

Klasifikasi Penumpang Kereta Api DAOP 6 Yogyakarta Berdasarkan Kelas Stasiun Menggunakan KNN

Ichsanudin¹, Supatman²

^{1,2}Informatika, Universitas Mercu Buana Yogyakarta; Gg. Jemb. Merah No.84C, Soropadan, Condong Catur Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55283, telp. (0274) 5635, Faks (0274) 550703

³Informatika, Fakultas Teknologi Informasi, Yogyakarta

e-mail: *¹211110125@student.mercubuana-yogya.ac.id, ²supatman@mercubuana-yogya.ac.id,

Abstrak

Perusahaan Transportasi terus melakukan adaptasi terhadap perkembangan teknologi untuk meningkatkan layanan kepada pengguna jasa. Kereta Api adalah transportasi masal yang paling dimintasi pengguna jasa saat ini. Karena tingkat ketepatan waktu, kenyamanan dan bebas macet sehingga Kereta Api menjadi moda transportasi andalan para pengguna jasa. Semakin banyak pengguna jasa tentunya kereta api harus meningkatkan layanan guna meningkatkan pelayanan. Dalam penelitian ini dilakukan klasifikasi penumpang kereta api untuk menganalisa jumlah penumpang di stasiun. Penelitian ini dilakukan menggunakan metode K-Nearest Neighbor dalam menentukan jumlah penumpang berdasarkan kelas stasiun. Metode K-Nearest Neighbor adalah Teknik mencari anggota k target dalam data (training) yang terdekat pada data uji. Dataset dalam penelitian ini menggunakan data yang bersumber dari Unit Angkutan Penumpang DAOP 6 Yogyakarta PT. Kereta Api Indonesia dari volume naik dan turun penumpang dari tahun 2016 sampai tahun 2023. Hasil klasifikasi dengan metode K-Nearest neighbor memperoleh hasil yang sangat baik dengan tingkat akurasi 93%.

Kata kunci—Dataset, Klasifikasi, Kereta Api, K-Nearest Neighbor, DAOP 6 yogyakarta

Abstract

Transportation companies continue to adapt to technological developments to improve services to service users. The train is the most crossed mass transportation for service users today. Because of the level of timeliness, comfort and traffic-free so that the train becomes the mainstay mode of transportation for service users. The more service users, of course, the train must improve services to improve services. Therefor The author wants to conduct research on the classification of train passengers, the classification algorithm is used to analyze the number of passengers at the station. This research was conducted using the K-Nearest Neighbor method in determining the number of passengers based on the station class. The K-Nearest Neighbor method is a technique for finding the k target members in the data (training) that are closest to the test data. The dataset in this study uses data sourced from the Passenger Transport Unit DAOP 6 Yogyakarta PT Kereta Api Indonesia from the volume of up and down passengers from 2016 to 2023. The classification results with the K-Nearest neighbor method obtained very good results with an accuracy rate of 93%.

Keywords—Dataset, Classification, Train, K-Nearest Neighbor, Transportation

1. PENDAHULUAN

Transportasi adalah proses memindahkan orang, barang dari suatu tempat ke tempat yang lain [1]. Transportasi meliputi transportasi darat, laut, dan udara. Kereta api adalah transportasi darat dengan perpindahan orang atau barang bersifat masal. Kereta api terus meningkatkan layanan dan adaptif terhadap perkembangan teknologi.

Kereta Api Indonesia telah bertransformasi menjadi transportasi yang nyaman dan diminati masyarakat, kereta api yang dahulu penumpang berdesak-desakan, penjual asongan yang tidak terkontrol, tingkat keamanan yang kurang, fasilitas stasiun dan sarana kereta yang kurang nyaman itu sudah tidak ditemukan lagi. Kereta api di semua kelas sekarang sudah terpasang AC, penumpang sesuai tempat duduk yang tersedia, sehingga tidak ada lagi penumpang yang berdiri apalagi berdesakan. Fasilitas di stasiun juga ditingkatkan, toilet yang bersih, pembangunan peron tinggi untuk memudahkan naik turun penumpang, pembangunan fasilitas stasiun, semua dilakukan oleh PT.KAI untuk meningkatkan pelayanan terhadap pelanggan.

