

CLUSTERING DENGAN ALGORITMA K-MEANS UNTUK MENINGKATKAN PENJUALAN PADA PT PULAU BINTANG EMAS BATAM

Putri Aprilianti Ali¹, Sasa Ani Arnomo²

¹Mahasiswa Program Studi Sistem Informasi, Universitas Putera Batam

²Dosen Program Studi Sistem Informasi, Universitas Putera Batam

email: pb181510073@upbatam.ac.id

ABSTRACT

This study was conducted to study K-Means Clustering in knowledge retrieval. The purpose of this study is to obtain the results of the analysis of the application of clustering with the K-Means algorithm in determining the number of tourist interests in each travel agent that cooperates with PT Pulau Bintang Emas in Batam. The data used to support marketing strategy decisions include gender, age and travel agent, promos which are analyzed using the K-Means Clustering Algorithm so that the right customer cluster is known for the promotional approach. The data will be processed using Rapidminer data processing software with the K-Means Clustering Algorithm and a value of $K = 3$ so that the data will produce three clusters. The total data that has been tested using Rapidminer software is 398 records with 3 clusters obtained, namely: Cluster 0 totaling 130 items, Cluster 1 totaling 111 items and Cluster 2 totaling 157 items . With a computing speed of 0.01 seconds.

Keywords: Cluster, Data mining, K-Means Algorithm, Marketing

PENDAHULUAN

Secara geografis, Kota Batam terletak di daerah berbukit dengan garis pantai yang mengelilingi wilayahnya. Karena lokasinya yang strategis di selat Singapura dan berbatasan dengan Malaysia serta Singapura, kota ini menjadi tujuan menarik tidak hanya untuk investasi tetapi juga untuk wisata. Batam memiliki banyak potensi wisata menarik, seperti pantai berpasir, hutan mangrove yang lebat, berbagai kuliner, serta tempat-tempat bersejarah dan edukatif. PT. Pulau Bintang Emas (Kepri Coral Resort), yang didirikan pada tahun 2016, adalah sebuah perusahaan yang berkembang di sektor pariwisata. Kepri Coral Resort menawarkan pengalaman unik bagi setiap tamu dengan lokasi

yang memukau. Perusahaan ini sedang membangun sebuah resort di Pulau Pengalap, yang merupakan destinasi wisata bahari terbaru di selatan Pulau Batam. Dengan investasi mencapai Rp11 triliun dan rencana pembangunan dalam 20 tahun, proyek ini diharapkan dapat menciptakan lapangan kerja untuk sekitar 10 ribu orang. PT Pulau Bintang Emas, yang berdiri sejak tahun 2016, telah sukses sebagai pengelola resmi yang berkembang di Kota Batam. Evaluasi kunjungan wisata didasarkan pada laporan penjualan tahunan. Fasilitas dan promosi yang disediakan disesuaikan dengan kapasitas penjualan di Kepri Coral Resort. Untuk mempertahankan atau meningkatkan daya tarik perusahaan, strategi promosi

yang efektif diperlukan untuk menjamin kunjungan wisatawan lokal dan mancanegara tetap berkelanjutan. Salah satu strategi yang digunakan adalah penggunaan algoritma clustering K-Means untuk mengelompokkan jenis promosi dan paket wisata yang paling diminati berdasarkan demografi pengunjung serta mengklasifikasikan paket atau pax yang ditawarkan oleh PT Pulau Bintang Emas. Teks ini menyoroti beberapa aspek pariwisata di Kota Batam serta penerapan teknologi dalam pengembangan industri pariwisata. Berikut adalah poin-poin utama dari teks tersebut. Kota Batam memiliki daya tarik signifikan sebagai destinasi wisata dengan jumlah kunjungan wisatawan mancanegara mencapai 2,8 juta pada tahun 2019, menurut data Kementerian Pariwisata. Pemerintah mendukung pembentukan Kawasan Ekonomi Khusus (KEK) Pariwisata Batam untuk memfasilitasi pertumbuhan sektor pariwisata dan menciptakan iklim investasi yang kondusif di Kepulauan Riau, termasuk insentif-insentif dan regulasi yang mendukung pengembangan infrastruktur pariwisata dan investasi di wilayah tersebut.

