

**KLASIFIKASI DIAGNOSA PENYAKIT DEMAM BERDARAH DENGUE PADA ANAK
MENGUNAKAN METODE *K-NEAREST NEIGHBOR*
STUDI KASUS RUMAH SAKIT PKU MUHAMMADIYAH UJUNG PANGKAH
GRESIK**

Rafika Amilia¹, Eko Prasetyo²

¹Rumah Sakit PKU Muhammadiyah Ujung Pangkah
Jalan Raya Dandels No.21, Sekapuk, Ujungpangkah, Doudo, Kec. Panceng, Kabupaten Gresik,
Jawa Timur 61154

²Teknik Informatika Universitas Muhammadiyah Gresik
Fakultas Teknik, Universitas Muhammadiyah Gresik
Jl. Sumatra 101 Gresik Kota Baru (GKB), Randuagung, 61121 Telp
E-mail: rafikaamilia877@gmail.com¹

ABSTRAK

Demam Berdarah Dengue (DBD) merupakan penyakit yang disebabkan oleh gigitan nyamuk *Aedes* yang terinfeksi salah satu dari empat tipe virus dengue dengan manifestasi klinis demam, nyeri otot dan atau nyeri sendi yang di sertai leukopenia, ruam, trombositopenia dan diathesis hemoragik. Ada 5 kriteria untuk menentukan seorang pasien dikategorikan positif atau negatif DB, yaitu variabel Usia pasien, Jenis kelamin pasien, peningkatan hemoglobin (Hb), peningkatan trombosit dan peningkatan hematokrit (Ht) berdasarkan kriteria tersebut maka akan dilakukan klasifikasi data pasien positif dan negatif DBD dengan metode *K-Nearest Neighbor* berbasis web. metode *K-Nearest Neighbor* (KNN) memiliki kelebihan menghasilkan nilai error yang kecil. Algoritma KNN menggunakan klasifikasi ketetangga sebagai nilai prediksi dari sampel uji yang baru. Dekat atau jauhnya nilai tetangga biasanya dihitung berdasarkan jarak *euclidean*. Berdasarkan dari hasil penelitian dan pembahasan yang dilakukan, klasifikasi diagnose penyakit demam berdarah ini menggunakan algoritma KNN dengan variabel usia, jenis kelamin, hemoglobin, trombosit, hematokrit.

Kata kunci : *Klasifikasi. Diaenosa. K-Nearest Neighbor. demam berdarah dengue*

I. PENDAHULUAN

Rumah sakit PKU Muhammadiyah Ujung Pangkah adalah salah satu unit pusat kesehatan masyarakat yang berada di Kecamatan Ujung Pangkah Kabupaten Gresik dengan menyelenggarakan upaya kesehatan yang bersifat menyeluruh, terpadu, merata dan dapat diterima oleh masyarakat. Dengan perkembangan zaman yang semakin maju, maka banyak pula penyakit yang dapat menyerang manusia terutama penyakit-penyakit yang berbahaya

Salah satu penyakit berbahaya yang sering dijumpai adalah penyakit demam berdarah dengue (DBD). Dengan demikian maka perlu dibuatkan suatu sistem yang dapat mendiagnosa penyakit demam berdarah dengue (DBD) sehingga dapat memudahkan dokter dalam memberikan informasi atau arahan.

Permasalahan yang dihadapi selama ini adalah banyaknya orang yang tidak mengetahui gejala dari penyakit demam berdarah dengue (DBD) sehingga Dokter harus cepat tanggap mengenali gejala penyakit demam berdarah dengue (DBD). Sebagian dari pasien yang mengalami demam berdarah akan berakhir dengan sakit yang serius. Pada penelitian sebelumnya yang pernah dibuat oleh (Hasibuan A.R, 2017) untuk tugas akhir dengan judul “Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue (DBD) Menggunakan *Support Vector Machine* (SVM)”. Teknik klasifikasi diagnose menggunakan metode *Support Vector Machine* (SVM) salah satu algoritma pembelajaran mesin yang di perkenalkan oleh vapnik yaitu bagaimana memisahkan dua kelas dengan suatu fungsi yang di peroleh dari data training yang tersedia. Konsep klasifikasi SVM dapat di jelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas data pada input *space*.

