



Implementasi Data Mining Untuk Klasifikasi Kunjungan Wisatawan Riau Menggunakan K-NN

Rifqi Muzaki Fhardana¹, Muhammad Siddik^{2*}

^{1,2}Institut Bisnis dan Teknologi Pelita Indonesia, Jalan Ahmad Yani No.78-88 Pekanbaru
e-mail: ¹rifqi.muzaki@student.pelitaindonesia.ac.id · ²siddik@lecturer.pelitaindonesia.ac.id

Abstrak

Pariwisata merupakan sektor penting dalam perekonomian daerah, termasuk pada Provinsi Riau yang memiliki lebih dari 200 destinasi wisata. Dengan banyaknya destinasi wisata, tidak semua wisata memiliki infrastruktur yang memadai dan baik. Hal ini dikarenakan, tidak adanya pengklasifikasian destinasi wisata yang setara dan membuat pembangunan kurang merata. Penelitian ini bertujuan untuk mengklasifikasikan destinasi wisata yang ada pada Provinsi Riau menjadi beberapa kategori. Dengan klasifikasi yang tepat, Dinas Pariwisata Riau dapat memberikan dana untuk pembangunan tempat wisata yang lebih merata. K-Nearest Neighbor digunakan pada penelitian ini karena dirasa cukup ampuh dalam melakukan klasifikasi. Dengan menggunakan data kunjungan wisatawan yang terdapat pada periode 2021 hingga 2023, data yang digunakan sudah masuk ke dalam tahapan preprocessing, pemberian label menggunakan label encoding, dan pemberian kategori dengan min-max scaling. Pada pembagian data menggunakan presentase 80 persen data latih dan 20 persen data uji, didapatkan akurasi sebesar 93,75%, akurasi label ramai sebesar 100%, akurasi label cukup ramai sebesar 75%, dan akurasi tidak ramai sebesar 96%. Dengan demikian, model dinyatakan cukup ampuh dalam mengklasifikasikan data dengan kategori tidak ramai, sementara untuk kategori ramai dan cukup ramai, model tidak terlalu baik dalam mengklasifikasikan data dikarenakan data yang didapat memiliki perbandingan kategori yang kurang merata

Kata kunci: Data Mining, K-Nearest Neighbor, Klasifikasi, Kunjungan Wisatawan

Abstract

Tourism is an important sector in the regional economy, including in Riau Province, which has more than 200 tourist destinations. With so many tourist destinations, not all of them have adequate and good infrastructure. This is because there is no equal classification of tourist destinations, resulting in uneven development. This study aims to classify existing tourist destinations in Riau Province into several categories. With proper classification, the Riau Tourism Department can allocate funds for more equitable development of tourist sites. K-Nearest Neighbor was used in this study as it is considered sufficiently effective for classification. Using visitor data from the period 2021 to 2023, the data underwent preprocessing, labeling using label encoding, and categorization with min-max scaling. When dividing the data into 80% training data and 20% testing data, an accuracy of 93.75% was achieved, with 100% accuracy for the "busy" label, 75% accuracy for the "moderately busy" label, and 96% accuracy for the "not busy" label. Thus, the model is deemed sufficiently effective in classifying data into the "not crowded" category, while for the 'crowded' and "moderately crowded" categories, the model is not as effective in classifying data due to the uneven distribution of categories in the data obtained.

Keywords: Data Mining, K-Nearest Neighbor, Classification, Tourist Visits

1. Pendahuluan

Pariwisata menjadi salah satu sektor unggulan di berbagai negara dan memberikan dampak positif bagi pembangunan pada banyak setkor, baik itu sektor budaya atau pertumbuhan dari suatu negara [1]. Dengan banyaknya keanekaragaman budaya, keindahan alam, dan warisan sejarah yang ada di Indonesia, ini semua dapat menjadi daya tarik tersendiri bagi wisatawan lokal atau mancanegara untuk menjadikan Indonesia sebagai salah satu destinasi wisata mereka, terutama di Provinsi Riau. Provinsi Riau sendiri memiliki berbagai macam destinasi wisata, mulai dari wisata kebudayaan, wisata alam, atau wisata buatan manusia yang tersebar di setiap Kabupaten/Kota yang ada di Provinsi Riau. Berdasarkan data yang tercatat pada Dinas Pariwisata Provinsi Riau, terdapat kurang lebih 350 tempat wisata yang ada pada masing-masing Kabupaten/Kota di Provinsi Riau pada tahun 2023. Dengan banyaknya objek wisata tersebut, Provinsi Riau kini sedang gencar-gencarnya mengembangkan potensi wisata pada seluruh daerah di Riau. Dengan pengembangan yang gencar ini, Provinsi Riau akan mendapatkan dampak positif dan negatif. Dampak negatif dari pembangunan objek wisata tentu saja alam sekitarnya yang akan dikorbankan, namun pengorbanan tersebut akan setimpal karena akan ada lapangan pekerjaan baru yang akan terbentuk dan akan mengurangi tingkat pengangguran yang ada pada daerah tersebut [2]. Dengan banyaknya objek wisata yang akan dibangun, Dinas Pariwisata Provinsi Riau memerlukan strategi yang mendukung untuk pembangunan yang ada, maka dari itu dibutuhkan klasifikasi yang tepat pada setiap tempat wisata agar pembangunan lebih didahulukan pada daerah yang tertinggal.