Kereta api Indonesia DAOP 6 Yogyakarta merupakan daerah operasi yang terletak di Jl. Lempuyangan No.1, Yogyakarta. Wilayahnya meliputi ujung barat yaitu stasiun jenar, wates, yogyakarta, lempuyangan, Solo, sragen, dan ujung timur yaitu stasiun Kedung Banteng. DAOP 6 Yogyakarta termasuk daerah operasi dengan kepadatan penumpang yang tinggi, hal ini tentunya fasilitas stasiun yang berada di DAOP 6 Yogyakarta harus ditingkatkan, salah satu fasilitasnya adalah Face Recognition yaitu sistem boarding dengan deteksi wajah yang terintegrasi dengan data tiket penumpang, proses ini hanya memerlukan waktu 1 detik saja sehingga mengurangi antrian saat boarding gate. Fasilitas ini tersedia di Stasiun Yogyakarta, Solo Balapan, dan stasiun Lempuyangan. Fasilitas berikutnya adalah cetak tiket mandiri, sehingga calon penumpang bisa mencetak sendiri tiket yang sudah dipesan sebelumnya melalui online. Semua fasilitas tersebut adalah bentuk peningkatan layanan kepada penumpang yang tinggi sehingga antrian saat boarding bisa berkurang.

Dengan banyaknya penumpang saat weekend atau hari libur nasional, PT.KAI DAOP 6 Yogyakarta mengantisipasi dengan menjalankan Kereta api tambahan baik keberangkatan dari stasiun Yogyakarta maupun stasiun Solo Balapan. Diantaranya adalah Kereta Api Taksaka Tambahan, KLB penumpang tambahan, KA lodaya tambahan.

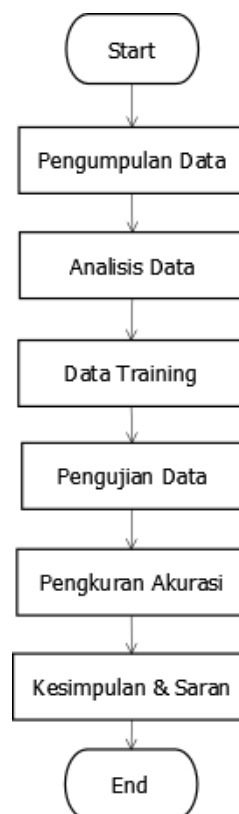
Perumusan masalah pada penelitian ini mengacu pada klasifikasi penumpang Kereta Api di Daop 6 Yogyakarta berdasarkan kelas stasiun. Penulis terlebih dahulu menyiapkan data penumpang dengan mendatangi kantor PT.Kereta Api Indonesia DAOP 6 Yogyakarta unit Angkutan penumpang untuk mendapatkan data penumpang di wilayah DAOP 6 yaitu rentan waktu 2016 sampai 2023. Perhitungan menggunakan K-Nearest Neighbor yaitu metode klasifikasi data berdasarkan kedekatan data uji dengan data lama. Dalam penelitian ini data akan diberikan label berdasarkan kelas stasiun yaitu, stasiun kelas 1, kelas 2, kelas 3, kelas 4, dan kelas 5. Untuk menghasilkan penelitian yang baik penulis menambahkan referensi penelitian terdahulu sebagai acuan dalam penelitian ini.

Penelitian terdahulu yang dilakukan oleh Widodo, Supatman dalam klasifikasi transaksi obat puskesmas bagelan menggunakan algoritma KNN dengan $k=7$ pada percobaan ke 3 menghasilkan akurasi 96% [2]. Penelitian kedua oleh Anton Prasetyo, Arita witanti pada tahun 2024 dengan judul implementasi metode k- nearest neighbor dalam menentukan waktu optimal penarikan pesanan driver ojol dengan akurasi mencapai 99%. Penelitian ke tiga dilakukan oleh Verdi Alvian Dwi Hidayatullah, Agung Nilogiri, Habibatul Azizah Al Faruq pada tahun 2022 dengan judul klasifikasi siswa berprestasi menggunakan Algoritma K-Nearest Neighbor di SMA N 2 Situbondo menghasilkan akurasi 94%, presisi 91%, dan recall 91% dengan $k=5$ [3]. Selanjutnya penelitian ke empat oleh Aisyah dan Siska pada tahun 2022 dengan judul analisis penerapan metode K-nearest neighbor (KNN) pada dataset citra penyakit malaria menggunakan $k=6$ menghasilkan akurasi 84%, presisi 88%, dan recall 92%. Penelitian ke lima dilakukan oleh

Bina Sukma Adicahya, Sri Wulandari, dan Donny Avianto dengan judul metode *neural network* dalam prediksi jumlah penumpang kereta api berbasis web menghasilkan akurasi 80,33%. Berdasarkan penelitian terdahulu yang telah disebutkan berfokus pada penumpang kereta api di Pula Jawa secara umum dan data yang digunakan hanya rentang satu atau dua tahun. Namun belum ada penelitian khusus di wilayah DAOP 6 Yogyakarta menggunakan metode KNN. Dengan demikian, dapat memberikan acuan kepada penulis untuk melakukan penelitian tentang klasifikasi penumpang Kereta Api di DAOP 6 Yogyakarta berdasarkan kelas stasiun menggunakan metode KNN.