Pembuatan sistem klasifikasi diagnosa yang akan di lakukan pada penelitian ini menggunakan metode K-Nearest neighbor metode ini nantinya akan menghasilkan sebuah perhitungan yang diharapkan dapat mengklasifikasi diagnosa penyakit demam berdarah. Metode *K-Nearest Neighbor* tepat digunakan untuk salah satu kasus klasifikasi dimana nilai K yang digunakan disini menyatakan jumlah tetangga terdekat yang dilibatkan dalam penentuan prediksi label kelas data uji. Kelas dengan jumlah suara tetangga terbanyaklah yang diberikan sebagai label hasil diagnosa pada data uji tersebut. Dari beberapa penelitian yang dilakukan dengan metode *K-Nearest Neighbor* (KNN) dapat dijadikan literature pada skripsi ini yang menggunakan metode klasifikasi diagnosa dengan *K-Nearest Neighbor*.

Pemilihan metode tersebut dikarenakan merupakan salah satu metode klasifikasi dimana data hasil dari sampel uji yang baru di klasifikasikan berdasarkan mayoritas dari kategori KNN. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Dekat atau jauhnya nilai tetangga biasanya dihitung berdasarkan jarak *euclidean*. Berdasarkan latar belakang tersebut, maka akan dilakukan penelitian masalah yang akan ditulis dalam bentuk tugas akhir dengan judul “**Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue (DBD) Pada Anak Menggunakan Metode K-NN (Studi Kasus : RS. PKU Muhammadiyah Ujung Pangkah Gresik)**”.

2. LANDASAN TEORI

2.1 Definisi Sistem

Definisi sistem berkembang sesuai konteks dimana pengertian sistem itu digunakan. Berikut akan diberikan definisi sistem secara umum:

1. Kumpulan dari bagian-bagian yang bekerja sama untuk mencapai tujuan yang sama (Jogiyanto, H.M. 2005).
2. Sistem adalah kumpulan dari elemen-elemen yang berinteraksi untuk mencapai suatu tujuan tertentu (Jogiyanto, H.M. 2005).
3. Suatu sistem adalah jaringan kerja dari prosedur-prosedur yang saling berhubungan, berkumpul bersama-sama untuk melakukan suatu kegiatan atau untuk menyelesaikan suatu sasaran yang tertentu (Jogiyanto, H.M. 2005).

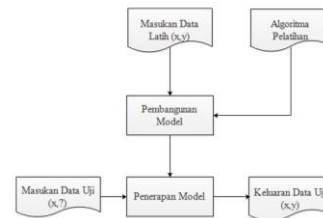
4. Sistem merupakan kumpulan elemen-elemen yang saling terkait dan bekerja sama untuk memproses masukan (*input*) yang ditujukan kepada sistem tersebut dan mengolah masukan tersebut sampai menghasilkan keluaran (*output*) yang diinginkan (Kadir A, Triwahyuni TCH. 2003).

Suatu sistem memiliki karakteristik atau sifat-sifat tertentu, yaitu:

1. Komponen sistem (*components*)
2. Batas sistem (*boundary*)
3. Lingkungan luar sistem (*environments*)
4. Penghubung sistem (*interface*)
5. Masukan sistem (*input*)
6. Keluaran sistem (*output*)
7. Pengolah sistem
8. Sasaran sistem (*objective*)

2.2 Klasifikasi

Menurut (Prasetyo, E, 2012), Klasifikasi dapat didefinisikan secara detail sebagai pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target f yang memetakan setiap vektor (set fitur) x ke satu dari sejumlah label kelas y yang tersedia. Model dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, di mana ada suatu model yang menerima masukan, kemudian mampu melakukan pemikiran terhadap masukan tersebut, dan memberikan jawaban sebagai keluaran dari hasil pemikirannya. Kerangka kerja (*framework*) klasifikasi ditunjukkan pada Gambar 2.1. Pada gambar tersebut, disediakan sejumlah data latih (x,y) untuk digunakan sebagai data pembangunan model, kemudian menggunakan model tersebut untuk memprediksi kelas dari data uji $(x,?)$ sehingga data uji $(x,?)$ diketahui kelas y yang sesungguhnya.



