
Klasterisasi Objek Wisata Menggunakan *Jaccard Similarity Coefficient* Berdasarkan *Attraction, Accessibility, Amenity* dan *Ancillary Service*

Candra Agustina¹ *, Eka Rahmawati²

¹ *Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika Indonesia*

² *Sistem Informasi, Universitas Bina Sarana Informatika Indonesia*

E-mail: *candra.caa@bsi.ac.id, eka.eat@bsi.ac.id

Abstract

The tourism industry is back on track after being quiet for a long time due to the Covid-19 pandemic. Currently tourism activists are intensively promoting. In various regions, new tourist destinations have emerged, both managed by large companies and groups of people. In order to realize a good and sustainable tourism industry, it is necessary to carry out coaching and coaching in all of these tourist destinations. In developing the tourism industry, a component known as 6A is known, namely Attraction, Accessibility, Activities, Ancillary Service and Available Packages. Each component has sub-indicators that form the basis of data collection. Of the 6 components, tourist destinations will be clustered, the results of the clustering can be used for various purposes, such as development plans by related agencies, or for recommendations to tourists to determine destinations. The dataset obtained will be processed using the CBR (Case Base Reasoning) technique using the Jaccard Similarity Coefficient algorithm. From framework 4A, an application will be created to input new data that will be clustered. The output of the application is a collection of data that has been input. The results of data processing produce 5 clustering groups.

Keyword : *Jaccard Similarity Coefficient, Tourism, 4A Component*

Abstrak

Industri pariwisata kembali menggeliat setelah sekian lama sepi karena pandemi Covid-19. Saat ini penggiat pariwisata sedang gencar untuk melakukan promosi. Pada berbagai wilayah, muncul destinasi wisata baru, baik dikelola perusahaan besar maupun sekelompok masyarakat. Agar terwujud industri pariwisata yang baik dan berkelanjutan maka perlu dilakukan pemantauan dan pembinaan pada seluruh destinasi wisata tersebut. Dalam mengembangkan industri pariwisata, dikenal komponen yang dinamakan 4A, yaitu *Attraction, Accessibility, Amenity* dan *Ancillary Service*. Masing-masing komponen mempunyai sub indikator yang menjadi dasar untuk pengumpulan data. Dari ke 4 komponen tersebut destinasi wisata akan di klasterisasi, hasil klasterisasi dapat digunakan untuk berbagai kepentingan semisal rencana pembinaan oleh dinas terkait, ataupun untuk rekomendasi kepada wisatawan untuk menentukan tujuan. Dataset yang diperoleh akan diolah dengan Teknik CBR (*Case Base Reasoning*) menggunakan algoritma

Jaccard Similarity Coefficient. Dari framework 4A, akan dibuat aplikasi untuk menginput data baru yang akan diklasterisasi. Hasil pengolahan data menghasilkan 5 kelompok klasterisasi.

Kata Kunci : *Jaccard Similarity Coefficient*, Pariwisata, Komponen 4A

1. Pendahuluan

Bidang pariwisata mengalami perkembangan yang sangat pesat, khususnya setelah dunia mengalami masa pandemi. Selama hampir 3 tahun (2020-2022) masyarakat dihadapkan masa-masa karantina, dimana kegiatan diluar rumah sangat dibatasi. Oleh karena itu ketika masa pandemi telah terlewati, berwisata merupakan salah satu kegiatan yang banyak dilakukan oleh masyarakat. Peluang itu juga dimanfaatkan oleh pelaku usaha wisata, ditandai dengan kemunculan banyak spot / destinasi wisata baru. Untuk mempermudah dinas pariwisata dalam melakukan pemantauan dan pembinaan, hendaknya setiap wilayah melakukan klasterisasi pada destinasi-destinasi yang berada diwilayahnya.

Peningkatan kunjungan destinasi wisata dapat dilakukan dengan analisis yang didasarkan pada komponen 4A (Attraction, Accessibility, Amenity dan Ancillary Service). Pengelompokan destinasi wisata berdasarkan sub indikator dari 4 komponen tersebut akan diolah dengan Teknik CBR (Case Base Reasoning). Teknik tersebut diimplementasikan menggunakan algoritma *Jaccard Similarity Coefficient*. Agar mempermudah stakeholder dalam

melakukan klasterisasi secara mandiri, maka dibuat aplikasi klasterisasi.

