IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI PENYAKIT GAGAL GINJAL KRONIS MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Nisa Fatimah Indrianti, Ade Kania Ningsih, Ridwan Ilyas
Teknik Informatika, Universitas Jenderal Achmad Yani
Jalan Terusan Jend. Sudirman, Cimahi, Jawa Barat, Indonesia
nisaindrianti24@gmail.com

ABSTRAK

Ginjal adalah salah satu organ tubuh paling penting yang berfungsi untuk menjaga kandungan yang ada pada darah dengan cara mencegah penumpukan limbah dan mengendalikan keseimbangan cairan yang ada di dalam tubuh. Penyakit ginjal adalah kelainan penyakit yang menyerang organ ginjal yang disebabkan karena berbagai faktor, misalnya pola hidup yang tidak sehat, bertambahnya usia, ataupun karena faktor turunan. Penelitian ini bertujuan untuk mengklasifikasi resiko penyakit Gagal Ginjal Kronis Metode yang dapat digunakan untuk mendeteksi PGK salah satu nya adalah metode K-Nearest Neighbor (KNN). Metode K-Nearest Neighbor (KNN) ini dapat mengklasifikasikan suatu data yang telah ada ke dalam beberapa kelas. Algoritma K-Nearest Neighbor (KNN) ini ialah algoritma yang dapat menentukan nilai suatu jarak yang berada pada pengujian data testing dengan data training berdasarkan nilai terkecil dari nilai ketetanggaan terdekat. Penelitian ini diharapkan memberikan kontribusi di bidang kesehatan dan di bidang informatika dalam mengklasifikasi penyakit Gagal Ginjal Kronis. Berapa akurasi yang dapat diperoleh dari klasifikasi K-Nearest Neighbor (KNN) dalam menentukan gejala dan resiko penyakit Gagal Ginjal Kronis. penelitian ini dibuat menggunakan algoritma K-Nearest Neighbor. Dalam penelitian ini menggunakan 75% dari data latih dan 25% dari data uji yang digunakan. Hasil klasifikasi penyakit ginjal kronis menunjukkan bahwa sistem berhasil melakukan klasifikasi dengan nilai akurasi mencapai 92,59%, 89,85% untuk nilai presisi, 87,32% untuk nilai recall dan f1-score mencapai nilai 88,57%.

Kata Kunci *: Gagal ginjal kronis, K-Nearest Neighbor, KNN, klasifikasi.*

1. PENDAHULUAN

Ginjal adalah salah satu organ tubuh paling penting yang berfungsi untuk menjaga kandungan yang ada pada darah dengan cara mencegah penumpukan limbah dan mengendalikan keseimbangan cairan yang ada di dalam tubuh. Ginjal berperan untuk melakukan proses penyaringan darah dan menghilangkan racun dari dalam tubuh. Organ ginjal dapat mengalami penurunan fungsi kinerja ataupun tidak dapat lagi bekerja sama sekali untuk menyaring pembuangan elektrolit tubuh, menjaga keseimbangan cairan dan zat kimia tubuh seperti kalium dan natrium pada darah ataupun dapat mengalami ketidaknormalan produksi urin. Pada saat fungsi ginjal menurun dapat diartikan sebagai kerusakan yang terjadi pada ginjal [1].

Penyakit ginjal kronis merupakan penyakit dimana seseorang mengalami penurunan fungsi pada ginjal. Penyakit ginjal kronis merupakan salah satu penyakit yang sering terjadi pada masyarakat berdasarkan pola hidup yang tidak sehat, bertambahnya usia, karena faktor turunan dan lain – lain [2]. Penderita seringkali tidak merasakan gejala yang begitu terasa pada awal menderita penyakit ini. Sehingga penderita baru dapat terdiagnosa penyakit ginjal kronis ketika melakukan tes darah dengan indikasi-indikasi penyakit ginjal kronis yang lebih terlihat [1].