KAJIAN TEORI

2.1 Knowledge Discovery in Database (KDD)

Pada tahun 1989, istilah Knowledge Discovery in Database pertama kali diperkenalkan untuk menggambarkan potensi yang dapat diperoleh melalui proses penambangan data dalam sebuah basis data, dengan tujuan untuk mengekstraksi informasi penting dari data tersebut (Shirazi, Baziyad, & Karimi, 2019). Ada lima tahapan dalam proses KDD (Dunham, 2018):

1. Seleksi Data

Pada tahap ini, dilakukan pemilihan data dari seluruh dataset yang akan dianalisis lebih lanjut.

2. Pra-pemrosesan Data

Pada tahap ini, data akan disaring dan dibersihkan untuk mempersiapkan data yang akan digunakan dalam proses perhitungan.

3. Transformasi

Tahap ini melibatkan perubahan format data ke format yang dapat diproses dalam proses penambangan data.

4. Penambangan Data

Tahap ini melibatkan proses perhitungan data sesuai dengan algoritma yang telah ditentukan.

5. Evaluasi

Pada tahap ini, hasil dari proses pengolahan data dievaluasi menggunakan aplikasi pengolah untuk memverifikasi kevalidan hasil yang diperoleh.

2.2 Data Mining

Data mining adalah proses untuk mengidentifikasi pola dan hubungan yang tersembunyi dalam sejumlah besar data dengan tujuan melakukan klasifikasi, estimasi, prediksi, aturan asosiasi, pengelompokan, deskripsi, dan visualisasi. Secara umum, data mining dapat dibagi menjadi dua kategori utama:

a. *Descriptive mining* yang bertujuan untuk mengungkapkan karakteristik penting dari data yang tersimpan dalam suatu basis data.

b. *Predictive mining* yang fokus pada pengembangan model untuk menemukan pola dari data dan memprediksi variabel lain di masa depan.

2.3 Clustering

Clustering adalah proses memisahkan atau memecah data ke dalam beberapa kelompok berdasarkan karakteristik

tertentu. Pada *clustering*, label untuk setiap data belum diketahui pada awalnya, namun diharapkan bahwa setelah proses *clustering* selesai, kelompok data dapat diberi label sesuai keinginan. Algoritma ini dirancang untuk memudahkan pengelompokan data berdasarkan variabel tertentu tanpa memerlukan proses pelatihan (training) terlebih dahulu.

2.4 K-Means

K-mean adalah metode pengelompokan non-hirarkis yang bertujuan untuk membagi data menjadi dua atau lebih kelompok. Metode ini membagi data ke dalam kelompok-kelompok (cluster) sedemikian rupa sehingga data dengan karakteristik serupa dikelompokkan bersama, sementara data yang memiliki karakteristik berbeda ditempatkan dalam kelompok yang berbeda (Prasetyo, 2018).

Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan *Rumus Euclidean* yang dirumuskan sebagai berikut:

$$(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2}$$

$D(i, j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Data ke j pada atribut data ke k

Rumus 2. 1 Rumus Euclidean Distance

2.5 Rapid Miner

Aplikasi Rapid Miner adalah aplikasi open-source yang dirancang memiliki framework tersendiri dalam proses penganalisaan data dan dapat dibaurkan ke aplikasi lainnya untuk melakukan kegiatan data mining (Kovács & Ghous, 2020).

2.6 Travel Agent

Travel agent adalah perusahaan yang menyediakan layanan perjalanan wisata beserta segala kelengkapannya (Faidah, 2022).

