

ANALISIS SENTIMEN PADA REVIEW WISATA PAGAR ALAM DENGAN ALGORITMA *K-NEAREST NEIGHBOR* (K-NN)

Desi Puspita¹, Riduan Syahri²

¹²Program Studi Teknik Informatika, Institut Teknologi Pagaralam, Sumatera Selatan

e-mail: *¹desiofira1@gmail.com, ²syahririduan@gmail.com,

*Corresponding Author

Abstrak

Kota Pagar Alam memiliki banyak pilihan wisata yang indah, udara yang segar dan dingin, hingga budaya dan kuliner yang khas menjadi daya Tarik sendiri. Sehingga menjadi daerah yang banyak dikunjungi oleh wisatawan lokal maupun manca negara. Dari banyak pengunjung yang datang, tidak sedikit dari mereka meninggalkan kesan-kesan berupa review atau ulasan terhadap tempat yang telah dikunjungi. Tujuan dari penelitian ini adalah untuk mengetahui akurasi yang dihasilkan dengan metode *K-Nearest Neighbor* (K-NN). Metode *K-Nearest Neighbor* (K-NN) digunakan untuk melakukan klasifikasi data sentiment pada wisata Pagar Alam. Pengujian yang dilakukan untuk mendapatkan hasil akurasi yang baik serta mengevaluasi menggunakan *confusion matrix*. Penelitian ini menghasilkan sebuah sistem analisis sentimen yang dapat mengidentifikasi dan mengklasifikasi ulasan pengunjung wisata Pagar Alam dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN) dengan hasil didapat akurasi yang terbesar dengan nilai $k=3$ dengan akurasi 99%, K_0 mendapatkan *precision* 98%, *recall* 100 dan *fi-score* 99%, untuk k_1 *precision* 100% *recall* 89% dan *fi-score* 92% sedangkan K_2 *precision* 100%, *recall* 100% *f1-score* 100%.

Kata kunci— Analisis sentimen, Review wisata, *K-Nearest Neighbor*, *confusion matrix*

Abstract

Pagar Alam City has many beautiful tourist options, fresh and cold air, to distinctive culture and culinary become its own attraction. So that it becomes an area that is visited by many local and foreign tourists. Of the many visitors who came, not a few of them left impressions in the form of reviews or reviews of the places they had visited. The purpose of this study was to find out the accuracy produced by the *K-Nearest Neighbor* (K-NN) method. The *K-Nearest Neighbor* (K-NN) method is used to classify sentiment data on Pagar Alam tourism. Tests conducted to obtain good accuracy results and evaluate using *confusion matrix*. This study resulted in a sentiment analysis system that can identify and classify reviews of Pagar Alam tourist visitors using the *K-Nearest Neighbor* (K-NN) algorithm with the results obtained the greatest accuracy with a value of $k = 3$ with an accuracy of 99%, K_0 getting a *precision* of 98%, *recall* 100 and a *fi-score* of 99%, for k_1 *precision* 100% *recall* 89% and *fi-score* 92% while K_2 *precision* 100%, *recall* 100% *f1-score* 100%.

Keywords— Synthetic analysis; Travel review; *K-Nearest Neighbor*; *confusion matrix*

I. PENDAHULUAN

Banyaknya review mengenai tempat wisata yang telah dikunjungi, memudahkan pengunjung lain yang ingin berkunjung ke tempat tersebut mengetahui keadaan di tempat tersebut dan menjadikan masukan

untuk pengunjung lain untuk memutuskan ke tempat tersebut atau tidak (Sari, 2020).

Kota Pagar Alam termasuk dalam wilayah administratif Provinsi Sumatera Selatan dan merupakan daerah hulu yang berada pada ketinggian 600– 3195 mdpl (meter dari permukaan laut) dengan luas

wilayah 63.366 hektare yang membuat daerah ini masih begitu sejuk dan masih asri (Pratama et al., 2021). Sektor wisata merupakan sektor penting yang tidak terpisahkan dari Kota Pagar Alam karena telah tertuang di dalam visi dari pembangunan Kota Pagar Alam dari tahun 2005 sampai 2025 yaitu “Pagar Alam sebagai Kota Jasa Berbasis Agrowisata Terdepan di Sumatera Bagian Selatan (SUMBANGSEL) tahun 2025” (RENSTRA BAPPEDA Kota Pagar Alam 2016).

