

## Klasifikasi Risiko Hipertensi Menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NWKNN)

Bayu Laksana Yudha<sup>1</sup>, Lailil Mufflikhah<sup>2</sup>, Randy Cahya Wihandika<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: <sup>1</sup>bayudhabay@gmail.com, <sup>2</sup>laililmf@gmail.com, <sup>3</sup>rendicahya@ub.ac.id

### Abstrak

Hipertensi atau disebut darah tinggi memiliki risiko kematian tinggi di Indonesia. Penyakit ini dapat memicu timbulnya penyakit lain seperti serangan jantung dan gagal ginjal. Menurut WHO, sebanyak 30% dari masyarakat Indonesia adalah penderita hipertensi, Perhimpunan Dokter Hipertensi Indonesia juga menyatakan bahwa 76% kasus dari hipertensi ini tidak dapat didiagnosis secara dini dan maka dari itu hipertensi disebut sebagai *silent killer*. Cara untuk meanggulangi hipertensi secara dini yaitu dengan melakukan deteksi dini hipertensi dalam bentuk Sistem Kewaspadaan Dini (SKD). Pada penelitian ini akan dilakukan klasifikasi risiko hipertensi berdasarkan rekam medis dengan menggunakan metode klasifikasi *Neighbor Weighted K-Nearest Neighbor* (NWKNN). Metode ini merupakan perkembangan dari metode KNN. Pada NWKNN terdapat proses pembobotan di setiap kelas risiko hipertensi. Di dalam penelitian ini dilakukan klasifikasi hipertensi menjadi 4 risiko yaitu Normal, Pre Hipertensi, Stadium 1 dan Stadium 2. Hasil dari penelitian ini menunjukkan bahwa metode NWKNN dapat melakukan klasifikasi risiko hipertensi dengan baik saat data latih yang digunakan sebanyak 100 data dan data uji sebanyak 25 data, nilai  $K=10$ , dan nilai  $E=4$  dengan hasil akurasi mencapai 88%.

**Kata kunci:** klasifikasi, hipertensi, NWKNN

### Abstract

*Hypertension or called high blood pressure caused high risk of death in Indonesia. This could trigger sustainable effect to other diseases such as heart attack and kidney failure. According to WHO, as many as 30% of Indonesians are sufferers hypertension, Indonesian Hypertension Doctor Association also said that 76% of cases of hypertension can not diagnosed earlier and therefore hypertension is called a silent killer. The way to handling hypertension earlier is by early detection of hypertension in form of Early Alertness System (SKD). In this research will classified risk of hypertension based on medical record using Neighbor Weighted K-Nearest Neighbor (NWKNN) classification method. This method is the development of the KNN method. In NWKNN there is a weighting process in each class of hypertension risk. In this study, the classification of hypertension into 4 risks that is Normal, Pre Hypertension, Stage 1 and Stage 2. The results of this research shows that the NWKNN method is able to classify the hypertension risk well when tested on 100 training data, 25 testing data, K score=10, and E score=4 with accuracy result that reached 88%.*

**Keywords:** classification, hypertension, NWKNN

### 1. PENDAHULUAN

Hipertensi merupakan penyakit darah tinggi yang memiliki salah satu risiko kematian tertinggi di Indonesia. Penyakit ini dapat memicu timbulnya penyakit lain yang lebih berbahaya daripada hipertensi itu. Beberapa penyakit tersebut antara lain gagal ginjal, stroke dan jantung (Yogiantoro, 2009). Hipertensi dapat dideteksi oleh petugas medis menggunakan alat bernama tensimeter dengan cara mengukur tekanan darah manusia.

Hipertensi timbul akibat tekanan darah yang meningkat secara abnormal, baik tekanan sistolik maupun diastolik. Seseorang dapat dikatakan menderita hipertensi jika tekanan darah sistolik/diastolik lebih dari 140/90 mmHg (tekanan darah normal 120/80 mmHg). Hipertensi sangat terkait dengan perubahan gaya hidup, konsumsi makanan yang berlemak tinggi, kolesterol, kurangnya aktifitas olahraga dan stress (Herwati & Wiwi, 2011).

