



Application of the K-Nearest Neighbor (KNN) Algorithm for the Classification of Indihome Products Sales in the Riau Region

Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Klasifikasi Penjualan Produk Indihome Daerah Riau

Mulya Permadi^{1*}, Febby Febriani², Martin Jonathan Panggabean³, Rahmaddeni⁴

^{1,2,3,4}Teknik Informatika, STMIK Amik Riau, Pekanbaru

E-Mail: ¹2010031802063@sar.ac.id, ²2010031802111@sar.ac.id,
³2010031802107@sar.ac.id, ⁴rahmaddeni@sar.ac.id

Corresponding Author: Mulya Permadi

Abstract

PT's IndiHome is a well-known product. Services like internet services, interactive internet (UseTV cable), Triple Play, and internet voicete are examples of IndiHome's offerings in the telecommunications, information, media, and education sectors. one of PT's strategies for marketing Telkom, specifically, by developing a marketing strategy that, to the fullest extent possible, calls for an efficient and effective marketing mix. This company places an emphasis on marketing by providing direction in relation to a number of variables, including target, positioning, market segmentation, and elements of the marketing mix. Operational variables are used as supporting variables in this study for both internal and external analysis. For the kind of research that was done, which was descriptive research. Through observation, interviews, and documentation, primary data are obtained directly from research subjects. Based on sales data from the previous year, the large number of requests from customers for service products indicates that it is necessary to classify sales of the best-selling service products in order to facilitate the company's service planning. Using the K-Nearest Neighbor algorithm, a classification method is used to determine the sales of the most popular service products. The classification of sales of the best-selling service products—Internet Only, Internet+TV, Internet+TV+Phone, and Gamer Packages—from the four types of packages sold is what this study found. On the best-selling product sales, the best accuracy value is 85.71 percent using the number $k = 2$.

Keyword: Gamers, Internet, K-Nearest Neighbor, Telecommunication.

Abstrak

PT IndiHome adalah produk yang terkenal. Layanan seperti layanan internet, internet interaktif (kabel UseTV), Triple Play, dan internet voicete adalah contoh penawaran IndiHome di sektor telekomunikasi, informasi, media, dan pendidikan. Salah satu strategi PT dalam memasarkan Telkom, khususnya dengan mengembangkan strategi pemasaran yang semaksimal memerlukan bauran pemasaran yang efisien dan efektif. Perusahaan ini menitikberatkan pemasaran dengan memberikan arahan terkait sejumlah variabel, antara lain target, positioning, segmentasi pasar, dan elemen bauran pemasaran. Variabel operasional digunakan sebagai variabel pendukung dalam penelitian ini baik untuk analisis internal maupun eksternal. Untuk jenis penelitian yang dilakukan yaitu penelitian deskriptif. Melalui observasi, wawancara, dan dokumentasi, data primer diperoleh langsung dari subjek penelitian. Berdasarkan data penjualan tahun sebelumnya, banyaknya permintaan dari pelanggan akan produk jasa mengindikasikan perlunya dilakukan klasifikasi penjualan produk jasa yang paling laris guna memudahkan perencanaan jasa perusahaan. Dengan menggunakan algoritma K-Nearest Neighbor, digunakan metode klasifikasi untuk menentukan penjualan produk jasa yang paling diminati. Klasifikasi penjualan produk layanan terlaris—Internet Only, Internet+TV, Internet+TV+Phone, dan Gamer Packages—dari empat jenis paket yang dijual adalah temuan studi ini. Pada penjualan produk terlaris nilai akurasi terbaik adalah 85,71 persen dengan menggunakan angka $k = 2$.

Kata Kunci: Gamers, Internet, Telekomunikasi, K-Nearest Neighbor.

1. PENDAHULUAN

Sebagai salah satu pelayanan telekomunikasi yang paling diminati di Indonesia, Telkom berkewajiban memberikan layanan yang terbaik kepada seluruh penggunanya. Penyedia layanan kini dapat menawarkan layanan produk dengan berbagai multi-layanan seperti telepon, internet, dan (IP-TV) dengan layanan triple-play berkat pesatnya perkembangan teknologi akses jaringan dan teknologi berbasis internet. IndiHome Telkom adalah layanan Triple Play yang menawarkan IPTV melalui UseeTV Cable, telepon (telepon rumah), dan Internet melalui Fiber atau High Speed Internet [1].