Kota Pagar Alam memiliki banyak pilihan wisata yang indah, udara yang segar dan dingin, hingga budaya dan kuliner yang khas menjadi daya Tarik sendiri. Sehingga menjadi daerah yang banyak dikunjungi oleh wisatawan lokal maupun manca negara. Dari banyak pengunjung yang datang, tidak sedikit dari mereka meninggalkan kesan-kesan berupa review atau ulasan terhadap tempat yang telah dikunjunginya. Ulasan ini bisa menjadi informasi yang sangat penting bagi pengelola wisata maupun bagi wisatawan lainnya. Bagi pengelola wisata informasi ini sebagai bahan evaluasi dan perbaikan terhadap kekurangan-kekurangan yang ada guna meningkatkan kualitas tempat wisata (Sibyan & Hasanah, 2022).

Banyaknya jumlah ulasan dan beragamnya ulasan yang ada membuat sulit untuk menyimpulkan isi dari ulasan-ulasan tersebut karena harus membaca satu per satu yang tentunya akan membutuhkan waktu yang cukup lama. Sehingga diperlukan sebuah alat bantu analisis yang bisa digunakan untuk menyimpulkan ulasan dengan cepat yaitu dengan analisis sentiment (Sibyan & Hasanah, 2022).

Analisis sentimen sering juga disebut sebagai opinion mining. Saat ini, pengguna sangat senang mengekspresikan perasaannya melalui platform *online*, seperti media sosial, *e-commerce*, dan

website. Oleh karena itu, sentiment analysis dilakukan di platform-platform tersebut. Salah satu keunggulan dari analisis sentimen adalah menghemat waktu dan tenaga. Kini, analisis sentimen dapat dilakukan secara otomatis (Permana et al., 2021). Metode yang digunakan dalam penelitian adalah *K-Nearest Neighbor* disingkat K-NN.

K-Nearest Neighbor (K-NN) merupakan salah satu metode pengambilan keputusan yang dimana hasil dari data masukan yang baru diklasifikasi berdasarkan terdekat dalam data nilai (Permana et al., 2021). Metode K-NN juga didasarkan pada klasifikasi terhadap objek berdasarkan data pelatihan yang jaraknya paling dekat dengan objek tersebut (Firdaus, 2022).

Terdapat penelitian mengenai analisis sentimen sebelumnya menggunakan algoritma *K-Nearest Neighbor* (K-NN). Penelitian sebelumnya yaitu analisis Sentimen pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma *K-Nearest Neighbor* (K-NN) (Sari, 2020). Penerapan algoritma *K-Nearest Neighbor* pada sentimen review agen travel (Siti Ernawati, 2018)

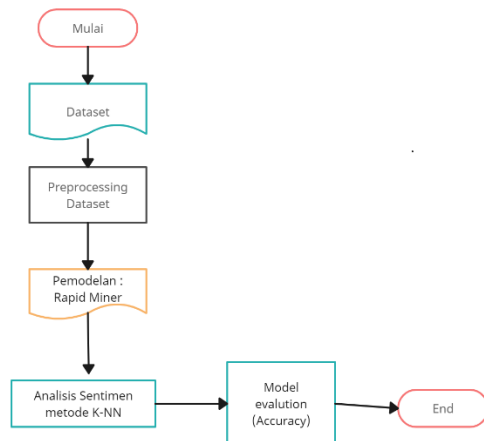
Dari hasil penelitian sebelumnya yang berjudul analisis Sentimen pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma *K-Nearest Neighbor* didapatkan akurasi sebesar 77.01% dengan nilai K 7 menggunakan 10 *fold cross validation* (Sari, 2020).

Penelitian dengan judul analisis sentimen pada review objek wisata Pagar Alam dengan algoritma *K-Nearest Neighbor* (K-NN) bertujuan untuk mengetahui akurasi yang dihasilkan.

