

KLASIFIKASI PENYAKIT GINJAL KRONIS BERDASARKAN *CHRONIC KIDNEY DISEASE DATASET* MENGGUNAKAN METODE *FUZZY K-NEAREST NEIGHBOR (FK-NN)*

Septyan Teguh Mahendra, M. Tanzil Furqon, S.Kom, M.CompSc, Randy Cahya W, S.ST., M.kom

Program Studi Teknik Informatika
Jurusan Teknik Informatika
Fakultas Ilmu Komputer
Universitas Brawijaya, Malang
Jl. Veteran No.8 Malang, Gedung A FILKOM - UB
e-mail: hendrahuget@gmail.com

ABSTRAK

Penelitian ini membahas tentang pengimplementasian algoritma *Fuzzy k-Nearest Neighbor* untuk melakukan pengklasifikasian penyakit ginjal kronis. Data yang digunakan dalam pembuatan sistem ini adalah dataset *chronic kidney disease* yang didapatkan dari *UCI Machine Learning Repository*. *Fuzzy k-Nearest Neighbor (Fk-NN)* merupakan metode klasifikasi yang menggabungkan teknik *Fuzzy* dengan *k-Nearest Neighbor classifier*. Dalam penelitian ini dilakukan beberapa pengujian untuk mengetahui tingkat akurasi sistem, pengujian yang pertama adalah pengaruh jumlah data latih terhadap hasil akurasi sistem, dan pengujian yang kedua adalah pengaruh nilai *k* terhadap hasil akurasi sistem. Pada pengujian pertama hasil akurasi tertinggi yang didapatkan sebesar 96.67%, dan pada pengujian kedua didapatkan hasil akurasi tertinggi sebesar 96.67%.

Kata kunci: klasifikasi, penyakit ginjal kronis, *Fk-NN*

ABSTRACT

This study discusses about implementation of Fuzzy k-Nearest Neighbor algorithm to perform chronic kidney disease classification. The data used in this system development is the chronic kidney disease dataset, which is obtained from the UCI Machine Learning Repository. Fuzzy k-Nearest Neighbor (Fk-NN) is a classification method that combines fuzzy techniques with k-Nearest Neighbor classifier. In this study conducts several tests to determine the system accuracy level, the first test is the effect of training data amount to the system accuracy results, and the second testing is the effect of the k value to system accuracy results. In the first test, the highest accuracy results obtained is in the amount of 96.67%, and in the second test obtained highest accuracy in the amount of 96.67%.

Key word: classification, chronic kidney disease, *Fk-NN*

1. Pendahuluan

1.1 Latar belakang

Satu dari sembilan orang atau setara 20 juta lebih penduduk Amerika adalah penderita *chronic kidney disease (CKD)* atau penyakit ginjal kronis (Lange, 2008). Penyakit ginjal kronis adalah suatu keadaan terjadinya kerusakan ginjal atau laju filtrasi glomerulus (LFG) < 60 mL/menit dalam waktu 3 bulan atau lebih (Nahas, 2003; Vijayakumar, et al, 2007; Parmar, 2002). Dalam kondisi tersebut, secara berangsur dan *irreversible*, ginjal mengalami penurunan fungsi dan akan terus berkembang hingga terjadi gagal ginjal terminal. Terjadinya kerusakan ginjal tersebut bisa dilihat dari kelainan yang terjadi pada darah atau urin, pencitraan, dan biopsi ginjal. Tidak hanya gagal ginjal, dalam perkembangannya, penyakit ginjal kronis juga akan mengakibatkan

terjadinya komplikasi lain karena fungsi ginjal yang mengalami penurunan dan juga penyakit kardiovaskular (Sharon, 2006; Levey, et al, 2003).

Penelitian terdahulu mengenai klasifikasi penyakit ginjal kronis pernah dilakukan oleh Abhinandan Dubey pada tahun 2015 dari Birla Institute of Applied Sciences dengan judul penelitian "*A Classification of CKD Cases Using MultiVariate K-Means Clustering*" (Dubey, 2015).