Mempertahankan barang dan jasa yang mereka hasilkan adalah salah satu masalah yang dihadapi bisnis yang menjual jasa. Diharapkan dengan mengklasifikasikan produk, Telkom dapat mengetahui apakah pelanggannya puas dengan layanan produk khususnya IndiHome. Suatu layanan bias dinilai memuaskan apabila banyak pelanggan tertarik dan ingin menggunakan layanan IndiHome sesuai dengan spesifikasi yang mereka butuhkan [2].

Proses mengevaluasi objek data untuk mengklasifikasikannya ke dalam salah satu dari banyak kelas yang tersedia dikenal sebagai klasifikasi [3]. Membangun model dari data pelatihan yang ada dan kemudian menggunakan model untuk mengklasifikasikan data baru adalah cara kerja klasifikasi. Pekerjaan pelatihan atau pembelajaran fungsi target di setiap set atribut (fitur) ke satu nomor label kelas yang tersedia juga dapat diartikan sebagai klasifikasi itu sendiri [4].

Metode K-Nearest Neighbor menggunakan algoritma terawasi di mana sebagian besar label kelas pada K-NN digunakan untuk mengklasifikasikan hasil instance kueri baru. Tujuan algoritma KNN adalah menggunakan atribut dan data pelatihan untuk mengklasifikasikan objek baru. Prediksi juga dapat dilakukan dengan menggunakan K-NN [5].

Penelitian yang dilakukan oleh (Sihananto dan Mahmudy, 2017) melakukan prediksi curah hujan dengan memanfaatkan jaringan saraf dengan backpropagation. Pengklasifikasi k-nearest neighbor (k-nn) dan naive Bayes classifier adalah dua pendekatan pembelajaran mesin yang menjadi subjek penyelidikan ini. Proses klasifikasi atau prediksi berbeda untuk masing-masing metode tersebut [6].

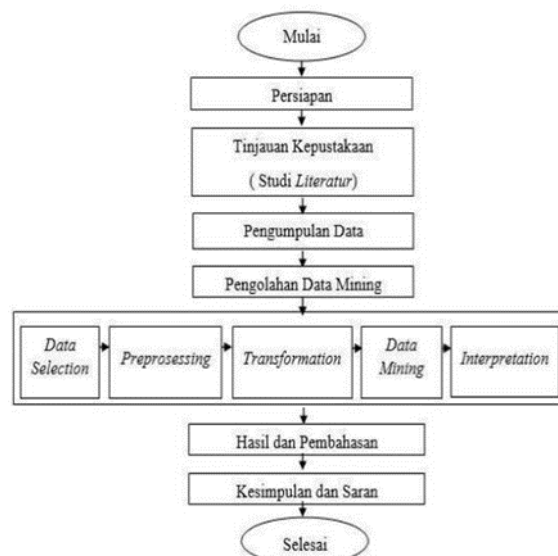
Dalam karya sebelumnya (Januarsyah, Zuhairi, & Malik, 2019), pendekatan Naive Bayes dan Random Forest dikontraskan. Metode Naive Bayes dengan akurasi nya 49,06 persen, metode Random Forest dengan akurasi 74,28%, C4.5 dengan akurasi 57,53%, metode Bayesian Network dengan akurasi 48,07% , dan Decision Stump yang memiliki akurasi 49,95% semuanya lebih baik dari metode Random Forest [7].

Menurut penelitian Chandel, Kunwar, Sabitha, Choudhury, & Mukherjee tahun 2017, “Studi perbandingan deteksi penyakit tiroid dengan menggunakan teknik klasifikasi Knearest Neighbor dan Naive Bayes” akurasi metode KNN adalah 93,44 persen, sedangkan akurasi metode Naive Bayes adalah 22,56% [8].

Peneliti tertarik untuk melakukan penelitian dengan judul “Penerapan Algoritma K-Nearest Neighbor Pada Klasifikasi Penjualan Produk Indihome Di Daerah Riau” Berdasarkan permasalahan tersebut diatas. Hal ini dilakukan dengan mengolah data dan memanfaatkan data yang datang sebelumnya untuk menampilkan informasi serta rangkaian pola yang dapat digunakan untuk memprediksi variabel lain yang belum diketahui nilai nya.