II. METODOLOGI PENELITIAN

Tahapan penelitian analisis sentiment ini menggunakan metode *K-Nearest*

Neighbor (K-NN) dengan *Flowchart* alur penelitian sebagai berikut :



Gambar 1. *Flowchart* alur Penelitian

Dari *flowchart* diatas dijelaskan beberapa tahapan, yaitu :

- 1) Dataset adalah data penelitian yang terkumpul dari data pengunjung wisata Pagar Alam yang didapat dari Dinas Pariwisata Kota Pagar Alam.
- 2) Proses *preprocessing* tahap awal teknik pada data mining dimana akan mengubah data mentah atau yang biasa di kenal dengan raw data. Proses ini bisa juga di sebut dengan langkah awal untuk mengambil informasi dari data yang telah diambil dengan cara membersihkan, memfilter, dan mengabungkan data-data tersebut (Amalia et al., 2021).
- 3) Pemodelan, Pada penelitian ini proses eksperimen menggunakan *RapidMiner*. Data training yang digunakan adalah data pengunjung wisata.
- 4) Metode *K-Nearest Neighbor* (K-NN) digunakan untuk melakukan klasifikasi data sentimen. Penentu klasifikasi berdasarkan contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, tetapi bergantung pada label kategori yang melekat pada dokumen pelatihan mirip dengan dokumen tes adalah algoritma *K-Nearest Neighbor* (KNN) (Deviyanto &

Wahyudi, 2018)(Yoga Pratama et al., 2021). Aplikasi menemukan k tetangga terdekat antara dokumen pelatihan pada tes dokumen. Rata kesamaan setiap dokumen tetangga terdekat dokumen uji digunakan sebagai berat kelas dokumen tetangga (Putri Fitrianti et al., 2019).

- 5) Adapun Evaluasi atau pengujian hasil klasifikasi K-NN menggunakan *Confussion Matrix* yang merupakan alat yang dapat digunakan untuk melakukan analisis terhadap seberapa baik klasifikasi yang telah dihasilkan dan dapat mengenali tuple dari kelas yang berbeda (Tanggu Mara et al., 2021). Dalam *confusion matrix* akan dihitung *acuracy*, *precision*, *recall* dan *f-measure* yang dirumuskan pada persamaan (Furqan et al., 2022).

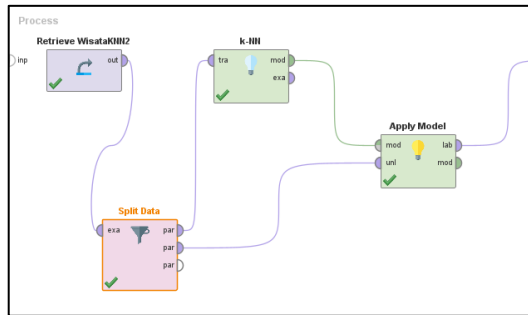
Tabel 1. Confusion Matrix

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP (True Positive) Correct result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (True Negative) Correct absence of result

III. HASIL DAN PEMBAHASAN

Dalam penelitian ini data yang digunakan adalah data yang di dapatkan dari dinas pariwisata dan objekk wisata dimana setelah melakukan pengumpulan data didapatkan total data ada sebanyak 20 data yang dibagi kedalam 12 bulan dan total wisata yang terdokumentasi jumlah pengunjungnya ada sebanyak 20 wisata dari jumlah tersebut maka didapatkan sebanyak 240 (12*20) data yang akan di olah.

1. Desain *Rapid Minner*



Gambar 2. Desain Rapidminer

Gambar diatas menunjukkan desain *rapid miner* yang terdiri dari dataset, *split data*, algoritma knn, dan *aply model*

2. ExampleSet (Dataset)

Row No.	Ket	predictionK	Wisata	Pengunjung1	Pengunjung2	Jumlah	Bulan
1	2	2	4	60	23443	23503	12
2	0	0	14	60	4654	4614	12
3	0	0	8	58	2455	2513	9
4	0	0	18	58	2454	2512	9
5	0	0	5	58	2121	2177	8
6	1	1	15	58	5452	5508	8
7	2	2	13	53	34329	34373	12
8	0	0	3	53	758	809	12
9	2	2	6	50	43240	43280	12
10	2	2	11	50	23443	23493	9
11	0	0	18	50	755	805	10
12	0	0	8	50	435	485	10
13	0	0	10	50	340	390	12
14	0	0	1	50	234	284	9

