



Artikel

Penerapan Metode *K-Means* Clustering dalam Mengelompokkan Jumlah Wisatawan Asing di Jawa Barat

Ana Neva¹, Fatihanursari Dikananda²

^{1,2}STMIK IKMI Cirebon, Manajemen Informatika, Cirebon, Indonesia

SUBMISSION TRACK

Received: Jan 22, 2023
Final Revision: March 12, 2023
Available Online: March 24, 2023

KEYWORD

Pariwisata, wisatawan, *K-Means*, *Clustering*

KORESPONDENSI

Phone: 083147734518
E-mail: ananeva702@gmail.com

A B S T R A K

Provinsi Jawa Barat merupakan provinsi yang memiliki daya pesona wisata alam yang cukup diminati oleh wisatawan lokal maupun warga wisatawan asing. Dibeberapa tempat wisata alam di Jawa Barat, jumlah wisatawan relatif mengalami kenaikan. Pemerintah provinsi belum melakukan analisa pemetaan wisatawan asing terhadap kunjungan wisata di daerah Jawa Barat. Data wisatawan asing dapat menjadi acuan pemerintah sebagai pemetaan kunjungan wisatawan. Tahapan penelitian menggunakan *knowledge discovery in database*. Pengolahan data menggunakan metode *K-Means*. *K-Means* merupakan metode data mining yang memberikan deskripsi kelompok sebuah item. *Tools* yang digunakan dalam pengolahan data adalah aplikasi *RapidMiner*. Tujuan dari penelitian ini untuk mengelompokkan jumlah kedatangan wisatawan asing di tempat wisata alam di Jawa Barat. Hasil pengelompokan didapat 3 cluster yaitu, cluster 0 (rendah) sebanyak 445 tempat wisata alam, cluster 2 (sedang) sebanyak 880 tempat wisata, dan cluster 1 (tinggi) sebanyak 880 tempat wisata. Tempat wisata alam yang termasuk dalam cluster rendah dapat dijadikan kontribusi untuk Pemerintah Provinsi Jawa Barat dalam hal peningkatan fasilitas yang ada di tempat wisata, agar wisatawan asing yang berkunjung akan semakin meningkat dimasa yang akan datang.

PENDAHULUAN

Indonesia adalah negara yang memiliki keindahan alam dan sektor wisata yang cukup banyak. Sektor pariwisata di Indonesia merupakan salah satu penyumbang devisa negara yang cukup besar. Pada tahun 2019 sebanyak 16,1 juta wisatawan asing datang ke Indonesia [1]. Provinsi Jawa Barat merupakan salah satu provinsi dengan daya tarik wisata alam yang cukup diminati baik wisatawan lokal atau wisatawan asing. Jumlah wisatawan asing yang berkunjung ke tempat wisata alam di Jawa Barat ini relatif meningkat setiap tahunnya. Untuk mempertahankan atau meningkatkan jumlah kunjungan diperlukan peningkatan daya tarik dan fasilitas tempat wisata.

Dalam melakukan penelitian klusterisasi wisatawan asing di Jawa Barat tahun 2016-2020 penulis menggunakan metode *knowledge discovery in database* (KDD). Algoritma yang digunakan adalah algoritma *K-Means*. Berdasarkan penelitian terdahulu yang diteliti oleh Edy Satria, Heru Satria Tambunan, Ilham Syahputra Saragih, Irfan Sudahri Damanik, dan Fany Than Ervina Sitanggang (2019), topik yang diangkat dalam penelitian ini adalah klusterisasi jumlah kunjungan wisatawan asing menurut kebangsaan dan bulan kedatangannya. [2].

Dilakukannya penelitian ini dikarenakan pemerintah daerah belum melakukan analisa pemetaan wisatawan asing terhadap kunjungan wisata alam di Jawa Barat. Data pengelompokan wisatawan asing dapat menjadi acuan pemerintah sebagai pemetaan kunjungan wisatawan. Dengan dilakukannya penelitian ini dapat diketahui tempat wisata mana yang masih perlu banyak dilakukannya peningkatan, baik dalam hal fasilitas maupun daya tarik wisatanya.

I. METODE

1.1. Data Mining

Data mining berisi pencarian trend atau pola yang diinginkan dalam database yang besar untuk membantu pengambilan keputusan di waktu yang akan datang [3].