2. METODE PENELITIAN

Untuk memperoleh hasil akurasi yang maksimal dalam penelitian ini, penulis telah menetapkan alur penelitian sebagai berikut: pemilihan data, preprocessing data, membagi data menjadi data *testing* dan data *training*, pengujian dengan metode *K-Nearest Neighbor*, analisa hasil akurasi, presisi, dan recall, kesimpulan, dan saran. *Flowchart* penelitian dapat dilihat pada Gambar 1.



Gambar 1. *Flowchart* Penelitian

2.1 Pengumpulan Data

Pengumpulan data dilakukan oleh penulis dengan mengunjungi unit angkutan penumpang DAOP 6 yogyakarta, kemudian penulis memperoleh data penumpang kereta api di wilayah DAOP 6 Yogyakarta tahun 2016 sampai 2023. Penulis kemudian melakukan rekap data yang diperlukan.

2.2 Analisis Data

Data yang diperoleh dilakukan rekap ulang kemudian dirapikan dalam rekap bulanan dan tahunan sehingga diperoleh data penumpang naik dan turun di stasiun wilayah DAOP 6 Yogyakarta periode 2016 sampai 2021. Data tersebut kemudian dibagi menjadi dua yaitu data traing dan data testing(uji). Selanjutnya data diberikan label menjadi 5 kelas berdasarkan kelas stasiun. Kelas klasifikasi dapat dilihat pada Tabel 1.

Tabel 1. Kelas Klasifikasi

No	Jumlah penumpang (1 Tahun)	Kelas Stasiun
1	0-5000	Kelas 5
2	5000-100.000	Kelas 4
3	100.000-500.000	Kelas 3
4	500.000-1.000.000	Kelas 2
5	>1.000.000	Kelas 1

Untuk menentukan jarak dalam penelitian ini penulis menggunakan perhitungan *Euclidean distance*.

2. 2.1 Data Training

Data training adalah kumpulan dataset yang digunakan sebagai pembelajaran model untuk memprediksi data baru [4].Selanjutnya data akan dilakukan seleksi. Sample dataset dari data training dapat dilihat pada Tabel 2.

Tabel 2 Sample Data Training

Delimiter:

	NO	JAN	FEB	MAR	APR	MEI	JUN	JUL	AGT	SEPT	OKT	NOV	DES	KELAS
1	1	0	0	0	0	0	57	224	141	194	209	211	327	5
2	2	4804	4240	4591	4378	5454	4567	7355	4925	4774	4502	4070	5877	4
3	3	32223	25558	27584	26381	31837	25150	50098	29136	33231	29455	30317	41093	3
4	4	142717	127322	126946	118667	145458	113139	121948	144616	134961	126973	121307	163294	1
5	5	21950	19242	21164	20262	22079	20726	22392	21622	21609	20229	19509	21241	3
6	6	792	391	564	364	416	455	927	275	306	364	271	701	4
7	7	83217	72370	76571	71782	82917	67292	93069	73837	72338	70512	69621	87479	2
8	8	8124	6086	6598	6182	7422	6595	18607	7230	8242	6867	7506	12971	4
9	9	1500	1365	1454	990	1455	724	1808	678	507	1029	785	1694	4
10	10	2112	1820	1863	2265	2815	1612	3160	1633	2288	2099	2228	3338	4
11	11	113597	100624	104082	99695	122510	93709	144219	115199	115310	108343	104481	146077	1
12	12	6813	5906	6181	5402	5142	4779	8911	5712	5719	5415	5538	5904	4
13	13	2512	2287	2263	2357	1625	756	1832	737	982	1981	1819	2433	4
14	14	4781	4449	4661	4701	4359	2870	4235	2142	2373	3626	3316	5311	4
15	15	10007	8643	9064	8743	11234	8121	16526	10038	11747	10045	9683	13939	3
16	16	188658	157028	159786	155728	208251	147824	230996	201067	194567	186278	181647	263070	1
17	17	0	0	0	0	0	38	166	217	76	145	57	205	5
18	18	909	823	937	816	1083	1223	2331	1368	1438	1348	1225	1913	4
19	19	20312	17429	19378	18491	23494	23447	34117	18486	23303	18406	19499	29455	3
20	20	104246	109511	103134	97767	124854	84264	115722	125594	114415	105865	100897	143708	1
21	21	4095	4969	5020	4939	4252	4014	5723	4764	4787	5124	4599	5566	4
22	22	271	140	189	57	82	125	125	63	99	53	36	83	5
23	23	41310	36546	38681	37024	47442	39823	58008	40657	44723	40796	38281	54354	2