Metode yang dapat digunakan salah satu nya adalah metode K-Nearest Neighbor (KNN). Metode K-Nearest Neighbor (KNN) ini dapat mengklasifikasikan suatu data yang telah ada ke dalam beberapa kelas, maka dari itu K-Nearest Neighbor (KNN) dapat mengklasifikasikan suatu data dari pasien PGK hingga mendapatkan kelas penderita PGK dan kelas yang tidak menderita PGK [3]. Pada algoritma K-Nearest Neighbor (KNN) ini ialah algoritma yang dapat menentukan nilai suatu jarak terdekat yang berada pada pengujian data testing dengan data training berdasarkan nilai terkecil dari nilai ketetanggaan terdekat [4].

Klasifikasi merupakan sebuah proses dalam mendapatkan model atau fungsi yang menjelaskan maupun membedakan konsep ataupun kelas pada data yang bertujuan dalam memprediksi kelas dari suatu objek pada labelnya yang tidak atau belum diketahui. Tujuan tersebut dapat dicapai jika pada proses klasifikasi dapat membuat sebuah model yang dapat memperlihatkan beda dari data ke dalam kelas-kelas berbeda berdasarkan fungsi ataupun aturan tertentu. K-Nearest Neighbor merupakan salah satu algoritma yang dapat digunakan untuk melakukan klasifikasi[1]. K-Nearest Neighbor (KNN) merupakan salah satu dari berbagai metode dalam proses klasifikasi suatu objek yang didasari pada informasi data yang mempunyai jarak sangat dekat dengan suatu objek tertentu. Algoritma K-Nearest Neighbor (KNN) bertujuan untuk melakukan pengklasifikasian objek menurut suatu atribut data uji dan data latih [5].

Penelitian terdahulu terkait menggunakan metode Naïve Bayes pada bidang kesehatan salah satunya adalah hasil yang diperoleh menunjukkan bahwa Naïve Bayes merupakan pengklasifikasi yang

paling akurat dengan akurasi 100% jika dibandingkan dengan JST yang memiliki akurasi 72,73%. Dalam studi penelitian terdahulu ini, beberapa faktor yang dipertimbangkan adalah usia, diabetes, tekanan darah, jumlah RBC dan lain-lain[6]. Pada penelitian ini akan menambahkan penggunaan data diet, data diet yang dimaksud adalah informasi tentang asupan protein, garam, natrium dan kalium dari pasien yang dapat digunakan untuk mengklasifikasikan resiko untuk penderita penyakit ginjal kronis.

Penelitian ini mengklasifikasikan penyakit gagal ginjal kronis menggunakan metode K-Nearest Neighbor (KNN) dengan menggunakan atribut seperti umur (age), tekanan darah (bp), berat jenis urin hasil tes urinalysis (sg), protein (al), gula darah dalam tubuh (su), gula darah setelah makan (bgr), kadar urea pada darah (bu), kadar kreatinin serum pada darah (sc), kadar sodium pada darah (sod), dan jumlah hematocrit pada darah (pvc). Penelitian ini diharapkan memberikan kontribusi keilmuan, di bidang kesehatan dan di bidang informatika dalam mengklasifikasi penyakit gagal ginjal kronis. Penelitian ini diharapkan agar dataset yang telah dibagi menjadi data latih dan data uji dapat diklasifikasikan dengan baik menggunkan metode K-Nearest Neighbor (KNN) dan akurasi yang dihasilkan dengan pengujian Confusion Matrix menghasilkan nilai yang tinggi.

2. TINJAUAN PUSTAKA

2.1. Data mining

Data mining, sering dikenal juga sebagai KDD (Knowledge Discovery in Databases), adalah kegiatan yang bertujuan untuk memproses himpunan data dengan tujuan menggali dan menganalisis informasi, pola, dan hubungan yang tersembunyi dalam dataset berukuran besar. Proses data mining ini bertujuan untuk mendapatkan pengetahuan baru yang bermanfaat dari data yang ada. Hasil dari proses data mining ini dapat digunakan untuk pengambilan keputusan di masa depan. Dengan menganalisis data secara mendalam, dapat diidentifikasi keteraturan dan tren yang dapat membantu pengambil keputusan dalam memahami situasi, membuat perkiraan, atau menemukan pola – pola yang berharga untuk meningkatkan performa atau efisiensi [4].