2. METODOLOGI PENELITIAN

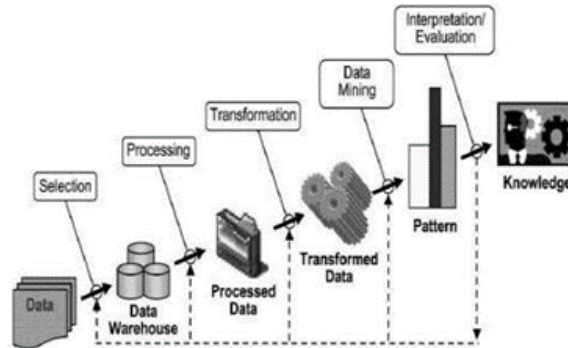
Pada gambar 1 akan dijelaskan alur penelitian sebagai berikut.



Gambar 1. Tahapan Penelitian

2.1 Data Mining

Studi tentang teknik untuk mengekstraksi pengetahuan atau pola dari data dikenal sebagai data mining. Data mining juga dapat didefinisikan sebagai proses pengolahan data untuk menemukan pola yang tersembunyi di dalam data [9]. Berikut ini adalah penjelasan komprehensif tentang pengetahuan baru yang diharapkan dari output serta kontribusi peningkatan yang diharapkan:



Gambar 2. Tahapan Data Mining

2.2 Algoritma K-Nearest Neighbor

Salah satu metode klasifikasi berbasis instance adalah K-NN, yang memilih objek pelatihan dengan properti tetangga terdekat [10]. Nilai kemiripan atau ketidaksamaan digunakan untuk menghitung sifat ketetanggaan ini. Nilai ketidaksamaan (Euclidian, Manhattan, Square Euclidian, dll.) dihitung dengan K-NN. Dengan melihat jumlah instance kelas pada K-neighbours yang dipilih, KNN akan memilih K-closest neighbor untuk menentukan hasil klasifikasi [11]. Kelas yang dihasilkan dari klasifikasi akan menjadi kelas yang paling sering muncul.

$$\sqrt{\sum_k^d} = 1(ak - bk)^2 \tag{1}$$

Pada K-NN, tahapan proses pelatihan hanya menyimpan unit vektor fitur dan mengklasifikasikan sampel data pelatihan. Fitur yang sama dari data uji (yang klasifikasinya tidak diketahui) juga dihitung pada tahap ini. K buah terdekat dipilih setelah menghitung jarak antara vektor baru dengan semua vektor sampel latih dari prosedur sebelumnya [12].

2.3 Dataset

Kumpulan data disebut sebagai dataset. Setiap kolom tabel, baris dari kumpulan data yang relevan mewakili variabel tertentu. Setiap kumpulan data memiliki nilai untuk setiap variabel, seperti tinggi dan berat objek. Datum adalah nama lain untuk setiap nilai. Sejumlah dokumen atau file juga disertakan dalam kumpulan data [9]. Di PT, peneliti menggunakan informasi penjualan produk tahun 2022. Hingga 367 line Telkom Indonesia tersedia di Pekanbaru.

Tabel 1. Dataset Pengguna IndiHome

NO	TANGGAL	NAMA PELANGGAN	PEKERJAAN	PENGHASILAN PERBULAN	PAKET
1	01/01/2022	AHMAD NAIL	SALES MOBIL	3500000	1P
2	02/01/2022	RIZKA HABIBAH	WIRASWASTA	3200000	2P
3	03/01/2022	NINGRUM SYAKILA	KARYAWAN SWASTA	2400000	1P
4	04/01/2022	DITO RAHMAYADI	TATA USAHA	2800000	2P
5	05/01/2022	RIZKA SYAFITRI	GURU	3000000	3P
6	06/01/2022	WANTO SUPRIADI	GURU	3100000	1P
7	07/01/2022	DISKA SURYA HARTATI	DOSEN	4200000	3P
8	08/01/2022	MUHAMMAD SYAUQI	KARYAWAN SWASTA	2500000	2P
9	09/01/2022	ROMY PARYASARADI	GURU	2700000	2P
10	10/01/2022	NOFRY PRIANTO	KARYAWAN BUMN	5100000	1P
...
359	23/12/2022	Freddy Ignatius Katuari	BURUH	2300000	2P
360	24/12/2022	Freddy Sumartono Santoso	GURU	2500000	2P
361	25/12/2022	Freddy Sutjipto	PENGUSAHA	6000000	3P