2. Bahan dan Metode

A. Komponen 4A Pariwisata

Salah satu kriteria yang bisa diterapkan adalah menggunakan framework 4A. Dalam framework terdapat beberapa komponen yang mempengaruhi kesuksesan dari destinasi wisata. Adapun 4 komponen tersebut yaitu[3]:

1. Attractions /Atraksi

Komponen atraksi di sebuah destinasi wisata dapat berupa wisata alam ataupun buatan manusia seperti taman hiburan ataupun wisata budaya yang dikemas dalam pertunjukan kesenian daerah yang tidak dimiliki oleh daerah lainnya [1]. Komponen atraksi juga dapat berupa acara di suatu desa. Setiap destinasi wisata mempunyai karakteristik tersendiri yang menjadi daya tarik untuk wisatawan [2].

2. Accessibility / Akses

Jarak tempuh dan kondisi jalan menjadi bagian dari komponen akses. Selain itu, komponen akses juga mencakup tersedianya sarana transportasi untuk menuju suatu destinasi wisata.

3. Amenities

Beberapa hal yang termasuk dalam komponen Amenities diantaranya penginapan, tempat makan dan biro perjalanan.

4. Ancillary Service

Termasuk dalam komponen ini yaitu adanya jaringan internet, pusat informasi, dukungan untuk pembayaran non tunai serta adanya ATM.

Komponen 4A digunakan untuk melihat partisipasi masyarakat dalam mengembangkan destinasi wisata di wilayahnya. Pengembangan dianalisis berdasarkan komponen-komponen tersebut, yang menghasilkan kesimpulan bahwa masyarakat di Dusun Serut telah aktif untuk memenuhi komponen *Attractions*, *Amenitis*, *Accesbility* dan *Ancillary*[4].

B. Case-Based Reasoning

Case Base Reasoning adalah Teknik untuk memecahkan masalah dengan meniru/mengadopsi cara yang dilakukan untuk kasus atau masalah yang mirip. Kemudian kasus baru akan disimpan kemudian digunakan untuk acuan pemecahan kasus baru berikutnya[5].

C. Jaccard Simlarity Coefficient

Algoritma *Jaccard Similarity Coefficient* digunakan untuk menghitung tingkat kemiripan kasus baru dengan kasus lama. Tingkat kesamaan dinyatakan dalaam persentase antara 0% sampai 100%.

Semakin tinggi presentasinya berarti semakin mirip[6].

Rumus *Jaccard Similarity Index*

$$\text{Jaccard (A,B)} = \frac{|A \cap B|}{|A \cup B|}$$

Contoh implementasi *Jaccard Similarity*[7]

$A = \{1,2,3,4\}$, $B = \{1,2,4\}$, dan $C = \{1,2,4,5\}$

$A \cap B = \{1,2,4\}$;

$|A \cap B| = 3$

$A \cup B = \{1,2,3,4\}$; $|A \cup B| = 4$

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = 3/4$$

$$J(B,C) = \frac{|A \cap B|}{|A \cup B|} = 3/4$$

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = 3/5$$

D. Mengukur tingkat kesamaan

Tingkat kesamaan diukur berdasarkan beberapa klasifikasi persentase hasil sebagai berikut:

Hasil 0% dikatakan dua objek tidak memiliki kesamaan.

Hasil <15% memiliki sedikit kesamaan

Hasil 15-50% ada kesamaan

Hasil 50% Objek memiliki kesamaan

Hasil 100% Sama Persis[8]

Untuk kasus klasifikasi dikatakan sama atau terdapat dalam 1 kelompok jika memiliki kesamaan tertinggi[9].

E. Unified Modeling Language

UML untuk menggambarkan proses bisnis dalam bidang layanan kesehatan anak di beberapa negara[10].

Diagram yang pertama dibuat adalah diagram use case, diagram ini akan

mendeskripsikan proses dan komponennya. Use case menggunakan skenario untuk menjelaskan langkah yang dilakukan user terhadap sistem[11].