Data mining merupakan suatu proses pencarian informasi berharga dari kumpulan data besar dengan menerapkan pendekatan interdisipliner yang relatif baru. Proses ini melibatkan analisis mendalam terhadap data dan penemuan pengetahuan berharga yang tersembunyi dalam basis data, serta mengadopsi pendekatan multifaset. Data mining memiliki beberapa model proses yang digunakan untuk memandu realisasi data mining. Model proses yang umum digunakan adalah knowledge discovery database (KDD), CRISP-DM dan SEMM [2].

Saat ini data mining sangat sering digunakan dalam memperoleh hasil diagnostik. Industri perawatan kesehatan mengumpulkan sebagian besar data yang belum dikumpulkan untuk mengetahui informasi tersembunyi yang dapat didiagnosis dan pengambilan keputusan yang efektif. Penambangan data adalah proses penggalian informasi tersembunyi dari kumpulan data besar dan mengklasifikasikan pola yang valid dan unik dalam data. Ada banyak teknik data mining seperti clustering, klasifikasi, analisis asosiasi, regresi, dan lain – lain [6].

Ada beberapa tahapan – tahapan yang ada di dalam data mining adalah sebagai berikut :

a. Pengumpulan Data
   Pengumpulan data dilakukan untuk menentukan data yang akan diproses. Dalam langkah ini, semua informasi yang dibutuhkan akan dikumpulkan menjadi satu set data, termasuk atribut yang diperlukan untuk proses tersebut. Dengan cara mencari informasi yang sudah tersedia dan juga mencari informasi tambahan yang dibutuhkan.
b. Data Cleaning
   Proses data cleaning dilakukan untuk menghilangkan noise dan data lainnya yang tidak diperlukan yang sering kali membuat hasil data tidak akurat.
c. Data Selection
   Proses seleksi digunakan untuk mempersempit fokus analisis pada subset data yang memiliki informasi yang paling berharga dan dapat mendukung tujuan bisnis yang ditentukan.
d. Transformasi Data
   Ada beberapa transformasi data yaitu :
   - Smoothing (menghilangkan noise pada data)
   - Aggregation (mekapitulasi tahapan opeperbandingannal diterapkan pada data)
   - Normalization (menyesuaikan ukuran data)
   - Discretization (mengganti data numerik dengan interval, misalnya dengan usia)

2.2. Klasifikasi

Klasifikasi adalah pengelompokan suatu objek ke dalam kategori tertentu. Berbagai situasi yang erat kaitannya dengan suatu pengelompokan objek dapat diatasi dengan menerapkan teknik klasifikasi [4]. Klasifikasi adalah proses menemukan suatu model dan dapat dideskripsikan atau dibedakan dari kelas - kelas data yang telah ada, atau bagaimana mengklasifikasikan suatu data ke dalam beberapa kategori yang telah ditentukan [10].

Klasifikasi adalah kegiatan yang dapat dilakukan ketika ingin melakukan pengukuran objek data yang dikelompokkan ke dalam kelas-kelas tertentu. Didalam klasifikasi, ada dua hal yang dapat dilakukan, yaitu acuan untuk pembuatan prototype lalu menyimpannya di dalam memori dan penggunaan model untuk memperkenalkan/ mengklasifikasikan/memprediksi data sehingga dapat diketahui keberadaan objek data tersebut berasal darimana dilihat dari model yang disimpan [11].

