JDSP

Jurnal Data Science Penusa

Implementasi Teknologi Data Mining Untuk Analisis Data Promosi Universal English

Swita Natalia[1)*], Amran Sitohang[2)]
[1,2)] Teknologi Informasi, STMIK Pelita Nusantara, Medan
[1,2)] swita@gmail.com amran89@gmail.com

Received: 10 Agustus 2024 Accepted: 08 September 2024 Published: 27 Desember 2024
Pp. 822-831

OPEN ACCESS

***Penulis Koresponden**

Kata Kunci: data mining, k-means clustering, analisis data

JDSP: Jurnal Data Science Penusa is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Universal English merupakan sebuah program pembelajaran bahasa Inggris yang menghasilkan data promosi dalam jumlah besar, mencakup respons peserta, preferensi, dan tingkat partisipasi. Mengelola data tersebut secara manual menjadi tidak efisien, sehingga dibutuhkan teknologi yang dapat mengekstrak wawasan dengan cepat dan objektif. Penelitian ini bertujuan untuk mengimplementasikan data mining menggunakan metode K-Means Clustering dalam menganalisis data promosi Universal English. Metode ini digunakan untuk mengatasi kompleksitas pengolahan data dan membantu dalam pengambilan keputusan yang lebih baik. Hasil clustering disajikan dalam bentuk iterasi yang ditampilkan di halaman website, memungkinkan analisis lebih mendalam terhadap data yang dikelompokkan. Perancangan aplikasi berbasis website dipilih karena kemudahannya dalam penggunaan serta keunggulan dalam keringanan aplikasi. Dengan ini, Universal English dapat memanfaatkan data promosi secara lebih efektif dan strategis.

PENDAHULUAN

Teknologi komputer pada saat ini sangat berguna dalam berbagai bidang. Teknologi mampu mengoptimalkan strategi promosi yang dibutuhkan dalam perusahaan. Analisis data yang mendalam dapat membantu mengidentifikasi tren pembelajaran, kebutuhan individu, dan preferensi dalam proses pengelompokan. Teknologi dengan data dapat dirancang strategi promosi yang lebih terarah dan efektif, meningkatkan partisipasi peserta, dan meningkatkan keberhasilan pembelajaran.

Universal English suatu institusi atau program pembelajaran bahasa Inggris. Kegiatan promosi menghasilkan data yang cukup besar yang melibatkan berbagai aspek seperti respons peserta, preferensi, dan tingkat partisipasi. Mengelola dan menganalisis data semacam itu secara manual menjadi tidak efisien, sehingga diperlukan teknologi yang mampu mengekstrak wawasan dengan cepat dan objektif. Berdasarkan jumlah data yang banyak maka perlu untuk diklasifikasikan sehingga lebih mudah untuk memperoleh informasi penting dan dijadikan sebagai dasar pengambilan keputusan (Wibowo & Jananto, 2020). Melalui perkembangan teknologi saat ini data transaksi tersebut dapat diolah untuk menghasilkan strategi yang bisa dijadikan kekuatan dalam bersaing di bidang usaha ritel.

Beberapa permasalahan yang ada dalam kasus ini yatu kompleksitas memproses dan menganalisis data, serta pengelolaan sumber daya yang diperlukan pada analisis data promosi. Permasalahan yang relevan dalam hal ini termasuk keragaman data promosi yang harus dihadapi, kecukupan data yang tersedia, serta relevansi dan akurasi informasi yang dihasilkan dari analisis maka dibutuhkan Metode *K-Means Clustering*: Penggunaan metode *K-Means Clustering* dalam analisis data memunculkan beberapa permasalahan potensial, seperti penentuan jumlah cluster yang optimal, pemilihan atribut yang tepat untuk analisis, serta evaluasi hasil clustering yang dihasilkan.