Secara garis besar, data mining dikelompokkan menjadi 2 kategori utama, yaitu :

1. *Descriptive* mining yaitu proses untuk mencari karakteristik yang penting dari data dalam suatu database.
2. *Predictive* yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variable lain di masa depan [4].

Data mining merupakan bagian dari proses *knowledge discovery in database*, meliputi algoritma yang mengeksplor data, membangun model dan menemukan pola yang belum diketahui. KDD bersifat otomatis, dan dapat didefinisikan sebagai pengorganisasian proses untuk mengidentifikasi yang benar, berguna dan penemuan pola dari kumpulan data yang kompleks dan besar[5]. Ada 5 tahapan dalam KDD yaitu:

1. *Data selection*
Dalam proses *data selection* terjadi pemilihan data yang relevan dan dapat dilakukan analisis dari data operasional.
2. *Processing*
Pada tahap *processing* terjadi proses pembersihan data bertujuan untuk menghilangkan data *missing value* atau tidak memiliki nilai yang mengakibatkan akurasi pada proses *data mining* menjadi tidak maksimal.
3. *Transformation*
Proses transformasi yaitu merubah data kedalam bentuk format tertentu sehingga data tersebut sesuai untuk proses data mining. Transformasi data dilakukan dengan memberikan inialisasi terhadap data yang memiliki nilai nominal menjadi bernilai numerik atau angka [6].
4. *Data mining*
Data mining merupakan proses mencari bentuk atau informasi yang menarik menggunakan algoritma tertentu. Pada tahap data mining dilakukan penerapan algoritma atau metode pencarian pengetahuan. Tahap adalah langkah penting dimana teknik kecerdasan

diterapkan untuk mengekstrak pola informasi yang berpotensi berguna dari data yang sudah dipilih [7].

5. Interpretasi/evaluasi

Interpretasi atau evaluasi yaitu menyimpulkan pola-pola dari hasil data mining. Menampilkan pola informasi yang dihasilkan dari tahap data mining, interpretasi ini memaparkan hasil sehingga menjadi informasi yang mudah dipahami oleh pemilik data [4].

1.2. Clustering

Clustering adalah salah satu metode dari *data mining* yang mempunyai sifat *unsupervised* atau tidak memiliki arahan. Ada dua jenis data pengelompokan yang dapat diaplikasikan dalam proses pengelompokan yaitu, data hirarki dan data *non* hirarki [8]. *Clustering* berbeda dengan klasifikasi yaitu tidak adanya variable target yaitu tidak digunakan untuk memprediksi nilai dari target, melainkan digunakan untuk melakukan proses pembagian atau pengelompokan terhadap data menjadi kelompok – kelompok yang memiliki kemiripan [9]

1.3. K-Means

K-Means adalah metode yang dikembangkan oleh Mac Queen pada tahun 1967 [8]. *K-Means* merupakan salah satu algoritma dalam pengelompokan. *K-Means* berfungsi sebagai metode menganalisa data atau metode data mining yang melakukan proses pemodelan *unsupervised* dan merupakan salah satu metode yang melakukan pengelompokan data menjadi beberapa kelompok [10]. Adapun tahapan-tahapan pada algoritma *K-Means* yaitu sebagai berikut :

Tentukan jumlah kelompok dan tetapkan pusat kelompok secara acak.

1. Hitung jarak setiap data ke pusat kelompok menggunakan jarak *euclidean*

$$d(i,k) = \sqrt{\sum_i^m (C_{ij} - C_{kj})^2}$$

2. Kelompokkan data ke dalam kelompok dengan jarak yang paling pendek.

3. Hitung pusat kelompok baru dengan menggunakan persamaan

$$C_{kj} = \frac{\sum_k^i x_{ij}}{p}$$

Dengan : $X_{ij \in}$ Kelompok ke k p = banyaknya anggota kelompok ke – k. Ulangi langkah dua sampai empat, sampai sudah tidak ada data yang berpindah dari satu kelompok ke kelompok yang lain [11].

