

KLASIFIKASI DATA REKAM MEDIS BERDASARKAN *STATISTICAL CLASSIFICATION OF DISEASES (ICD-10)* MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR (K-NN)*

Mohamad Farhan Azahry¹, Martaleli Bettiza², Nola Ritha³
farhanazahry@gmail.com

Program studi Teknik Informatika, Fakultas Teknik, Universitas Maritim Raja Ali Haji

Abstract

hospital is a health service institution that provides complete individual health services that provide inpatient, outpatient and emergency services. In general, the benefits of medical records in hospitals are as a support for orderly administration in an effort to improve health services in hospitals. However, after the administrative goals are met medical records of patients going into the archive to be stored in accordance with Permenkes 269 / Menkes / Per / III / 2008 in Chapter IV, Article 8 provides that the medical records of inpatients in the hospital must be kept at the lack of for a period of 5 (five) years from the last date the patient was treated or discharged. In the end, the medical record data that grows every day becomes a pile of data. Therefore, the author wishes to create a system that can process hospital medical record data in the form of disease classification based on the International Statistical Classification of Disease and Related Health Problems (ICD-10). The study used pre-existing medical record data with a total of 1000 data. Where the experimental data is divided into 3 variations for the system experiment, namely 21 classes of ICD-10, 11 classes of ICD-10, and 3 classes of ICD-10. The results of this study get the highest accuracy of 73.13% on experimental data 80%: 20%. with 3 ICD-10 classes.

Keywords: *Medical record, Classification, K-Nearest Neighbor, ICD-10*

I. Pendahuluan

Rumah sakit merupakan institusi pelayanan kesehatan yang menyelenggarakan pelayanan kesehatan perorangan secara paripurna yang menyediakan pelayanan rawat inap, rawat jalan dan gawat darurat. Dalam memberikan pelayanan, rumah sakit selalu berusaha untuk memberikan yang terbaik bagi masyarakat seperti dengan mencatat semua tindakan pelayanan medis yang diberikan rumah sakit terhadap pasien serta semua data sosial dan riwayat kesehatan pasien. Data pasien tersebut dicatat dalam sebuah berkas atau dokumen yang disebut dengan rekam medis. Secara umum manfaat rekam medis pada rumah sakit yaitu sebagai penunjang tertib administrasi dalam rangka upaya peningkatan pelayanan kesehatan di rumah sakit. Namun, setelah tujuan administratif tersebut terpenuhi data rekam medis pasien akan menjadi arsip untuk disimpan. data rekam medis yang bertambah setiap harinya menjadi tumpukan data. Padahal data rekam medis dapat diolah kembali agar menjadi suatu representasi berupa pengetahuan yang dapat digunakan sebagai rujukan bagi rumah sakit dalam membuat kebijakan maupun dalam penanganan kesehatan seperti pola kecenderungan penyakit yang diderita oleh sekelompok masyarakat. Pada penelitian ini dengan judul “klasifikasi data rekam medis berdasarkan *International Statistical Classification of Disease (ICD-10)*” menggunakan data yang ada di Rumah sakit umum Raja Ahmad Tabib berupa data rekam medis. Dengan adanya data *mining* algoritma *K-Nearest Neighbor* diharapkan dapat mengklasifikasikan kelas berdasarkan data *training* dan *data testing*.

II. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data rekam medis pada rumah sakit umum Raja Ahmad Tabib Tanjungpinang 1 tahun dimulai dari januari 2019 hingga desember 2019. Data tersebut memiliki beberapa atribut yang digunakan dalam penelitian yaitu alamat (kecamatan), umur, jenis kelamin, bulan kunjungan pasien, kode ICD. Sebelum diproses menggunakan algoritma *K-Nearest Neighbor*, data ditransformasikan agar data dapat diolah.