K-Nearest Neighbor yang merupakan salah satu algoritma klasifikasi dirasa cocok untuk melakukan klasifikasi tempat wisata yang ada di Provinsi Riau. Hal ini didasari oleh penelitian yang dilakukan pada SMA Negeri 15 Pekanbaru dalam mengklasifikasikan jurusan siswa. Peneliti menggunakan algoritma *K-Nearest Neighbor* berdasarkan nilai ujian psikotes dan nilai raport saat SMP. Dengan menggunakan $K = 3$, didapatkan akurasi sebesar 93,52% yang berarti *K-Nearest Neighbor* cukup baik dalam klasifikasi jurusan siswa di SMA Negeri 15 Pekanbaru [3]. Selain itu, pada penelitian lain yang dilakukan untuk mengklasifikasikan penerima beasiswa di SMA, dengan menggunakan data latih dan data uji yang didapat dari data penerima beasiswa tahun 2017 hingga 2019. Didapatkan hasil akurasi sebesar 90,5% yang berarti *K-Nearest Neighbor* dapat memberikan klasifikasi yang akurat untuk seleksi penerimaan beasiswa [4]. Pada penelitian lain yang bertujuan untuk mengklasifikasikan data iris bunga, digunakan 135 data iris bunga dengan 4 karakteristik utama. Pengujian dilakukan dengan beberapa variasi nilai K , dan didapati akurasi mencapai 100% yang berarti *K-NN* efektif dalam klasifikasi data iris bunga [5]. Pada penelitian yang membahas pencemaran udara di Jakart, digunakan algoritma *K-NN* berdasarkan data Indeks Standar Pencemaran Udara (ISPU). Dengan menggunakan 304 data latih dan satu data uji serta nilai $K = 7$, didapatkan akurasi sebesar 95,78%. Hasil tersebut membantu memberikan informasi kepada masyarakat dan Dinas Lingkungan Hidup mengenai polusi udara di Jakarta [6]. Pada penelitian yang bertujuan untuk memprediksi penerima bantuan desa, digunakan algoritma *K-NN* dengan *dataset* yang terdiri dari 159 data. Dengan menggunakan beberapa nilai K , menunjukkan bahwa $K = 15$ dan $K = 30$, data baru dikategorikan tidak layak dengan akurasi 100%, sedangkan $K = 45$, $K = 60$, dan $K = 75$ dikategorikan layak dengan akurasi 81,25% [7]. Tujuan dari penelitian ini adalah untuk melakukan klasifikasi kunjungan wisatawan di Provinsi Riau berdasarkan data kunjungan wisatawan pada tahun 2021 hingga 2023. *K-Nearest Neighbor* akan digunakan dalam mengklasifikasikan kunjungan wisatawan Provinsi Riau karena terbukti unggul dalam menangani data numerik dan menghasilkan klasifikasi yang akurat.

2. Landasan Teori

2.1. Data Mining

Data mining merupakan serangkaian proses untuk mencari sebuah data dari kumpulan data yang banyak dan besar yang sudah ada sejak masa lampau. *Data mining* sendiri masuk ke dalam langkah dari *Knowledge Discovery in Database* (KDD) [8]. Terdapat 5 tahapan yang harus

dilakukan pada proses *data mining* seperti: data selection, data preprocessing, transformation, data modeling, evaluasi.

2.1.1. Data Selection

Proses pertama ini merupakan proses untuk mencari data yang relevan. Jika penelitian berfokus pada kunjungan wisata, maka diperlukan data kunjungan wisatawan [9].

2.1.2 Data Preprocessing

Data yang sudah ditemukan, akan diolah kembali agar data yang digunakan menjadi bersih. Dapat dilakukan integrasi data atau penggabungan seluruh data, membersihkan data yang tidak lengkap, atau melakukan penanganan *outliers* agar tidak ada ketimpangan data, sehingga proses perhitungan menjadi lebih akurat [9].

2.1.3. Transformation

Tahapan ini ketika seluruh data sudah diintegrasikan. Data yang belum memiliki atribut kategori akan diberikan kategori pada proses ini agar nantinya data dapat digunakan dalam proses klasifikasi [9].

2.1.4. Data Modeling

Pada tahapan ini, data akan digunakan dan diimplementasikan pada metode yang ada. Ada banyak metode yang dapat digunakan, yaitu *K-Nearest Neighbor*, *Naive Bayes*, dan lain sebagainya [9].

2.1.5. Evaluasi

Evaluasi merupakan tahapan terakhir, yang di mana pada proses ini akan ada proses menerjemahkan pola-pola dari *data mining*. Menguji apakah pola atau informasi yang ditemukan sesuai dengan fakta atau tidak [9].

2.2. Machine Learning

Machine learning merupakan sebuah komponen dari kecerdasan buatan yang digunakan untuk mengatasi berbagai masalah kompleks dengan berbagai pendekatan. *Machine learning* memiliki berbagai kategori, yang pertama merupakan *supervised learning* untuk memanfaatkan data berlabel dalam melatih model. Ada *unsupervised learning* yang digunakan untuk memanfaatkan data tidak berlabel dalam melatih model. Lalu juga ada *reinforcement learning* yang bekerja dengan memanfaatkan *feedback* dari lingkungan kerja untuk meningkatkan pengambilan keputusan agar semakin optimal. Semua kategori tersebut tidak hanya relevan, tetapi akan mengalami perkembangan berkelanjutan seiring berjalannya waktu [10].

