

PENERAPAN METODE K-MEANS UNTUK PENGELOMPOKAN DATA KUNJUNGAN WISATA PADA DINAS KEBUDAYAAN DAN PARIWISATA PROVINSI JAMBI

Lucy Simorangkir¹, Ezrifal Sany², Muhammad Feraldi. N³

¹²³Program Studi Teknik Informatika, Universitas Nurdin Hamzah, Jambi

E-mail: ¹lucy.simorangkir@yahoo.co.id, ²ezrifalsany@gmail.com, ³feraldi1703@gmail.com

Abstract - Tourism is a vital economic sector in Indonesia, including in Jambi City, contributing significantly to regional income and job creation. However, the management of tourist visitation data at the Department of Culture and Tourism of Jambi Province is still conducted manually, making it challenging to obtain accurate information on tourist trends and preferences. This issue stems from the lack of modern information technology and the limited application of data analysis methods such as clustering, which hampers the effectiveness of strategic decision-making. This study proposes the application of the K-Means clustering method to group tourism data based on characteristics such as tourist origin, purpose of visit, and type of destination. The Orange data analysis tool is utilized to expedite the analysis and visualization process, providing in-depth and strategic insights. By integrating Orange and Python, data analysis becomes more flexible and adaptable to specific needs. This approach is expected to improve the management of tourist visitation data, facilitate the identification of tourist patterns, and support more effective tourism development in Jambi Province.

Keywords: Tourism, K-Means, clustering, Python, Orange.

I. PENDAHULUAN

1.1. Latar Belakang

Pariwisata menjadi sektor ekonomi vital bagi Indonesia, termasuk Kota Jambi, yang memberikan kontribusi signifikan terhadap pendapatan daerah dan peningkatan lapangan kerja (P. Sari dan T. Y. Syah, 2021). Dalam hal ini, pengelolaan data kunjungan wisata menjadi krusial untuk memahami tren wisatawan dan merumuskan strategi pengembangan pariwisata yang efektif. Dinas Kebudayaan dan Pariwisata Provinsi Jambi merupakan instansi pemerintah yang mengatur dan menjalankan keperluan pemerintahan tentang kebudayaan dan kepariwisataan Provinsi Jambi yang terletak di Kecamatan Kota Baru Kota Jambi.

Dinas Kebudayaan dan Pariwisata Provinsi Jambi mempunyai tugas membantu Gubernur dalam rangka melaksanakan sebagian urusan pemerintahan bidang kebudayaan dan pariwisata berdasarkan azas otonomi dan tugas pembantuan sesuai ketentuan peraturan perundang-undangan yang dipimpin oleh seorang Kepala Dinas. Dinas Kebudayaan dan Pariwisata Provinsi Jambi menghadapi tantangan dalam pengelolaan data kunjungan wisata yang masih dilakukan secara manual, menyebabkan data seringkali terfragmentasi dan sulit dianalisis secara menyeluruh. Hal ini menghambat kemampuan untuk mendapatkan informasi yang akurat tentang profil wisatawan dan preferensi mereka terhadap destinasi pariwisata lokal.

Proses pengelolaan data kunjungan wisata yang masih tergantung pada sistem manual di Dinas Kebudayaan dan Pariwisata Provinsi Jambi menyulitkan upaya untuk mengidentifikasi pola kunjungan wisatawan secara efisien. Kurangnya

integrasi teknologi informasi yang canggih juga menjadi kendala utama dalam menyajikan data kunjungan wisata yang dapat dipercaya dan digunakan untuk pengambilan keputusan strategis.

Kendala dalam pengelolaan data kunjungan wisata mencakup keterbatasan akses terhadap teknologi informasi yang modern dan kurangnya pemahaman dalam menerapkan metode analisis data seperti *clustering* untuk mengoptimalkan penggunaan data yang ada. Dalam hal ini, diperlukan pendekatan baru yang lebih terstruktur dan teknologi yang lebih maju untuk meningkatkan efektivitas pengelolaan data kunjungan wisata.