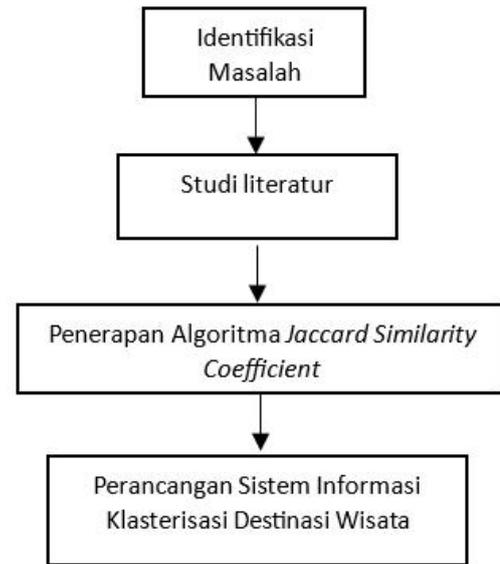
Diagram berikutnya adalah Activity diagram, menurut sukamto dan salahudin dalam Syarif dan Nugraha menjelaskan bahwa *activity diagram* menjelaskan aliran kerja dari sebuah proses bisnis[12]. Dalam diagram tersebut juga menampilkan menu yang dibutuhkan dalam perangkat lunak yang akan dibuat.

F. Desain antar muka

Rancangan antarmuka dibuat untuk memenuhi kebutuhan pengguna, dengan menganut prinsip *accessibility*, *identity*, *navigation* dan *content*. Syarat lainnya adalah desain antarmuka saat ini dapat dijadikan referensi untuk pengembangan ke depan. Untuk itu dokumentasi yang baik juga sangat dibutuhkan.

Untuk mengetahui penerimaan pengguna terhadap desain antarmuka perlu menggunakan alat ukur. Hasil dari pengukuran tersebut dapat dijadikan juga sebagai acuan untuk pengembangan lebih lanjut[13].

3. Metode Penelitian



Gambar 1 Metode Penelitian

1. Identifikasi Masalah
Dalam pengelolaannya, industri pariwisata harus terus dikembangkan agar eksistensinya terus terjaga sehingga kunjungan masyarakat terus meningkat. Upaya dalam pengembangan destinasi wisata belum dilakukan secara maksimal.
2. Studi Literatur
Penelitian sebelumnya tentang penerapan komponen 4A dalam
3. Penerapan Algoritma *Jaccard Similarity Coefficient*
4. Perancangan Sistem Informasi Klasterisasi Destinasi Wisata

3.1 Variabel Penelitian :

Adapun variabel penelitian adalah sebagai berikut:

1. Atraksi
Atribut dalam komponen ini adalah alam, buatan, kesenian
2. Akses
Komponen akses meliputi ada atau tidaknya kendaraan umum ke objek tersebut, kondisi jalan dan lokasi mudah dijangkau atau tidak[14].

3. Amenities

Komponen amenitis meliputi tempat makan, toilet, area parkir[15].

4. Ancillary Service

Fasilitas pendukung dapat berupa tersedianya ATM, Klinik[3] dan pusat informasi[14].

3.2 Penyajian Data

Data disajikan dalam tabel, lalu diklasterisasi berdasarkan komponen 4A. Penyajian data dilakukan seperti pada Tabel 1.

Tabel 1. Penyajian Data

Objek	A11	A12	A13	A21	A22	A23	A31	A32	A33	A41	A42	A43
1	1	0	1	1	1	0	0	1	0	1	0	1
2	1	1	0	1	1	0	1	1	0	1	1	1
ds
t												

4. Hasil Dan Pembahasan

Sebelumnya dipersiapkan terlebih dahulu data yang akan diolah, Tabel 2 menunjukkan komponen 4A dengan nilai 1 jika fasilitas tersedia di objek wisata tersebut dan 0 jika tidak tersedia. Adapun data terdapat pada Tabel 2.

Tabel 2 Data

Objek	A11	A12	A13	A21	A22	A23	A31	A32	A33	A41	A42	A43
1	1	0	1	1	1	0	0	1	0	1	0	1
2	1	1	0	1	1	0	1	1	0	1	1	1
3	1	0	1	1	1	0	0	1	1	1	0	1
4	1	1	0	1	1	0	1	1	0	1	1	1
5	0	0	1	0	1	1	0	0	0	1	0	0
6	0	1	0	0	0	1	0	0	0	0	0	1
7	1	0	1	1	0	1	1	0	1	0	1	1
8	1	1	0	0	1	0	1	1	0	1	0	1