2. 2.2 Data Testing

Data testing adalah bagian dari dataset yang digunakan untuk menguji performa suatu model dari pembelajaran data training tersebut [5]. Berikut adalah data testing yang digunakan dalam klasifikasi penumpang di DAOP 6 Yogyakarta,

Tabel 3. Data *Testing*

Delimiter:														
	NO	JAN	FEB	MAR	APR	MEI	JUN	JUL	AGT	SEPT	OKT	NOV	DES	KELAS
1	225	0	0	0	0	0	0	0	0	0	0	0	0	5
2	226	399	254	148	65	85	79	78	37	93	51	80	228	5
3	227	23191	17359	19680	32234	27508	24059	32780	23111	24653	25593	23815	32742	3
4	228	105523	86119	79115	97417	102031	116578	132399	115908	111497	114957	119141	154472	1
5	229	19514	16253	18517	29012	24406	24563	27572	22251	24565	24108	25266	30874	3
6	230	6722	5497	5948	11039	8597	9040	9200	7473	8193	8120	9519	12315	3
7	231	350	237	306	825	454	703	1223	793	820	931	889	1344	4
8	232	61334	51672	63533	77366	67402	88419	100119	81642	87050	90761	89738	116649	2
9	233	1884	1647	1768	1785	2944	2585	1879	1141	1581	1838	1092	2255	4
10	234	6010	4632	5075	9341	8118	6529	8863	6450	6934	7254	6972	8920	4
11	235	91	63	88	128	114	19	28	11	16	21	10	25	5
12	236	9666	7617	8496	14443	11535	11410	14545	10201	10716	11596	10840	15307	3
13	237	0	0	0	0	0	0	0	0	0	0	0	0	5
14	238	203754	163989	168160	179389	191156	228698	275843	225746	213448	229180	229647	291200	1
15	239	0	0	0	0	0	0	0	0	0	0	0	0	5
16	240	416	182	44	87	81	71	74	41	98	49	86	226	5
17	241	17939	15938	18621	34952	20664	26055	27857	21474	23531	22902	22487	34263	3
18	242	98202	89331	77642	86541	111008	111543	129713	119145	116010	111040	119246	151420	1
19	243	17907	15185	17169	30084	21325	21964	25613	21558	23655	22836	25091	28618	3
20	244	5597	5398	5912	11248	7769	8354	8603	6990	7357	6694	8115	11728	4
21	245	248	195	266	716	309	575	798	561	657	681	659	1033	4
22	246	63390	58921	68691	88454	69337	96606	106358	91161	97451	98911	100932	126638	1
23	247	1436	1117	2089	1333	2410	3134	2101	956	1202	2366	1158	2263	4
24	248	4959	4087	4731	10791	6058	6843	7710	6018	6572	6766	6680	9115	4
25	249	104	95	110	163	134	41	37	30	29	43	19	56	5
26	250	8985	7681	9305	15199	10146	13262	14262	10931	11291	11582	11015	16212	3
27	251	0	0	0	0	0	0	0	0	0	0	0	0	5
28	252	186762	172227	168729	179379	196886	223155	263292	224396	219068	218214	229832	288244	1

2.4 Metode K-Nearest Neighbor

Metode K-Nearest Neighbor adalah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data sebelumnya yang telah dilakukan klasifikasi. Metode ini termasuk dalam supervised learning [6].

Metode ini bekerja berdasarkan minimum jarak terhadap data uji ke data training untuk menentukan KNNnya [7]. Langkah- Langkah untuk menghitung metode K-Nearest Neighbor adalah sebagai berikut:

- Menghitung K
- Menghitung jarak antara data *training* dan data *testing*, rumus *euclidean distance* dapat dilihat pada persamaan 1.

$$euc = \sqrt{\left(\sum_{i=1}^n (p_i - q_i)^2 \right)} \quad (1)$$

P_i = Data training

q_i = Data testing

i = Variabel data

n = dimensi data

- Menghitung jarak
- Menentukan jarak terdekat
- Pemasangan kelas yang sesuai
- Mencari jumlah kelas dari tetangga terdekat

2.4.1 Akurasi

Akurasi adalah tingkat kedekatan nilai aktual dengan nilai prediksi [8]. Dapat dilihat pada persamaan 2.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

2.4.1 Presisi

Presisi adalah kecocokan antara hasil yang diperoleh relevan dengan yang dicari [9]. Rumus presisi dapat dilihat pada persamaan 3.