2.1 Algoritma *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* merupakan algoritma supervised learning dimana hasil yang didapat diklasifikasikan berdasarkan mayoritas dari jarak tetangga yang terdekat. Tujuan algoritma ini untuk mengklasifikasikan objek baru didapat berdasarkan atribut dan sampel dari data pelatihan. Pada umumnya nilai K menggunakan jumlah ganjil agar tidak terdapat jarak yang sama dalam proses klasifikasi. Jauh atau dekatnya tetangga dihitung menggunakan persamaan jarak Euclidean. Algoritma K-NN adalah metode untuk mengklasifikasikan objek berdasarkan kedekatan objek dalam set pelatihan. Algoritma ini mengambil data pasien dan membandingkannya dengan tabel pelatihan pasien dengan berbagai kondisi medis dan atribut medis per kondisi. Algoritma tersebut kemudian menggunakan teknik K-NN untuk mengklasifikasikan pasien dalam kondisi tertentu atau tidak. Hal ini disebabkan oleh berbagai keunggulan metode K-NN dibandingkan metode *supervised learning* lainnya. Misalnya, K-NN lebih cocok untuk proyek ini daripada Naive Bayes karena sifat non-parametrik. K-NN dapat beradaptasi dengan set pelatihan yang relatif beragam, mudah diterapkan, dan secara alami menangani kasus multi-class. K-NN juga memiliki sejarah tingkat keberhasilan yang lebih tinggi di bidang medis.

$$d(x, y) = \sqrt{(\sum_{i=1}^n (x_i - y_i)^2)}$$

Keterangan:

x_i = data sampel

y_i = data training

i = variabel data

$d(x, y)$ = jarak antara x dengan y

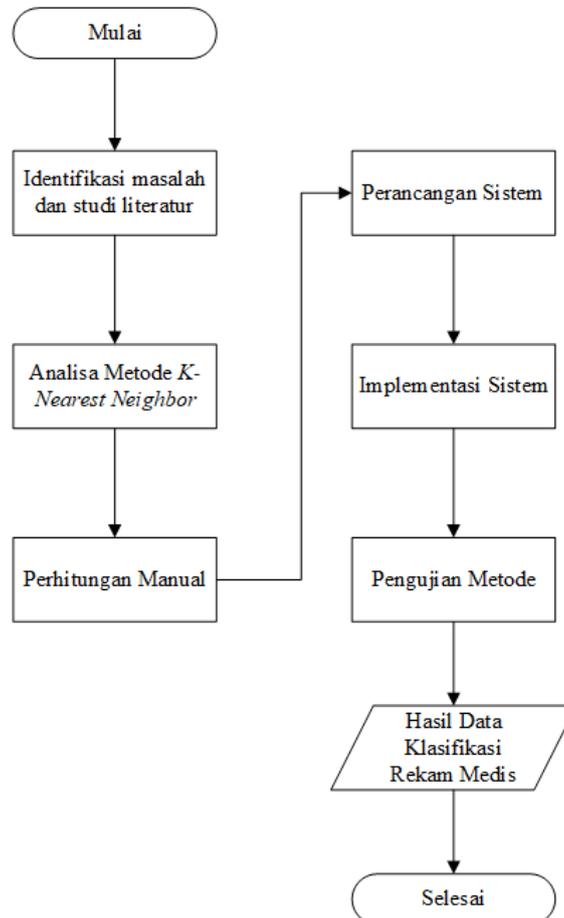
n = dimensi data

Secara garis besar ada beberapa tahapan cara kerja algoritma *K-Nearest Neighbor* yaitu (Tayeb et al., 2018).

1. Tentukan nilai untuk k
2. Temukan k titik terdekat pada set pelatihan dibandingkan dengan titik data uji
3. Klasifikasikan titik data uji berdasarkan yang paling populer

2.2 Flowchart Metode Penelitian

Metode penelitian pada klasifikasi algoritma *K-Nearest Neighbor* memiliki beberapa tahapan yang dilakukan, urutan ini merupakan gambaran dari langkah-langkah yang dikerjakan selama proses penelitian berlangsung. Berikut ini adalah tahapan kerangka pikir penelitian yang dapat dilihat pada Gambar 1.