Menurut Putri Mai Sarah Tarigan,dkk (2022) Data mining merupakan serangkaian proses untuk

menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data Mining digunakan untuk mengambil pola dan informasi yang digunakan antara data dalam pengelompokkan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. *K-Means Clustering* merupakan klasifikasi tanpa pengawasan danmerupakan proses partisi sekumpulan objek data dari satu set menjadi beberapa kelas. Hal ini dapat dilakukan dengan menerapkan berbagai persamaan dan langkah-langkah mengenai jarak algoritma, yaitu dengan *Euclidean Distance* (Nurzahputra dkk., 2018), masalah yang terjadi pada perlu memberikan informasi lebih lanjut tentang karakteristik, tujuan, dan metode pengajaran yang digunakan oleh Universal English. Keterbatasan informasi dapat menghambat pemahaman tentang konteks dan relevansi penelitian. Analisis data promosi harus memperhitungkan tanggapan dan masalah yang mungkin timbul melibatkan sejauh mana data yang dapat diakses, tingkat respons peserta, dan bagaimana mengatasi bias yang mungkin muncul dalam dataset. Diperlukan pemahaman yang komprehensif tentang strategi promosi yang digunakan oleh objek. Data mining adalah proses ekstraksi pola atau pengetahuan yang bermanfaat dari suatu himpunan data besar atau kompleks. Metode data mining merujuk pada teknik-teknik atau pendekatan yang digunakan dalam proses ini. Beberapa penelitian sebelumnya telah melakukan penerapan Algoritma Data Mining untuk mengolah data transaksi hanya saja penelitian tersebut sampai pada analisis pola pembelian , seperti yang dilakukan oleh Wibowo dan Jananto (2020) yang menganalisa pola pembelian alat tulis kantor pada perusahaan ritel.

TINJAUAN LITERATUR

Data Mining

Proses yang dikenal sebagai *data mining* adalah menemukan pola data dan pengetahuan yang menarik dari kumpulan data yang sangat besar. Sumber datanya termasuk *database*, *data warehouse*, situs web, *repository*, dan data yang dialirkan ke dalam sistem dinamis. *Data mining* adalah proses ekstraksi yang digunakan untuk mendapatkan informasi penting yang belum diketahui dan implisit. Selain itu, *data mining* berhubungan dengan banyak bidang seperti *statistik, machine learning* (pembelajaran mesin), *pattern recognition, computing algorithms, database technology,* dan *high performance computing* (Muslim et al., 2019).

Beberapa faktor yang menjadi alasan utama mengapa *data mining* diguanakan (Febriani & Sulistiani, 2021) adalah :

1. Banyaknya data yang terkumpul sehingga memerlukan waktu yang sangat lama dan tenaga ahli yang cukup banyak untuk menganalisisnya.
2. Komputer menjadi salah satu pilihan utama karena kemampuannya dalam kecepatan, ketepatan, tidak pernah lelah dan mudah dioperasikan.
3. Tekanan dari kompetisi bisnis yang terus menguat sehingga menjadikan informasi menjadi sangat penting dan harus segera dimiliki.
4. Mampu menemukan suatu pola yang tidak terpikirkan sama sekali.

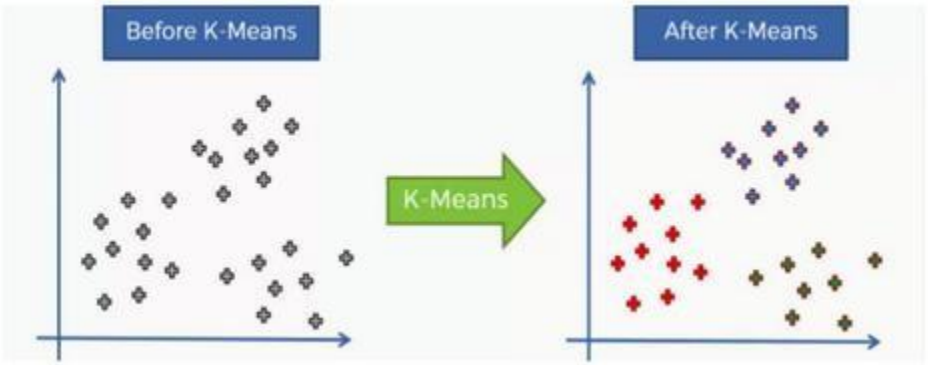
Data mining dibagi menjadi beberapa tahapan (Marlina & Bakri, 2021), antara lain :

1. Pembersihan data: untuk membuang data yang tidak konsisten dan *noise*.
2. Integrasi data: untuk menggabungkan data dari beberapa sumber.
3. Transformasi data : untuk mengubah data menjadi bentuk yang sesuai untuk di *mining*.
4. Aplikasi teknik *data mining*.
5. Evaluasi pola yang ditemukan : untuk menemukan informasi yang menarik atau pun bernilai.