2.7 Promo

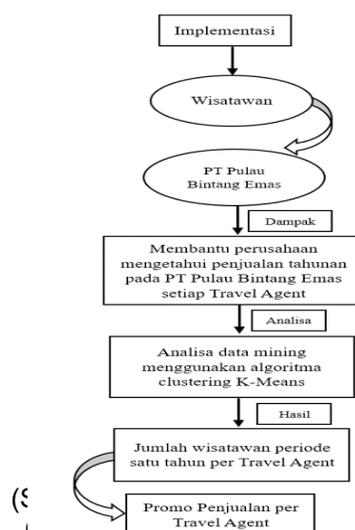
Secara sederhana, promo adalah suatu penawaran yang menguntungkan. Definisi promo merujuk pada penawaran yang sering kali digunakan dalam kegiatan pemasaran untuk memperkenalkan produk kepada masyarakat luas

2.8 Wisatawan

wisatawan adalah individu atau kelompok yang melakukan perjalanan untuk tujuan beristirahat, berbisnis, berobat, melakukan kunjungan keagamaan, atau perjalanan studi (Sugiana, 2018).

METODE PENELITIAN

Penelitian ini menggunakan metode kualitatif. Penjelasan mengenai metode penelitian yang diterapkan oleh penulis dapat dilihat pada gambar di bawah ini.



HASIL DAN PEMBAHASAN

Hasil perhitungan menggunakan algoritma k-means clustering menunjukkan segmentasi penjualan di PT Pulau Bintang Emas yang dibagi ke dalam tiga klaster. Klaster-klaster ini telah ditentukan sebelumnya oleh peneliti sebelum algoritma k-means clustering diterapkan. Klaster yang terbentuk adalah: klaster 0, yang berisi promosi dengan penjualan yang menengah; klaster 1, yang mencakup promosi dengan penjualan yang tidak laris; dan klaster 2, yang juga berisi promosi dengan penjualan yang laris. Berdasarkan perhitungan manual, klaster 0 terdiri dari 130 data, klaster 1 terdiri dari 111 data, dan klaster 2 juga terdiri dari

157 data. Tahap awal penelitian melibatkan pra-pemrosesan data dengan menarik sampel yang akan dianalisis. Setelah itu, data yang telah dipra-pemrosesan akan dibersihkan untuk menghapus data yang tidak relevan, menyisakan data yang akan digunakan dalam perhitungan. Setelah pembersihan, dilakukan seleksi data untuk menentukan data yang akan digunakan dalam proses perhitungan dengan algoritma k-means clustering. Sebelum tahap data mining, data akan mengalami tahap transformasi. Dalam penelitian ini, data sudah dalam format numerik sehingga tidak memerlukan perubahan. Hasil dari semua tahapan sebelumnya dapat dilihat pada tabel di bawah ini.

Tabel 1. Data Penelitian

No.	Jenis Kelamin	Usia	Pax	Travel Agent
1	1	3	1	9
2	1	3	2	9
3	1	3	1	2
4	1	2	1	8
5	1	3	3	7
6	2	1	3	6
7	2	1	3	9
8	1	3	2	7
9	2	1	3	5
10	2	1	2	6
...
...
398	1	1	3	1

Setelah tahap transformasi data, langkah berikutnya adalah menentukan titik centroid. Titik centroid ditetapkan dengan memilih data yang mewakili tingkat penjualan menengah (C0), terendah (C1), dan tertinggi (C2), sehingga diperoleh titik centroid pada data nomor 150 (C0), 150

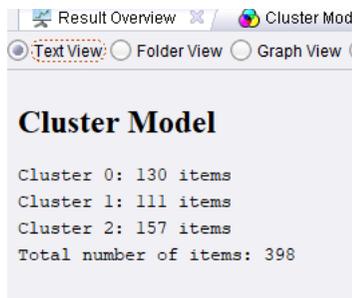
(C1), dan 350 (C2). Berdasarkan pembagian ketiga klaster tersebut, klaster 0 terdiri dari 130 orang, klaster 1 terdiri dari 111 orang, dan klaster 2 terdiri dari 157 orang. Perhitungan dilakukan menggunakan rumus Euclidean Distance dan dilakukan secara iteratif hingga

hasilnya konvergen. Dalam penelitian ini, perhitungan dilakukan sebanyak 6 kali untuk memastikan hasil yang konsisten, dan nilai titik centroid yang telah stabil dapat dilihat pada tabel di bawah ini.