Untuk meminimalisir kendala dalam pengelolaan data kunjungan wisata yang optimal, maka perlu dibangun suatu alat bantu *data mining* dengan metode K-Means. Penerapan metode K-means *clustering* diharapkan dapat mengatasi tantangan dalam pengelolaan data kunjungan wisata di Dinas Kebudayaan dan Pariwisata Provinsi Jambi. Metode ini memungkinkan pengelompokan data berdasarkan karakteristik seperti asal wisatawan, tujuan kunjungan, dan jenis destinasi yang dikunjungi, sehingga memberikan wawasan yang lebih mendalam untuk pengambilan keputusan strategis. Aplikasi yang digunakan dalam membantu pengolahan data tersebut adalah Orange.

Penggunaan alat analisis data seperti Orange memungkinkan percepatan proses analisis dan visualisasi data kunjungan wisata, sehingga membantu dalam pengambilan keputusan strategis yang lebih tepat. Dengan fitur-fitur yang mudah digunakan, Orange mampu mengelompokkan data secara efisien menggunakan metode K-Means. Selain itu, integrasinya dengan bahasa pemrograman Python memberikan fleksibilitas lebih dalam analisis

yang sesuai dengan kebutuhan spesifik. memodifikasi dan mengembangkan

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah: “Bagaimana Penerapan Metode K-Means untuk Pengelompokan Data Kunjungan Wisata pada Dinas Kebudayaan dan Pariwisata Provinsi Jambi?”.

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Untuk membangun Penerapan Metode K-means untuk Pengelompokan Data Kunjungan Wisata pada Dinas Kebudayaan dan Pariwisata Provinsi Jambi.
2. Untuk memudahkan Dinas Kebudayaan Provinsi Jambi dalam melakukan pengolahan data kunjungan wisata.

1.4 Manfaat Penelitian

Adapun manfaat penelitian ini adalah dapat Meningkatkan perancangan strategi yang lebih tepat dalam mempromosikan destinasi pariwisata lokal, serta Meningkatkan efisiensi dalam analisis data kunjungan wisata.

1.5 Metode Penelitian

Adapun metode penelitian yang digunakan adalah *Waterfall*. Penerapan metode *Waterfall* mulai dari tahapan analisis, kebutuhan sistem, perancangan, implementasi, dan pengujian dapat menghasilkan rancangan sistem aplikasi yang dibutuhkan oleh pihak Dinas Kebudayaan dan Pariwisata Provinsi Jambi.

Berikut ini Gambar 1 menunjukkan tahapan penelitian mengikuti alur metode *Waterfall*.



Gambar 1 Tahapan Pengembangan Sistem Aplikasi

II. TINJAUAN PUSTAKA

2.1 K-Means

K-Means adalah metode matematika yang digunakan untuk membagi data menjadi *cluster* atau kelompok menggunakan pusat atau rata-rata (A. F.

Al Shammari, 2024). Metode ini bekerja dengan cara membagi dataset kedalam sejumlah *cluster* yang telah ditentukan sebelumnya, di mana setiap *cluster* terdiri dari data yang memiliki atribut serupa. Analisis *cluster* digunakan untuk mengelompokkan data berdasarkan kesamaan atau kedekatan data dalam suatu kelompok ke kelompok lain. Ketika *item* dikelompokkan, kedekatan dapat ditunjukkan dengan jarak atau berdasarkan koefisien korelasi dan ukuran asosiasi serupa (Savitri, N., Pranata, R., Clara, A. N. M., and Rahajeng, O. S, 2022). Metode K-Means *Clustering* 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 (Z. N. Khalisah, 2022).

Sehingga dari berbagai pengertian atau definisi para ahli dapat disimpulkan bahwa metode K-Means adalah metode analisis *cluster* yang membagi data ke dalam sejumlah kelompok atau berdasarkan kesamaan atribut, di mana setiap data dalam satu kelompok memiliki karakteristik yang mirip satu sama lain dan berbeda dari kelompok lainnya, dengan pengelompokan ini dilakukan menggunakan pusat atau rata-rata untuk menentukan kedekatan antar data.