Objek	A11	A12	A13	A21	A22	A23	A31	A32	A33	A41	A42	A43
9	1	0	1	1	1	1	1	0	1	0	1	1
10	0	1	1	0	1	1	0	0	0	1	0	0

Data pada Tabel 2 dihitung menggunakan algoritma *Jaccard Similarity Index* sebagai berikut:

Pseudocode

Input O1 = {1,0,1,1,1,0,0,1,0,1,0,1}

Input O2 = {1,1,0,1,1,0,1,1,0,1,1,1}

Rumus Jaccard = $\frac{6}{10} = 0,6$

Input O3 = {1,0,1,1,1,0,0,1,1,1,0,1}

Jaccard O3,O1 = $\frac{7}{8} = 0,88$

Jaccard O3,O2 = $\frac{6}{11} = 0,55$

Input O4 = {1,1,0,1,1,0,1,1,0,1,1,1}

Jaccard O4,O1 = $\frac{6}{10} = 0,6$

Jaccard O4,O2 = $\frac{9}{9} = 1$

Jaccard O4,O3 = $\frac{6}{11} = 0,54$

Perhitungan dilakukan sampai data terakhir, setelah dilakukan pada 10 sampel menghasilkan perhitungan sebagai berikut:

Tabel 3 Hasil Hitungan Jaccard Similarity Index

	O1	O2	O3	O4	O5	O6	O7	O8	O9
O1	-								
O2	0,6								
O3	0,8	0,5							
O4	0,6	1	0,5						
O5	0,3	0,1	0,3	0,18					
O6	0,1	0,2	0,1	0,2	0,16				
O7	0,36	0,4	0,4	0,4	0,2	0,2			
O8	0,55	0,78	0,5	0,78	0,2	0,2	0,2		
O9	0,45	0,5	0,5	0,5	0,3	0,2	0,889	0,33	
O10	0,3	0,2	0,3	0,2	0,8	0,1	0,18	0,33	0,2

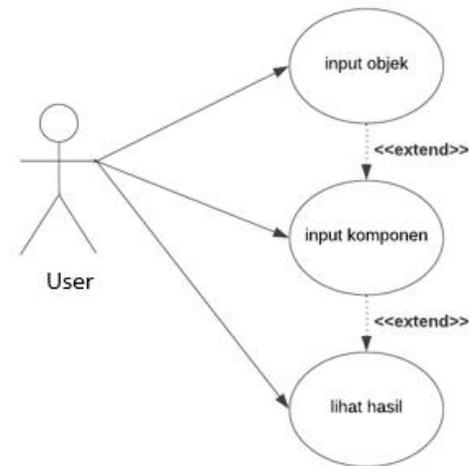
Tabel 3 menunjukkan hasil dari perhitungan menggunakan algoritma *Jaccard Similarity Index*. Untuk menentukan kelompoknya adalah dengan melihat nilai terbesar dan lebih dari 0,75. Jika perhitungan aplikasikan, akan diperoleh hasil seperti dalam Tabel 4:

Tabel 4 Hasil Pengelompokan

Objek	A11	A12	A13	A21	A22	A23	A31	A32	A33	A41	A42	A43	Kelompok
1	1	0	1	1	1	0	0	1	0	1	0	1	1
2	1	1	0	1	1	0	1	1	0	1	1	1	2
3	1	0	1	1	1	0	0	1	1	1	0	1	1
4	1	1	0	1	1	0	1	1	0	1	1	1	2
5	0	0	1	0	1	1	0	0	0	1	0	0	3
6	0	1	0	0	0	1	0	0	0	0	0	1	4
7	1	0	1	1	0	1	1	0	1	0	1	1	5
8	1	1	0	0	1	0	1	1	0	1	0	1	2
9	1	0	1	1	1	1	1	0	1	0	1	1	5
10	0	1	1	0	1	1	0	0	0	1	0	0	3

Dari Table 4, Objek pertama merupakan kelompok 1 karena belum ada pembandingan, Berikutnya Objek 2 dibandingkan dengan objek 1, dengan nilai 0,6, karena kurang dari 0,75 maka tidak masuk kedalam kelompok 1 tapi membantu kelompok 2. Objek 3 masuk kedalam kelompok 1 karena memiliki kesamaan dengan Objek 1 (Kelompok 1) sebesar 0,8 Demikian proses diulang sampai data terakhir.