Gambar 3. ExampleSet

Gambar 3 menunjukkan dataset yang sudah mempunyai hasil prediksi untuk klasifikasi data yang di dapatkan dari *rappid minner* kemudian selanjutnya di olah menggunakan *python*

3. Dataset Python

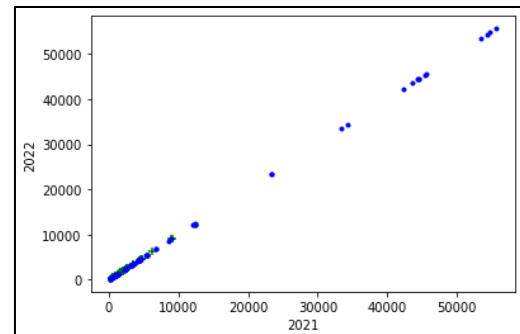
	Wisata	Pengunjung1	Pengunjung2	Jumlah	Bulan	Ket
0	10	9210	12302	21512	7	2
1	20	9020	14200	23220	7	2
2	10	6249	21065	27314	1	2
3	10	4700	6110	10810	8	2
4	20	4700	6110	10810	8	2
...	...	...	...	...	...	...
235	15	0	345	345	10	0
236	9	0	345	345	12	0
237	9	0	224	224	9	0
238	9	0	109	109	8	0
239	5	0	36	36	10	0

Gambar 4. Dataset

Dari gambar diatas dapat dilihat bahwa jumlah data yang di olah adalah sebanyak 240 data yang terdiri dari 6 kolom, dimana dari 240 data di bagi

kedalam 2 variabel yaitu X_train dan X_test, jumlah X_test adalah sebesar 30% (72) sedangkan X_train 70%(168)

4. Hasil perbandingan



Gambar 5. Perbandingan

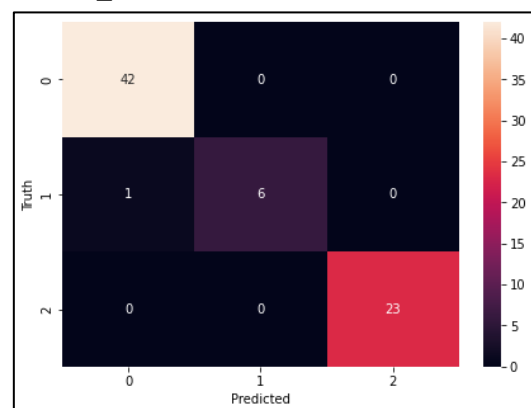
Gambar 6 menunjukkan perbandingan data pengunjung antara tahun 2021 dengan tahun 2022 dimana untuk data tahun 2021 ditandai dengan (+) plus berwarna hijau sedangkan tahun 2022 ditandai dengan (.) titik berwarna biru.

5. Confusion Matriks

Tabel 2. Confusion matriks

	0	1	2
0	42	0	0
1	1	6	0
2	0	0	23

Tabel diatas merupakan confusion matriks yang didapatkan dari pengelolaan data yang dilakukan di python dimana variabel yang diambil adalah variable (y_test) dan variable (y_pred) y_pred didapatkan dari prediksi KNN x_test.



Gambar 6. Confusion matriks

6. Clasification Report

	precision	recall	f1-score	support
0	0.98	1.00	0.99	42
1	1.00	0.86	0.92	7
2	1.00	1.00	1.00	23
accuracy			0.99	72
macro avg	0.99	0.95	0.97	72
weighted avg	0.99	0.99	0.99	72

Gambar 7. Laporan Klasifikasi

Gambar diatas menunjukkan hasil laporan klasifikasi yang di dapatkan di *python* dimana akurasi yang di dapatkan adalah sebesar 99%, kemudian klasifikasi 0 mendapatkan precision 98% recall 100% dan *f1-score* 99%, Klasifikasi 1 mendapatkan *precision* 100%, *recall* 86% dan *f1-score* 92%, kemudian klasifikasi 2 menndapatkan *precision* 100% *recall* 100% dan *f1-score* 100%