2.3. K-Nearest Neighbor

Algoritma K-Nearest Neighbor adalah salah satu algoritma yang paling sering digunakan untuk klasifikasi. K-Nearest Neighbor (KNN) ialah salah

satu metode dalam proses pengklasifikasian objek berdasarkan data informasi jarak jauh objek tertentu. Algoritma K-Nearest Neighbor (KNN) memiliki tujuan untuk mengklasifikasikan objek sesuai dengan atributnya dan sampel data pelatihan. Namun, algoritma K-Nearest Neighbor (KNN) memiliki kekurangan dalam memilih atau menentukan nilai k [2].

Data yang memiliki ukuran q, pada algoritma K-Nearest Neighbor (KNN), dapat menghitung suatu jarak data satu ke data yang lainnya. K - Nearest Neighbor termasuk dalam kelompok instance-based learning. K-Nearest Neighbor merupakan contoh algoritma yang dapat dipelajari dan sering digunakan, dimana data set pelatihan (training) disimpan, sehingga klasifikasi untuk data dan rekaman baru yang tidak diklasifikasi didapatkan dengan membandingkan data dan rekaman yang paling mirip dengan training set [10].

Algoritma K-Nearest Neighbor (KNN) dilakukan degan tahapan awal yaitu menghitung jarak dari data latih, lalu mengatur catatan pelatihan berdasarkan jarak (pemilihan K tertangga terdekat), dan tahapan pada algoritma KNN ini adalah memiliki mayoritas antara K tetangga terdekat. K-Nearest Neighbor (KNN) dapat mengklasifikasikan objek berdasarkan data yang telah dilatih dan yang paling dekat dengan sasaran. Euclidean distance adalah perhitungan jarak dari dua buah titik dalam euclidean space untuk mempelajari hubungan antara sudut dan jarak[2]. Metode Euclidean Distance digunakan dalam penelitian sebagai metode analisa untuk memecahkan masalah, karena euclidean distance adalah suatu metode pencarian kedekatan nilai jarak dari dua buah variabel. Jarak dekat atau jauh dapat dihitung dengan jarak acuan Euclidean menggunakan persamaan yang banyak digunakan/umum yang ada pada rumus, seperti rumus dibawah: [11].

$$d\,(x,y) = \sqrt{\sum(x_i - y_i)^2} \quad \ldots (1)$$

Keterangan:
Xi= data latih
Yi= data uji
d = jarak [2]

Penelitian ini bertujuan untuk mengklasifikasi penyakit gagal ginjal kronis dengan gejala yang dapat dilihat secara spesifik dan data yang lengkap menggunakan metode K-Nearest Neighbor (KNN).

2.4. Confusion Matrix

Confusion matrix adalah ukuran kinerja untuk masalah klasifikasi machine learning yang hasilnya bisa berupa dua kelas atau lebih. Confusion matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Confusion matrix berisi empat term yang mewakili hasil proses klasifikasi, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). True Negative (TN) adalah jumlah data negatif yang dikenali dengan benar, sedangkan False Positive (FP) adalah data negatif tetapi dikenali sebagai data positif. Sedangkan True Positive (TP) adalah data positif yang dikenali dengan benar. False Negative (FN) adalah kebalikan dari True Positive sehingga datanya positif, tetapi diakui sebagai data negative. Dapat dilihat pada gambar dibawah.

		True Values	
		True	False
Prediction	True	TP Correct Result	FP Unexpected Result
	False	FN Missing Result	TN Correct Absence of Result

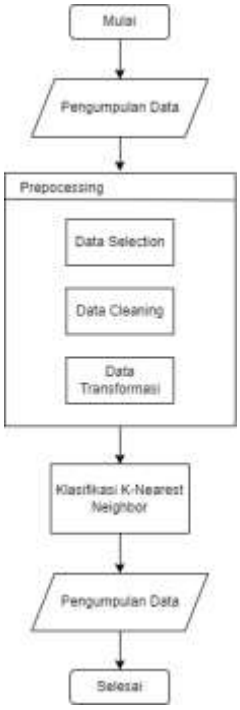
Gambar 1. *Confusion matrix*

Dari hasil proses klasifikasi pada confusion matrix, dihitung tiap nilai dengan menghitung accuracy.