2.3. Website

Website merupakan kumpulan halaman yang terdiri dari beberapa halaman yang isinya berupa informasi dalam bentuk digital, baik itu dalam bentuk gambar, teks, atau animasi yang disediakan melalui jaringan internet. Pada awalnya website merupakan suatu layanan sajian informasi yang menggunakan konsep hyperlink, hal ini dapat memudahkan pengguna internet untuk melakukan pencarian informasi di internet. Informasi yang disajikan pada website biasanya menggunakan konsep multimedia, informasi dapat berupa teks, gambar, animasi, suara, atau film. Dengan demikian, website dapat lebih menjangkau semua umur dan pengguna yang ada di internet tanpa harus mengeluarkan biaya yang besar dalam penggunaannya [11].

2.5. Klasifikasi

Klasifikasi merupakan salah satu teknik yang sangat penting dalam *data mining* yang berfungsi untuk mempelajari model atau fungsi yang bertujuan untuk mengaitkan setiap himpunan atribut dari suatu objek dengan salah satu label kelas yang sudah ditentukan sebelumnya. Terdapat beberapa algoritma yang dapat digunakan dalam melakukan klasifikasi, di antaranya adalah regresi logistik, pohon keputusan, *random forest*, *Support Vector Machine (SVM)*, *K-Nearest Neighbor (K-NN)*, dan *Naive Bayes* [12]. Proses dari klasifikasi sendiri terdiri dari 2 proses, yaitu melakukan *training data* dan melakukan *testing data*. Yang di mana pada proses *training data* akan menggunakan data yang sudah diberikan label, sementara pada *testing data*, yang yang sudah dilatih akan diuji untuk menentukan akurasi dari model yang sudah diciptakan layak atau tidak layak, maka dari itu diperlukan proses *testing data* untuk melakukan perhitungan akurasi [13].

2.6. K-Nearest Neighbor

K-Nearest Neighbor merupakan algoritma yang dapat dijadikan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat engan objek tersebut. *K-Nearest Neighbor* bekerja ketika ada sebuah data yang diinputkan tetapi tidak memiliki label, maka *K-Nearest Neighbor* akan mencari nilai K buah pada data pembelajaran yang jaraknya paling dekat dengan data yang diinputkan [14]. Tujuan utama dari algoritma *K-Nearest Neighbor* adalah melakukan klasifikasi terhadap objek baru berdasarkan atribut dan data pembelajaran. Algoritma *K-Nearest Neighbor* menggunakan klasifikasi tetangga terdekat sebagai nilai prediksi pada data uji yang baru. Dekat atau jauhnya tetangga ini dihitung berdasarkan metrik jarak [15]. Terdapat beberapa metrik yang dapat digunakan, salah satunya ialah *euclidean distance*. Berikut merupakan persamaan *euclidean distance*:

$$d_e = \sqrt{\sum_{k=1}^m (fd_{i,k} - k_j)^2} \quad (2.1)$$

Dimana: d_e = Jarak *euclidean*
 fd_i = Data latih
 k = Data uji
 m = Jumlah data

3. Metode Penelitian

Berikut merupakan beberapa tahapan yang dilakukan pada penelitian ini. Terdapat beberapa tahapan, antara lain:

3.1. Pengumpulan Data

Pengumpulan data merupakan proses awal dan krusial dalam penelitian yang bertujuan untuk mendapatkan informasi yang relevan dan akurat dari sumber yang dilakukan. Data yang digunakan pada penelitian ini didapat melalui proses wawancara dengan Dinas Pariwisata Provinsi Riau, data yang didapat berupa data kunjungan wisatawan yang tersebar di Provinsi Riau mulai dari tahun 2021 hingga 2023, meliputi jumlah kunjungan wisatawan setiap bulan, nama kota, dan nama tempat wisata yang dikunjungi.

3.2. Preprocessing Data

Data yang sudah didapat akan masuk ke dalam tahapan preprocessing data. Pada tahapan ini, seluruh data akan diintegrasikan menjadi satu data. Setelah menjadi satu, akan dilakukan cleaning pada data yang hilang (*missing values*), mengoreksi data yang salah, dan mendeteksi outliers yang nantinya akan ditangani agar akurasi menjadi lebih baik. Setelah data dibersihkan, akan kembali dilakukan validasi pada data yang ada apakah sudah sesuai atau belum untuk perhitungan yang akan dilakukan.

3.3. Transformasi Data

Pada tahapan ini akan dilakukan proses pemberian kategori, hal ini dikarenakan pada data yang digunakan tidak memiliki kategori sama sekali. Maka dari itu, untuk melatih model nantinya diperlukan kategori agar model dapat melakukan klasifikasi. Selain itu, beberapa data yang masih memiliki format kategorikal akan diubah menjadi format numerik agar perhitungan dapat berjalan dengan baik.

