

KLASIFIKASI PENGUNJUNG WISATA DI KOTA PAGAR ALAM DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)

Riduan Syahri¹, Desi Puspita², Dedi Setiadi³

^{1,2,3}Program Studi Informatika, Institut Teknologi Pagar Alam

e-mail: *syahririduan@gmail.com, desiofiral@gmail.com, dedisetiadi1212@gmail.com

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 dikunjunginya. Tujuan dari penelitian ini adalah untuk menentukan klasifikasi dan untuk mengetahui akurasi yang dihasilkan dengan metode K-Nearest Neighbor (K-NN). Metode K-Nearest Neighbor (K-NN) digunakan untuk melakukan klasifikasi data pengunjung pada wisata Pagar Alam. Pengujian yang dilakukan untuk mendapatkan hasil akurasi yang baik serta mengevaluasi menggunakan confusion matrix. Penelitian ini menghasilkan sebuah sistem klasifikasi yang dapat mengidentifikasi dan mengklasifikasi 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% recal 89% dan fi-score 92% sedangkan K_2 precision 100%, recal 100% f1-score 100%.

Kata kunci: Terdiri dari 3 – 5 kata kunci yang disusun sesuai urutan alfabetis.

Abstract

The city of Pagar Alam has many beautiful tourist options, fresh and cold air, and unique culture and culinary delights. So that it becomes an area that is visited by many local and foreign tourists. Of the many visitors who come, not a few of them leave impressions in the form of reviews of the places they have visited. The purpose of this study is to determine the classification and to determine the accuracy produced by the K-Nearest Neighbor (K-NN) method. The K-Nearest Neighbor (K-NN) method is used to classify visitor data on Pagar Alam tours. Tests carried out to get good accuracy results and evaluate using a confusion matrix. This research produces a classification system that can identify and classify Pagar Alam tourism visitors using the K-Nearest Neighbor (K-NN) algorithm with the results obtained the greatest accuracy with a value of $k = 3$ with 99% accuracy, K_0 gets 98% precision, recall 100 and a fi-score of 99%, for k_1 precision 100%, the recal is 89% and the fi-score is 92%, while for K_2 the precision is 100%, the recal is 100%, the f1-score is 100%.

Keywords— Sentiment analysis, tourist review, K-Nearest Neighbor, confusion matrix

1. 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.

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 and Hasanah 2022).

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini Sedangkan menurut Han dan Kamber, "Data mining adalah proses menambang (mining) pengetahuan dari sekumpulan data yang sangat besar". Data mining merupakan suatu langkah dalam *knowledge discovery in database* (KDD)(Ginting, Kusri, and Taufiq 2020)

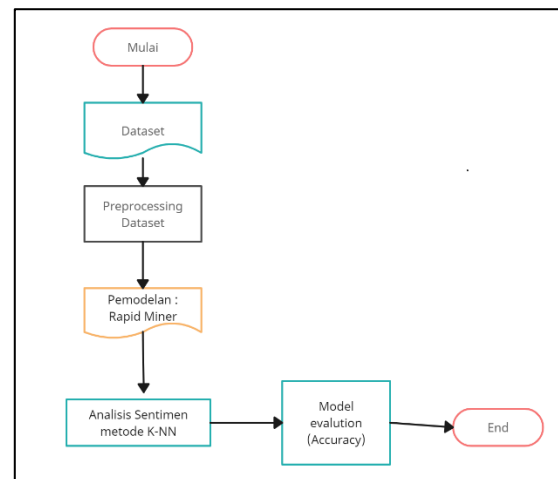
Analisis sentimen sering juga disebut sebagai opinion mining. Saat ini, pengguna sangat senang mengekspresikan perasaannya melalui platform online, seperti media sosial, *e-commerce*, dan website. Oleh karena itu, sentiment analysis dilakukan di platform-platform tersebut. Salah satu keunggulan dari analisis sentimen adalah menghemat waktu dan tenaga. Kini, analisis sentimen dapat dilakukan secara otomatis (Pratama, Soekmadi, and Makalew 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 (Pratama, Soekmadi, and Makalew 2021). Metode K-NN juga didasarkan pada klasifikasi terhadap objek berdasarkan data pelatihan yang jaraknya paling dekat dengan objek tersebut (Firdaus 2022).

Tujuan dari penelitian ini untuk menghasilkan tingkat akurasi dalam penentuan objek wisata Pagar Alam dengan menerapkan Algoritma *K-Nearest Neighbor* (K-NN)

2. METODE

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 peneliti 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.
- 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 and Wahyudi 2018). 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 Confusion Matrix yang merupakan alat yang dapat digunakan untuk melakukan analisis terhadap seberapa baik klasifikasi yang telah dihasilkan dan dapat mengenali tuple dari kelas yang Dalam confusion matrix akan dihitung accuracy, precision, recall dan f-measure yang dirumuskan pada persamaan (Furqan, Sriani, and Sari 2022).