Untuk penelitian kali ini, metode yang digunakan adalah *Fuzzy* dengan gabungan metode untuk pengklasifikasian data pada *data mining*. Ada beberapa metode dalam *data mining* yang dapat digunakan dalam pengklasifikasian data, antara lain metode *k-Nearest Neighbor*, *Bayesian*, ID3 dan C45, serta masih ada pula beberapa metode lainnya, dimana

terdapat target variable kategori dalam klasifikasi (Kusrini, 2009).

Sesuai uraian diatas, salah satu metode dalam *data mining* yang dapat digunakan untuk mengklasifikasikan data adalah metode *k-Nearest Neighbor*. Metode ini dapat digabungkan dengan metode lain, salah satunya adalah digabungkan dengan metode *Fuzzy*, sehingga menjadi metode *Fuzzy k-Nearest Neighbor (Fk-NN)*. Dalam metode ini, teknik *Fuzzy* digabungkan dengan teknik *data mining* (Zhang et al., 2009). Pada dasarnya algoritma *Fk-NN* menetapkan nilai keanggotaan sebagai fungsi jarak vektor dari *k-NN* dan keanggotaan tetangganya pada kelas-kelas yang memungkinkan. Algoritma *Fk-NN* tidak menempatkan vektor pada kelas tertentu dalam implementasinya, algoritma ini memberikan nilai keanggotaan kelas pada vektor sampel. Sehingga, nilai-nilai keanggotaan vektor akan memberikan tingkat jaminan terhadap hasil pengklasifikasian.

Metode *Fk-NN* sudah banyak diterapkan pada berbagai penelitian antara lain pada tahun 2013 penelitian yang dilakukan Rahmi Amiratus untuk menentukan kualitas rendemen tanaman tebu, tingkat akurasi tertinggi yang diperoleh adalah 98%. Adi Yuistiyanto pada tahun 2016 menggunakan metode *Fk-NN* untuk pengklasifikasian penyakit kelainan tulang, akurasi yang dihasilkan sebesar 83,30%. Dan masih banyak lagi penelitian lain yang menerapkan metode *Fk-NN*, namun dari penelitian yang dilakukan Rahmi Amiratus dan Adi Yuistiyanto tersebut, dapat dilihat bahwa hasil akurasi yang didapatkan cukup tinggi. Selain karena tingkat akurasi yang cukup tinggi, metode ini juga sederhana dan mudah digunakan.

Berdasarkan uraian di atas, maka penelitian ini diberi judul **“KLASIFIKASI PENYAKIT GINJAL KRONIS BERDASARKAN CHRONIC KIDNEY DISEASE DATASET MENGGUNAKAN METODE FUZZY K-NEAREST NEIGHBOR (FK-NN)”**. Data yang digunakan untuk klasifikasi pada *Chronic Kidney Disease Dataset* yaitu *age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema*, dan *anemia*.

1.2 Rumusan masalah

Berdasarkan apa yang telah disampaikan dalam latar belakang penelitian tentang klasifikasi penyakit ginjal kronis berdasarkan *chronic kidney disease dataset* menggunakan metode *Fuzzy k-Nearest*

Neighbor (Fk-NN), maka berikut ini adalah rumusan masalah dari penelitian yang akan dilakukan.

1. Bagaimana melakukan klasifikasi penyakit ginjal kronis berdasarkan *chronic kidney disease dataset* dengan menerapkan metode *Fuzzy k-Nearest Neighbor (Fk-NN)*.
2. Bagaimana tingkat akurasi hasil klasifikasi penyakit ginjal kronis berdasarkan pada *chronic kidney disease dataset* menggunakan metode *Fuzzy k-Nearest Neighbor (Fk-NN)*.