2.2 Clustering

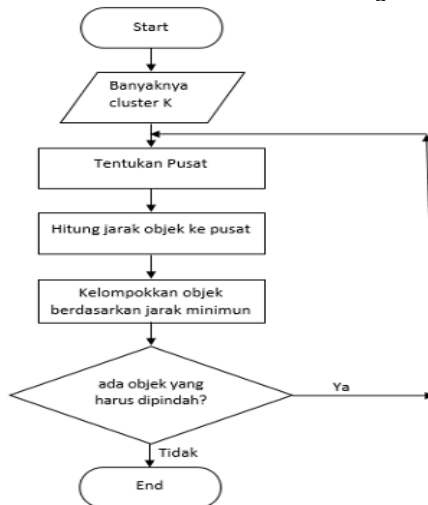
Clustering adalah alat yang berguna dalam ilmu data. Ini adalah metode untuk menemukan struktur *cluster* dalam kumpulan data yang ditandai dengan kesamaan terbesar dalam klaster yang sama dan ketidaksamaan terbesar di antara *cluster* yang berbeda (K. P. Sinaga and M. -S. Yang, 2020). Data *Clustering* adalah pemrosesan data mentah untuk menemukan klaster atau kelompok data yang serupa. Dalam setiap *cluster*, anggota-anggotanya memiliki kesamaan tipe data. Prinsip dari data *clustering* adalah menemukan nilai dari hasil yang serupa dan mengaitkan setiap anggota ke kelompok yang sama dengan anggota lain yang memiliki hasil yang serupa atau sama (S. Yadav and S. Sharma, 2024). *Clustering* data, juga disebut sebagai klasifikasi *Unsupervised*, didefinisikan sebagai teknik pembuatan kelompok objek, sehingga objek dalam satu klaster sangat identik dan objek dalam klaster yang berbeda relatif berbeda. Tujuan utama dari pengelompokan adalah untuk menemukan sekumpulan pola, titik, atau objek dari pengelompokan alami (J. Oyelade *et al*, 2019).

Sehingga dari berbagai pengertian atau definisi para ahli dapat disimpulkan bahwa *Clustering* dalam ilmu data adalah metode untuk mengelompokkan data berdasarkan kesamaan antar anggota dalam kelompok yang sama dan perbedaan dengan kelompok lainnya. Teknik ini, yang dikenal juga sebagai klasifikasi tidak diawasi (*unsupervised*), bertujuan untuk menemukan pola

mining data. Pembersihan data berisi beberapa kegiatan yang bertujuan untuk melakukan pengenalan dan perbaikan pada data yang akan diteliti. Perbaikan pada data perlu dilakukan karena data mentah cenderung tidak siap untuk di-mining.

3.2 Flowchart K-Means Clustering

Pada *Flowchart* proses ini dijelaskan bagaimana proses dari pengolahan data kunjungan wisata dengan menggunakan metode K-Means clustering untuk menghasilkan Pengelompokan kunjungan wisata pada tempat wisata. Untuk lebih jelasnya *Flowchart* proses ini dapat dilihat pada gambar 3 *Flowchart K-Means clustering* berikut ini



Gambar 3 Flowchart K-Means Clustering

3.3 Implementasi

Implementasi tools aplikasi program python untuk pengelompokan data kunjungan wisata menggunakan metode K-Means clustering.

1. Implementasi Antarmuka Input

Implementasi antarmuka input pada model program ini adalah beberapa input yang dilakukan pada program, seperti input *library*, input data training.

a. Input Library

Untuk mengambil dan memproses input dari pengguna atau sumber lain, menjalankan operasi tertentu, baik melalui konsol, file, atau argumen baris perintah, dapat dilihat pada gambar 4 berikut ini :

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
```

Gambar 4 Input Library

a. Input Data Training

```
[2]: df= pd.read_csv('olahdata1.csv', delimiter=';')
df.head()
```

Gambar 5 Input Data Training

b. Input Algoritma K-Means Clustering

Untuk dilakukan dalam tools Python agar data training bisa di proses dan dilakukan perhitungan, dapat dilihat pada gambar 6 berikut ini :