3.4. Implementasi K-Nearest Neighbor

Setelah data sudah dibersihkan dan diberikan kategori pada proses transformasi data, data akan dapat diolah menggunakan algoritma K-Nearest Neighbor. Sebelum data diolah, akan terdapat proses pembagian data, yang di mana perbandingan yang dapat digunakan antara data latih dan data uji bisa 80:20, 70:30, atau 60:40. Klasifikasi akan dilakukan menggunakan perhitungan metrik *euclidean distance* yang di mana akan mencari K tetangga terdekat. Metode ini dipilih karena kesederhanaannya dan kemampuannya dalam penanganan data dengan pola yang tidak linier.

3.5. Evaluasi Model

Setelah model terbaik dipilih, maka langkah berikutnya ialah melakukan evaluasi terhadap model yang sudah dibangun. Proses ini bertujuan untuk meninjau kinerja model yang ada dalam melakukan klasifikasi data. Metode yang digunakan dalam evaluasi model adalah confusion matrix yang memiliki beberapa metrik perhitungan, antara lain metrik akurasi, metrik presisi, metrik recall, dan metrik f1-score.

3.6. Interpretasi Hasil

Hasil dari interpretasi hasil akan dikelompokkan berdasarkan 3 kategori, yaitu kategori ramai, kategori cukup ramai, atau kategori tidak ramai.

4. Hasil dan Pembahasan

Hasil dari penelitian merupakan sebuah *website* yang berfungsi untuk melakukan klasifikasi terhadap data kunjungan wisatawan yang sudah dilakukan proses *preprocessing data*. Pada sistem ini, tidak ada proses *preprocessing data*, jadi data yang digunakan merupakan data yang sudah dibersihkan melalui program pengolahan data seperti Microsoft Excel. Sebelum menjelaskan sistem yang akan dirancang, di sini peneliti akan menjelaskan cara kerja algoritma *K-Nearest Neighbor* menggunakan *euclidean distance*.

4.1. Perhitungan Manual

Perhitungan manual akan menggunakan 10 data latih dan 1 data uji yang akan diklasifikasikan masuk ke kelas mana. Datanya dapat di lihat pada tabel di bawah ini.

Tabel 1 : Data Perhitungan

No	Kota	Jenis	2021	2022	2023	Kategori
1	0	0.600	0.208	0.236	0.484	1
2	0	0.200	0.877	0.420	0.397	2
3	0	0.200	0.103	0.122	0.111	0
4	0	0.200	0.108	0.111	0.042	0
5	0.364	0.800	0.098	0.667	0.289	1
6	0.364	0.800	0.398	0.636	0.374	2
7	0.909	0.700	0.051	0.010	0.136	0
8	0.909	0.700	0.576	0.375	0.230	1
9	0.909	0.300	0.117	0.455	1	2
10	0.636	0.700	0.588	0.430	0.354	2
11	1	0.700	0.413	1	0.403	?

Data tersebut terdiri dari 11 data tempat wisata, yang di mana:

1. Harationica Waterpark Duri,
2. Pantai Indah Selatbaru,
3. Pantai Ketapang Rupert,
4. Pantai Pesona Pulau Rupert,
5. Ulu Kasok,
6. Puncak Kompe,
7. Hutan Kota,
8. Kahati,
9. Pasar Modern,
10. Taman Bunga Impian Okura,
11. Wisata Embung Terpadu.

Data nomor 1 hingga 10 akan digunakan sebagai data latih karena sudah memiliki kategori, sementara data nomor 11 akan digunakan sebagai data uji karena belum diketahui apa kategorinya. K yang digunakan pada perhitungan ini ialah $K = 3$, yang di mana akan diambil 3 data terdekat dengan data uji yang sudah dihitung menggunakan *euclidean distance*. Berikut merupakan hasil perhitungan jarak dengan *euclidean distance*.

1. $\sqrt{(0-1)^2 + (0,6-0,7)^2 + (0,208-0,413)^2 + (0,236-1)^2 + (0,484-0,403)^2} = 1,282$
2. $\sqrt{(0-1)^2 + (0,2-0,7)^2 + (0,877-0,413)^2 + (0,420-1)^2 + (0,397-0,403)^2} = 1,342$
3. $\sqrt{(0-1)^2 + (0,2-0,7)^2 + (0,103-0,413)^2 + (0,122-1)^2 + (0,111-0,403)^2} = 1,484$
4. $\sqrt{(0-1)^2 + (0,2-0,7)^2 + (0,108-0,413)^2 + (0,111-1)^2 + (0,042-0,403)^2} = 1,505$
5. $\sqrt{(0,364-1)^2 + (0,8-0,7)^2 + (0,098-0,413)^2 + (0,667-1)^2 + (0,289-0,403)^2} = 0,799$
6. $\sqrt{(0,364-1)^2 + (0,8-0,7)^2 + (0,398-0,413)^2 + (0,636-1)^2 + (0,374-0,403)^2} = 0,741$
7. $\sqrt{(0,909-1)^2 + (0,7-0,7)^2 + (0,051-0,413)^2 + (0,010-1)^2 + (0,136-0,403)^2} = 1,091$
8. $\sqrt{(0,909-1)^2 + (0,7-0,7)^2 + (0,576-0,413)^2 + (0,375-1)^2 + (0,230-0,403)^2} = 0,674$
9. $\sqrt{(0,909-1)^2 + (0,3-0,7)^2 + (0,117-0,413)^2 + (0,455-1)^2 + (1-0,403)^2} = 0,953$
10. $\sqrt{(0,636-1)^2 + (0,7-0,7)^2 + (0,588-0,413)^2 + (0,430-1)^2 + (0,354-0,403)^2} = 0,700$

Berikut data hasil perhitungan *euclidean distance* yang dimasukkan ke dalam tabel.