IV. KESIMPULAN

Berdasarkan dari penelitian yang sudah dilaksanakan, dapat disimpulkan bahwa penelitian ini review wisata Pagar Alam menggunakan K-Nearest Neighbor dapat mnegidentifikasi dan mengkalsifikasi ulasan pengunjung wisata. Hasil dari pengujian dengan akurasi yang terbesar dengan nilai $k=3$ dengan akurasi sebesar 99%, kemudian klasifikasi 0 mendapatkan precision 98% recall 100% dan *f1-score* 99%, Klasifikasi 1 mendapatkan precision 100%, recall 86% dan *f1-score* 92%, kemudian klasifikasi 2 menndapatkan precision 100% recall 100% dan *f1-score* 100%

V. SARAN

Berdasarkan dari hasil penelitian dimana data yang diperoleh dari lapangan. Penelitian ini terlaksana dengan baik. Saran peneliti hendaknya pada penelitian selanjutnya dapat mencoba metode lain selain K-NN yang akan digunakan untuk menguji nilai akurasi sebagai perbandingan.

VI. DAFTAR PUSTAKA

- Amalia, B. S., Umaidah, Y., & Mayasari, R. (2021). Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritma Support Vector Machine Dan K-Nearest Neighbor. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, 19(1), 28–34.
- Deviyanto, A., & Wahyudi, M. D. R. (2018). Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 3(1), 1. <https://doi.org/10.14421/jiska.2018.31-01>
- Firdaus, A. (2022). Aplikasi Algoritma K-Nearest Neighbor pada Analisis Sentimen Omicron Covid-19. *Jurnal Riset Statistika*, 85–92. <https://doi.org/10.29313/jrs.v2i2.1148>
- Furqan, M., Sriani, S., & Sari, S. M. (2022). Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia. *Techno.Com*, 21(1), 51–60. <https://doi.org/10.33633/tc.v21i1.5446>
- Permana, K., Putra, Z. P., & Nugroho, A. (2021). Analisa Sentimen Pengunjung Hotel Dengan K-Nearest Neighbor Studi Kasus Hotel Pop! Surabaya. *Jurnal Ilmu Komputer Dan Bisnis*, 12(2), 159–166. <https://doi.org/10.47927/jikb.v12i2.170>
- Pratama, Y. Z., Soekmadi, R., & Makalew, A. D. (2021). Potensi Wisata Alam Di Kota Pagar Alam, Provinsi Sumatera Selatan Berdasarkan Penawaran, Permintaan Dan Daya Dukung. *Jurnal TENGKAWANG*, 11(1), 14–24. <https://doi.org/10.26418/jt.v11i1.41546>
- Putri Fitrianti, R., Kurniawati, A., Agusten, D., Sistem Informasi, J., & Ilmu Komputer dan Teknologi Informasi,

- F. (2019). Implementasi Algoritma K-Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, 1907–5022.
- Sari, R. (2020). Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn). *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1), 10–17. <https://doi.org/10.31294/evolusi.v8i1.7371>
- Sibyan, H., & Hasanah, N. (2022). Analisis Sentimen Ulasan Pada Wisata Dieng Dengan Algoritma K-Nearest Neighbor (K-Nn). *Jurnal Penelitian Dan Pengabdian Kepada Masyarakat UNSIQ*, 9(1), 38–47. <https://doi.org/10.32699/ppkm.v9i1.2218>
- Tangu Mara, A., Sedyono, E., & Purnomo, H. (2021). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba. *Jointer - Journal of Informatics Engineering*, 2(01), 24–31. <https://doi.org/10.53682/jointer.v2i01.30>
- Yoga Pratama, A., Umaidah, Y., Voutama, A., Informatika, T., Ilmu Komputer, F., Singaperbangsa Karawang Ds Paseurjaya, U., Telukjambe Timur, K., Karawang, K., & Barat, J. (2021). Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja). *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 897–910.
- .