NO	TANGGAL	NAMA PELANGGAN	PEKERJAAN	PENGHASILAN PERBULAN	PAKET
362	26/12/2022	Gada Prakasa	DOKTER	5500000	3P
363	27/12/2022	Gary Dino Ridwan Sjah	PERAWAT	2000000	2P
364	28/12/2022	Gary Neal Christenson	SALES MOBIL	3000000	1P
365	29/12/2022	Gary Phair	PEDAGANG	3500000	PAKET GAMER
366	30/12/2022	Gatot Kariyoso Wiroyudo	POLISI	4300000	2P
367	31/12/2022	Fong Felix	KARYAWAN SWASTA	3800000	2P

2.4 Split Data

Proses pembagian dataset menjadi dua bagian—set pelatihan ("data pelatihan") dan set pengujian ("data uji")—disebut sebagai data terpisah atau pembagian data. Data dibagi 90:10 dalam penelitian ini. Akurasi terbaik dalam hasil adalah tujuan dari pemisahan data ini [13].

Tabel 2. Data Latih

NO	TANGGAL	NAMA PELANGGAN	PEKERJAAN	PENGHASILAN PERBULAN	PAKET
1	01/01/2022	AHMAD NAIL	SALES MOBIL	3500000	1P
2	02/01/2022	RIZKA HABIBAH	WIRASWASTA	3200000	2P
3	03/01/2022	NINGRUM SYAKILA	KARYAWAN SWASTA	2400000	1P
4	04/01/2022	DITO RAHMAYADI	TATA USAHA	2800000	2P
5	05/01/2022	RIZKA SYAFITRI	GURU	3000000	3P
6	06/01/2022	WANTO SUPRIADI	GURU	3100000	1P
7	07/01/2022	DISKA SURYA HARTATI	DOSEN	4200000	3P
8	08/01/2022	MUHAMMAD SYAUQI	KARYAWAN SWASTA	2500000	2P
9	09/01/2022	ROMY PARYASARADI	GURU	2700000	2P
10	10/01/2022	NOFRY PRIANTO	KARYAWAN BUMN	5100000	1P
...
324	19/11/2022	Fahmi Idris	DOSEN	4000000	1P
325	20/11/2022	Fahyudi Djaniatmadja	KARYAWAN SWASTA	3200000	2P
326	21/11/2022	Faisal Panggabean Law Firm	TEKNISI PLN	4300000	2P
327	22/11/2022	Faisol Soleh Masjkoer	KARYAWAN SWASTA	2800000	1P
328	23/11/2022	Faiz Shahab	KARYAWAN SWASTA	3000000	2P
329	24/11/2022	Faizan Abdul Rahan	KARYAWAN SWASTA	3200000	PAKET GAMER
330	25/11/2022	FAJAR SUPRAYETNO	BURUH	2100000	2P

Tabel 3. Data Uji

NO	TANGGAL	NAMA PELANGGAN	PEKERJAAN	PENGHASILAN PERBULAN	PAKET
331	26/11/2022	Fan Chiang, Teng-Chang	GURU	2400000	1P
332	27/11/2022	Farah Liza	DOSEN	5000000	2P
333	28/11/2022	FARINA TADJOEDIN	PEDAGANG	1800000	2P
334	29/11/2022	Farouk Rais	NOTARIS	3500000	1P
335	30/11/2022	FAS Consulting Asia Pacific Pte Ltd	GURU	3000000	1P
336	01/12/2022	Fauzi Ezeddin	PENGACARA	4500000	1P
337	02/12/2022	Fauzi Jahja	BIDAN	3200000	2P
338	03/12/2022	Fauzi Jurnalis	DOKTER	6000000	PAKET GAMER
339	04/12/2022	Febrina Sari Bak	KARYAWAN SWASTA	2700000	1P
340	05/12/2022	FITRI DWI WAHYUNI SAPUTRI	PERAWAT	1500000	1P
...
361	19/11/2022	Fahmi Idris	DOSEN	4000000	1P
362	20/11/2022	Fahyudi Djaniatmadja	KARYAWAN SWASTA	3200000	2P
363	21/11/2022	Faisal Panggabean Law Firm	TEKNISI PLN	4300000	2P
364	22/11/2022	Faisol Soleh Masjkoer	KARYAWAN SWASTA	2800000	1P
365	23/11/2022	Faiz Shahab	KARYAWAN SWASTA	3000000	2P
366	24/11/2022	Faizan Abdul Rahan	KARYAWAN SWASTA	3200000	PAKET