Tabel 2 : Tabel Perhitungan Euclidean Distance

No	Jarak
1	1.282
2	1.342
3	1.484
4	1.505
5	0.799
6	0.741
7	1.091
8	0.674
9	0.953
10	0.700

Berikut merupakan hasil pengurutan jarak dari kecil hingga besar.

Tabel 3 : Tabel Pengurutan Jarak

Urutan	Data	Jarak
1	8	0.674
2	10	0.700
3	6	0.741
4	5	0.799
5	9	0.953
6	7	1.091
7	1	1.282
8	2	1.342
9	3	1.484
10	4	1.505

Dari pengurutan data tersebut, dapat diketahui 3 tetangga terdekat dan akan menjadi penentuan untuk kategori pada data uji atau data nomor 11, datanya dapat dilihat sebagai berikut.

Tabel 4 : Tabel Penentuan Kategori

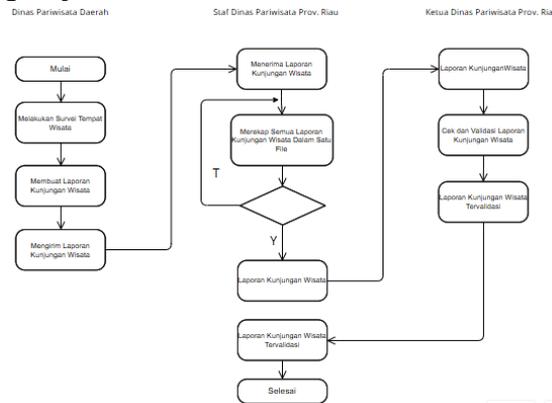
Urutan	Data	Jarak	Kategori
1	8	0.674	1
2	10	0.700	2
3	6	0.741	2

Dari tabel tersebut, terlihat 3 data dengan jarak terdekat, yaitu data nomor 8 yang merupakan tempat wisata Kahati, data nomor 10 yang merupakan tempat wisata Taman Bunga Impian Okura, dan data nomor 6 yang merupakan tempat wisata Puncak Kompe. Data nomor 8

memiliki kategori 1 atau cukup ramai, sedangkan data nomor 10 dan 6 memiliki kategori 2 atau ramai. Dengan demikian, kategori 2 atau ramai muncul sebanyak 2 kali yang dan menjadi kategori yang dominan untuk nilai $K = 3$, maka kategori data nomor 11 atau Wisata Embung Terpadu merupakan kategori 2 atau ramai.

4.2. Diagram Aliran Sistem Informasi Berjalan

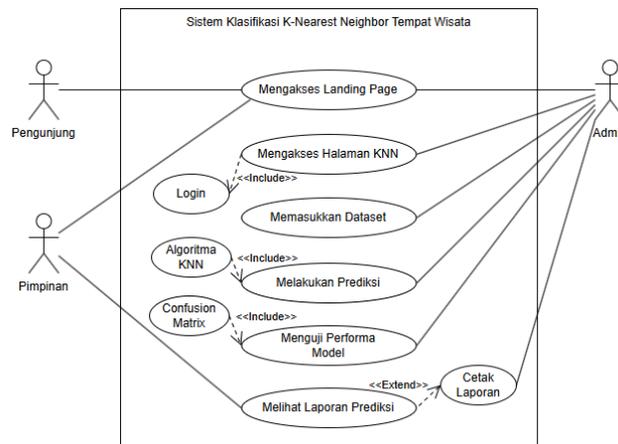
Berikut merupakan alur proses data kunjungan wisatawan diterima berdasarkan peran *stakeholders* yang sedang berjalan.



Gambar 1. Alur Proses Penerimaan Data Kunjungan Wisatawan

4.3. Use Case Diagram

Untuk menggambarkan interaksi antar aktor dengan sistem yang sedang dikembangkan perlu digambarkan dengan *use case*, berikut *use case* sistem yang dirancang.



Gambar 2. Use Case Diagram

Use case tersebut terdiri dari 3 aktor dan 7 *use case*, yang di mana aktornya adalah pengunjung, pimpinan, dan *admin*. *Landing page* sendiri dapat diakses oleh 3 aktor yang ada, namun lebih ditujukan kepada pengunjung. Sementara untuk mengakses halaman *K-NN* hanya dapat diakses oleh *admin* selaku petugas yang ada. Untuk masuk ke dalam halaman *K-NN* diperlukan akses *login*. Pada halaman *K-NN* *admin* dapat memasukkan *dataset* yang akan digunakan untuk perhitungan. Setelah *dataset* dimasukkan, maka *admin* akan dapat melakukan prediksi menggunakan algoritma *K-NN*. Untuk menguji model yang ada, disediakan halaman untuk menguji performa model, untuk pengujian performa model, sistem menggunakan *confusion matrix*. Dari semua kegiatan yang ada, hasil prediksi dapat dicetak dalam bentuk PDF yang nantinya petugas dapat memberikannya kepada pimpinan. Berikut merupakan penjabaran aktor pada *use case* dan definisi *use case* yang berkaitan dengan sistem yang dirancang:

Tabel 5 : Definisi Aktor Use Case Diagram

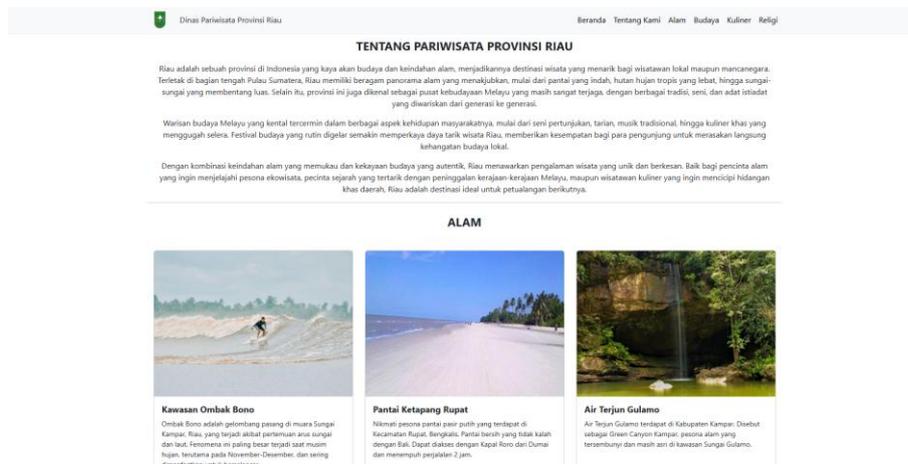
No	Nama Aktor	Deskripsi
1	Admin	Aktor yang mengelola <i>website</i> , dapat melakukan beberapa tugas utama yang ada di <i>website</i> , yaitu melakukan klasifikasi.
2	Pengunjung	Aktor yang hanya dapat mengakses <i>landing page</i> yang berisi informasi mengenai tempat wisata di Provinsi Riau.
3	Pimpinan	Aktor yang menjadi atasan <i>admin</i> , dapat menerima laporan hasil klasifikasi, menjadi pengambil keputusan terhadap perkembangan pariwisata Provinsi Riau berdasarkan hasil klasifikasi.

Tabel 6 : Definisi Use Case Diagram

Kode	Use case	Deskripsi
UC_1	Mengakses <i>landing page</i>	<i>Landing page</i> merupakan halaman yang berisi informasi tempat wisata yang ada di Prov. Riau. Pengunjung dapat mengaksesnya untuk mendapatkan informasi tempat wisata yang ada di Provinsi Riau.
UC_2	Melakukan <i>Login</i>	Proses di mana <i>admin</i> akan melakukan proses masuk ke dalam halaman klasifikasi dengan <i>KNN</i> .
UC_3	Mengakses halaman <i>KNN</i>	Halaman yang menjadi fungsi utama dari sistem yang dibangun. Dapat diakses oleh admin untuk melakukan klasifikasi dengan <i>KNN</i> .
UC_4	Memasukkan <i>dataset</i>	<i>Usecase</i> yang dilakukan untuk memasukkan <i>dataset</i> . Dengan <i>dataset</i> yang ada, dapat dilakukan untuk melakukan klasifikasi kunjungan wisatawan.
UC_5	Melakukan prediksi	<i>Usecase</i> yang dilakukan untuk melakukan prediksi kunjungan wisatawan. Prediksi dilakukan menggunakan algoritma <i>KNN</i> .
UC_6	Menguji performa model	<i>Usecase</i> yang dilakukan untuk menguji performa model <i>KNN</i> yang sudah dibangun. Pengujian menggunakan metrik pengujian, seperti <i>confusion matrix</i> .
UC_7	Melihat laporan prediksi	<i>Usecase</i> yang dilakukan untuk melihat data laporan hasil prediksi. Laporan terlebih dahulu dicetak oleh <i>admin</i> , yang nantinya akan diserahkan ke pimpinan.

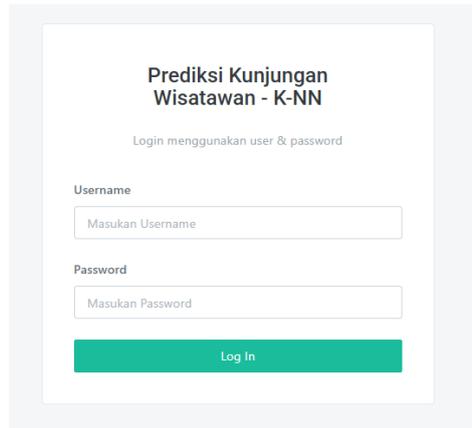
4.4. Implementasi Sistem

Dari sistem yang sudah dirancang, berikut merupakan hasil akhir sistemnya sesuai dengan kebutuhan pada *use case diagram* dengan menerapkan algoritma *K-Nearest Neighbor*.