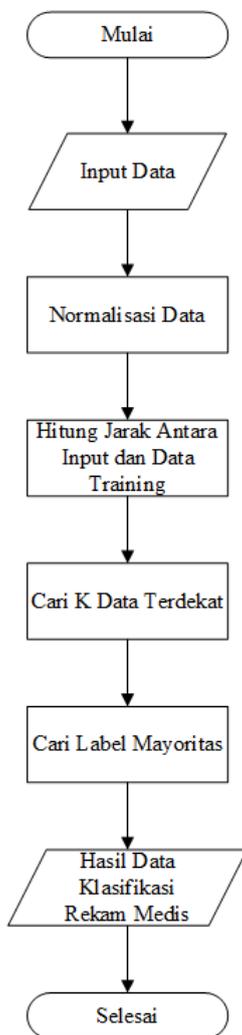


Gambar 1. Flowchart Metode Penelitian

Berdasarkan *flowchart* diatas, menunjukkan urutan tahapan penelitian. Dimulai dari mengidentifikasikan masalah dan studi literatur untuk mengetahui dan mendapatkan informasi terkait dengan metode, studi kasus, parameter yang akan dijadikan sebagai acuan dalam pengumpulan data yang sesuai penelitian ini. Kemudian melakukan analisa metode *K-Nearest Neighbor* untuk digunakan dalam penelitian. Data yang diperoleh akan digunakan untuk diolah dalam proses perhitungan manual metode *K-Nearest Neighbor* sehingga dapat dilakukan perancangan sistem dan implementasi sistem. Selanjutnya dilakukan pengujian metode untuk mendapatkan hasil data klasifikasi rekam medis sehingga mendapatkan hasil dan kesimpulan dalam penelitian.

2.3 Flowchart Algoritma K-Nearest Neighbor

Adapun *flowchart* algoritma *K-Nearest Neighbor* dalam menentukan klasifikasi menggunakan data rekam medis dapat dilihat pada gambar 2.



Gambar 2. *Flowchart* algoritma *K-Nearest Neighbor*

Klasifikasi rekam medis dengan menggunakan Algoritma K- Nearest Neighbor harus melalui beberapa tahapan yaitu dengan menginputkan data terlebih dahulu. Kemudian melakukan normalisasi data, setelah normalisasi data menghitung jarak antara data training dan data testing. Setelah proses tersebut maka dicari label mayoritas dari kelas-kelas yang ada. Kemudian setelah mencari label mayoritas maka akan didapatkan hasil klasifikasi data rekam medis.

III. Hasil dan Pembahasan

Tabel Hasil Proses Sistem

Hasil uji coba sistem pada penelitian ini dilakukan sesuai dengan percobaan yang telah dijelaskan. Hasil dari setiap percobaan dikelompokkan dalam bentuk tabel. Hasil akurasi percobaan dapat dilihat pada tabel berikut :

1. Hasil Akurasi Percobaan ke-1

Berikut merupakan pengujian data terhadap 21 kelas ICD-10 dapat dilihat pada Tabel 1.

Tabel 1. Hasil akurasi percobaan ke-1

Kelas ICD-10	K	70% : 30%	80% : 20%	90% : 10%
21	3	22,00%	26,50%	19,00%
	5	26,67%	30,50%	21,00%
	7	29,00%	29,00%	27,00%
	9	28,33%	31,50%	29,00%
	11	29,67%	30,50%	36,00%

2. Hasil Akurasi Percobaan ke-2

Berikut merupakan pengujian data terhadap 11 kelas ICD-10 dapat dilihat pada Tabel 2.

Tabel 2. Hasil akurasi percobaan ke-2

Kelas ICD-10	K	70% : 30%	80% : 20%	90% : 10%
11	3	31,89%	38,10%	26,87%
	5	36,76%	44,44%	37,31%
	7	36,76%	43,65%	35,82%
	9	42,70%	40,48%	44,78%
	11	41,62%	40,48%	40,30%

3. Hasil Akurasi Percobaan ke-3

Berikut merupakan pengujian data terhadap 3 kelas ICD-10 dapat dilihat pada Tabel 3.