a. Accuracy, merupakan perbandingan prediksi True (positif dan negatif) dengan keseluruhan data. Akurasi menggambarkan seberapa akurat model yang digunakan dalam klasifikasi.
   Rumus: ((TP+TN) / (TP+FP+TN+FN))*100
b. Precission, merupakan perbandingan prediksi true positif dibandingkan dengan keseluruhan hasil yang diprediksi positf. Precission menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model.
   Rumus: (TP / (TP+FP))*100
c. Recall, merupakan perbandingan prediksi true positif dibandingkan dengan keseluruhan data yang benar positif. Recall menggambarkan keberhasilan model dalam menemukan kembali informasi.
   Rumus: (TP / (TP+FN))*100
d. F-1 score, merupakan perbandingan rata-rata presisi dan recall yang dibobotkan menggambarkan perbandingan rata-rata precision dan recall yang dibobotkan.
   Rumus: 2 * precision * recall / (precision+recall).

3. METODE PENELITIAN

Metode penelitian adalah langkah dalam melakukan penelitian sehingga mampu menjawab permasalahan dan tujuan penelitian. Pada metode penelitian terdapat beberapa tahapan dalam penelitian ini. Alur metode penelitian yang dilakukan untuk mengklasifikasi Penyakit Gagal Ginjal Kronis adalah sebagai berikut :



Gambar 2 Tahapan penelitian

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Atribut Data

Sebelum memasuki tahap proses data mining, dataset ini perlu dijelaskan tiap atribut yang ada pada data tersebut, agar dapat menentukan atribut yang digunakan dari data ini. Data diambil selama periode 2 bulan di India dengan 25 fitur (misalnya jumlah sel darah merah, jumlah sel darah putih, dll). Targetnya adalah 'klasifikasi', yaitu 'ckd' atau 'notckd' – ckd = penyakit ginjal kronis. Data perlu dibersihkan karena memiliki NaN dan fitur numerik harus dipaksa untuk mengapung, yang dimana setiap baris jika hanya memiliki satu NaN saja akan dihapus. Data yang telah diperoleh untuk melakukan proses klasifikasi harus melalui tahap normalisasi data terlebih dahulu. Atribut yang digunakan antara nya adalah umur (*age*), tekanan darah (*bp*), berat jenis urin hasil tes urinalysis (*sg*), protein (*al*), gula darah dalam tubuh (*su*), gula darah setelah makan (*bgr*), kadar urea pada darah (*bu*), kadar kreatinin serum pada darah (*sc*), kadar sodium pada darah (*sod*), dan jumlah hematocrit pada darah (*pvc*). Deskripsi atribut dapat dilihat pada Tabel 1 berikut :

Table 1 Deskripsi atribut data

No.	Atribut Data	Deskripsi
1	age	Umur
2	bp	Tekanan Darah
3	sg	Berat jenis urin hasil tes urinalysis
4	al	Albumin, protein hasil sintesis hati
5	su	Gula darah dalam tubuh
6	bgr	Gula darah setelah makan
7	bu	Kadar urea pada darah
8	sc	Kadar kreatinin serum pada darah
9	sod	Kadar sodium pada darah
10	pcv	Jumlah hematocrit pada darah
11	class	1 = terkena gagal ginjal 2= tidak terkena gagal ginjal

4.2. Dataset Gagal Ginjal Kronis

Berikut adalah sampel dari data penjakit ginjal kronis dengan jumlah atribut 14 yang dapat dilihat pada Tabel 2 berikut.