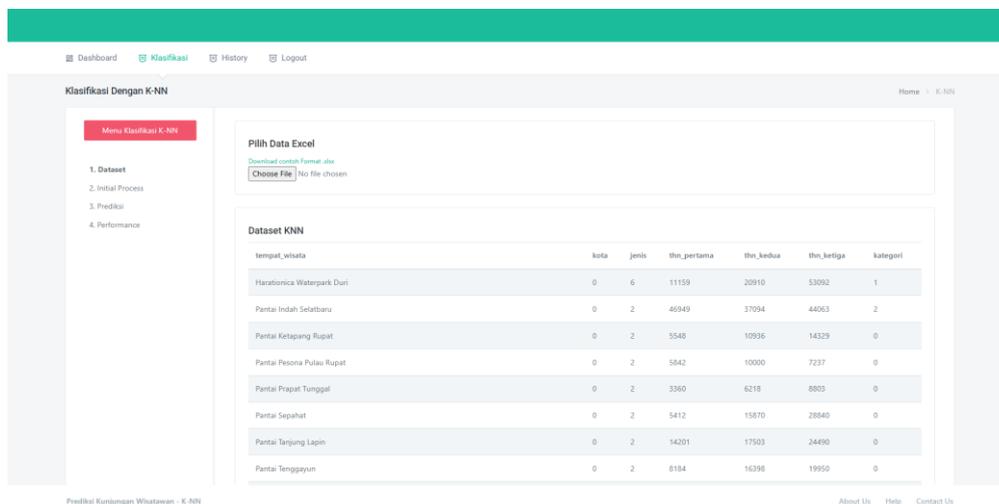
Gambar 3. Landing Page

Gambar 3 merupakan Landing page yang berguna untuk menampilkan informasi tempat wisata yang ada di Provinsi Riau.



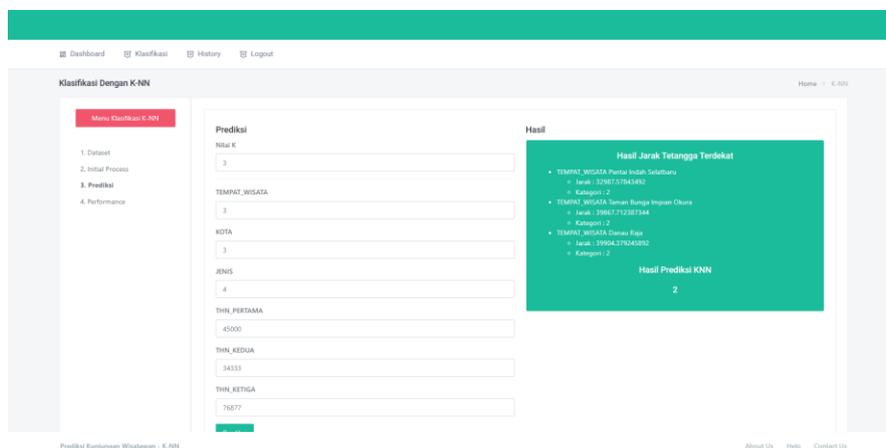
Gambar 4. Halaman *Login*

Gambar 4 merupakan halaman *login*, terdapat beberapa kotak teks yang harus diisi agar pengguna dapat masuk ke halaman utama klasifikasi.



Gambar 5. Halaman Memasukkan Dataset

Gambar 5 merupakan halaman pengguna untuk memasukkan *dataset* yang akan digunakan pada proses klasifikasi menggunakan algoritma *K-Nearest Neighbor*.



Gambar 6. Halaman Prediksi

Gambar 6 merupakan halaman prediksi, pengguna dapat melakukan prediksi pada data yang belum memiliki kategori berdasarkan *dataset* yang sudah dimasukkan.

Performa Confusion Matrix

Label: Tidak Ramai >> PREDICT Tidak Ramai: 12 PREDICT Cukup Ramai: 0 PREDICT Ramai: 0

Label: Cukup Ramai >> PREDICT Tidak Ramai: 1 PREDICT Cukup Ramai: 1 PREDICT Ramai: 0

Label: Ramai >> PREDICT Tidak Ramai: 0 PREDICT Cukup Ramai: 0 PREDICT Ramai: 2

Label: Tidak Ramai
Accuracy: 96
Precision: 100
Recall: 92.307692307692
F1-Score: 96

Label: Cukup Ramai
Accuracy: 75
Precision: 50
Recall: 100
F1-Score: 66.666666666667

Label: Ramai
Accuracy: 100
Precision: 100
Recall: 100
F1-Score: 100

Gambar 7. Halaman Evaluasi Model

Gambar 7 merupakan halaman evaluasi model, terdapat hasil akurasi model yang sudah dibangun. Dengan menggunakan pembagian antara data latih dan data uji sebesar 80:20, terlihat bahwa akurasi untuk kategori tidak ramai sebesar 96%, akurasi kategori cukup ramai sebesar 75%, dan akurasi kategori ramai sebesar 100%.

Dashboard | Klasifikasi | History | Logout

Home | Dashboard

History Prediksi

Start Date: End Date:

Data ID	Tempat Wisata	Kota	Jenis	Tahun Pertama	Tahun Kedua	Tahun Ketiga	Kategori	Create Date
3	3	3	4	45000	34333	76877	2	31 Jan 2025 13:55:15
4	3	3	4	45000	34333	76877	2	06 Feb 2025 22:49:40

Prediksi Kunjungan Wisatawan - K-NN About Us | Help | Contact Us

Gambar 8. Halaman Laporan Hasil Prediksi

Gambar 8 merupakan halaman yang berisi hasil prediksi yang pernah dilakukan, data ini dapat dicetak menjadi file PDF yang bisa diberikan ke Pimpinan.