Tabel 3. Hasil akurasi percobaan ke-3

Kelas ICD-10	K	70% : 30%	80% : 20%	90% : 10%
3	3	65,42%	73,13%	67,57%
	5	63,55%	71,64%	62,16%
	7	63,55%	70,15%	62,16%
	9	69,16%	70,15%	59,46%
	11	70,09%	70,15%	67,57%

Setelah dilakukan uji coba terhadap Metode *K-Nearest Neighbor*, nilai akurasi tertinggi terdapat pada **Tabel 3** dengan 3 kelas ICD-10, pengujian k-3 sebesar 73,13%. Setelah dilakukan uji coba terhadap Metode *K-Nearest Neighbor* pada data training 50% dan data testing 50%, nilai akurasi tertinggi terdapat pada pengujian k=5 dengan akurasi 72,83%. Kemudian uji coba pada data training 60% dan data testing 40% nilai akurasi tertinggi terdapat pada pengujian k=9 dengan akurasi 72,73%. Uji coba pada data training 70% dan data testing 30% nilai akurasi tertinggi terdapat pada pengujian k=11 dengan akurasi 70,09%. Uji coba pada data training 80% dan data testing 20% nilai akurasi tertinggi terdapat pada pengujian k=3 dengan akurasi 73,13%. Pengujian terakhir pada data training 90% dan data testing 10% nilai akurasi tertinggi terdapat pada pengujian k=11 dengan 67,57% . Dengan pengujian tersebut memberikan hasil yang baik pada 3 kelas ICD-10 data rekam medis, sedangkan untuk pengujian dengan 21 kelas ICD-10 memberikan hasil akurasi yang rendah, sehingga tingkat akurasi yang dilakukan saat percobaan dibawah 50%.

IV. Kesimpulan

Berdasarkan pengujian pada sistem yang telah dilakukan , diperoleh beberapa kesimpulan sebagai berikut :

1. Algoritma *K-Nearest Neighbor* yang digunakan ini mampu mengklasifikasi data rekam medis kedalam 21 kelas sesuai dengan pedoman klasifikasi penyakit yang dibuat oleh WHO yaitu *International Statistical Classification of Disease*.
2. Pengujian untuk menemukan nilai akurasi terbaik dilakukan pada 3 jumlah kategori ICD-10. Hasil akurasi tertinggi didapat pada percobaan ke-3 dengan akurasi 73,13%. Hal ini menunjukkan bahwa tingkat akurasi Algoritma *K-Nearest Neighbor* akan lebih baik jika menggunakan kategori yang lebih sedikit.

V. Daftar Pustaka

- Fiandra, Y.A., Defit,S., & Yuhandri, 2017, Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10), *Jurnal Rekayasa Sistem dan Teknologi Informasi*. 1(2): 82-89.
- Hand, D.J., 2007, *Principles of Data Mining*, Springer, London.
- Hutagalung, N., 2018, Klasifikasi Tipe Kepribadian Pengguna Sosial Media Berdasarkan Teori Big Five Menggunakan K-Nearest Neighbor, *Skripsi, Universitas Sumatera Utara*.
- Ina,W.T, 2013, Klasifikasi Data Rekam Medis Berdasarkan Kode Penyakit Internasional Menggunakan Algoritma C4.5, *Jurnal Media Elektro*. Vol 2, No2.
- Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. 2013, *Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm, Procedia Technology*, 10, 85–94.
- Munas, R. A., 2014, Sistem Pendukung keputusan Untuk Diagnosa Dini Gangguan Pemusatan Perhatian dan Hiperaktivitas (GPPH) Pada Anak Menggunakan Metode Case-Based Reasoning (CBR), *Skripsi, UIN Malik Ibrahim*
- Park, J., & Lee, D. H. 2018, *Privacy preserving k-nearest neighbor for medical diagnosis in e-health cloud, Journal of Healthcare Engineering*, 2018.
- Republik Indonesia, 2008, Peraturan Menteri Kesehatan No. 269 Tahun 2008 Tentang Rekam Medis, Sekretariat Negara, Jakarta.
- S.Raikwal, J., & Saxena, K. 2012, *Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set, International Journal of Computer Applications*, 50(14), 35–39.
- Tayeb, S., Pirouz, M., Sun, J., Hall, K., Chang, A., Li, J., Song, C., Chauhan, A., Ferrar, M., Sager, T., Zhan, J., & Latifi, S., 2018, *Toward Predicting Medical Conditions Using K-Nearest Neighbors, IEEE International Conference on Big Data* 3897– 3903.
- WHO, 2008, *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Volume 2, Instruction Manual, Geneva, World Health Organization*.
- WHO, 2016, *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Volume 1, Tabular List, Geneva, World Health Organization*.