1.3 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini ialah sebagai berikut:

1. Menerapkan metode *Fuzzy k-Nearest Neighbor (Fk-NN)* untuk klasifikasi penyakit ginjal kronis berdasarkan pada *chronic kidney disease dataset*.
2. Menguji tingkat akurasi hasil klasifikasi penyakit ginjal kronis berdasarkan pada *chronic kidney disease dataset* menggunakan metode *Fuzzy k-Nearest Neighbor (Fk-NN)*.

1.4 Manfaat

Setelah penelitian ini diselesaikan diharapkan dapat membantu masyarakat umum atau bahkan para tenaga medis dalam mengklasifikasikan penyakit ginjal kronis, sehingga untuk kedepannya bisa segera dilakukan tindakan medis untuk pasien yang membutuhkan.

1.5 Batasan masalah

Berikut adalah batasan masalah yang telah ditentukan dalam penelitian ini.

1. Batasan subjek penelitian berupa *dataset* dari *UCI Machine Learning Repository* yang dapat diakses pada alamat website berikut ini [https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease], *dataset* yang digunakan adalah *Chronic Kidney Disease Dataset* yang memiliki dua kelas yaitu *ckd* dan *notckd*.
2. Parameter yang digunakan untuk menentukan kelas yaitu *age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema*, dan *anemia*. Dengan jumlah total data yang digunakan sebanyak 158 data.

2. Metodologi

2.1 Langkah penelitian

Langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut.

1. Mempelajari sumber-sumber yang mendukung untuk mendalami materi mengenai penyakit ginjal kronis dan metode *Fuzzy k-Nearest Neighbor*.
2. Mengumpulkan data-data dari *Chronic Kidney Disease Dataset*.
3. Menganalisis data-data yang terdapat pada *Chronic Kidney Disease Dataset*.
4. Mengimplementasikan sistem untuk membuat perangkat lunak berdasarkan analisis dan perancangan yang telah dilakukan dengan bahasa pemrograman *Java*.
5. Melakukan uji coba terhadap perangkat lunak.
6. Mengevaluasi uji coba terhadap perangkat lunak.

2.2 Data penelitian

Dalam penelitian ini penulis menggunakan *Chronic Kidney Disease Dataset* tahun 2015 yang diperoleh dari website *UCI Machine Learning Repository*. Pada dataset tersebut, terdapat 24 atribut yang digunakan, yaitu *age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema*, dan *anemia*. Dengan 2 kelas output yaitu 1 (*ckd*) dan 2 (*notckd*).

2.3 Analisa dan perancangan sistem

Pada dasarnya sistem yang dibangun adalah perangkat lunak yang menerapkan metode *Fuzzy k-Nearest Neighbor (Fk-NN)* untuk pengklasifikasian penyakit ginjal kronis atau *chronic kidney disease*. Selain itu perangkat lunak yang dibangun juga akan melakukan pengujian terhadap tingkat akurasi terhadap hasil pengklasifikasian penyakit ginjal kronis yang telah dilakukan oleh sistem.

Yang menjadi batasan sistem dalam penelitian mengenai klasifikasi penyakit ginjal kronis ini adalah input sistem berupa *dataset* pasien penderita penyakit ginjal kronis atau *chronic kidney disease*.

Berikut adalah langkah-langkah pada tahap klasifikasi yang diterapkan dalam sistem pengklasifikasian penyakit ginjal kronis dengan metode *Fk-NN*.

1. Menginputkan data latih dan data uji penderita penyakit ginjal kronis.
2. Menormalisasi setiap atribut pada data yang diinputkan. Persamaan yang digunakan adalah persamaan (2.1) sebagai berikut.

$$v' = \frac{v - \min_a}{\max_a - \min_a} \quad (2.1)$$

3. Memproses data yang telah dinormalisasi dengan perhitungan *k-NN*. Pada tahap ini dilakukan proses perhitungan untuk mencari jarak terdekat antara data uji dengan data latih dengan menggunakan persamaan (2.2) sebagai berikut.