Gambar 2.1 Proses kerja klasifikasi

Model yang sudah dibangun pada saat pelatihan kemudian dapat digunakan untuk memprediksi label kelas data baru yang belum diketahui label kelasnya. Dalam pembangunan model selama proses pelatihan tersebut diperlukan adanya suatu algoritma untuk membangunnya yang disebut sebagai algoritma pelatihan (*learning algorithm*). Kerangka kerja seperti yang ditunjukkan pada **Gambar 2.1** meliputi dua langkah proses yaitu induksi dan deduksi. Induksi merupakan suatu langkah untuk membangun model klasifikasi dari data latih yang diberikan, disebut juga proses pelatihan. Sedangkan deduksi merupakan suatu langkah untuk menerapkan model tersebut pada data uji sehingga data uji dapat diketahui kelas yang sesungguhnya atau disebut juga proses prediksi.

2.3 Normalisasi Data

Normalisasi data linier adalah proses penskalaan nilai atribut data sehingga bisa jatuh pada range tertentu. Tujuan dari normalisasi data adalah untuk mempersempit atau mengecilkan nilai range pada data tersebut. Normalisasi yang digunakan pada penelitian ini adalah *min-max normalization* yang merupakan proses transformasi nilai dari data yang dikumpulkan pada range *value* antara 0.0 dan 1.0, dimana nilai terkecil (*min*) adalah 0.0 dan nilai tertinggi (*max*) adalah 1.0 (Chandrasekhar, Thangavel dan Elayaraja, 2011).

$$\text{normalisasi } x_{ik} = \frac{x - \min(x_k)}{\max(x_k) - \min(x_k)} \dots\dots\dots(2.1)$$

Keterangan :

- x_{ik} = nilai hasil normalisasi
- x = nilai x sebelum normalisasi
- $\min(x_k)$ = nilai minimum dari fitur
- $\max(x_k)$ = nilai maksimum dari fitur

2.4 Metode K-Nearest Neighbor (KNN)

K-Nearest Neighbors (KNN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. KNN termasuk dalam golongan *supervised learning*, dimana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN. Nantinya kelas yang baru dari suatu data akan dipilih berdasarkan grup kelas yang paling dekat jarak vektornya.

Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah k obyek atau (titik *training*) yang paling dekat dengan titik *query*. Klasifikasi menggunakan *voting* Algoritma *K-Nearest Neighbors* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru. Metode *K-Nearest Neighbors* sangatlah sederhana,

bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan KNN- nya. *Training sample* diproyeksikan ke ruang berdimensi banyak, dimana masing- masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi *training sample*. Sebuah titik pada ruang ini ditandai jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance*.

Jarak *Euclidean* paling sering digunakan dalam menghitung jarak karena sangat cocok untuk menggunakan jarak terdekat (lurus) antara dua data. Jarak *euclidean* berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua obyek yang direpresentasikan sebagai berikut:

$$D(a,b) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \dots\dots\dots(2.2)$$

Keterangan :

- D (a ,b) = Jarak *Euclidean* Data a dan Data b
- X = Koordinat titik X
- Y = Koordinat titik Y

3. ANALISIS DAN PERANCANGAN SISTEM

3.1 Analisis Sistem

Analisis sistem merupakan suatu proses melihat keseluruhan masalah dengan cara sistematis, menetapkan tujuan sistem, mengidentifikasi hambatan untuk mengidentifikasi pemecahan masalah tersebut. Analisis ini diperlukan sebagai dasar bagi tahapan perancangan sistem. Sistem yang akan dibuat merupakan sistem diagnosa penyakit demam berdarah dengue. Dengan adanya sistem ini maka diharapkan dapat mempermudah dokter dalam mendiagnosa beberapa pasien yang terkena atau tidak nya penyakit demam berdarah dengue.

Banyaknya masyarakat yang tidak menyadari dengan gejala penyakit demam berdarah mengakibatkan dinas kesehatan mengalami kesulitan dalam mengetahui banyaknya masyarakat yang terkena DBD. Pengendalian penyakit DBD dapat dilakukan dengan melakukan pendekatan faktor resiko dari penyakit demam berdarah.