4. Kesimpulan

Pengimplementasian metode *K-Nearest Neighbor* menggunakan sistem berbasis *website* dapat mengolah *dataset* kunjungan wisatawan dengan sangat baik. Tidak terdapat kesalahan yang banyak saat sistem dibangun. Evaluasi hasil menggunakan *confusion matrix* untuk menguji akurasi metode *K-Nearest Neighbor* juga menunjukkan hasil yang baik. Di mana dengan menggunakan perbandingan data latih dan data uji sebesar 80:20, mendapatkan akurasi untuk

label tidak ramai sebesar 96%, label cukup ramai sebesar 75%, dan label ramai sebesar 100%. Untuk label cukup ramai, model sedikit sulit mengklasifikasikannya karena data yang digunakan memiliki ketimpangan yang cukup signifikan.

Daftar Pustaka

- [1] M. H. Aziz, "Model Pariwisata Digital dalam Pengembangan Pariwisata Indonesia," *J. Ilm. Univ. Batanghari Jambi*, vol. 22, no. 3, p. 2279, 2022, doi: 10.33087/jiubj.v22i3.2246.
- [2] L. V Tobing, "Strategi Dinas Pariwisata Provinsi Riau Dalam Meningkatkan Kunjungan Wisata Di Provinsi Riau," 2020, [Online]. Available: <https://repository.uin-suska.ac.id/40711/>
- [3] Q. A. A'yuniyah and M. Reza, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru," *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 39–45, 2023, doi: 10.57152/ijirse.v3i1.484.
- [4] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.
- [5] P. Putra, A. M. H. Pardede, and S. Syahputra, "Analisis Metode K-Nearest Neighbour (Knn) Dalam Klasifikasi Data Iris Bunga," *J. Tek. Inform. Kaputama*, vol. 6, no. 1, pp. 297–305, 2022.
- [6] Nurjanah Siti, Siregar Amril Mutoi, and Kusumaningrum Dwi Sulistya, "Penerapan Algoritma K – Nearest Neighbor (KNN) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta," *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 2, pp. 71–76, 2020, [Online]. Available: <http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/14>
- [7] R. L. Hasanah, M. Hasan, W. E. Pangesti, F. F. Wati, and W. Gata, "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor)," *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 1–6, 2019, doi: 10.33480/techno.v16i1.25.
- [8] R. Syahri and D. Puspita, "Klasifikasi Pengunjung Wisata Di Kota Pagar Alam Dengan Menggunakan Algoritma K-Nearest Neighbor (K-Nn)," *J. Ilm. Teknosains*, vol. 9, no. 2, 2023.
- [9] A. Muhadi and A. Octaviano, "Penerapan Data Mining Untuk Prediksi Hasil Keuntungan Lelang Mesin X-Ray Tahun 2020 Dengan Metode K-Nearest Neighbor (Studi Kasus : PT.Ramadika Mandiri)," *J. Inform. MULTI*, vol. 1, no. 2, pp. 2985–8860, 2023, [Online]. Available: <https://jurnal.publikasitecno.id/index.php/jim126>
- [10] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
- [11] I. Arthalita and R. Prasetyo, "Penggunaan Website Sebagai Sarana Evaluasi Kegiatan Akademik Siswa Di Sma Negeri 1 Punggur Lampung Tengah," *JIKI (Jurnal Ilmu Komput. Informatika)*, vol. 1, no. 2, pp. 93–108, 2020, doi: 10.24127/jiki.v1i2.678.
- [12] R. P. Pridiptama, W. Wasono, and F. D. . Amijaya, "Perbandingan Algoritma Support Vector Machine dan NaïveNaïve Bayes pada Klasifikasi Penyakit Tekanan Darah Tinggi (Studi Kasus: Klinik Polresta Samarinda)," *Basis*, vol. 3, no. 1, pp. 1–16, 2024, [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/Basis>
- [13] A. David Imanuel, N. Nawaningtyas Pusparini, and A. Sani, "Klasifikasi Untuk Memprediksi Tingkat Kelulusan Mahasiswa Stmik Widuri Menggunakan Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 12, no. 01, pp. 1–7, 2024, doi: 10.33884/jif.v12i01.8201.
- [14] A. Ghani Muttaqin, K. Auliasari, and F. Santi Wahyuni, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Berbasis Web Pada Pt.Wika Industri Energy," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 4, no. 2, pp. 1–6, 2020, doi: 10.36040/jati.v4i2.2728.
- [15] L. Suryadi, N. Ngajiyanto, N. E. Pratiwi, F. Ardhy, and P. Riswanto, "Penerapan Data

Mining Prediksi Penjualan Mebel Terlaris Menggunakan Metode K-Nearest Neighbor(K-Nn) (Studi Kasus : Toko Zerita Meubel),” *JUSIM (Jurnal Sist. Inf. Musirawas)*, vol. 7, no. 2, pp. 174–184, 2022, doi: 10.32767/jusim.v7i2.1697.