Tabel 2. Titik *Centroid* Iterasi keenam

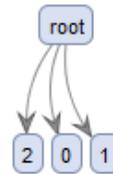
C0	1.43	1.84	1.96	2.02
C1	1.41	1.73	2.02	4.91
C3	1.40	1.86	2.03	8.53

Untuk memverifikasi keabsahan hasil perhitungan manual, digunakan perangkat lunak RapidMiner sebagai alat bantu perhitungan. Hasil dari perhitungan yang dilakukan dengan RapidMiner dapat dilihat di bawah.



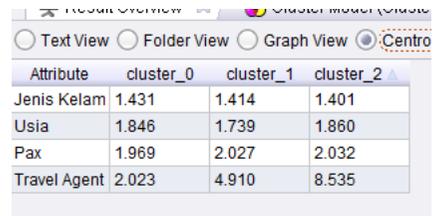
Gambar 2. *Text View*

Perbedaan urutan kluster antara perhitungan manual dan aplikasi disebabkan oleh ketidaksesuaian titik centroid yang digunakan dalam perhitungan manual dengan urutan data. Dalam aplikasi, kluster 0 adalah kluster dengan promosi yang terjual dengan tingkat penjualan menengah, kluster 1 adalah kluster dengan promosi yang terjual tidak laku, dan kluster 2 merupakan kluster promo yang terjual sangat laku.



Gambar 3. *Graph View*

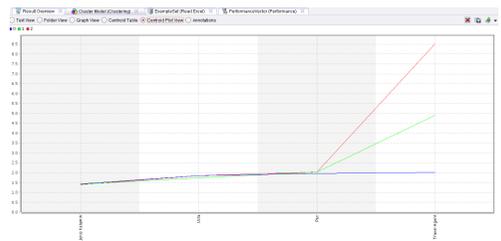
Pada tampilan *Graph View*, ditunjukkan jumlah kluster hasil perhitungan yang berjumlah tiga kluster.



Attribute	cluster_0	cluster_1	cluster_2
Jenis Kelam	1.431	1.414	1.401
Usia	1.846	1.739	1.860
Pax	1.969	2.027	2.032
Travel Agent	2.023	4.910	8.535

Gambar 4. *Centroid Table*

Pada *Centroid Table*, ditampilkan nilai-nilai titik centroid yang telah konsisten dari hasil perhitungan.



Gambar 5. *Centroid Plot View*

Grafik yang ditampilkan di *Centroid Plot View* menunjukkan lokasi titik-titik nilai yang tercantum dalam *Centroid Table*.

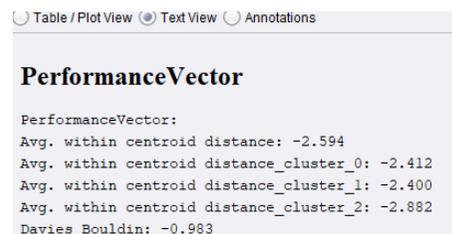


Table / Plot View Text View Annotations

PerformanceVector

PerformanceVector:
Avg. within centroid distance: -2.594
Avg. within centroid distance_cluster_0: -2.412
Avg. within centroid distance_cluster_1: -2.400
Avg. within centroid distance_cluster_2: -2.882
Davies Bouldin: -0.983

Gambar 6. Performance vector

Pada bagian Performance Vector, ditampilkan nilai rata-rata antar centroid serta nilai optimal dari perhitungan, yaitu sebesar 0,983.

SIMPULAN

Setelah penelitian selesai, penulis dapat menarik kesimpulan sebagai berikut:

1. Pengelompokan penjualan menggunakan metode data mining dengan algoritma k-means clustering dapat digunakan untuk memberikan segmentasi pada promosi yang dijual di PT Pulau Bintang Emas. Hasil pengujian menunjukkan pembagian kluster promosi menjadi tiga kategori: menengah, tidak laris, dan sangat laris.

2. Berdasarkan perhitungan dengan algoritma k-means clustering, diperoleh kluster 0 yang terdiri dari 130 data promosi yang terjual menengah, kluster 1 yang berisi 111 data promosi yang terjual tidak laris, dan kluster 2 yang mencakup 157 data promosi yang terjual sangat laris.