1.4. Wisata

Wisata dapat dikategorikan sebagai kebutuhan manusia untuk melepas lelah, pikiran penat, stress dan masalah. Jenis wisata berkembang berdasarkan banyak hal seperti kriteria individu, motivasi, *trend* dan kebutuhannya [12]. Saat ini destinasi wisata dapat dibedakan menjadi 3 (tiga) yaitu: destinasi wisata berbasis ekowisata atau alam, destinasi wisata berbasis sosial budaya dan destinasi wisata minat khusus [13]. Terdapat beberapa faktor yang mempengaruhi tingkat kedatangan wisatawan di suatu tempat wisata. Salah satu faktor yang mempengaruhi minat berkunjung wisatawan adalah daya tarik wisata. Jika daya tarik meningkat, maka minat wisatawan untuk berkunjung juga semakin meningkat [14]. Selain daya tarik wisata, tentunya fasilitas juga mempengaruhi jumlah kedatangan pengunjung. Kualitas Pelayanan dan Fasilitas berpengaruh positif dan signifikan terhadap kepuasan wisatawan [15].

1.5. Wisatawan

Wisatawan adalah orang yang melakukan perjalanan dari tempat kediamannya dan tidak menetap ditempat yang didatanginya, atau hanya untuk sementara waktu yang ditempat tinggal didatanginya. Mereka yang dianggap sebagai wisatawan adalah orang yang melakukan kesenangan. Wisatawan adalah orang-orang yang melakukan kegiatan wisata [2]. Ada 2 jenis wisatawan yaitu:

1. Wisatawan asing
wisatawan asing, menurut Badan Pusat Statistik adalah setiap pengunjung yang datang ke suatu

negara di luar tempat tinggalnya, dikarenakan memiliki beberapa keperluan tanpa bermaksud memperoleh penghasilan di tempat yang dikunjungi dan lamanya kunjungan tersebut tidak lebih dari 12 (dua belas) bulan [1].

2. Wisatawan lokal

Wisatawan lokal adalah seorang warga negara yang melakukan kunjungan wisata dalam wilayah negaranya sendiri dengan tidak melewati batas negaranya sendiri. [16].

1.6. Sumber Data

Data yang digunakan dalam penelitian tugas akhir ini bersumber dari *website* <https://opendata.jabarprov.go.id/>. Data memuat 8 *field* atau atribut yaitu, id, kode Provinsi, nama provinsi, pengelola kawasan, jenis wisatawan, jumlah, satuan, dan tahun. Data yang digunakan yaitu data pada tahun 2016-2020, dengan jumlah record data 1760 data.

Tabel 1 Sampel Data Wisatawan di Jawa Barat

| Pengelola Kawasan | Jenis Wisatawan | jumlah | tahun |
|---------------------------|-----------------|---------|-------|
| Balai Besar Tn.Gn. | Wisatawan | 161.154 | 2016 |
| Gedepangran go | Lokal | | |
| Balai Besar Tn.Gn | Wisatawan | 1.030 | 2016 |
| Gedepangran go | Asing | | |
| Tn.Gn. Gede Pangrango | Wisatawan | 0 | 2016 |
| Tn.Gn. Gede Pangrango | Lokal | | |
| Balai Tn Gn Halimun-Salak | Wisatawan | 0 | 2016 |
| Balai Tn Gn Halimun-Salak | Asing | 131.073 | 2016 |
| Tn.Gn.Halimun Salak | Wisatawan | 208 | 2016 |
| Tn.Gn.Halimun Salak | Lokal | | |
| Tn.Gn.Halimun Salak | Asing | 0 | 2016 |
| Tn.Gn.Halimun Salak | Lokal | | |