```
[4]: #Terapkan K-Means dengan 3 cluster
km = KMeans(n_clusters=3, random_state=42)
df['CLUSTER'] = km.fit_predict(x)
df['CLUSTER']

[5]: df['CLUSTER']
df.head()
```

Gambar 6 Input Algoritma K-Means Clustering

Pada kode input algoritma K-Means clustering ini dilakukan pemrosesan data training yaitu proses pengelompokan dan perhitungan probabilitas setiap kelas, yang dimana proses ini dilakukan agar pada saat pengujian data testing baru bisa dilakukan pengelompokan dengan algoritma K-Means clustering.

c. Tampilan Output Membaca Data Training

Output membaca data training dalam model ini mengacu pada proses di mana data training diimport dan disiapkan untuk digunakan dalam model algoritma K-Means clustering. Data ini disimpan dalam format seperti CSV atau Excel. Output dari langkah ini adalah sekumpulan data yang telah berhasil dimuat dan diatur, mencakup informasi data kunjungan wisata yang diperlukan untuk melatih model. Output data training dapat dilihat pada gambar 7 berikut ini :

```
[5]: NAMA OBIEK_WISATA JANUARI FEBRUARI MARET APRIL MEI JUNI JULI AGUSTUS SEPTEMBER OKTOBER NOVEMBER DESEMBER CLUSTER
0 Taman Tingo Rajo 3500 3000 4900 6500 3500 3900 16000 4200 5300 6350 6200 2
1 Danau Sigombak 219 279 268 489 321 290 335 469 495 438 396 726 0
2 Jembatan Ilham 813 3685 7500 765 3858 2110 3875 4250 3287 2987 3124 4018 2
3 Makam Haji Ahmad 87 113 430 362 412 350 434 1500 432 789 730 763 0
4 Air Panas Geragai 70 73 37 120 59 37 35 40 55 37 87 91 0
```

Gambar 7 Input Data Training

d. Tampilan Output Visualisasi Data

Visualisasi ini menggambarkan pengelompokan objek wisata berdasarkan hasil clustering menggunakan metode K-Means dengan tiga kategori: "Rendah" (lingkaran hijau), "Sedang" (segitiga biru), dan "Tinggi" (kotak merah). Sebagian besar objek wisata termasuk dalam kategori "Rendah," sementara hanya satu objek wisata berada di kategori "Tinggi," dan beberapa lainnya masuk dalam kategori "Sedang." Visualisasi ini menunjukkan distribusi tingkat popularitas atau kinerja objek wisata, dengan mayoritas memiliki performa rendah dibandingkan kategori lainnya. Hasil output visualisasi ini dapat dilihat pada gambar 8 berikut ini :



Gambar 8 Tampilan Output Hasil Visualisasi

IV. PENUTUP

4.1 Kesimpulan

Berdasarkan hasil pembahasan di atas, maka dapat disimpulkan sebagai berikut:

1. Proses klasifikasi penyakit diabetes dilakukan menggunakan metode Naïve Bayes, yang diimplementasikan melalui aplikasi *Microsoft Excel* dan, sebagai perbandingan, menggunakan aplikasi *Python*.
2. Data training yang digunakan sebagai sampel memiliki pengaruh yang signifikan terhadap akurasi model program.
3. Hasil evaluasi dari klasifikasi ini dapat digunakan sebagai bahan pertimbangan dalam proses diagnosis awal penyakit bagi pasien dengan gejala-gejala yang mengarah pada diabetes mellitus, sehingga memudahkan pihak Klinik Pratama dan Apotek U.K Jambi.

4.2 Saran

Adapun saran-saran yang dapat dikemukakan adalah:

1. Pengembangan dataset disarankan untuk penelitian selanjutnya agar memperluas dataset yang digunakan dengan menambah jumlah sampel data pasien dan variabel yang relevan agar dapat meningkatkan akurasi model klasifikasi.
