam sistem pengklusteran atau pengelompokan buah berdasarkan kadar vitamin, sample data buah buahan tersebut diambil dari beberapa desa yang ada di Kabupaten Karo. Data buah buahan tersebut akan dijadikan data untuk pengklusteran dan pengelompokannya, kemudian pada tampilan awal program ini juga berisi penjelasan tentang metode K-Medoids yang digunakan dalam penelitian ini. Tampilan utama yang ada pada aplikasi ini cukup user friendly karena mudah memahami penggunaan aplikasi pada tampilannya. Sebelum masuk ke menu dashboard pengguna aplikasi ini harus terlebih dahulu login dengan user dan password yang telah ditentukan pada aplikasi ini.

No	Nama Buah	Aksi
1	Jeruk	
2	Nanas	
3	Pepaya	
4	Pisang	
5	Semangka	
6	Apel	
7	Pir	

Gambar 4. Data Buah Buah

Buah adalah hasil reproduksi antara putik dan serbuk sari pada tumbuhan. Buah termasuk organ pada tumbuhan berbunga yang merupakan perkembangan lanjutan dari bakal buah (ovarium). Buah biasanya membungkus dan melindungi biji. Aneka rupa dan bentuk buah tidak terlepas kaitannya dengan fungsi utama buah, yakni sebagai pemencar biji tumbuhan. Oleh karena itu, untuk membedakannya, buah menurut pengertian botani biasa disebut buah sejati. Buah sering kali memiliki nilai ekonomi sebagai bahan pangan maupun bahan baku industri karena di dalamnya disimpan berbagai macam produk metabolisme tumbuhan, mulai dari karbohidrat, protein, lemak, vitamin, mineral, alkaloid, hingga terpena dan terpenoid. Ilmu yang mempelajari segala hal tentang buah dinamakan pomologi. Pada menu ini terdapat sebelas jenis buah yang akan diklusterkan atau dikelompokkan ke dalam kategori masing masing buah yang memiliki kadar vitamin yang sama dengan memanfaatkan hal lain yang perlu diperhatikan.

KESIMPULAN

1. Implementasi Algoritma K-Medoids dalam Menentukan pengelompokkan buah berdasarkan kadar vitamin telah berhasil dibangun dan dapat digunakan untuk menentukan buah berdasarkan kadar vitaminnya.
2. Untuk menentukan tingkat pengelompokkan nya harus ditentukan nilai variasinya.
3. Penggunaan Pricipal Component Analysis cukup berpengaruh dalam peningkatan hasil uji validasi menggunakan Silhoutte Index. Terlihat dari hasil penyebaran data di setiap cluster terbaik menggunakan atribut 3 dan 4 terbaik dibandingkan dengan menggunakan tanpa Pricipal Component Analysis.

REFERENSI

- Boiculese, V.L., Dimitru, G., & Moscalu, M. 2013. Improving Recall of K-Nearest Neighbor Algorithm for Classes of Uneven Size. The 4th IEEE International Conference on E-Health and Bioengineering-EHB: 1-4.
- Haviluddin, "Memahami Penggunaan UML (Unified Modelling Language)", Jurnal Informatika Mulawarman, Vol. 6, No. 1, Februari 2011.
- Jaafar, H., Mukahar, N., & Ramli, D.A. 2016. Methodology of Nearest Neighbor: Design and Comparison of Biometric Image Database. IEEE Student Conference on Research



- and Development (SCORED) : 1-6.
- Pan, D., Zhao, Z., Zhang, L., & Tang, C. 2017. Recursive Clustering K-Nearest Neighbors Algorithm and the Application in the Classification of Power Quality Disturbance. IEEE Conference on Energy Internet and Energy System Integration (EI2) : 1-5.
- Priyanto Hidayatullah, 2012, "Visual Basic .NET Membuat Aplikasi Database dan Program Kreatif". Bandung: Informatika.
- Preeti, Arora., Dr. Deepali., Shipra, V. (2016). Analysis of K-Means and K- Medoids Algorithm For Big Data. *Procedia Computer Science* 507 – 512.
- Sangga, V. A. P. (2018). Perbandingan Algoritma K-Means Dan Algoritma K-Medoids Dalam Pengelompokan Komoditas Peternakan Di Provinsi Jawa Tengah Tahun 2015. Skripsi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia: Yogyakarta.
- Sundari, Siti., Irfan, S.D., Agus, P.W., Heru, S.T., Jalaluddin., Anjar, W. (2019). Analisis K-Medoids Clustering Dalam Pengelompokkan Data Imunisasi Campak Balita di Indonesia. *Prosiding Seminar Nasional Riset Information Science*, hal:687-696.