NO	TANGGAL	NAMA PELANGGAN	PEKERJAAN	PENGHASILAN PERBULAN	PAKET
367	25/11/2022	FAJAR SUPRAYETNO	BURUH	2100000	GAMER 2P

2.5. Confusion Matrix

Saat menentukan apakah suatu objek memiliki nilai benar atau salah, matriks kebingungan sering diterapkan pada evaluasi model klasifikasi. Matriks memprediksi yang akan dibandingkan dengan bagian input asli, yang berisi informasi tentang nilai aktual dan prediksi klasifikasi [14].

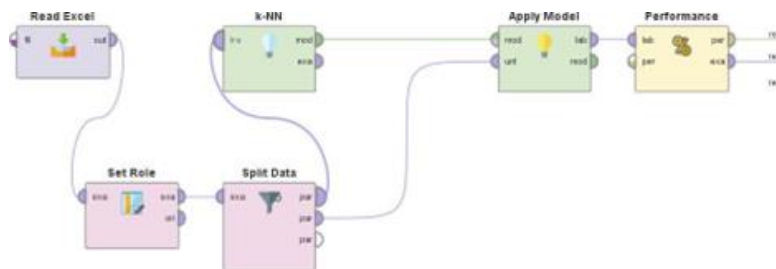
2.6. Rapidminer Tool

Lingkungan terintegrasi untuk pembelajaran mesin, pembelajaran mendalam, penambahan teks, dan analitik prediktif disediakan oleh perangkat lunak RapidMiner. Semua aspek yang ada di dalam machine learning, misalnya persiapan data, hasil visualisasi, validasi, dan pengoptimalan didukung dalam perangkat lunak ini. Ini juga digunakan untuk penelitian, pendidikan, pelatihan, pembuatan prototipe cepat, dan pengembangan aplikasi. Model inti terbuka digunakan dalam pembuatan RapidMiner [15].

3. HASIL DAN PEMBAHASAN

Data penjualan layanan produk IndiHome yang digunakan dalam penelitian ini adalah data sepanjang Tahun 2022. Variabel yang akan di pakai dalam penelitian diperoleh setelah data diolah. Ini adalah variabel independen yang digunakan adalah Career dan Pendapatan bulanan.

3.1 Algoritma K-Nearest Neighbor



Gambar 3. Design pada Rapidminer

Gambar diatas merupakan proses desain pada rapidminer dimulai dari tahapan menginput dataset dengan format excel, memilih target untuk label, pemilihan algoritma k-NN, pembagian data 90:10 hingga melihat hasil akurasi dengan confusion matrix.

accuracy: 85.71%

	true 1P	true 2P	true 3P	true PAKET GAMER	class precision
pred. 1P	13	2	0	0	86.67%
pred. 2P	0	10	0	1	90.91%
pred. 3P	1	0	5	1	71.43%
pred. PAKET GAMER	0	0	0	2	100.00%
class recall	92.86%	83.33%	100.00%	50.00%	