Penderita hipertensi di Indonesia mencapai lebih dari 30 persen menurut WHO. Sedangkan

pada tahun 2014 menurut Perhimpunan Dokter Hipertensi Indonesia, sebanyak 76 persen kasus penyakit ini tidak dapat didiagnosis sejak dini. Jumlah kasus penyakit hipertensi di masyarakat meningkat karena banyak penderita yang belum mendapatkan pengobatan yang layak ataupun pengobatan yang belum tuntas sehingga tekanan darah belum mencapai batas normal (Yogiantoro, 2009). Hal seperti ini terjadi karena akses untuk mendapatkan pengobatan yang masih terbatas dan pengetahuan di dalam masyarakat tentang penyakit hipertensi yang masih rendah.

Tahun 2007 diadakan sebuah riset kesehatan oleh Riset Kesehatan Dasar Indonesia, hasilnya penyebab kasus kematian hipertensi menempati urutan ke 3. Riset ini dilakukan setiap 6 tahun sekali, dan pada tahun 2013 kasus kematian dengan penyakit yang sama meningkat sebanyak 1,9 persen dibanding tahun 2007. Hal ini menyebabkan penanganan hipertensi menjadi prioritas nasional di bagian penyakit tidak menular. Banyak lembaga kesehatan yang sudah melakukan penelitian di dalam masyarakat untuk mengukur faktor risiko penyakit yang tidak menular, salah satunya yaitu hipertensi. Seperti hasil riset di atas, hipertensi perlu mendapat perhatian dan penanganan yang serius dikarenakan semakin meningkatnya angka kematian terkait penyakit ini. Salah satu upaya penanggulangannya yaitu dengan melakukan deteksi dini hipertensi dalam bentuk Sistem Kewaspadaan Dini (SKD). Hasil deteksi ini selanjutnya dapat digunakan untuk instansi kesehatan dalam penanggulangan hipertensi di Indonesia (Nindy, 2015).

Berdasar penjelasan di atas maka dibutuhkan suatu sistem cerdas yang dapat melakukan klasifikasi dini hipertensi berdasarkan kondisi pasien. Sistem cerdas dapat dengan cepat dan akurat dalam mendeteksi suatu kesalahan atau penyakit didalam sistem maupun tubuh manusia, hal ini dikarenakan sistem cerdas memakai suatu algoritma perhitungan untuk menyelesaikan sebuah masalah. Hal ini dapat membantu para petugas medis dalam pencegahan secara dini terhadap kasus hipertensi. Klasifikasi merupakan cara untuk mengelompokan suatu data atau benda menurut standar yang telah ditetapkan. Klasifikasi erat hubungannya dengan Data Mining karena banyak metode klasifikasi di dalamnya. Di dalam pengklasifikasian data, terkadang ada data yang tidak seimbang, berarti data yang distribusinya tidak merata. Pemecahan terkait dengan data

yang tidak seimbang dapat menggunakan metode Neighbor Weighted K-Nearest Neighbor (NWKNN) yang dikenalkan oleh Sangbo Tan pada tahun 2005 (Valerian, 2015).

Penelitian yang memanfaatkan sistem cerdas sudah cukup banyak dewasa ini, baik dalam bidang kesehatan maupun pada bidang lain. Pertama yaitu penelitian tentang klasifikasi penyakit ADHD. Tujuan penelitian ini yaitu untuk melakukan klasifikasi terhadap anak usia dini yang terserang penyakit Attention Deficit Hyperactivity Disorder (ADHD). Metode yang digunakan yaitu Neighbor Weighted K-Nearest Neighbor (NWKNN). Penulis menggunakan 80 data latih, 20 data uji, nilai  $K=10$  dan nilai  $E=4$  agar mencapai hasil yang optimal. Penyakit ADHD sendiri diklasifikasikan kedalam 4 kelas yaitu Hyperactivity, Impulsif, Inattention dan Tidak ADHD. Hasil akurasi yang didapatkan mencapai 95% dan memiliki peningkatan sebanyak 2% daripada klasifikasi menggunakan metode KNN (Putri, et al., 2016).