Sistem klasifikasi diagnosa penyakit DBD pernah dilakukan penelitian sebelumnya dengan menggunakan metode klasifikasi *support vector machine* (SVM) yaitu dengan memisahkan dua kelas dengan suatu fungsi yang diperoleh dari data training yang tersedia. Konsep klasifikasi dengan SVM dapat dijelaskan secara sederhana sebagai usaha mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah kelas data pada input space. Karena SVM terbukti memilih kinerja yang lebih unggul, karena telah mampu serratus persen mengklasifikasi data berdasarkan kelas yang tepat kemampuan generalisasi SVM sangat baik yaitu 95,4 %.

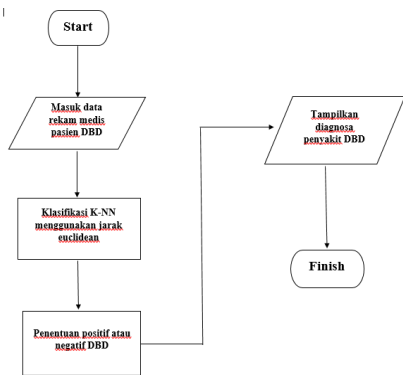
terdapat dua proses pengenalan pada metode *support vector machine* (SVM) yaitu proses pembelajaran dan proses pengujian. Sehingga untuk memudahkan dalam mengklasifikasi diagnosa penyakit DBD akan digunakan dengan metode *K-Nearest Neighbor* dimana pada metode tersebut termasuk kedalam golongan *supervised learning*, dimana hasil query *instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam *K-Nearest Neighbor* nantinya kelas yang baru dari suatu data akan dipilih berdasarkan grup kelas yang paling dekat jarak vektornya.

Data yang digunakan sebagai atribut yaitu data yang berhubungan dengan penyakit demam berdarah dengue seperti usia, jenis kelamin, hemoglobin, trombosit, hematokrit. Oleh karena itu pada sistem klasifikasi diagnosa ini akan dibagi kedalam kategori kelas positif dan kelas negatif. Dengan demikian dokter dapat memberikan informasi atau arahan kepada pasien agar terhindar atau meminimalkan dari serangan penyakit demam berdarah.

3.2 Hasil Analisis

Hasil analisis yang didapat adalah sistem klasifikasi ini dapat membantu dokter untuk mendukung keputusan yang tepat mengenai bagaimana penanganan yang sesuai berdasarkan tingkat pasien untuk mempercepat proses penyembuhan dan mencegah kemungkinan terburuk bagi pasien dengan mengklasifikasikan tingkat pasien yang positif dan negatif terkena demam berdarah dengue. Pembuatan aplikasi klasifikasi metode *K-Nearest Neighbor* dibutuhkan data pembelajaran, data tersebut didapatkan dari data rekam medis di Rumah Sakit PKU Muhammadiyah Ujung Pangkah Gresik.

Data tersebut akan diolah menggunakan metode *K-Nearest Neighbor* yaitu berupa keluaran positif atau negatif. Sistem yang akan dibangun termasuk kedalam sistem pengelompokan (*clustering*). Sistem yang dibuat harus mampu mendiagnosa pasien penderita demam berdarah berdasarkan data dari rekam medis menggunakan teknik klasifikasi data mining dengan metode *K-Nearest Neighbor*. Sistem yang akan dibangun ditujukan untuk digunakan dokter dalam menentukan diagnosa pasien penyakit demam berdarah dengue. Sistem ini akan menghasilkan nilai keluaran berupa kategori positif atau negatif. Terdapat beberapa atribut yang dibutuhkan untuk mengklasifikasikan pasien terkena atau tidaknya diantaranya usia, jenis kelamin, hemoglobin, trombosit, hematokrit. Pada **Gambar 3.1** akan menjelaskan bagaimana proses alur sistem pada aplikasi sistem klasifikasi diagnosa penyakit demam berdarah dengue.