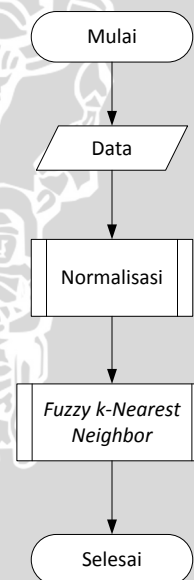
$$d_i = \sqrt{\sum_{j=1}^p (x_{2i} - x_{1i})^2} \quad (2.2)$$

4. Mentransformasikan hasil perhitungan *k-NN* kedalam data *Fuzzy* dengan menggunakan persamaan (2.3) sebagai berikut.

$$u_i \times = \frac{\sum_{j=1}^k u_{ij} \frac{1}{1 + (x - x_j)^2} \frac{1}{m-1}}{\sum_{j=1}^k \frac{1}{1 + (x - x_j)^2} \frac{1}{m-1}} \quad (2.3)$$

5. Menampilkan hasil klasifikasi kelas penyakit ginjal kronis.

Gambar 2.1 dibawah ini menunjukkan alur dari tahap klasifikasi.



Gambar 2.1 Alur proses klasifikasi

3. Hasil

3.1 Lingkungan implementasi

Dalam pembuatan sistem yang dibangun dalam penelitian ini digunakan beberapa perangkat keras seperti yang disebutkan dibawah ini.

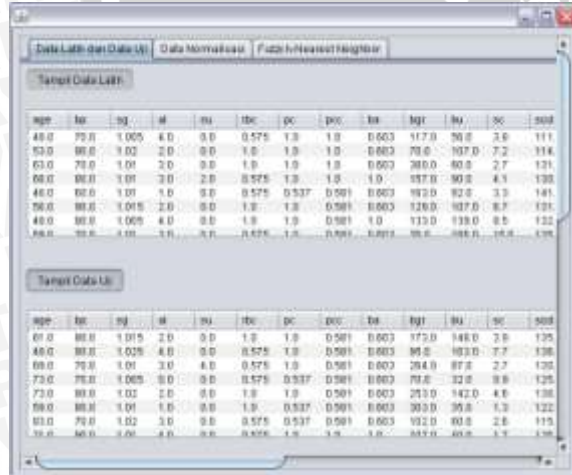
1. Processor : Intel(R) Atom(TM) CPU N475 @1.83GHz (2 CPUs)
2. Memory : 1 GB
3. Harddisk : 150 GB

Kebutuhan perangkat lunak yang digunakan untuk membangun sistem dalam penelitian ini disebutkan pada rincian di bawah ini.

1. Sistem Operasi yang digunakan Windows XP Professional (5.1, Build 2600) Service Pack 3
2. Aplikasi pembangunan GUI dan code menggunakan NetBeans IDE 6.8
3. Bahasa pemrograman yang dipakai yaitu bahasa pemrograman java
4. Komponen java yang digunakan yaitu Java SE Development Kit 7 update 17

3.2 Implementasi program

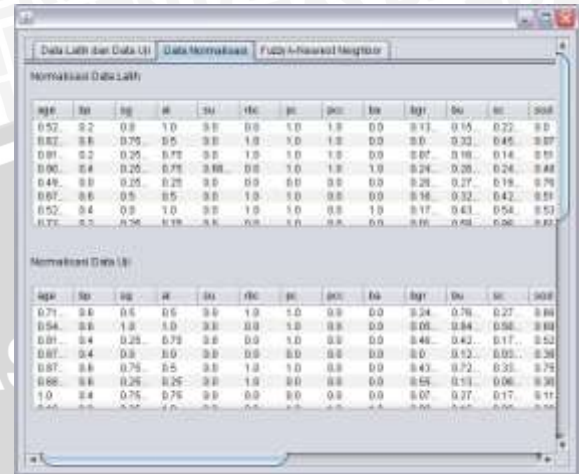
Screenshot program saat menampilkan data latih dan data uji akan ditunjukkan pada gambar 3.1 di bawah ini.