Kedua yaitu penelitian tentang klasifikasi risiko hipertensi. Penelitian ini mempunyai tujuan untuk mengklasifikasi risiko hipertensi menjadi 3 kelas yaitu Tinggi, Rendah dan Sedang. Metode yang digunakan adalah Fuzzy Tsukamoto dan Algoritma Genetika dengan 75 data. Untuk mencapai hasil yang maksimal, peneliti menggunakan kombinasi nilai  $cr=0,2$  dan nilai  $mr=0,8$  untuk data laki-laki dengan akurasi sebesar 87%. Untuk kombinasi nilai  $cr$  dan  $mr$  untuk data perempuan adalah  $cr=0,4$  dan  $mr 0,4$  dengan hasil akurasi tertinggi 93% (Nindy, 2015). Namun di dalam penelitian ini data yang digunakan yaitu data tentang kebiasaan hidup sehari-hari dari pasien, bukan data dari laboratorium rumah sakit (Nindy, 2015).

Ketiga yaitu penelitian terkait klasifikasi hipertensi dengan menggunakan algoritma Naïve Bayes. Di dalam penelitian ini terdapat empat kelas risiko hipertensi yaitu normal, pre hipertensi, stadium 1 dan stadium 2. Metode Naïve Bayes mampu dengan cukup baik mengklasifikasi risiko hipertensi dengan data masukan sebanyak 15 dan mendapatkan akurasi sebesar 90% (Nyoman, 2016).

Pada penelitian ini penulis menggunakan data terkait risiko hipertensi yang di beberapa kelas ada data yang tidak seimbang. Metode yang dapat digunakan dalam penyelesaian data yang tidak seimbang yaitu Neighbor Weighted K-Nearest Neighbor (NWKNN), maka berdasar uraian di atas penulis mengusulkan penelitian dengan judul "Klasifikasi Risiko Hipertensi

Menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NWKNN)". Penelitian ini diharapkan dapat memberi informasi yang akurat terkait dengan risiko hipertensi.

**2. TINJAUAN PUSTAKA**

**2.1 Hipertensi**

Penyakit Hipertensi merupakan penyakit kardiovaskular yang berarti peningkatan abnormal pada tekanan darah baik sistolik maupun diastolik. Seseorang dapat dikatakan menderita hipertensi jika tekanan darah sistolik/diastolik lebih dari 140/90 mmHg (tekanan darah normal 120/80 mmHg). Hipertensi sangat terkait dengan perubahan gaya hidup, konsumsi makanan yang berlemak tinggi, kolesterol, kurangnya aktifitas olahraga dan stres (Herwati & Wiwi, 2011). Penderita penyakit ini di tidak akan merasakan tanda-tanda seperti penyakit pada umumnya dikarenakan sulit untuk dideteksi. Namun ketika dilakukan pemeriksaan terkait dengan penyakit yang berhubungan dengan hipertensi seperti stroke dan diabetes baru akan terdeteksi. Penyakit ini dikenal dengan nama silent killer dikarenakan penjelasan di atas (Ari & Rika, 2013). Hipertensi menjadi perhatian utama pemerintah diakrenakan masih banyak penderita yang belum mendapatkan pertolongan medis sehingga risiko kematian meningkat. Hipertensi yang tidak diketahui penyebab penyakitnya didefinisikan sebagai hipertensi esensial atau hipertensi primer. Sedangkan hipertensi sekunder penyebab kemunculannya jelas diketahui (Nindy, 2015). Hipertensi dapat menimbulkan kecacatan yang permanen dan bahkan keamtian yang mendadak. Pencegahan dan penanggulangan hipertensi di dalam masyarakat dapat dilakukan dengan meningkatkan kesadaran dalam merubah pola hidup kearah yang lebih sehat (Herwati & Wiwi, 2011). Gejala pada hipertensi yaitu:

- Pusing
  - Sering gelisah
  - Tenguk pegal
  - Telinga sering berdengung
  - Mimisan
  - Sesak nafas
  - Mata berkunang-kunang
- Cara paling dasar dalam mengetahui hipertensi yaitu dengan mengukur tekanan darah, namun

kadang penyakit ini tidak bias disimpulkan dengan hanya mengukur tekanan darah (Wisnu, 2016).

**2.1.1 Klasifikasi Hipertensi**

Penelitian yang dilakukan The Seventh Report of The Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7) tekanan darah pada orang dewasa diklasifikasikan menjadi 4 kelompok seperti yang ditunjukkan pada Tabel 1

Tabel 1 Tabel Klasifikasi Tekanan Darah Menurut JNC 7

Klasifikasi Tekanan Darah	TDS (mmHg)		TDD (mmHg)
Normal	<120	dan	<80
Prahipertensi	120-139	atau	80-89
Hipertensi derajat 1	140-159	atau	90-99
Hipertensi derajat 2	≥160	atau	≥100

Keterangan:

TDS: Tekanan Darah Sistolik

TDD: Tekanan Darah Diastolik

Klasifikasi terkait tekanan darah juga dilakukan oleh World Health Organization (WHO), dan International Society of Hypertension (ISH). Namun klasifikasi JNC 7 merupakan klasifikasi yang paling umum digunakan (Nindy, 2015).