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining disebut juga dengan knowledge-discovery in database (KDD) ataupun pattern recognition. Istilah KDD atau disebut penemuan pengetahuan data karena tujuan utama data mining adalah untuk memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan informasi baru yang berguna. Sedangkan istilah pattern recognition atau disebut pengenalan pola mempunyai tujuan pengetahuan yang akan digali dari dalam bongkahan data yang sedang dihadapi (Purwa Hasan Putra et al., 2022).

Metode K-Means Clustering

Pada algoritma *K-Means Clustering*, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dahulu target kelasnya. Masukan yang diterima adalah data atau objek dan *k* buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek kedalam *k* buah kelompok tersebut. Kemudian pada setiap *cluster* terdapat titik pusat (*centroid*) yang mempresentasikan *cluster* tersebut. Secara sederhana algoritma *k-means* dapat dijelaskan sebagai algoritma *data mining* yang digunakan untuk menyelesaikan masalah pengelompokkan (*clustering*). Clustering merupakan bidang penelitian dalam analisis dan data mining. Pada clustering partisi dari objek data yang mempunyai karakteristik sama akan dikelompokkan padasatu kelompok dan data yang memiliki karakteristik berbeda akan dikelompokkan pada kelompok yang lainnya. Salah satu metode pada clustering yang terkenal daripada algoritma clustering lainnya adalah K-means karena kesederhanaan algoritma dan efisiensinya . (Nova Novitasari dkk, 2023).



Gambar 1. K-Means Clustering

Proses Perhitungan Algoritma *K-Means Clustering*

K-means adalah salah satu algoritma pengelompokan yang bekerja berdasarkanprinsip partitioned clustering. Prinsip kerja dari partitioned clustering yaitu mengelompokkan item secara acak karena dipengaruhi centroid. SedangkanPrinsip kerja dari pengelompokan hierarchical clusteringdilakukan secara bertahap. Dan disetiap iterasi dari pengelompokan hierarchical clustering hanya akan ada satu pemilihan penggabungan suatu item terhadap itemlainnya

Berikut tahapan-tahapan dalam melakukan clustering dengan menggunakan algoritma K-Means Clustering sebagai berikut:

1. Tentukan jumlah cluster K. Inisialisasi K pusat cluster. Ini dapat dilakukakn dengan berbagai cara. Paling banyak digunakan adalah dengan menentukan secara random.cluster diberi nilai awal dengan angka random.
2. Alokasikan semua data/objek ke dalam cluster terdekat. Kedekatan dua objek bisa didapatkan berdasarkan jarak kedua objek tersebut. Untuk dapatmenghitung jarak semua data terhadap titik pusat cluster maka dapat menggunakan teori jarak Euclidean yang dirumuskan dibawah ini:

$V_{ij} = \frac{1}{Ni} ?_{k=0} Ni\ Xkj$ .....(2.1)

Keterangan

Vij	= Centroid rata-rata *Cluster* ke-I untuk variabel ke-j
Ni	= Jumlah anggota *cluster* ke-i
i, k	= Indeks dari *Cluster*

j = Indeks dari Variabel
Xkj = Nilai data ke-k variabel ke-j untuk *cluster*

3. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusatcluster adalah rata-rata dari semua objek/data dalam cluster tertentu. Jikadikehendaki bisa juga menggunakan median dari clustr tersebut. Jadi rata-rata (mean) bukan satu satunya ukuran yang bisa dipakai, seperti rumus dibawah ini.

dimana :

$D_e = \sqrt{(xi - si)^2 + (yi - ti)^2}$ .....(2.2)

Keterangan
De = Eulidean Distance
I = Banyaknya Objek
(x,y) = Koordinat Objek
(s, t) = Koordinat Centroid

4. Tugaskan lagi tiap objek memakai pusat cluster yang baru. Jika pusat clustertidak berubah lagi maka proses clustering selesao.atau kembali ke angkah 3.