Table 2 Sampel dataset gagal ginjal kronis

age	bp	sg	al	su	bgr	bu	sc	sod	pcv	class
48	8	1.02	1	2	121	36	1.2	135	44	Ckd
7	5	1.02	4	2	99	18	0.8	135	38	Ckd
62	8	1.01	2	3	423	53	1.8	135	31	Ckd
…	…	…	…	…	…	…	…	…	…	…
4	8	1.025	1	2	14	1	1.2	135	48	Notckd
23	8	1.025	1	2	7	36	1	15	52	Notckd
45	8	1.025	1	2	82	49	0.6	147	46	Notckd

4.3. Cleaning Data

Proses data cleaning dilakukan untuk menghilangkan noise dan data lainnya yang tidak diperlukan yang sering kali membuat hasil data tidak akurat. Oleh karena itu, dalam penelitian ini, proses pembersihan data dilakukan dengan mengidentifikasi nilai kosong pada atribut bp, sg, al, su, bgr, bu, sc, sod dan pcv. Untuk mengatasi nilai kosong tersebut, dilakukan penghapusan baris data untuk yang mengandung nilai 0 atau missing value.

4.4. Transformasi Data

Tahap ini melibatkan transformasi data setelah proses pembersihan data selesai. Dalam tahapan ini, variabel data diubah menjadi bentuk numerik atau format data yang sesuai. Dataset yang digunakan dalam penelitian ini terdiri dari 11 atribut yang mencakup age, bp, sg, al, su, bgr, bu, sc, sod, pcv dan class. Dari dataset tersebut, 10 atribut memiliki jenis data numerik sedangkan 1 atribut memiliki jenis data nominal. Untuk mengatasi hal ini dilakukan transformasi data, transformasi ini dilakukan untuk atribut class, yang mana class merupakan bentuk data nominal, menjadi data numerik dengan menerapkan teknik one-hot encode. Teknik ini mewakili bentuk data kategori menjadi numerik. Pada penelitian ini, untuk atribut Class dengan nilai “ckd” menjadi 1, dan “notckd” menjadi 2. hasil transformasi data dapat dilihat seperti pada Tabel 3.

Table 3 Tahapan setelah transformasi data

age	bp	sg	al	su	bgr	bu	sc	sod	pcv	class
48	8	1.02	1	2	121	36	1.2	135	44	1
7	5	1.02	4	2	99	18	0.8	135	38	1
62	8	1.01	2	3	423	53	1.8	135	31	1
…	…	…	…	…	…	…	…	…	…	…
4	8	1.025	1	2	14	1	1.2	135	48	2
23	8	1.025	1	2	7	36	1	15	52	2
45	8	1.025	1	2	82	49	0.6	147	46	2

Penelitian ini dibuat menggunakan algoritma K-Nearest Neighbor. Dalam penelitian ini menggunakan 75% dari data latih dan 25% dari data uji yang digunakan. Hasil pengujian akurasi merupakan tahapan untuk menguji hasil akurasi yang didapatkan pada model yang digunakan. Pengujian ini menggunakan metode *confusion matrix*. Pengujian hasil klasifikasi yang dilakukan oleh model akan diuji menggunakan metode *confusion matrix*. Berikut adalah nilai *confusion matrix* yang dapat dilihat dibawah ini.

Table 4. *Confusion matrix*

True Positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)
62	7	9	138

Rumus :

$$akurasi = \left(\frac{(TP+TN)}{TP+FP+TN+FN}\right) \times 100\%$$

Hasil :

$$akurasi = \left(\frac{(62+138)}{62+7+138+9}\right) \times 100\% = 92.59\%$$

Hasil performa klasifikasi penyakit ginjal kronis menggunakan metode *K-Nearest Neighbor* ditampilkan pada Tabel 5.

Table 5. Hasil performa klasifikasi

Akurasi	Presisi	Recall	F1-score
92,59	89,85	87,32	88,57

Hasil klasifikasi penyakit ginjal kronis menunjukkan bahwa sistem berhasil melakukan klasifikasi dengan nilai akurasi mencapai 92,59%, 89,85% untuk nilai presisi, 87,32% untuk nilai recall dan f1-score mencapai nilai 88,57%.