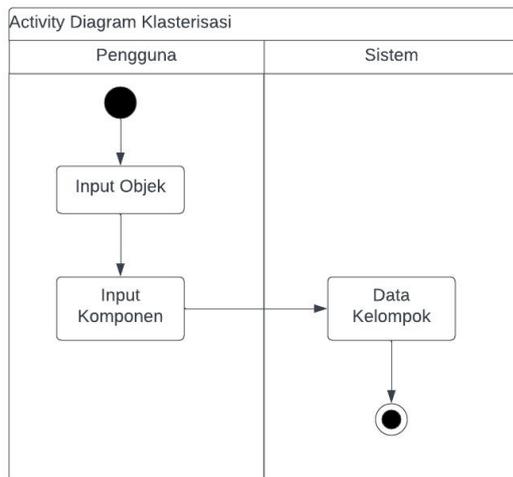
Adapun *Usecase diagram* untuk klasterisasi adalah sebagai berikut:



Gambar 2 Use Case Diagram Klasterisasi

<i>Use Case Name</i>	Klasterisasi
<i>Requirements</i>	Untuk melakukan input data klasterisasi
<i>Goal</i>	Dapat melakukan input data dan muncul data klaster baru
<i>Pre-Conditions</i>	Membuka halaman klasterisasi, mengisi data komponen
<i>Post-Conditions</i>	Berhasil menyimpan data komponen
<i>Failed end Condition</i>	Gagal memuat kelompok data baru
<i>Actors</i>	Pengguna
<i>Main flow/ Basic path</i>	<ol style="list-style-type: none"> 1. Pengguna membuka halaman klasterisasi 2. Mengisi data komponen 3. Klik lihat hasil 4. Memperoleh kelompok data baru
<i>Alternate Flow/ Invariant A</i>	A.1 Sistem menampilkan halaman klasterisasi
<i>Invariant B</i>	B.1 Pengguna membuka halaman klasterisasi B.2 Mengisi data komponen B.3 Klik lihat hasil B.4 Sistem menampilkan pesan “data belum lengkap”

Alur dari klasterisasi digambarkan dengan *Activity Diagram* pada Gambar 3.



Gambar 3 *Activity Diagram* Klasterisasi

Proses dimulai ketika pengguna memasukan objek baru. Setelah itu, pengguna memasukan data komponen. Saat semua data komponen sudah dimasukan, maka data kelompok akan muncul sebagai akhir dari proses klasterisasi.

Agar mempermudah proses klasterisasi, maka dibuat aplikasi klasterisasi objek. Halaman klasterisasi tersebut terdapat pada Gambar 4.

The screenshot shows a web application interface titled "Klasterisasi Objek". It has two main sections: "Objek Baru" (New Object) and "Data Objek" (Object Data). Under "Objek Baru", there is a form with "Nama Objek" (Object Name) set to "Objek 1". Under "Data Objek", there is a section for "Masukan Data Komponen" (Enter Component Data) with a grid of 14 input fields labeled A11 through A46. Below the grid is a button labeled "Lihat Hasil Klaster" (View Clustering Results). At the bottom, a text box displays the result: "Objek 1 masuk dalam Kelompok 1" (Object 1 is in Group 1).

Gambar 4 Antarmuka Halaman Klasterisasi

Halaman klasterisasi akan memberikan informasi terkait kelompok dari objek yang telah di *input*.

5. Kesimpulan

Penerapan algoritma *Jaccard Coefficient Index* untuk klasifikasi objek wisata termasuk dalam kelompok *unsupervised learning*. Hasil dari implementasi algoritma *Jaccard Coefficient Index*, menghasilkan 5 kelompok klasterisasi objek wisata.

Jumlah kelompok yang terbentuk dapat terus bertambah saat datanya juga bertambah. Penelitian selanjutnya sebaiknya menggunakan data sebenarnya yang didapatkan dari hasil observasi dan dengan objek yang lebih banyak.