***Unifed Modelling Language* (UML)**

Menurut (Destriana, 2021) UML merupakan bahasa buat memvisualisasi, membangun, menspesifikasi serta mendokumentasikan artifacts(bagian dari data yang dipakai buat ditampilkan oleh mekanisme pembuatan piranti lunak, artifact tersebut bisa berbentuk model, uraian ataupun piranti lunak) dari sistem piranti lunak, seperti untuk pemodelan bisnis serta pola non piranti lunak yang lain. Tidakhanya itu UML merupakan bahasa pemodelan yang memakai konsep orientadi object. Dibawah bendera Rasional Software Corps UML di dirikan oleh James Rumbaugh, Ivar Jacobson, serta Grady Booch. UML sediakan notasi- notasi yang menolong memodelkan sistem dari bermacam prespektif. UML bukan cuma dipakai dalam pemodelan piranti lunak, tetapi nyaris dalam seluruh bidang yang memerlukan pemodelan.

Menurut (Deimas et al., 2021) UML merupakan salah satu tools atau modelyang digunakan untuk membuat suatu aplikasi pengembangan software dengan berbasis object-oriented. Selain itu juga mampu menghasilkan tulisan standar padasistem blue print seperti kerangka bisnis proses, penulisan kelas yang spesifik, skema database, serta 27 berbagai komponen yang dibutuhkan pada sistem tersebut. Unified Modeling Language merupakan set standar pada teknik untuk membuat diagram. Diagram UML yang paling berguna diantaranya adalah Use Case Diagram, Statechart Diagram, Sequence Diagram, Component Diagram, Class Diagram, Activity Diagram, dan Deployment Diagram.

HTML

HTML singkatan dari *Hypertext Markup Language*. HTML lebih merujuk ke Bahasa *mark up* yang diperuntukkan untuk menentukan format atau *style* dari teks yang ditandai. Dokumen HTML merupakan dasar dari semua konten yang muncul di *World Wide Web* (WWW), terdiri dari dua bagian penting: konten informasi dan satu set instruksi yang memberi tahu komputer Anda cara menampilkan konten tersebut. Setiap dokumen HTML harus mengandung minimalempat elemen:

<html> ... </html>

<head> ... </head>

<title> ... </title>

<body> ... </body>

Elemen-elemen yang menentukan bagian penting dari dokumen HTML yaitudokumen itu sendiri, bagian judul, bagian judul, dan badan. Keempat elemen harusdisertakan meskipun tidak menyertakan konten apa pun. Setiap HTML elemen ditentukan oleh satu atau dua tag, biasanya tag awal dan tag akhir. Tag selalu diapit oleh kurung sudut: <…>. Tag akhir dimulai dengan garis miring (/). (Mira Orisa, Ahmad Faisol, Mochammad Ibrahim Ashari, 2023)

PHP

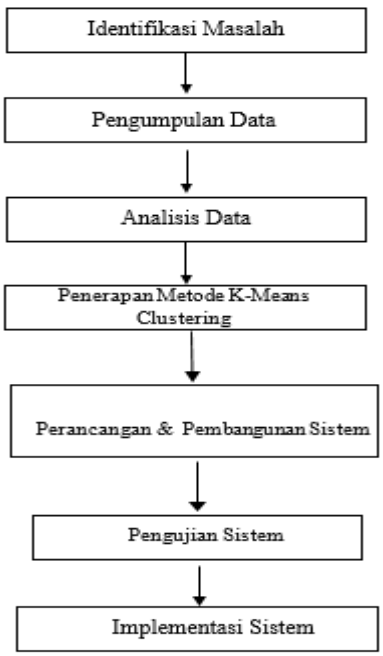
PHP adalah singkatan dari *hypertext Preprocessor* yang merupakan bahasa pemrograman yang berjalan disisi server. Bahasa pemrograman PHP pertama kali diperkenalkan oleh Rasmus Lerdorf sebagai aplikasi *opensource*. Kemudian para pengembang mulai mengembangkan Bahasa pemrograman ini hingga tahun 1997 lebih dari 500.000 *website* didunia menggunakan PHP pada sistem yang digunakanbaik untuk koneksi ke basisdata, menampilkan konten yang dinamis dan banyak hallainnya. (Mira Orisa, Ahmad Faisol, Mochammad Ibrahim Ashari, 2023). PHP adalah bahasa pemrograman yang bersifat open source. Adapun fungsi PHP adalah sebagai berikut : PHP dapat mengatur user pengguna. PHPberfungsi sebagai bahasa pemrograman terbesar di dunia hingga saat ini. PHP dapat mengamankan suatu data. (Wibowo, 2018)