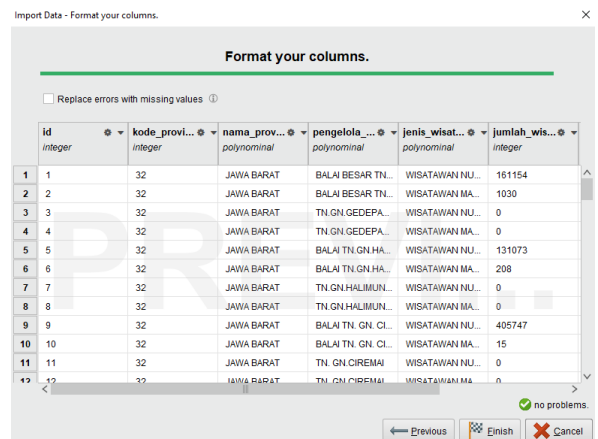
| Pengelola Kawasan | Jenis Wisatawan | jumlah | tahun |
|----------------------|-----------------|---------|-------|
| Tn.Gn.Halimun Salak | Wisatawan | 0 | 2016 |
| Balai Tn.Gn. Ciremai | Wisatawan | 405.747 | 2016 |
| Balai Tn.Gn. Ciremai | Lokal | | |
| Balai Tn.Gn. Ciremai | Asing | 15 | 2016 |
| ... | ... | ... | ... |
| Batu Bubut | Wisatawan | 0 | 2020 |
| Batu Bubut | Lokal | | |
| Batu Bubut | Asing | 0 | 2020 |

Sumber <https://opendata.jabarprov.go.id/>

II. HASIL

2.1 Data Selection

Terdapat 2 tahapan pada proses *data selection*. Pertama melakukan *import* data menggunakan *read excell* pada *RapidMiner*. Operator *read excell* ini berfungsi untuk membaca data kunjungan wisata di Jawa Barat berdasarkan jenis wisatawannya. Pada tahap ini menggunakan data murni atau tidak melewati proses editing seperti, menghapus data dan menambahkan data.



Gambar 1 Preview Data Read Excel

Pada tahap kedua yaitu memasukan operator *select attributes*. Pada tahap ini dilakukan pembersihan atribut atau menyeleksi atribut yang tidak terpakai dalam proses data mining. Dari total 12 atribut dipilih 4 atribut, yaitu pengelola kawasan, jenis wisatawan, jumlah, dan tahun.

Gambar 2 Hasil Proses Select Atribut

2.2 Data Cleaning

Proses *data cleaning* bertujuan untuk membersihkan data *missing value* menggunakan operator *replace missing values*. Pada dataset yang digunakan tidak terdapat data missing. Dari total *record* 1760 data berjumlah tetap dan tidak berkurang.

Gambar 3 Hasil Proses Data Cleaning

2.3 Transformasi Data

Terdapat 2 tahapan dalam proses transformasi data. Tahap pertama yaitu mengubah bentuk data nominal menjadi *numerical* agar dapat diproses oleh algoritma *K-Means*, menggunakan operator *nominal to numerical*.

Gambar 4 Hasil Proses Nominal to Numerical

Tahap kedua yaitu melakukan normalisasi menggunakan operator *normalize*. Tujuan dilakukannya normalisasi adalah untuk memperkecil *range* data jumlah wisatawan agar proses *data mining* lebih optimal. Menggunakan model transformasi *range transformation*, yang merupakan metode dimana nilai jarak dapat ditentukan. Pada penelitian ini menggunakan *range* nilai *min* = 0.0 dan nilai *max* = 1.0.

Gambar 5 Hasil Proses Normalize

2.4. Data Maining

Pada tahapan ini dilakukan pemodelan data menggunakan data hasil transformasi. Metode yang digunakan dalam proses *data mining* ini adalah *clustering*, dengan menerapkan algoritma *K-Means*. Pada proses *data mining* ini dimana diketahui nilai *k* = 3, *max run* = 10 kali putaran, jenis *measure typenya* adalah *Euclidean Distance* dan *max optimization steps* berjumlah 100.

```

Cluster Model

Cluster 0: 445 items
Cluster 1: 435 items
Cluster 2: 880 items
Total number of items: 1760
    
```