$$Presisi = \frac{TP}{TP + FP} \quad (3)$$

2.4.1 Recall

Recall adalah ukuran tentang informasi yang relevan dipilih dan ditemukan [10]. Rumusnya dapat dilihat pada persamaan 4.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Keterangan:

TP = True Positif

TN = True Negatif

FP = False Positif

FN = False Negatif

2.4.1 F-Measure

Measure adalah nilai gabungan presisi dan recall untuk mengetahui tingkat kinerja suatu sistem [11]. Rumusnya dapat dilihat pada persamaan 5.

$$F - Measure = 2 \frac{Presisi \cdot Recall}{Presisi + Recall} \quad (5)$$

3. HASIL DAN PEMBAHASAN

Pengumpulan data adalah langkah awal dari penelitian ini, penulis mendatangi kantor PT.Kereta Api DAOP 6 Yogyakarta unit angkutan penumpang, penulis merekap data naik turun penumpang dari tahun 2016 sampai tahun 2023. Dataset berjumlah 242 dengan pembagian 224 data *training* dan 18 data *testing*. Selanjutnya data bulanan kami rekap menjadi data tahunan agar memudahkan dalam perhitungan, Berikut ini adalah data rekap tahunan penumpang yang naik di stasiun wilayah DAOP 6 YOGYAKARTA tahun 2016-2023.

Tabel 4. Sample Data Awal

PENUMPANGNAIK2016													
STASIUN	JAN	FEB	MAR	APR	MEI	JUN	JUL	AGT	SEPT	OKT	NOV	DES	Grand Total
BBN						57	224	141	194	209	211	327	1,363
JN	4,804	4,240	4,591	4,378	5,454	4,567	7,355	4,925	4,774	4,502	4,070	5,877	59,537
KT	32,223	25,558	27,584	26,381	31,837	25,150	50,098	29,136	33,231	29,455	30,317	41,093	382,063
LPN	142,717	127,322	126,946	118,667	145,458	113,139	121,948	144,616	134,961	126,973	121,307	163,294	1,587,348
MGW	21,950	19,242	21,164	20,262	22,079	20,726	22,392	21,622	21,609	20,229	19,509	21,241	252,025
PNT	792	391	564	364	416	455	927	275	306	364	271	701	5,826
PWS	83,217	72,370	76,571	71,782	82,917	67,292	93,069	73,837	72,338	70,512	69,621	87,479	921,005
SK	8,124	6,086	6,598	6,182	7,422	6,595	18,607	7,230	8,242	6,867	7,506	12,971	102,430
SKH	1,500	1,365	1,454	990	1,455	724	1,808	678	507	1,029	785	1,694	13,989
SLM	2,112	1,820	1,863	2,265	2,815	1,612	3,160	1,633	2,288	2,099	2,228	3,338	27,233
SLO	113,597	100,624	104,082	99,695	122,510	93,709	144,219	115,199	115,310	108,343	104,481	146,077	1,367,846
SR	6,813	5,906	6,181	5,402	5,142	4,779	8,911	5,712	5,719	5,415	5,538	5,904	71,422
STA	2,512	2,287	2,263	2,357	1,625	756	1,832	737	982	1,981	1,819	2,433	21,584
WNG	4,781	4,449	4,661	4,701	4,359	2,870	4,235	2,142	2,373	3,626	3,316	5,311	46,824
WT	10,007	8,643	9,064	8,743	11,234	8,121	16,526	10,038	11,747	10,045	9,683	13,939	127,790
YK	188,658	157,028	159,786	155,728	208,251	147,824	230,996	201,067	194,567	186,278	181,647	263,070	2,274,900
Grand Total	623,807	537,331	553,372	527,897	652,974	498,376	726,307	618,988	609,148	577,927	562,309	774,749	7,263,185

Tabel 5. Sample Data *Training*

Delimiter:

	NO	JAN	FEB	MAR	APR	MEI	JUN	JUL	AGT	SEPT	OKT	NOV	DES	KELAS
1	1	0	0	0	0	0	57	224	141	194	209	211	327	5
2	2	4804	4240	4591	4378	5454	4567	7355	4925	4774	4502	4070	5877	4
3	3	32223	25558	27584	26381	31837	25150	50098	29136	33231	29455	30317	41093	3
4	4	142717	127322	126946	118667	145458	113139	121948	144616	134961	126973	121307	163294	1
5	5	21950	19242	21164	20262	22079	20726	22392	21622	21609	20229	19509	21241	3
6	6	792	391	564	364	416	455	927	275	306	364	271	701	4
7	7	83217	72370	76571	71782	82917	67292	93069	73837	72338	70512	69621	87479	2
8	8	8124	6086	6598	6182	7422	6595	18607	7230	8242	6867	7506	12971	4
9	9	1500	1365	1454	990	1455	724	1808	678	507	1029	785	1694	4
10	10	2112	1820	1863	2265	2815	1612	3160	1633	2288	2099	2228	3338	4
11	11	113597	100624	104082	99695	122510	93709	144219	115199	115310	108343	104481	146077	1
12	12	6813	5906	6181	5402	5142	4779	8911	5712	5719	5415	5538	5904	4
13	13	2512	2287	2263	2357	1625	756	1832	737	982	1981	1819	2433	4
14	14	4781	4449	4661	4701	4359	2870	4235	2142	2373	3626	3316	5311	4
15	15	10007	8643	9064	8743	11234	8121	16526	10038	11747	10045	9683	13939	3
16	16	188658	157028	159786	155728	208251	147824	230996	201067	194567	186278	181647	263070	1
17	17	0	0	0	0	0	38	166	217	76	145	57	205	5
18	18	909	823	937	816	1083	1223	2331	1368	1438	1348	1225	1913	4
19	19	20312	17429	19378	18491	23494	23447	34117	18486	23303	18406	19499	29455	3
20	20	104246	109511	103134	97767	124854	84264	115722	125594	114415	105865	100897	143708	1
21	21	4095	4969	5020	4939	4252	4014	5723	4764	4787	5124	4599	5566	4
22	22	271	140	189	57	82	125	125	63	99	53	36	83	5
23	23	41310	36546	38681	37024	47442	39823	58008	40657	44723	40796	38281	54354	2

Tabel 6. Data *Testing*

Delimiter: <input type="text" value=","/> <input type="button" value="v"/>														
	NO	JAN	FEB	MAR	APR	MEI	JUN	JUL	AGT	SEPT	OKT	NOV	DES	KELAS
1	225	0	0	0	0	0	0	0	0	0	0	0	0	5
2	226	399	254	148	65	85	79	78	37	93	51	80	228	5
3	227	23191	17359	19680	32234	27508	24059	32780	23111	24653	25593	23815	32742	3
4	228	105523	86119	79115	97417	102031	116578	132399	115908	111497	114957	119141	154472	1
5	229	19514	16253	18517	29012	24406	24563	27572	22251	24565	24108	25266	30874	3
6	230	6722	5497	5948	11039	8597	9040	9200	7473	8193	8120	9519	12315	3
7	231	350	237	306	825	454	703	1223	793	820	931	889	1344	4
8	232	61334	51672	63533	77366	67402	88419	100119	81642	87050	90761	89738	116649	2
9	233	1884	1647	1768	1785	2944	2585	1879	1141	1581	1838	1092	2255	4
10	234	6010	4632	5075	9341	8118	6529	8863	6450	6934	7254	6972	8920	4
11	235	91	63	88	128	114	19	28	11	16	21	10	25	5
12	236	9666	7617	8496	14443	11535	11410	14545	10201	10716	11596	10840	15307	3
13	237	0	0	0	0	0	0	0	0	0	0	0	0	5
14	238	203754	163989	168160	179389	191156	228698	275843	225746	213448	229180	229647	291200	1
15	239	0	0	0	0	0	0	0	0	0	0	0	0	5
16	240	416	182	44	87	81	71	74	41	98	49	86	226	5
17	241	17939	15938	18621	34952	20664	26055	27857	21474	23531	22902	22487	34263	3
18	242	98202	89331	77642	86541	111008	111543	129713	119145	116010	111040	119246	151420	1
19	243	17907	15185	17169	30084	21325	21964	25613	21558	23655	22836	25091	28618	3
20	244	5597	5398	5912	11248	7769	8354	8603	6990	7357	6694	8115	11728	4
21	245	248	195	266	716	309	575	798	561	657	681	659	1033	4
22	246	63390	58921	68691	88454	69337	96606	106358	91161	97451	98911	100932	126638	1
23	247	1436	1117	2089	1333	2410	3134	2101	956	1202	2366	1158	2263	4
24	248	4959	4087	4731	10791	6058	6843	7710	6018	6572	6766	6680	9115	4
25	249	104	95	110	163	134	41	37	30	29	43	19	56	5
26	250	8985	7681	9305	15199	10146	13262	14262	10931	11291	11582	11015	16212	3
27	251	0	0	0	0	0	0	0	0	0	0	0	0	5
28	252	186762	172227	168729	179379	196886	223155	263292	224396	219068	218214	229832	288244	1