The screenshot shows two tables: 'Tampilkan Data Latih' and 'Tampilkan Data Uji'. Both tables have 14 columns: age, sex, bmi, glc, al, bu, rbc, pc, hct, ha, hgb, hu, sc, and sglb. The 'Tampilkan Data Latih' table contains 10 rows of data, and the 'Tampilkan Data Uji' table contains 10 rows of data.

Gambar 3.1 Screenshot program saat menampilkan data latih dan data uji

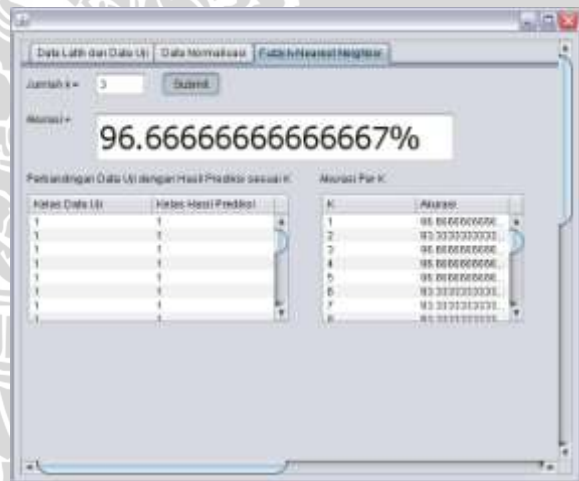
Screenshot program saat menampilkan hasil dari normalisasi data latih dan data uji ditunjukkan pada gambar 3.2 berikut ini.



The screenshot shows two tables: 'Normalisasi Data Latih' and 'Normalisasi Data Uji'. Both tables have 14 columns: age, sex, bmi, glc, al, bu, rbc, pc, hct, ha, hgb, hu, sc, and sglb. The 'Normalisasi Data Latih' table contains 10 rows of normalized data, and the 'Normalisasi Data Uji' table contains 10 rows of normalized data.

Gambar 3.2 Screenshot program saat menampilkan hasil normalisasi

Screenshot program saat menampilkan hasil perhitungan Fk-NN ditunjukkan pada gambar 3.3 di bawah ini.



The screenshot shows the result of the Fk-NN calculation. It displays a large percentage value: 96.66666666666667%. Below this, there are two tables: 'Perbandingan Data Uji dengan Hasil Prediksi sesuai K' and 'Matriks Per K'. The 'Perbandingan Data Uji dengan Hasil Prediksi sesuai K' table has 2 columns: 'Hasil Data Uji' and 'Hasil Hasil Prediksi'. The 'Matriks Per K' table has 2 columns: 'K' and 'Akurasi'.

Gambar 3.3 Screenshot program saat menampilkan hasil perhitungan Fk-NN

4. Pembahasan

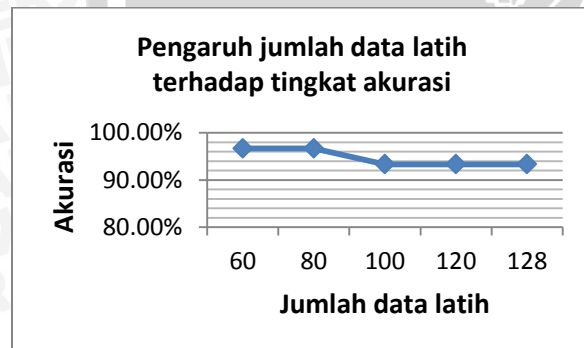
4.1 Pengujian

Dalam pengujian pengujian tingkat akurasi terhadap data latih, variasi jumlah data yang digunakan meliputi 60 data latih, 80 data latih, 100 data latih, 120 data latih, dan 128 data latih. Untuk data uji yang digunakan pada pengujian-pengujian tersebut adalah sebanyak 30 data uji dan merupakan data yang sama untuk setiap pengujian. Hasil dari pengujian ini ditunjukkan dalam tabel 4.1 di bawah ini, dengan menggunakan nilai $k = 3$.