**2.1.2 Faktor Risiko Hipertensi**

Faktor risiko hipertensi merupakan kebiasaan individu yang lebih umum dialami oleh penderita daripada orang lain yang normal. Atribut individu tersebut dapat berupa umur, jenis kelamin, atau riwayat penyakit tertentu. Sedangkan kebiasaan yang dapat menjadi faktor risiko dapat berupa kebiasaan merokok, penyalahgunaan narkoba, asupan makanan, dan kebiasaan olahraga. Pada penyakit hipertensi terdapat beberapa faktor risiko, di antaranya (Yogiantoro, 2009):

- Kebiasaan merokok. Perokok aktif memiliki probabilitas yang lebih tinggi untuk terkena hipertensi daripada orang yang bukan perokok.
- Obesitas. Tingkat obesitas bisa ditentukan dari Body Mass Index (BMI).
- Kurangnya aktifitas fisik seperti olahraga setiap hari.
- Riwayat penyakit diabetes mellitus dan penyakit ginjal. Penyakit ini umumnya menimbulkan hipertensi esensial pada penderita.
- Umur dan jenis kelamin. Laki-laki dengan

umur > 55 tahun dan perempuan dengan umur > 65 tahun memiliki kemungkinan yang besar untuk terkena penyakit hipertensi.

- Riwayat keluarga dengan penyakit jantung kardiovaskular prematur pada laki-laki berumur < 55 tahun dan perempuan berumur < 65 tahun.
- Genetis. Sebagian besar penderita hipertensi menurunkan penyakit ini pada keturunannya.

Pemeriksaan lebih lanjut terhadap hipertensi bisa dilakukan dengan uji laboratorium jika pasien mengalami resistensi terhadap suatu obat dan menemukan bentuk risiko hipertensi sekunder. Pemeriksaan lain dapat menggunakan elektrodiagram (EKG) untuk melihat gejala dari risiko hipertensi sekunder. Pemeriksaan ini terdiri dari:

- Tes darah
- Tes glukosa
- Pemeriksaan kolesterol HDL dan LDL
- Pemeriksaan kadar Trigliserida
- Pemeriksaan kadar asam urat
- Pemeriksaan kadar kreatinin
- Pemeriksaan kadar hemoglobin dan hematocrit

Pada beberapa pasien hipertensi diperlukan pemeriksaan secara berkala jika ada kerusakan organ atau risiko hipertensi sudah mencapai stadium 2, namun jika hanya ada keluhan pasien pemeriksaan tidak akan dilakukan secara berkala (Nyoman, 2016).

## 2.2 Klasifikasi

Klasifikasi adalah salah satu metode yang ada di dalam data mining. Di dalam klasifikasi, label dari setiap kelas sudah ditentukan terlebih dahulu. Proses klasifikasi sendiri merupakan proses untuk menemukan model atau membedakan kelas atau data yang bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Klasifikasi merupakan bentuk analisis data yang dapat menggambarkan ekstrak model dari suatu data yang penting (Jiawei, et al., 2012). Proses dalam klasifikasi terdiri atas:

- Learning model: di dalam fase ini data training akan dianalisa dengan algoritma klasifikasi, pada bagian ini terdapat bentuk rule klasifikasi yang mempunyai 2 atribut kelas yaitu loan\_decision dan classifier
- Klasifikasi: pada fase ini data testing digunakan untuk memperkirakan hasil

akurasi berdasarkan rule klasifikasi.