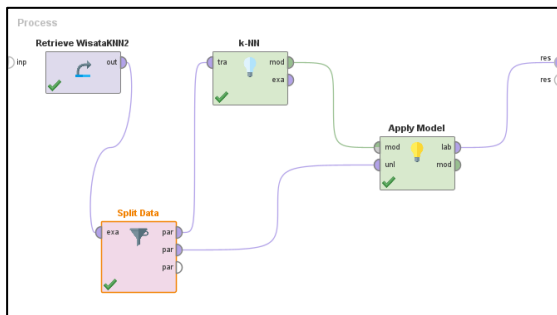
Tabel 1. Confusion Matrix

		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

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini data yang digunakan adalah data yang di dapatkan dari dinas pariwisata dan objek 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 Miner



Gambar 2. Desain Rapidminer

Gambar diatas menunjukan desai *rapid minner* 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	4854	4814	12
3	0	0	8	58	2455	2513	9
4	0	0	18	58	2454	2512	9
5	0	0	5	56	2121	2177	8
6	1	1	15	56	5452	5508	8
7	2	2	13	53	34320	34373	12
8	0	0	3	53	758	809	12
9	2	2	5	50	43240	43290	12
10	2	2	11	50	23443	23493	9
11	0	0	18	50	755	805	10
12	0	0	6	50	425	485	10
13	0	0	16	50	340	390	12
14	0	0	1	50	234	284	9

Gambar 3. ExampleSet

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

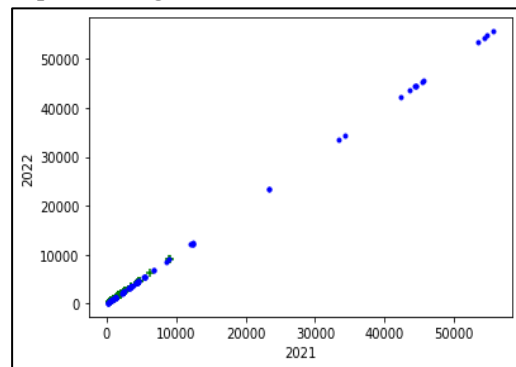
3. Dataset Python

	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

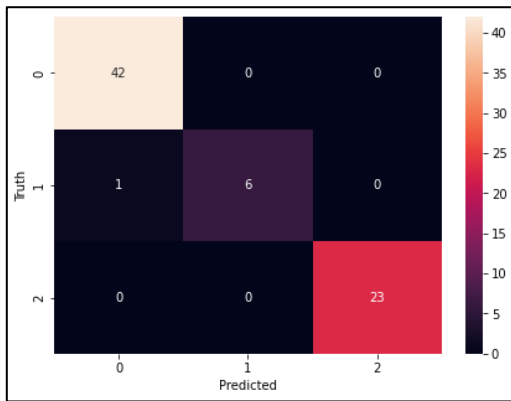
Gambar 6 menunjukan perbandingan data pengunjung antara tahun 2021 dengan tahun 2022 dimana untuk data tahun 2021 ditandai dengan (+) plus berwarna hijau sedangkan tahun 2022 ditandai dengan (.) titik berwarna biru.

5. Confusion Matrics

Tabel 2. Confusion matrics

	0	1	2
0	4 2	0	0
1	1	6	0
2	0	0	2 3

Tabel diatas merupakan confusion matrics 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 mendapatkan *precision* 100% *recall* 100% dan *f1-score* 100%.

4. SIMPULAN

Berdasarkan dari penelitian yang sudah dilaksanakan, dapat disimpulkan bahwa penelitian ini review wisata Pagar Alam menggunakan K-Nearest Neighbor dapat mnedidentifikasi 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%.

5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada tim peneliti dan Pusat Penelitian dan Pengabdian Masyarakat.

6. DAFTAR PUSTAKA

Deviyanto, Akhmad, and Muhammad Didik Rohmad Wahyudi. 2018. "Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor." *JISKA (Jurnal Informatika Sunan Kalijaga)* 3(1): 1.

Firdaus, Alfiari. 2022. "Aplikasi Algoritma K-Nearest Neighbor Pada Analisis Sentimen Omicron

Covid-19." *Jurnal Riset Statistika*: 85–92.

Furqan, Mhd, Sriani Sriani, and Susan Mayang Sari. 2022. "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia." *Techno.Com* 21(1): 51–60.

Ginting, Victor Saputra, Kusri Kusri, and Emha Taufiq. 2020. "Implementasi Algoritma C4.5 Untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python." *Inspiration: Jurnal Teknologi Informasi dan Komunikasi* 10(1): 36–44.

Pratama, Yogie Zulni, Rinekso Soekmadi, and Afra DN Makalew. 2021. "Potensi Wisata Alam Di Kota Pagar Alam, Provinsi Sumatera Selatan Berdasarkan Penawaran, Permintaan Dan Daya Dukung." *jurnal TENGGAWANG* 11(1): 14–24.

Putri Fitrianti, Risma et al. 2019. "Implementasi Algoritma K-Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia." *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*: 1907–5022.

Sari, Retno. 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.

Sibyan, Hidayatus, and Nur Hasanah. 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.