Tabel 4.1 Hasil pengujian tingkat akurasi terhadap data latih

Jumlah data latih	Akurasi
60	96.67%
80	96.67%
100	93.33%
120	93.33%
128	93.33%

Grafik dari hasil pengujian tingkat akurasi terhadap data latih di atas dapat dilihat pada gambar 4.1 berikut.



Gambar 4.1 Grafik hasil pengujian tingkat akurasi terhadap data latih

Dalam pengujian tingkat akurasi terhadap nilai k , data latih yang digunakan sebanyak 60 data. Dan yang digunakan sebagai data uji sebanyak 30 data. Hasil dari pengujian tingkat akurasi terhadap nilai k ini dapat dilihat pada tabel 4.2 di bawah ini.

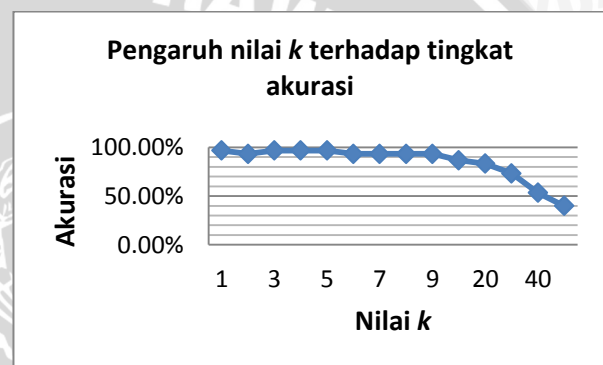
Tabel 4.2 Hasil pengujian tingkat akurasi terhadap nilai k

Nilai k	Akurasi
1	96.67%
2	93.33%
3	96.67%
4	96.67%
5	96.67%

Tabel 4.3 Hasil pengujian tingkat akurasi terhadap nilai k (lanjutan)

6	93.33%
7	93.33%
8	93.33%
9	93.33%
10	86.67%
20	83.33%
30	73.33%
40	53.33%
50	40.00%

Grafik dari hasil pengujian tingkat akurasi terhadap nilai k di atas dapat dilihat pada gambar 4.2 berikut.



Gambar 4.2 Grafik hasil pengujian tingkat akurasi terhadap nilai k

4.2 Analisis

Dari hasil pengujian pengaruh jumlah data latih terhadap tingkat akurasi yang telah dilakukan dapat dilihat bahwa hasil akurasi tertinggi ada pada pengujian dengan 60 dan 80 data latih yaitu dengan hasil akurasi 96.67%, tingkat akurasi mengalami sedikit penurunan ketika dilakukan pengujian dengan data latih yang lebih banyak. Padahal seharusnya semakin banyak data latih yang digunakan akan menghasilkan akurasi yang semakin tinggi persentasenya, karena dengan semakin banyaknya data latih yang digunakan, maka sistem dapat mengenali pola yang ada pada dataset yang digunakan dalam pengujian. Namun dalam penelitian ini dataset yang digunakan memiliki sebaran kelas yang tidak seimbang atau *imbalance class*, tingkat *imbalance* data mempengaruhi hasil klasifikasi, semakin tinggi tingkat *imbalance* sebuah data, maka nilai *precision*, *recall*, dan *f-measure* pada kelas mayor semakin meningkat sedangkan untuk nilai *precision*, *recall*, dan *f-measure* pada kelas minor cenderung mengalami penurunan.