Tabel 7. Jumlah Penumpang Tahunan

NO	STASIUN	JUMLAH PENUMPANGNAIK								KELAS
		2016	2017	2018	2019	2020	2021	2022	2023	
1	BBN	1363	4110	4112	5092	2042	0	0	0	5
2	JN	59537	73376	85398	78668	34877	548	795	1597	4
3	KT	382063	456762	502934	507626	202281	103223	229670	306725	3
4	LPN	1587348	1734769	1826764	1990993	688884	393177	995742	1335157	1
5	MGW	252025	253185	259258	293034	114960	0	0	0	5
6	PNT	5826	5264	5565	4091	1109	0	0	0	5
7	PWS	921005	820404	731968	759218	263832	76166	192675	286901	2
8	SK	102430	124309	132999	134830	32601	14319	58474	101663	3
9	SKH	13989	11211	10881	8797	3314	0	0	0	5
10	SLM	27233	28544	31002	28515	3073	1909	4506	8875	5
11	SLO	1367846	1851722	2238919	2236589	879283	209185	617839	975685	1
12	SR	71422	79752	87541	7399	24442	12605	55686	85098	3
13	STA	21584	14491	17269	85166	3785	0	0	0	5
14	WNG	46824	33861	41043	15871	9379	0	0	0	5
15	WT	127790	166026	197720	241772	85532	44241	117031	136372	3
16	YK	2274900	2928319	3443906	3669853	1163069	679282	2089118	2600210	1

Setelah melakukan preprocessing data, yaitu dengan menghilangkan kolom nama stasiun dan mengisi *missing value* dengan nilai 0. Data setelah *preprocessing* dapat dilihat pada Tabel 5.

Tabel 8. *Data Setelah Preprocessing*Delimiter:

	NO	JAN	FEB	MAR	APR	MEI	JUN	JUL	AGT	SEPT	OKT	NOV	DES	KELAS
1	1	0	0	0	0	0	57	224	141	194	209	211	327	5
2	2	4804	4240	4591	4378	5454	4567	7355	4925	4774	4502	4070	5877	4
3	3	32223	25558	27584	26381	31837	25150	50098	29136	33231	29455	30317	41093	3
4	4	142717	127322	126946	118667	145458	113139	121948	144616	134961	126973	121307	163294	1
5	5	21950	19242	21164	20262	22079	20726	22392	21622	21609	20229	19509	21241	3
6	6	792	391	564	364	416	455	927	275	306	364	271	701	4
7	7	83217	72370	76571	71782	82917	67292	93069	73837	72338	70512	69621	87479	2
8	8	8124	6086	6598	6182	7422	6595	18607	7230	8242	6867	7506	12971	4
9	9	1500	1365	1454	990	1455	724	1808	678	507	1029	785	1694	4
10	10	2112	1820	1863	2265	2815	1612	3160	1633	2288	2099	2228	3338	4
11	11	113597	100624	104082	99695	122510	93709	144219	115199	115310	108343	104481	146077	1
12	12	6813	5906	6181	5402	5142	4779	8911	5712	5719	5415	5538	5904	4
13	13	2512	2287	2263	2357	1625	756	1832	737	982	1981	1819	2433	4
14	14	4781	4449	4661	4701	4359	2870	4235	2142	2373	3626	3316	5311	4
15	15	10007	8643	9064	8743	11234	8121	16526	10038	11747	10045	9683	13939	3
16	16	188658	157028	159786	155728	208251	147824	230996	201067	194567	186278	181647	263070	1
17	17	0	0	0	0	0	38	166	217	76	145	57	205	5
18	18	909	823	937	816	1083	1223	2331	1368	1438	1348	1225	1913	4
19	19	20312	17429	19378	18491	23494	23447	34117	18486	23303	18406	19499	29455	3
20	20	104246	109511	103134	97767	124854	84264	115722	125594	114415	105865	100897	143708	1
21	21	4095	4969	5020	4939	4252	4014	5723	4764	4787	5124	4599	5566	4
22	22	271	140	189	57	82	125	125	63	99	53	36	83	5
23	23	41310	36546	38681	37024	47442	39823	58008	40657	44723	40796	38281	54354	2

selanjutnya membagi data yang berjumlah 242 data menjadi 228 data training dan 28 data uji, selanjutnya perhitungan dilakukan dengan menggunakan jupyter notebook mendapatkan hasil akurasi 93%, f1-score 88%.

```
[[5 5 3 1 3 3 4 2 4 4 5 3 5 1 5 5 3 1 3 4 4 1 4 4 5 3 5 1]]
[5 5 3 1 3 4 4 2 4 4 5 3 5 1 5 5 3 1 3 4 4 2 4 4 5 3 5 1]
precision    recall  f1-score   support

     1         1.00      0.80      0.89         5
     2         0.50      1.00      0.67         1
     3         1.00      0.86      0.92         7
     4         0.88      1.00      0.93         7
     5         1.00      1.00      1.00         8

 accuracy          0.93         28
 macro avg          0.88      0.93      0.88         28
weighted avg          0.95      0.93      0.93         28
```