Gambar 6 Cluster Model Data Wisatawan

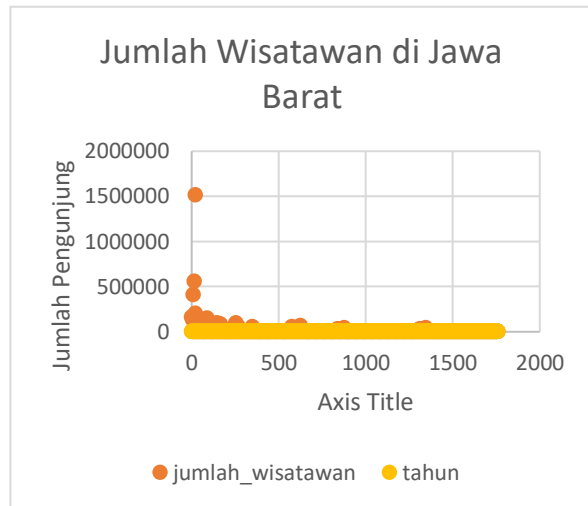
III. PEMBAHASAN

Untuk menghasilkan nilai k yang paling optimum proses akan dievaluasi menggunakan model *performance* untuk menghitung *davies bouldin index*. Nilai k yang optimum adalah nilai k yang *Davies Bouldin Index* nya mendekati 0. Peneliti melakukan 10 kali percobaan dengan nilai k yang berbeda dan diketahui nilai DBI yang paling optimal yaitu 0.684 dengan menggunakan $k=3$. Hasil percobaan dapat dilihat pada tabel 1.

Tabel 2 Hasil DBI

| Cluster (k) | Davies Bouldin Index |
|-------------|----------------------|
| 2 | 0.849 |
| 3 | 0.684 |
| 4 | 0.703 |
| 5 | 0.746 |
| 6 | 0.774 |
| 7 | 0.793 |
| 8 | 0.830 |
| 9 | 0.856 |
| 10 | 0.894 |

Pada hasil percobaan didapatkan nilai k optimum = 3, dimana dihasilkan 3 *cluster* yaitu, *cluster 0*, *cluster 1*, dan *cluster 2*. *Cluster 0* merupakan *cluster* rendah dengan anggota sebanyak 445 tempat wisata. *Cluster 2* merupakan *cluster* sedang, dengan banyak anggota 880 tempat wisata. *Cluster 1* merupakan *cluster* tinggi dengan anggota sebanyak 435 tempat wisata.



Gambar 7 Scater Plot Kunjungan Wisatawan di Jawa Barat

Pada *cluster 0* didominasi oleh wisatawan lokal sebanyak 194.330 wisatawan, pada rentan tahun 2019 sampai dengan 2020. Pada *cluster 2* didominasi oleh wisatawan mancanegara sebanyak 4.797.802 wisatawan, pada rentan tahun 2016-2020. Pada *cluster 1* didominasi wisatawan lokal sebanyak 4.958.956 wisatawan, pada rentan tahun 2016-2019.

IV. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Penggunaan teknik *data mining* menggunakan algoritma *K-Means* dapat diterapkan dalam mengelompokkan kunjungan wisata berdasarkan jenis wisatawannya di Jawa Barat. Dari hasil pengelompokan data kunjungan wisatawan di Jawa Barat berdasarkan jenis wisatawannya dari tahun 2016-2020 dapat dikelompokkan menjadi 3 kelompok. *cluster 0* merupakan tempat wisata yang jumlah kunjungan wisatawannya rendah, dengan anggota *cluster* sebanyak 445 tempat wisata. *Cluster 2* merupakan tempat wisata yang jumlah kunjungannya sedang dengan jumlah anggota sebanyak 880 tempat wisata. *Cluster 1* merupakan tempat wisata yang jumlah kunjungannya tinggi, dengan jumlah anggota *cluster* sebanyak 435 tempat wisata.

Penulis melakukan 10 kali percobaan yang telah tervalidasi menggunakan *cluster distance performance* dengan nilai k yang berbeda dan diketahui nilai DBI yang paling optimal yaitu 0.684 dengan menggunakan k=3.

4.2 Saran

1. Dapat dilakukan penelitian yang relevan menggunakan algoritma lain yang lebih

- baik sebagai pembandingan hasil dari penelitian.
2. Hasil pengelompokan dapat dikembangkan menjadi basis pengetahuan untuk sistem pendukung keputusan maupun rekomendasi peningkatan fasilitas dan daya tarik wisata untuk pemerintah Jawa Barat.