Pada pengujian pengaruh nilai k terhadap tingkat akurasi dapat dilihat hasil tingkat akurasi akan semakin menurun ketika pengujian dilakukan dengan nilai k yang semakin besar. Dalam hal ini kemungkinan penurunan tingkat akurasi terjadi juga karena tidak seimbangannya sebaran kelas pada dataset yang digunakan, 30 data yang digunakan sebagai data uji memiliki 20 data yang termasuk dalam kelas 1 dan 10 data yang termasuk dalam kelas 2, maka semakin besar nilai k , akan semakin banyak data termasuk dalam kelas mayor yang sebenarnya bukan merupakan tetangga terdekat yang mempengaruhi data yang termasuk dalam kelas minor, sehingga kelas minor akan memiliki peluang yang besar untuk salah diklasifikasikan.

5. Penutup

5.1 Kesimpulan

Berikut ini adalah kesimpulan yang dapat ditarik dari hasil penerapan metode *Fuzzy k-Nearest Neighbor* (*Fk-NN*) untuk pengklasifikasian penyakit ginjal kronis, dan juga hasil pengujian dan analisa hasil yang dilakukan dalam penelitian ini

1. Metode *Fuzzy k-Nearest Neighbor* (*Fk-NN*) bisa diterapkan dalam pengklasifikasian penyakit ginjal kronis. Langkah awal yang dilakukan adalah menormalisasi data latih dan data uji, lalu mencari tetangga terdekat dengan persamaan *Euclidean Distance*, hingga pada akhirnya hasil dari data yang telah dilakukan perhitungan ditransformasikan kedalam bentuk *fuzzy*, maka akan diperoleh hasil klasifikasi penyakit ginjal kronis dengan menggunakan metode *Fuzzy k-Nearest Neighbor*
2. Berdasarkan pengujian tingkat akurasi yang dipengaruhi oleh banyaknya data latih yang digunakan, didapatkan tingkat akurasi sebesar 96.67% dengan 60 dan 80 data latih dan $k = 3$. Untuk pengujian tingkat akurasi terhadap nilai k dihasilkan akurasi sebesar 96.67% dengan $k = 1$ dan menggunakan 60 data latih.

5.2 Saran

Berikut ini adalah saran-saran yang diharapkan bisa diperhatikan sebagai pertimbangan guna pengembangan penelitian-penelitian selanjutnya.

1. Diharapkan pada penelitian-penelitian selanjutnya digunakan metode lain atau modifikasi dari *Fk-NN* untuk mendapatkan hasil yang lebih maksimal.
2. Diharapkan pada penelitian selanjutnya untuk bisa menggunakan data yang lebih baik lagi, misalnya yang lebih minim *missing value* dan sebagainya,

agar pengujian bisa dilakukan dengan data yang lebih besar jumlahnya.

DAFTAR PUSTAKA

- Dubey, Abhinandan. 2015. A Classification of CKD Cases Using MultiVariate K-Means Clustering. Dalam: International Journal of Scientific and Research Publications, Volume 5, Issue 8. Tersedia di: < <http://www.ijsrp.org/research-paper-0815/ijsrp-p4437.pdf>> [Diakses 26 Desember 2015]
- Kusrini dan Luthfi, Emha Tufiq. 2009. *Algoritma Data Mining*. Yogyakarta: Penerbit ANDI.
- Levey AS, Coresh J, Balk E, Kautz T, Levin A, Steves M et al. 2003. National Kidney Foundation Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification. Ann Intern Med. USA : Annals of Internal Medicine.
- Nahas, ME. 2003. The Patient with Failing Renal Failure. Dalam: Cameron JS, Davison AM. Oxford Textbook of Clinical Nephrology. Edisi ke-3. Oxford University Press.
- Parmar, MS. 2002. Chronic Renal Disease. United Kingdom: BMJ.
- Sharon K. 2006. Chronic Kidney Disease. Critical Care Nurse.
- Vijayakumar M, Namalwar R, Prahlad N. 2007. Prevention of Chronic Kidney Disease in Children. Ind J of Nephrol. New Delhi: Indian Journal of Nephrology.
- Zhang, Juan, Yi Niu, Huabei Nie. 2009. "Web Document Classification Based on Fuzzy KNN Algorithm". International Conference on Computational Intelligence and Security.