2. Diharapkan untuk dilakukan perbandingan terhadap metode klasifikasi lain untuk mengetahui atau menemukan metode paling efektif dalam klasifikasi penyakit diabetes mellitus sehingga dapat meningkatkan akurasi dari peneliti sebelumnya.

DAFTAR REFERENSI

- P. Sari dan T. Y. Syah, "Pengelolaan Data Kunjungan Wisata Berbasis Clustering untuk Meningkatkan Promosi Destinasi Wisata," *J. Ilmiah Teknol. Inform.*, vol. 15, no. 2, pp. 78-83, 2021.
- A. F. AlShammari, "Implementation of Clustering using K-Means in Python," *International Journal of Computer Applications*, vol. 186, no. 40, pp. 12-17, 2024.
- Savitri, N., Pranata, R., Clara, A. N. M., and Rahajeng, O. S., "Pengelompokan

Kunjungan Wisata Kabupaten Kulon Progo Tahun 2019 Menggunakan K-Means Clustering," *Just IT: Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, vol. 12, no. 1, pp. 1-8, 2022.

- Z. N. Khalisah, "Study Comparison K-Means Clustering Dengan Algoritma Hierarchical Clustering: AHC, K-Means Clustering, Study Comparison," *Seminar Nasional Teknologi & Sains*, vol. 1, no. 1, 2022.
- K. P. Sinaga and M. -S. Yang, "Unsupervised K-Means Clustering Algorithm," in *IEEE Access*, vol. 8, pp. 80716-80727, 2020.
- S. Yadav and S. Sharma, "Study of existing methods and techniques of K-means clustering," *Educational Administration: Theory and Practice*, vol. 30, 2024.
- J. Oyelade et al., "Data clustering: Algorithms and its applications," in *2019 19th International Conference on Computational Science and Its Applications (ICCSA)*, 2019, pp. 71-81.
- J. Unpingco, "Python Programming for Data Analysis," *Springer Nature*, 2021.
- K. R. V. Kothapalli, M. A. Mohammed, R. Mohammed, and P. Pasam, "Python for Beginners: A Comprehensive Guide to Learning Python Programming," *Warta Saya*, 2024.
- A. Rawat, "A Review on Python Programming," *International Journal of Research in Engineering, Science and Management*, vol. 3, no. 12, pp. 8-11, 2020.
- Popchev and D. Orozova, "Algorithms for machine learning with Orange system," *International Journal of Online & Biomedical Engineering*, vol. 19, no. 4, 2023.
- J. Demšar et al., "Orange: Data mining toolbox in Python," *Journal of Machine Learning Research*, vol. 14, pp. 2349-2353, 2013.
- A. Ishak et al., "Orange software usage in data mining classification method on the dataset lenses," *IOP Conference Series: Materials Science and Engineering*, vol. 1003, no. 1, IOP Publishing, 2020.

IDENTITAS PENULIS

Nama : Lucy Simorangkir, M.Kom
 NIDN/NIK : 1028097801 / 78.11.2.0037
 TTL : Kuala Tungkal / 28 September 1978
 Golongan/Pangkat : III D
 Jabatan Fungsional : Lektor
 Alamat Rumah : Jl. Sermak Ishak Ahmad No. 24 RT 07 Mayang Jambi
 Telp. : 081366009242
 Email : lucy.simorangkir@yahoo.co.id

Nama : Ezrifal Sany, ST., M.Kom
NIDN/NIK : 1001068103 / 81.10.1.0029
TTL : Jambi / 01 Juni 1981
Golongan/Pangkat : III C
Jabatan Fungsional : Lektor
Alamat Rumah : Perumahan Mutiara Hijau
Blok G1
Telp. : 081277444644
Email : ezrifalsany@gmail.com

Nama : Muhammad Feraldi. N
NIDN / NIK : -
TTL : Bangko, 17 Februari 2003
Golongan/Pangkat : -
Jabatan Fungsional : -
Alamat Rumah : Jl. Imam Bonjol RT 37/RW
05 Pematang Kandis Bangko
Telp. : 0822279988114
Email : feraldi1703
@gmail.com