MySQL

MySQL merupakan aplikasi yang digunakan untuk mengelola 32 data yangakan tersimpan dalam database server bersifat multiuser dan multi-threaded. Structured Query Language (SQL) digunakan untuk mengelola data secara terstruktur dan pada MySQL digunakan untuk menambah, merubah dan menghapus data yang berada dalam database (Sri Endang Anjarwani & BaiqMaisum Jum, 2022)

METODE PENELITIAN

Metodologi penelitian merupakan kerangka kerja untuk melakukan suatu tindakan, atau suatu kerangka berpikir untuk menyusun ide atau gagasan yang terarah dan terkait dengan maksud dan tujuan. Kerangka kerja penelitian dapat di gambarkan dalam suatu alur kegiatan kerja penelitian seperti terlihat pada gambar di bawah ini:



Gambar 2 Kerangka Kerja Penelitian

HASIL PENELITIAN

Pengerjaan algoritma *k-means clustering* merupakan pengerjaan yang berulang. Sesuai dengan cara penyelesaian yang telah dijelaskan diawal. Setelah dilakukan proses perhitungan jarak setiap data dengan pusat cluster yang ditentukan di awal, Proses clustering dilanjutkan dilakukan melakukan perhitungan kembali terhadap hasil peng-clusteran awal untuk menghasilkan pusat cluster yang baru. Nilai pusat cluster yang baru dihitung dengan menjumlahkan nilai seluruh data yang termasuk dalam sebuah cluster, kemudian dibagi dengan jumlah anggota yang tergabung dalam sebuah cluster. Penjumlahan dilakukan sesuai dengan menjumlahkan nilai setiap kolom agar dapat kembali dijadikan pusat cluster. Selengkapnya dapat dilihat pada tabel dibawah ini:

Tabel 1

Penentuan Pusat Cluster Baru Pada Iterasi II Terhadap C1

No	Nama Pengajuaan Data Promosi	Nilai Kelompok A	Nilai Kelompok B	Cluster
1	Dongan Maria Sihombing	83,71428571	78,66666667	C1
2	Inez Deswanti Simarmata	82,71428571	79	C1
3	Laura G. Siringoringo	82,71428571	78,33333333	C1
4	Gracia Margareth Sinaga	82,28571429	80	C1
5	Valentina Sinaga	82	79,33333333	C1
6	Winda Serefina Sitohang	81,71428571	78	C1
7	Yohana Anggraini Sinaga	81,42857143	76,66666667	C1
8	Aurel Oktaviani Sinaga	81,28571429	92	C1
9	Hendra Gunawan Sinaga	83,85714286	79,33333333	C1
	Mayoryn Prodia Simbolon	81,42857143	78,66666667	C1
	Nicolas P. B.Simangunsong	83,28571429	80,33333333	C1
	Chyntia Mare C Sinaga	86,42857143	78,66666667	C1
	Eriatmo Felix Sirumapea	82,14285714	78	C1
	Glen Yuda Situmorang	86	81	C1
	Johan Frans Situmorang	82,14285714	77,66666667	C1
	Melani Nainggolan	83,71428571	77	C1
	Ray Febryanti G Silalahi	82,71428571	77,66666667	C1
	Rossaria CSirumapea	85,42857143	80,66666667	C1
	Rata-Rata Nilai Kelompok A	83,05555556		
	Rata-Rata Nilai Kelompok B		79,5	

Tabel 2

Perbandingan Hasil Iterasi I dan Iterasi II

Cluster	Iterasi I	Iterasi II	Keterangan
C1	18	16	Kelas A

C2	49	40	Kelas B
C3	69	79	Kelas C
C4	3	4	Kelas D

Pada tabel diatas dapat dilihat perbandingan hasil peng-clusteran pertama dan kedua, membandingkan hasil peng-*Clusteran* disini berfungsi untuk mengetahui apakah ada data yang berpindah *Cluster* atau tidak. Berdasarkan tabel diatas, masih terdapat data yang berpindah *Cluster*. karena masih ada data yang berpindah *Cluster*, maka perhitungan kembali dilanjutkan dengan kembali mencari pusat Cluster yang baru dengan menggunakan langkah dan rumus yang sama dengan perhitungan mencari pusat Cluster baru yang sebelumnya.