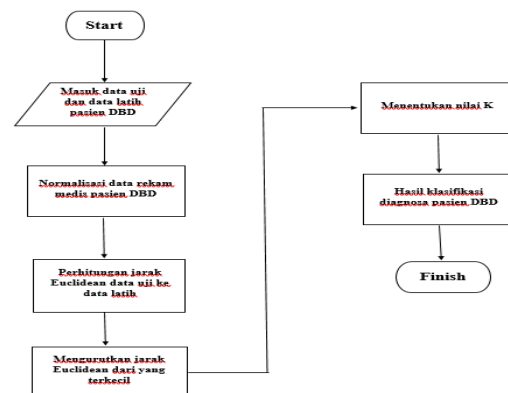


Gambar 3.1 Flowchart sistem klasifikasi diagnosa penyakit demam berdarah

Penjelasan untuk **Gambar 3.1** :

1. Pertama memasukkan data rekam medis pasien yang positif dan negatif terkena demam berdarah *dengue*.
2. Sistem akan mulai melakukan proses perhitungan klasifikasi dengan menggunakan metode *K-Nearest Neighbor* (KNN) dimana menggunakan perhitungan kedekatan jarak *Euclidean*, nanti nya kelas yang baru dari suatu data akan di pilih berdasarkan grup kelas yang paling dekat dengan jarak vektornya.
3. Setelah dilakukan proses klasifikasi dan telah menentukan hasil akhir, selanjutnya menentukan diagnosa pasien demam berdarah dengue berdasarkan hasil keluaran kelas baru dari proes perhitungan klasifikasi.
4. Selanjutnya sistem akan menampilkan keluaran hasil diagnosa pasien demam berdarah dengue yang masuk kedalam kategori positif (1) atau negatif (2).

Sedangkan untuk gambar diagram alir dari metode *K-Nearest Neighbor* dapat dilihat pada **Gambar 3.2**



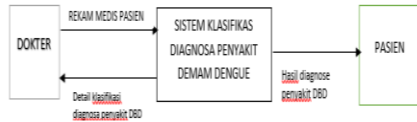
Gambar 3.2 Flowchart sistem metode *K-Nearest Neighbor*

INDEXIA: Informatic and Computational Inteligent Journal

Rafika Amilia, Eko Prasetyo

Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue Pada Anak Menggunakan Metode K-Nearest Neighbor Studi Kasus Rumah Sakit Pku Muhammadiyah Ujung Pangkah Gresik

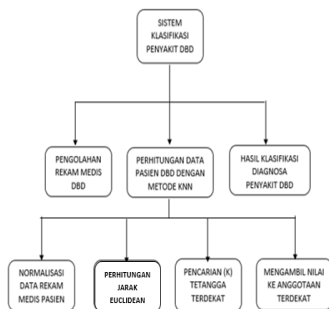
3.3 Diagram Konteks



Gambar 3.3 Diagram Konteks

Penjelasan dari **Gambar 3.3** terlihat bahwa *stackholder* atau *entity* yang terlibat dalam sistem ini adalah dokter dan pasien. Dokter memasukkan data rekam medis pasien penyakit demam berdarah dengue sebagai data uji dan data latih yang terdiri dari umur, jenis kelamin, *hemoglobin*, *trombosit*, *hematrokrit* Kemudian akan diproses didalam sistem klasifikasi penyakit demam berdarah dengue dengan menggunakan metode *K-Nearest Neighbor*. Selanjutnya dokter akan menerima hasil detail klasifikasi diagnosa pasien dari sistem dan laporan diagnosa penyakit demam berdarah dengue akan diberikan untuk pasien.

3.4 Diagram Berjenjang



Gambar 3.4 Diagram Berjenjang

Diagram berjenjang disajikan pada **Gambar 3.4** berikut penjelasannya:

Top Level : sistem klasifikasi penyakit demam berdarah dengue di Rumah sakit PKU Muhammadiyah Ujung pangkah Gresik

Level 0 :

1. Pengelolahan data rekam medis
2. Perhitungan diagnose pasien penyakit demam berdarah dengue dengan metode K-Nearest Neighbor.
3. Laporan klasifikasi diagnosa pasien demam berdarah dengue.

Level 1 :

- 2.1 Normalisasi data rekam medis pasien
- 2.2 Perhitungan Jarak *Euclidean*
- 2.3 Pencarian K tetangga terdekat
- 2.4 Mengambil nilai keanggotaan terdekat.

4. IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1 Pengujian Satu Data

Data rekam medis yang didapat akan dilakukan satu pengujian data uji, contoh kasus dari permasalahan sebagai berikut:

1. Seorang pasien gagal ginjal kronis memiliki data rekam medis umur = 53, bun= 47.5, kreatinin = 3.93, hb = 11.1, lfg = 25.56, bb = 62 dan jk = laki-laki. Masukkan data tersebut pada menu pasien, kemudian lakukan uji klasifikasi data rekam medis dengan nilai K yang sudah ditentukan seperti pada **Gambar 6**.