Gambar 2. Hasil Pengujian

4. KESIMPULAN

Kesimpulan dari klasifikasi penumpang kereta api DAOP 6 Yogyakarta berdasar kelas stasiun adalah sebagai berikut: (1) Algoritma KNN dapat diimplementasikan dengan baik. (2) Hasil klasifikasi menghasilkan akurasi 93%, f1-score 88%, sehingga sistem ini dapat memberikan rekomendasi yang tepat untuk menentukan klasifikasi penumpang berdasarkan kelas stasiun. (3) Sebagai rekomendasi dalam pengambilan keputusan terkait penambahan fasilitas stasiun sesuai kebutuhan. (4) Dari hasil analisis klasifikasi penumpang di DAOP 6 Yogyakarta dapat sebagai acuan dalam peningkatan layanan kepada pelanggan Kereta Api.

5. SARAN

Berdasarkan hasil penelitian yang telah dilakukan, saran dari penulis untuk peneliti selanjutnya adalah faktor penggunaan data terbaru. Penelitian ini menggunakan data dari 2016 sampai 2023, sehingga peneliti selanjutnya agar menggunakan data tahun berikutnya agar hasilnya bisa relevan dengan kondisi pada saat itu.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada orang tua, isteri, anak, serta semua pihak yang telah memberi dukungan finansial terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Widodo., Supatman. 2024. *Klasifikasi Transaksi Obat Puskesmas Bagelan Menggunakan Algoritma KNN*. *Jurnal Zetroem*, vol 06(1), hal 79-85.
- [2] A. Prasetyo., A. Witanti. 2024. *Implementasi Metode K-Nearest Neighbor Dalam Menentukan Waktu Optimal Penarikan Pesanan Driver Ojol*. *Jurnal Teknik Informatika dan Sistem Informasi*, vol 11(2), hal 255-269.
- [3] V. Alvian., D. Hidayatullah., A. Nilogiri., H. Azizah., A. Faruq. 2022. *Klasifikasi Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor (KNN) Pada SMA Negeri 2 Situbondo Classification of Achieving Students Using K-Nearest neighbor (KNN) Method At SMA Negeri 2 Situbondo*. *Jurnal Smart Teknologi*, vol 01(1), hal 100-102.
- [4] A. Aisyah., S. Anraeni. 2022. *Analisis Penerapan Metode K-Nearest Neighbor (KNN) Pada Dataset Citra Penyakit Malaria*. *Indonesian Journal of Data and Sciense*, vol 03(1), hal 17-29.
- [5] B. S. Adicahya., S. Wulandari., D. Avianto. 2024. *Metode Neural Network Dalam Prediksi Jumlah Penumpang Kereta Api Berbasis Web*. *Journal of Information System Research*, vol 06(1), hal 302-314.
- [6] A. Yogiarto., A. Homaidi., Z. Fatah. 2024. *Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung*. *G-Tech: Jurnal Teknologi Terapan*, vol 08(3), hal 1720-1728.

- [7] E. Purwaningsih., E. Nurelasari. 2021. *Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa. Syntax: Jurnal Informatika*, vol 10 (1), hal 46-56.
- [8] N. F. Indrianti., A. K. Ningsih., R. Ilyas. 2024. *Implementasi Data Mining Untuk Klasifikasi Penyakit Gagal Ginjal Kronis Menggunakan Metode K-Nearest Neighbor.JATI (Jurnal Mahasiswa Teknik Informatika)*, vol 08(2), hal 2255-2260.
- [9] I. A. Khakim., A. Rozaq. 2022. *Klasifikasi Penyakit Dengue Menggunakan Algoritma K-Nearest Neighbors Berbasis Flask. Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol 06(3), hal 359-369.
- [10] H. Dharmawan., Tukino., S. S. Hilabi., I. Kaniawulan. 2023. *Sistem Rekomendasi Buku Dengan Menggunakan Metode K-Nearest Neighbor(K-NN) Pada Gramedia. ZONAsi; Jurnal Sistem Informasi*, vol 05(1), hal 16-25.
- [11] R. M. Candra., A. N. Rozana. 2020. *Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor.IT Journal Research and Development (ITJRD)*, vol 05(1), hal 45-52.