Iterasi Ke III

Untuk mencari pusat *Cluster* baru, dihitung dengan mengambil nilai rata-rata dari setiap *Cluster* dengan menjumlahkan data untuk tiap *Cluster* kemudian dibagi dengan jumlah data yang tergabung dalam *Cluster* tersebut uraian formula yang digunakan berikut ini :

a. Pusat *Cluster* 1 (C1) baru
   Rata-Rata Nilai Kelompok A = 1331,86/ 16 = 83,24
   Rata- Rata Nilai Kelompok B = 1249,75/ 16 = 78,11
   Maka pusat *Cluster* baru untuk *Cluster* 1 adalah 83,24 dan 78,11
b. Pusat *Cluster* 2 (C2) baru
   Rata-Rata Nilai Kelompok A = 3188,14/ 40 = 79,70
   Rata- Rata Nilai Kelompok B = 3058,00/ 40 = 76,45
   Maka pusat *Cluster* baru untuk *Cluster* 2 adalah 79,70 dan 76,45
c. Pusat *Cluster* 3 (C3) baru
   Rata-Rata Nilai Kelompok A = 6022,14/79= 76,23
   Rata- Rata Nilai Kelompok B = 5989,75/79=75,82
   Maka pusat *Cluster* baru untuk *Cluster* 1 adalah 76,23 dan 75,82
d. Pusat *Cluster* 4 (C4) baru
   Rata-Rata Nilai Kelompok A = 278,29/4=69,57
   Rata- Rata Nilai Kelompok B = 298,00/4=74,50
   Maka pusat *Cluster* baru untuk *Cluster* 1 adalah 69,57 dan 74,50

Uraian dari perhitungan tersebut diuraikan pada tabel berikut ini :

Tabel 3

Titik Pusat Iterasi Pada Iterasi III

Cluster	Rata-Rata Nilai Kelompok A	Rata-Rata Nilai Kelompok A
C1	83,24	78,11
C2	79,70	76,45
C3	76,23	75,82
C4	69,57	74,50

Selanjutnya kembali dilakukan penghitungan jarak setiap data ke pusat *Cluster* baru dengan menggunakan rumus yang sama dengan yang digunakan sebelumnya. Dimulai dengan menghitung jarak setiap data dengan

pusat Cluster, Dengan menghasilkan tabel berikut ini :

Tabel 4.

Jarak Antara Cluster dengan Titik Pusat Pada Iterasi III

No	Nilai Kelompok A	Nilai Kelompok B	C1	C2	C3	C4
1	80,57142857	77	2,89022	1,03048	4,49893	11,2819
2	80,57142857	77,25	2,80372	1,18296	4,57088	11,3399
3	83,71428571	77,25	0,98211	4,09322	7,61967	14,4091
4	80,42857143	77	3,02262	0,91286	4,36124	11,1426
5	82,71428571	78	0,5371	3,38946	6,84093	13,6023
6	81	76,25	2,91156	1,31529	4,78934	11,5632
7	82,71428571	77,25	1,00796	3,11864	6,64009	13,4289
8	79,28571429	77,5	4,00106	1,12877	3,48709	10,1683
9	80,14285714	76	3,74759	0,63137	3,917	10,6787
10	80,57142857	78	2,67084	1,77817	4,85802	11,5448

Dari uraian data tersebut berdasarkan jumlah data yang diuji, maka jumlah data promosi pada iterasi I, II, dan III pada tabel berikut ini :

Tabel 5

Pengelompokan Nilai Pada Iterasi I sampai dengan III

Cluster	Iterasi I	Iterasi II	Iterasi III	Penempatan Kelas
C1	18	16	16	Kelas A
C2	49	40	43	Kelas B
C3	69	79	76	Kelas C
C4	3	4	4	Kelas D

Informasi yang dapat diberikan dari perhitungan dengan memanfaatkan Algoritma K-means adalah:

1. Pada tahapan iterasi III memberikan rekomendasi penempatan kelas bahwa pada Kelas A berjumlah 16 orang, Kelas B berjumlah 43 orang, Kelas C berjumlah 76 orang serta kelas D berjumlah 4 orang.
2. Tabel diatas ditemukan perbedaan masing-masing iterasi dengan beberapa perbandingan bahwa iterasi 2 dan 3 memiliki jumlah data yang sama untuk cluster C1 dan C4
3. Cluster 1 dikategorikan kelas A, Cluster 2 dikategorikan kelas B, Cluster 3 dikategorikan Kelas C dan cluster 4 dikategorikan kelas D, berdasarkan hasil data promosi Universal English menggunakan metode K-Means Clustering

PEMBAHASAN

Kebutuhan sistem merupakan kondisi atau kemampuan yang harus dimiliki oleh perangkat lunak untuk memenuhi kebutuhan pengguna. Dalam membangun aplikasi data mining menggunakan algoritma K-Means, terdapat dua aspek penting yang harus diperhatikan: kebutuhan perangkat lunak (software) dan perangkat keras (hardware). Kebutuhan perangkat lunak mencakup semua program yang diperlukan untuk menjalankan aplikasi. Fungsinya adalah memproses data, menghubungkan pengguna dengan perangkat keras, mengidentifikasi program, serta menyediakan fungsi dasar bagi perangkat keras. Software yang digunakan dalam pengembangan aplikasi data mining ini meliputi Sistem Operasi Windows 10 Ultimate atau versi yang lebih baru, bahasa pemrograman PHP, dan PhpMyAdmin sebagai manajemen basis data yang diaktifkan melalui XAMPP. Sementara itu, kebutuhan perangkat keras merujuk pada spesifikasi fisik komputer yang diperlukan untuk menjalankan aplikasi secara optimal. Spesifikasi yang direkomendasikan antara lain Processor Intel Pentium dengan kecepatan 1.87 GHz**,** RAM minimal 2 GB**,** VGA Card minimal 64 MB, serta perangkat input seperti mouse dan keyboard. Perangkat keras ini mendukung aplikasi data mining dalam memproses data secara efisien dan menjalankan sistem secara lancar.

KESIMPULAN

Untuk mengimplementasikan data mining menggunakan metode K-Means Clustering pada data promosi Universal English, langkah pertama yang dilakukan adalah mengumpulkan data terkait promosi, seperti data calon siswa, respons terhadap promosi, dan faktor-faktor lain yang relevan. Setelah data terkumpul, dilakukan proses preprocessing untuk membersihkan dan menormalkan data. Selanjutnya, jumlah cluster (K) ditentukan, dalam hal ini K ditetapkan sebagai 4, untuk membagi data menjadi empat kelas: Kelas A, B, C, dan D. Algoritma K-Means kemudian digunakan untuk mengelompokkan data ke dalam empat cluster berdasarkan karakteristik yang relevan. Pada iterasi ketiga, ditemukan bahwa Kelas A terdiri dari 16 orang, Kelas B berjumlah 43 orang, Kelas C terdiri dari 76 orang, dan Kelas D berjumlah 4 orang. Setiap iterasi memberikan hasil yang berbeda, namun pada iterasi kedua dan ketiga, jumlah data pada cluster C1 dan C4 tetap sama. Cluster 1 dikategorikan sebagai Kelas A, Cluster 2 sebagai Kelas B, Cluster 3 sebagai Kelas C, dan Cluster 4 sebagai Kelas D. Hasil pengelompokan ini memberikan wawasan mengenai distribusi data promosi dan membantu dalam penargetan strategi promosi yang lebih tepat. Dalam merancang dan membangun teknologi data mining untuk analisis data promosi Universal English, perancangan sistem dimulai dengan menggunakan diagram UML, seperti *Use Case* Diagram, *Activity* Diagram, dan *Class* Diagram, yang menggambarkan alur proses mulai dari input hingga output sistem. *Use Case* Diagram menunjukkan interaksi antara pengguna dan sistem, sedangkan *Activity* Diagram menggambarkan alur kerja proses clustering. *Class* Diagram digunakan untuk menunjukkan hubungan antar objek dan struktur data dalam sistem. Sistem kemudian dibangun menggunakan bahasa pemrograman PHP dengan MySQL sebagai basis data untuk menyimpan data promosi dan hasil pengelompokan.

REFERENCES

[1] Dewi, A. O. P. (2020). Big Data di Perpustakaan dengan Memanfaatkan Data Mining. *ANUVA*, *4*(2), 8.