Referensi

- [1] D. Buhalis and A. Amaranggana, "Information and Communication Technologies in Tourism 2014," *Inf. Commun. Technol. Tour.* 2014, pp. 553–564, 2013, doi: 10.1007/978-3-319-03973-2.
- [2] S. F. Chaerunissa and T. Yuniningsih, "Analisis Komponen Pengembangan Pariwisata Desa Wisata Wonopolo Kota Semarang," *J. Public Policy Manag. Rev.*, vol. 9, no. 4, pp. 159–175, 2020.
- [3] S. Rosanto and V. Chainarta, "Analisa Aspek Kebijakan Pemerintah Terhadap Potensi Wisata Alam di Danau Sarantangan, Singkawang, Kalimantan Barat," vol. 1, no. 11, pp. 2805–2812, 2021.
- [4] R. A. Nugraha, H. Abdillah, S. T. Untoro, and A. Makruf, "Partisipasi Masyarakat Melalui Metode 4A Dalam Pengembangan Sektor Wisata Dusun Serut," *Mawa Izh J. Dakwah Dan Pengemb. Sos. Kemanus.*, vol. 13, no. 1, pp. 27–48, 2022, doi:

- 10.32923/maw.v13i1.2290.
- [5] M. Nugraheni, "Detection Coronavirus using Cased-Based Reasoning with Extended Jaccard Coefficient," *IJISTECH (International J. Inf. Syst. Technol.,* vol. 5, no. 1, p. 31, 2021, doi: 10.30645/ijistech.v5i1.112.
- [6] A. Gunawan, C. Suhery, and T. Rismawan, "Implementasi Metode Case-Based Reasoning Dan Similarity Jaccard Coefficient Dalam Identifikasi Kerusakan Laptop," *Coding J. Komput. dan Apl.,* vol. 09, no. 02, pp. 292–305, 2021.
- [7] S. Rianti and Ri. Adrianti Supono, "Perbandingan Algoritma Edit Distance, Levenshtein Distance, Hamming Distance, Jaccard Similarity Dalam Mendeteksi String Matching," *J. Sist. Inf. Univ. Suryadarma,* vol. 10, no. 1, pp. 305–314, 2023.
- [8] A. Yudhana, S. -, and A. Djalil, "Implementation of Pattern Matching Algorithm for Portable Document Format," *Int. J. Adv. Comput. Sci. Appl.,* vol. 8, no. 11, pp. 509–512, 2017, doi: 10.14569/ijacsa.2017.081162.
- [9] M. Nugraheni, Widodo, and I. P. Sari, "A Case-Based Reasoning for Detection Coronavirus (Covid-19) Using Cosine Similarity," *Proc. Conf. Broad Expo. to Sci. Technol. 2021 (BEST 2021),* vol. 210, no. Best 2021, pp. 178–183, 2022, doi: 10.2991/aer.k.220131.030.
- [10] F. Pecoraro and D. Luzi, "Using Unified Modeling Language to Analyze Business Processes in the Delivery of Child Health Services," *Int. J. Environ. Res. Public Health,* vol. 19, no. 20, 2022, doi: 10.3390/ijerph192013456.
- [11] L. Setiyani, "Desain Sistem: Use Case Diagram Pendahuluan," *Pros. Semin. Nas. Inov. Adopsi Teknol. 2021,* no. September, pp. 246–260, 2021.
- [12] M. Syarif and W. Nugraha, "Pemodelan Diagram UML Sistem Pembayaran Tunai Pada Transaksi E-Commerce," *J. Tek. Inform. Kaputama,* vol. 4, no. 1, p. 70 halaman, 2020, [Online]. Available: <http://jurnal.kaputama.ac.id/index.php/JTIK/article/view/240>
- [13] A. Mufti, "Rancangan Layar Sebagai Alat Bantu Pendewasa," *Fakt. Exacta,* vol. 8, no. 2, pp. 181–185, 2015.
- [14] Alfitriani, "Pengaruh Komponen 4A Terhadap Minat Kunjung Ulang Wisatawan Pada Destinasi Wisata Bayt Al-Qur'an Al-Akbar Kota Palembang," *Apl. Manaj. Bisnis,* vol. 1, no. 2, pp. 66–77, 2021, [Online]. Available: <https://jurnal.polsri.ac.id/index.php/JAMB>
- [15] M. Y. F. Salasa and T. Ismail, "Analisis Pengaruh Attraction, Accessibility, Amenities, dan Ancillary Terhadap Kepuasan Wisatawan Pantai Tiga Warna Malang," *J. Ilm. FEB,* vol. 7, no. 1, pp. 1–8, 2018.