DAFTAR PUSTAKA

Shirazi, S., Baziyad, H., & Karimi, H. (2019). An Application-Based Review of Recent Advances of Data Mining in Healthcare. *J Biostat Epidemiol.*, 5(4), 268–278.

Intermedia, Beon. (2020, Maret 17). *Data Mining : Definisi, Fungsi, Metode dan Penerapannya*. Diakses pada

14 Oktober 2020, dari <https://www.jagoanhosting.com/blog/apa-itu-data-mining/>

Prasetyo, E. (2018). *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: CV Andi Offset.

Ridwan, M., Suryono, H., & Sarosa, M. (2022). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *EECCIS*, 7(1), 59-64.

Kovács, L., & Ghous, H. (2020). Efficiency comparison of Python and RapidMiner. *Multidisziplináris Tudományok*, 10(3), 212–220. <https://doi.org/10.35925/j.multi.2020.3.26>

Pratama, M. (2020). *Penerapan Metode Clustering untuk Pengelompokan Potensi Wisata di Kabupaten Sumedang* (Doctoral dissertation, Universitas Komputer Indonesia).

Putri, D. L., & Santoso, H. A. (2019). Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien (Studi Kasus : Puskesmas Kaje).

Purnomo, B. S., & Prasetyaningrum, P. T. (2021). Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means. *Journal of Computer Science and Technology (JCS-TECH)*, 1(1), 27-32.

Rahmawati, E., & Sagirani, T. (2023). Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran. *Jurnal Nasional Teknologi dan Sistem Informasi*, 9(2), 141-149.

Sundari, S., Buatun, R., & Saragih, R. (2021). Clustering Kepuasan Layanan Pengguna Bus Trans Binjai Dengan Metode Cluster Data Mining Studi Kasus Dinas Perhubungan Kota Binjai. Seminar Nasional, 5.

Turban, E., Delen, D., & Sharda, R. (2018). Business intelligence, analytics, and data science: A managerial perspective. Harlow ; Munich: Pearson Prentice Hall.

Widyaningrum, S. R. (2016). Implementasi Data Mining Untuk Pengelompokan Data Siswa Menggunakan Algoritma K-Means Clustering (Studi Kasus: SMKN 1 Kediri). Artikel Skripsi Universitas Nusantara PGRI Kediri.

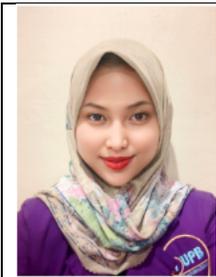
Ridwan, M., Suryono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. EECCIS, 7(1), 59-64.

Muningsih, E., & Kiswati, S. (2017). Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang. Jurnal Bianglala Informatika, 10-17.

Pattipeilohy, R. L., & Pakereng, M. A. I. (2023). Penerapan K-Means Clustering Pada Data Mahasiswa Fakultas Interdisiplin Program Studi D4 Destinasi Pariwisata Untuk Menentukan Strategi Promosi. J-SAKTI (Jurnal Sains Komputer dan Informatika), 7(1)

Rumahorbo, N. R. S. S., & Arnomo, S. A. (2020). Implementasi Data Mining Untuk Market Basket Analysis Menggunakan Asosiasi Rules Pada Minimarket 212 Mart Batam. Journal of Information System Research (JOSH), 1(4), 305-310.

Arnomo, S. A., & Yulia, Y. (2021). Clustering the potential bandwidth upgrade of FTTH broadband subscribers. ILKOM Jurnal Ilmiah, 13(1), 51-57.

	<p>Biodata Penulis Pertama, Putri Aprilianti Ali merupakan mahasiswa Prodi Sistem Informasi Universitas Putera Batam</p>
	<p>Biodata Penulis kedua, Sasa Ani Arnomo Sunargo, S.Kom., M.Si., Ph.D., merupakan Dosen Prodi Sistem Informasi Universitas Putera Batam. Penulis banyak berkecimpung di bidang sistem informasi</p>