DAFTAR PUSTAKA

- [1] E. Muningsih, N. Hasan, and G. B. Sulisty, “Penerapan Metode Principle Component Analysis (PCA) untuk Clustering Data Kunjungan Wisatawan Mancanegara ke Indonesia,” *Bianglala Inform.*, vol. 8, no. 1, pp. 58–62, 2020, doi: 10.31294/bi.v8i1.8470.
- [2] E. Satria, H. S. Tambunan, I. S. Saragih, and I. S. Damanik, “Penerapan Clustering dalam Mengelompokkan Jumlah Kunjungan Wisatawan Mancanegara,” no. September, pp. 462–471, 2019.
- [3] M. Wahyudi and S. Pujiastuti, “Penerapan Data Mining Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan,” *Pros. Semin. Nas. Ris. dan Inform. Sci.*, vol. 2, pp. 432–440, 2020.
- [4] F. M. Basysyar, Y. Arie Wijaya, I. Ali, and S. Anwar, “Clustering Data Disabilitas menggunakan Algoritma K-Means di Kabupaten Cirebon,” *JURSIMA (Jurnal Sist. Inf. dan Manajemen)*, vol. 9, no. 3, pp. 247–255, 2021, [Online]. Available: <https://ejournal.stmikgici.ac.id/index.php/jursima/article/view/305>
- [5] E. Muningsih and S. Kiswati, “Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan,” *Joutica*, vol. 3, no. 1, p. 117, 2018, doi: 10.30736/jti.v3i1.196.
- [6] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, “Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- [7] Khaerul Mukhlisin, “Penerapan Algoritma Naïve Bayes Pada Produk Brownies Tape Dalam Menentukan Tingkat Penjualan Laris Dan Kurang Laris,” *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 5, no. 2, pp. 51–56, 2021, doi: 10.32485/kopertip.v5i2.168.
- [8] M. A. Putri, N. Rahaningsih, F. M. Basysyar, and O. Nurdiawan, “Penerapan Data Mining Menggunakan Metode Clustering Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan,” vol. 5, no. 2, pp. 145–156, 2022.
- [9] Puji Rahayu, Ika Anikah, Dias Bayu Saputra, Tri Anelia, and Martanto, “Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Rotan,” *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 2, pp. 42–50, 2020, doi: 10.32485/kopertip.v4i2.118.
- [10] S. N. Br Sembiring, H. Winata, and S. Kusnasari, “Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.
- [11] L. Maulida, “Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means,” *Maulida, L. (2018). Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means. JISKA (Jurnal Inform. Sunan Kalijaga), 2(3), 167. https://doi.org/10.14421/jiska.2018.23-06JISKA (Jurnal, vol. 2, no. 3, p. 167, 2018, doi: 10.14421/jiska.2018.23-06.*
- [12] P. A. Kusuma *et al.*, “OF TOURISM,” vol. 5, no. 1, pp. 46–62, 2022.
- [13] I. A. Zuhriah, S. Alvianna, S. Hidayatullah, and R. G. Patalo, “Dampak Attraction , Accessibility , Amenity , Ancillary Terhadap Minat Berkunjung Wisatawan Di Destinasi Wisata Religi Makam Gus Dur Kabupaten Jombang,” *J. TeslaPerhotelan -Destinasi Wisata -Perjalanan Wisata*, vol. 2, no. 1, pp. 1–11, 2022.
- [14] I. Novitaningtyas, A. Giovanni, and C. A. Lionora, “Faktor-Faktor Yang Mempengaruhi

- Minat Berkunjung Wisatawan Di Kawasan Balkondes Borobudur,” *J. Pariwisata*, vol. 9, no. 1, pp. 28–36, 2022, doi: 10.31294/par.v9i1.12048.
- [15] N. N. A. Wiratini M, N. D. Setiawan, and N. N. Yuliarmi, “Analisis Faktor-Faktor Yang Mempengaruhi Niat Kunjungan Kembali Wisatawan Pada Daya Tarik Wisata Di Kabupaten Badung,” *E-Jurnal Ekon. dan Bisnis Univ. Udayana*, vol. 1, p. 279, 2018, doi: 10.24843/eeb.2018.v07.i01.p10.
- [16] S. Darma *et al.*, “Penerapan Metode K-Means Dalam Pengolompokan Jumlah Wisatawan Asing Di Indonesia,” *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 255–261, 2020, [Online]. Available: <https://www.bps.go.id>.

BIOGRAFI

Ana Neva lahir di Cirebon pada tanggal 02 Oktober 2000. Menyelesaikan pendidikan Diploma 3 (D3) pada tahun 2023 , pada Program Studi Manajemen Informatika di STMIK IKMI Cirebon.