Gambar 4. Report Accuracy pada Rapidiner

PerformanceVector

```

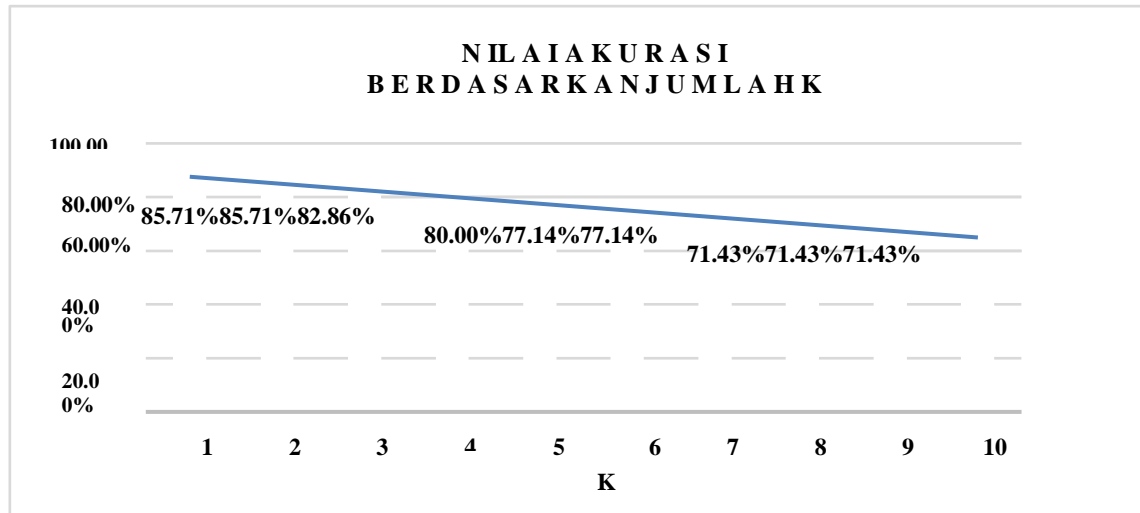
PerformanceVector:
accuracy: 85.71%
ConfusionMatrix:
True:   1P   2P   3P   PAKET GAMER
1P:    13   2   0   0
2P:     0  10   0   1
3P:     1   0   5   1
PAKET GAMER: 0   0   0   2
absolute_error: 0.143 +/- 0.256
root_mean_squared_error: 0.293 +/- 0.000
    
```

Gambar 5. Performance Vector

Pada Gambar 5. Dari hasil pengolahan dataset dengan menggunakan perangkat lunak RapidMiner didapatkan nilai dengan akurasi sebesar 85,71%.

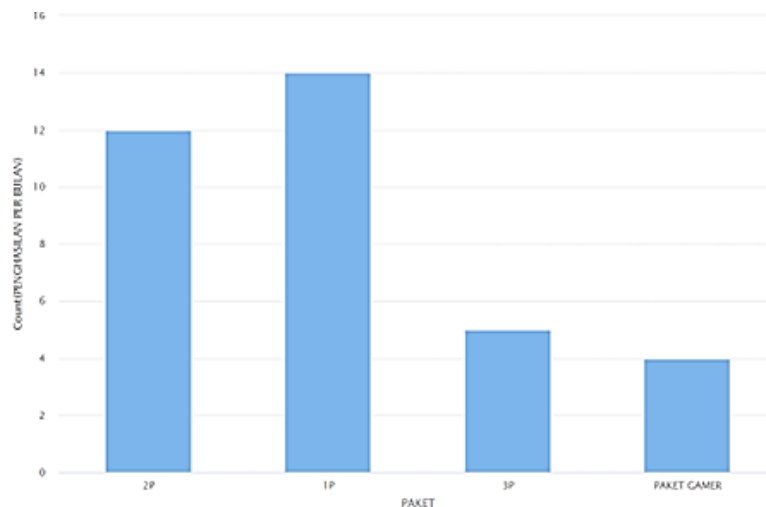
3.2 Analisis Hasil

Berikut adalah kesimpulan yang dapat ditarik dari pengujian yang dilakukan dengan software Rapidminer untuk mengklasifikasikan penjualan produk IndiHome menggunakan algoritma K-Nearest:



Gambar 6. Akurasi Berdasarkan Jumlah K

Hasil akurasi yang didapatkan seperti gambar diatas terlihat semakin besar jumlah k maka semakin kecil akurasi yang didapatkan.



Gambar 7. Hasil klasifikasi produk penjualan terlaris setiap bulan

Gambar 7. merupakan hasil klasifikasi dari produk penjualan setiap bulan nya dimana Paket 1P menjadi produk paling banyak diminati pelanggan IndiHome

4. KESIMPULAN

Sebanyak 367 baris data pengguna layanan IndiHome digunakan dalam pengujian, dan rapidminer mengembalikan nilai dengan akurasi 85,71 persen. Data ini dapat digunakan untuk membuat prediksi tentang langkah atau proses selanjutnya. Ketika algoritma K-Nearest Neighbor digunakan untuk penelitian, dari empat produk layanan, dihasilkan satu jenis produk dengan penjualan bulanan tertinggi. Metode K-Nearest Neighbor adalah satu-satunya yang dapat digunakan untuk mengolah data dan membuat prediksi yang paling akurat karena pengujian menunjukkan bahwa metode ini lebih akurat daripada metode lainnya.