5. KESIMPULAN DAN SARAN

Berdasarkan pada penelitian ini dibuat menggunakan algoritma *K-Nearest Neighbor*. Dalam penelitian ini menggunakan 75% dari data latih dan 25% dari data uji yang digunakan, hasil klasifikasi penyakit ginjal kronis menunjukkan bahwa sistem berhasil melakukan klasifikasi dengan nilai akurasi mencapai 92,59%, 89,85% untuk nilai presisi, 87,32% untuk nilai recall dan f1-score mencapai nilai 88,57%. Hasil akurasi yang diperoleh cukup tinggi dalam mengklasifikasi penyakit gagal ginjal sesuai dengan atribut yang digunakan pada data.

Kesimpulan dari penelitian ini adalah metode *K-Nearest Neighbor* merupakan metode klasifikasi yang baik untuk diterapkan pada bidang kesehatan terutama dalam mengklasifikasi penyakit gagal ginjal kronis yang dimana penyakit tersebut dikategorikan sebagai penyakit yang tingkat kematiannya tinggi. Adapun saran yang penulis berikan saat melakukan penelitian ini ialah dapat memperbanyak data yang digunakan agar hasil dari nilai akurasi semakin akurat dan semakin tinggi. Hal ini dikarenakan metode K-Nearest Neighbor sangat tergantung dari banyaknya data latih dan data uji yang digunakan.

DAFTAR PUSTAKA

[1] I. Fadilla, P. P. Adikara, and R. Setya Perdana, "Klasifikasi Penyakit Chronic Kidney Disease (CKD) Dengan Menggunakan Metode Extreme Learning Machine (ELM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3397–3405, 2018, [Online]. Available: https://www.researchgate.net/publication/323365845

[2] J. Snegha, V. Tharani, S. D. Preetha, R. Charanya, and S. Bhavani, "Chronic Kidney Disease Prediction Using Data Mining," *Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020*, pp. 1–5, 2020, doi: 10.1109/ic-ETITE47903.2020.482.

[3] A. Ariani and Samsuryadi, "Klasifikasi Penyakit Ginjal Kronis menggunakan K-Nearest Neighbor," *Pros. Annu. Res. Semin. 2019*, vol. 5, no. 1, pp. 148–151, 2019.

[4] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.

[5] K. A. Sugiarta, I. Cholissodin, and E. Santoso, "Optimasi K-Nearest Neighbor Menggunakan Bat Algorithm untuk Klasifikasi Penyakit Ginjal Kronis," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, no. 10, p. 964X, 2020.

[6] F. Yunita, "Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor (K-NN)," *Bappeda*, vol. 2, no. 1, pp. 223–230, 2016.

[7] R. Aji Pangestu, Taslim, Y. Yunefri, Kursiasih, and E. Sabna, "Optimasi Nilai k Pada Algoritma K-Nearest Neighbor Untuk Klasifikasi Pasien Covid-19 Yang Membutuhkan Ruangan ICU," *Jl. Tamtama*, vol. 7, no. 1, pp. 147–155, 2022, [Online]. Available: www.kaggle.com

[8] V. Kunwar, K. Chandel, A. S. Sabitha, and A. Bansal, "Chronic Kidney Disease analysis using data mining classification techniques," *Proc. 2016 6th Int. Conf. - Cloud Syst. Big Data Eng. Conflu. 2016*, pp. 300–305, 2016, doi: 10.1109/CONFLUENCE.2016.7508132.

[9] A. Yandi Saputra and Y. Primadasa, "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour Implementation of Classification Method to Predict Student Graduation Using K-Nearest Neighbor Algorithm," *Techno.Com*, vol. 17, no. 4, p. 9, 2018.

[10] U. Islam *et al.*, "Klasifikasi Dokumen Tugas Akhir Berbasis Text Mining menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor," *Semin. Nas. Teknol. Informasi, Komun. dan Ind. 11*, no. November, pp. 178–186, 2019.