The screenshot shows a software interface with several data tables. The first table, 'Tabel Min Max', lists minimum and maximum values for variables TB, HB, HT, U, and JK. Subsequent tables show the normalization of these values and the calculation of Euclidean distances for various data points. The final table, '8. Hasil Pembulatan', shows the rounded classification results for each data point, all of which are labeled as 'NEGATIF'.

Gambar 6 Hasil klasifikasi satu data

Hasil diagnosa dari semua K disajikan pada **Gambar 6** keluaran klasifikasi semua K hasil perhitungan yang sudah di tentukan dari data uji pasien tersebut termasuk dalam diagnosa penyakit demam berdarah dengue . Hasil klasifikasi ini tergantung dari data rekam medis pasien yang dimasukkan.

4.2 Keberhasilan Sistem Klasifikasi

Terdapat 1 kali skenario percobaan dimana skenario percobaan memiliki jumlah proporsi data latih **Tabel 2** menunjukkan hasil perbandingan pengujian dari skenario pertama dan skenario kedua menggunakan proses normalisasi data.

Tabel 1 Hasil skenario pengujian dengan proses normalisasi data

TOTAL DATA UJI = 75	K3	K5	K7	K4	K6	K8
SAMA	72	72	68	72	72	68
TIDAK SAMA	3	3	7	3	3	7
AKURASI RATA-RATA PER K	96.00%	96.00%	90.67%	0.96	0.96	0.9067
RATA-RATA NILAI K	94.22%					
LAJU EROR RATA-RATA PER K	4.00%	4.00%	9.33%	4.00%	4.00%	9.33%
LAJU EROR RATA-RATA NILAI K	5.78%					

5. KESIMPULAN dan SARAN

5.1 Kesimpulan

Hasil yang didapatkan dari penelitian ini dapat disimpulkan sebagai berikut:

1. Sistem ini menggunakan metode *K-Nearest Neighbor* (KNN) dengan parameter umur, jenis kelamin, hemoglobin, trombosit, hemotrokit yang akan digunakan untuk mendeteksi dini diagnosa penyakit dbd.
2. Berdasarkan hasil perhitungan *confusion matrix* untuk pengujian 150 data latih dan 75 data uji sistem ini memiliki nilai K terbaik yaitu 94,22 % pada K=3 K=5 K=7 K=4, K=6, dan K=8, nilai K terbaik diatas dapat digunakan untuk acuan klasifikasi selanjutnya.

5.2 Saran

Saran yang akan diberikan untuk mengembangkan penelitian ini adalah sebagai berikut:

1. Perlu adanya pengembangan pada desain tampilan antarmuka agar lebih menarik bagi para pengguna.
2. Perlu ditambahkan fasilitas *help* pada sistem dengan tujuan agar dapat membantu user dalam menggunakan sistem tersebut.
3. Perlu dibuatkan sistem *backup* agar data-data yang telah ada akan tersimpan dengan baik.

DAFTAR PUSTAKA

Kadir A, Triwahyuni TCH. 2003. *Pengenalan Teknologi Informasi*. Yogyakarta: Andi

Kristina, Isminah, Wulandari L. 2004. *Kajian Masalah Kesehatan Demam Berdarah Dengue*.

<http://www.litbang.depkes.go.id>

Lestari, M. 2014. *Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) Untuk Mendeteksi Penyakit Jantung*. Universitas Indraprasta, Jakarta Selatan.

Prasetyo, E. 2012. *Data Mining - Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

Prasetyo, E. 2014. *Data Mining – Mengelola Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

Suhendro *et al*, 2006. *Demam Berdarah Dengue*. Jakarta: Pusat Penerbitan Departemen Ilmu Penyakit Dalam. Fakultas Kedokteran Universitas Indonesia.

Tan, P. *et al*. 2006. *Introduction to Data Mining*. Boston: Pearson Education.

Tiaratuni dan Sudaryanto. 2014. *Pengertian Data Mining dan Sifat Data Mining*. Diakses 15 Februari 2017