[2] Handayani, M. T. (2021, Oktober 15). *Data mining: Pengertian, proses, manfaat, dan 3 contoh penerapannya*. Data mining: Pengertian, proses, manfaat, dan 3 contoh penerapannya. https://www.ekrut.com/media/data-mining-adalah

[3] Helma, S. S., Mustakim, Rustiyan R, R., & Normala, E. (2019). Clustering pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru MenggunakanAlgoritma K-Means. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)*, *11*, 7.

[4] Hussein, S. (2021, Oktober 12). *K-means Clustering: Pengertian, Metode Algoritma, Beserta Contoh*. GEOSPASIALIS. https://geospasialis.com/k- means-clustering/

[5] Indarti, I., Wahyuningsih, U., Yulistiana, Y., Suhartini, R., & Prihatina, Y. I. (2021). Faktor-Faktor Yang Mempengaruhi Kepuasan Belajar Jarak Jauh Mahasiswa Vokasi Di Masa Pandemi Covid-19. *Jurnal Pendidikan Teknologi dan Kejuruan*, *18*(2), 226. https://doi.org/10.23887/jptk- undiksha.v18i2.31147

[6] Indriyani, F., Yunita, Y., & Muthia, D. A. (2019). *Analisa Perancangan Sistem Informasi*. Graha Ilmu.

[7] Jannah, N., & Yulianto, T. (2018). *Mengelompokkan Siswa Berprestasi Akademik dengan Menggunakan Metode K Means Kelas VII MT*. *2*(2), 5.

[8] Marzuki, K., & Amir, R. (2019). Kepuasan Belajar Warga Belajar Pada Program Kesetaraan Paket C. *PROSIDING SEMINAR NASIONAL LP2M UNM*, 6.

[9] Nawawi, M. S., & Gustian, D. (2020). Algoritma K-Means Dalam Penentuan Tingkat Kepuasan Siswa Terhadap Pembelajaran Online Di Masa PandemiCovid 19. *SEMNASIF*, 11.

[10] Nurzahputra, A., Muslim, M. A., & Khusniati, M. (2018). Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa. *Techno.Com*, *16*(1), 17–24. https://doi.org/10.33633/tc.v16i1.1284

[11] Pawening, R. E., Fatmawati, & Aisyah, S. (2021). Algoritma K-Means untuk Mengukur Kepuasan Mahasiswa Menggunakan E-Learning. *JoTI*, *3*(1), 27–33. https://doi.org/10.37802/joti.v3i1.201

[12] Populix. (2021, Juni 22). *Apa itu Data Mining? Pengertian, Metode, Tahapan,dan Contoh Terbaru*. Populix. https://www.info.populix.co/post/data- mining-adalah

[13] Putra, I. D. G. R. D. (2019). Peran Kepuasan Belajar Dalam Mengukur Mutu Pembelajaran Dan Hasil Belajar. *Jurnal Penjaminan Mutu*, *5*(1), 22. https://doi.org/10.25078/jpm.v5i1.756

[14] Rohmah, A., Sembiring, F., & Erfina, A. (2021). Implementasi Algoritma K- Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung). *Seminar Nasional Sistem Informasi dan Manajemen Informatika*, 9.

[15] Sutanto, N. H., Setiawan, B. A., & Rakhman, G. F. (2020). Analisis Aspek-Aspek Kualitas Skema Basis Data (Studi Kasus: Analisis Jabatan Bagian Organisasi Kabupaten Balangan). *Jurnal Syntax Admiration*, *1*(7), 12.

[15] Sutrisno, S. (2020). Penerapan Algoritma Apriori Untuk Mencari Pola Penjualan Produk Dana Pada PT Bank Rakyat Indonesia (Persero) Tbk Kanca Jakarta Pasar Minggu. *Jurnal Sistem Informasi dan Informatika (Simika)*, *3*(1), 12–26. https://doi.org/10.47080/simika.v3i1.834

[17] Syafnidawaty. (2020, April 19). *K-means Clustering*. Universitas Raharja. https://raharja.ac.id/2020/04/19/k-means-clustering/

[18] Tabrani, M., & Priyandaru, H. (2021). Sistem Informasi Manajemen Berbasis Website Pada Unl Studio Dengan Menggunakan Framework Codeigniter. *Jurnal ILMIAH M-PROGRESS*, *11*(1), 9.