REFERENSI

- [1] M. M. Ulkhaq, M. Putri, and B. Barus, "Analisis Kepuasan Pelanggan dengan Menggunakan SERVQUAL: Studi Kasus Layanan IndiHome PT. Telekomunikasi Indonesia, Tbk, Regional 1 Sumatera," *Jurnal Sistem dan Manajemen Industri*, vol. 1, pp. 61–67, 2017.
- [2] P. R. Utami, "ANALISIS PERBANDINGAN QUALITY OF SERVICE JARINGAN INTERNET BERBASIS WIRELESS PADA LAYANAN INTERNET SERVICE PROVIDER (ISP) INDIHOME DAN FIRST MEDIA," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 2, pp. 125–137, 2020, doi: 10.35760/tr.2020.v25i2.2723.
- [3] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 81–86, Aug. 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [4] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 2, p. 437, Apr. 2020, doi: 10.30865/mib.v4i2.2080.
- [5] A. U. Haspriyanti and P. T. Prasetyaningrum, "Penerapan Data Mining Untuk Prediksi Layanan Produk Indihome Menggunakan Metode K-Nearest Neighbor 1," 2019.
- [6] A. N. , dan M. W. F. Sihananto, "Rainfall Forecasting Using Backpropagation Neural Network," *Journal of Information Technology and Computer Science*, 2017.
- [7] M. Fariz Januarsyah, E. Zuhairi, and R. Firsandaya Malik, *Perbandingan Algoritma Random Forest, Decision Stump, Naïve Bayes, Bayesian Network dan Algoritma C4.5 Untuk Prediksi Pola Kartu Poker*, vol. 5, no. 1. 2019. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Poker+Hand>.
- [8] V. K. S. S. T. C. dan S. M. K. Chandel and A. Sudrajat, "A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques," *CSI Trans. ICT*, 2019.
- [9] W. Puspita Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *Jurnal Informatika dan Teknologi*, vol. 3, no. 2, 2020.
- [10] M. Reza Noviansyah, T. Rismawan, D. Marisa Midyanti, J. Sistem Komputer, and F. H. MIPA Universitas Tanjungpura Jl Hadari Nawawi, "PENERAPAN DATA MINING MENGGUNAKAN METODE K-NEAREST NEIGHBOR UNTUK KLASIFIKASI INDEKS CUACA KEBAKARAN BERDASARKAN DATA AWS (AUTOMATIC WEATHER STATION) (STUDI KASUS: KABUPATEN KUBU RAYA)," 2018.
- [11] Danar P, "Ekstraksi Citra menggunakan Metode GLCM dan KNN untuk Identifikasi Jenis Anggrek (Orchidaceae)," 2019.
- [12] N. Tri Romadloni, I. Santoso, S. Budilaksono, and M. Ilmu Komputer STMIK Nusa Mandiri Jakarta, "PERBANDINGAN METODE NAIVE BAYES, KNN DAN DECISION TREE TERHADAP ANALISIS SENTIMEN TRANSPORTASI KRL COMMUTER LINE."
- [13] R. Rahmadden, M. K. Anam, Y. Irawan, S. Susanti, and M. Jamaris, "Comparison of Support Vector Machine and XGBSVM in Analyzing Public Opinion on Covid-19 Vaccination," *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, pp. 32–38, Apr. 2022, doi: 10.33096/ilkom.v14i1.1090.32-38.
- [14] D. Ardiansyah and W. Walim, "ALGORITMA C4.5 UNTUK KLASIFIKASI CALON PESERTA LOMBA CERDAS CERMAT SISWA SMP DENGAN MENGGUNAKAN APLIKASI RAPID MINER," Online, 2018.
- [15] Z. Muhammad, R. Rahmadhani, H. Rizqifaluthi, and M. A. Yaqin, "Process Mining Akademik Sekolah Menggunakan RapidMiner," *MATICS*, vol. 10, no. 2, p. 39, Mar. 2018, doi: 10.18860/mat.v10i2.5745.