## 2.3 Neighbor Weighted K Nearest Neighbor (NWKNN)

Metode NWKNN merupakan pengembangan dari metode KNN. Metode ini menggunakan prinsip pembobotan. Bobot akan diberi lebih sedikit ke jumlah k tetangga yang berasal dari kelas mayoritas, dan sebaliknya untuk kelas minoritas. Metode NWKNN mampu melakukan klasifikasi dengan baik, karena metode ini cocok untuk diimplementasikan ke data yang tidak terdistribusi secara rata (Indriati dan Ridok, 2016). Langkah algoritma pada metode NWKNN tidak jauh berbeda dengan langkah algoritma KNN, yang membedakan adalah adanya pembobotan untuk setiap jenis/kelas dan proses perhitungan skor untuk menentukan klasifikasi terhadap data uji (Faldy, 2014).

Langkah-langkah dalam algoritma NWKNN menurut D.A Adeniyi et al (2016) adalah sebagai berikut :

1. Menentukan nilai variabel K
2. Menghitung nilai kedekatan ketetanggan antara data uji terhadap data latih menggunakan Persamaan *Euclidean Distance* atau *Cosine Similarity* (*CosSim*)
- Menghitung ketetanggaaan terdekat yaitu dengan menghitung jarak antara data latih terhadap data uji menggunakan rumus *Euclidean Distance* dapat digunakan Persamaan 1 :

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_2 - x_1)^2} \quad (1)$$

Keterangan :

X<sub>1</sub> = nilai data latih

X<sub>2</sub> = nilai data uji

n = jumlah data

i = data ke-i

- Menghitung ketetanggaaan terdekat yaitu dengan menghitung kedekatan menggunakan rumus *Cosine Similarity* dapat digunakan Persamaan 2 :

$$(q, d_j) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \cdot |\vec{q}|} = \frac{\sum_{i=1}^m (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^m w_{ij}^2} \cdot \sqrt{\sum_{i=1}^m w_{iq}^2}} \quad (2)$$

Keterangan :

*Cosine Similarity* (*q, d<sub>j</sub>*) = nilai similaritas data uji q dengan data latih d<sub>j</sub>

q = data uji

d<sub>j</sub> = data latih

$\vec{d}_j \cdot \vec{q}$  = hasil total perkalian vektor antara data latih dengan data uji

$|\vec{d}_j| \bullet |\vec{q}|$  = hasil total perkalian vektor antara normalisasi data latih dengan data uji  
 $w_{ij}$  = bobot nilai i pada data latih j  
 $w_{iq}$  = bobot nilai i pada data uji q  
 $m$  = banyaknya jumlah nilai

3. Mengurutkan hasil perhitungan jarak atau kedekatan kedalam kelompok yang mempunyai kedekatan jarak atau *similarity*.
4. Mengumpulkan kategori klasifikasi *nearest neighbor*
5. Perhitungan bobot digunakan Persamaan 2.3

$$Weight_i = \frac{1}{\left(\frac{Num(c_i^d)}{\text{Min}\{Num(c_j^d) | j=1, \dots, K^*\}}\right)^{1/exp}} \quad (3)$$

Keterangan :

$Num(c_i^d)$  = banyaknya data latih d pada kelas i

$Num(c_j^d)$  = banyaknya data latih d pada kelas j, dimana j terdapat dalam himpunan k tetangga terdekat

Exp = eksponen (nilai exp lebih dari 1)

Setiap data yang telah dihitung nilai bobotnya akan digunakan untuk menghitung nilai skor. Dimana hasil nilai bobot akan dikalikan dengan Persamaan hasil skor. Rumus hasil skor dihitung dengan Persamaan 2.4 dan 2.5.

$$Skor(X, C_i) = Weight_i * \left( \sum_{d_{jKNN}(X)} \left( \sqrt{\sum_{i=1}^n (x_{2i} - x_{1i})^2} \right) \right) * \delta(d_j, C_i) \quad (4)$$

Atau

$$Skor(X, C_i) = Weight_i * \left( \sum_{d_{jKNN}(X)} (Sim(q, d_j)) \right) * \delta(d_j, C_i) \quad (5)$$

Keterangan :

$Weight_i$  = bobot jenis/kelas i  
 $d_{jKNN}(x)$  = data latih dj pada kumpulan tetangga terdekat dari data uji x

$\sqrt{\sum_{i=1}^n (x_{2i} - x_{1i})^2}$  = jarak antara data uji dan data latih

$\delta(d_j, C_i)$  = akan bernilai 1 jika nilai jarak  $\in C_i$  dan bernilai 0 jika nilai jarak  $\notin C_i$ .

$Sim(q, d_j)$  = nilai Cosine Similiraty antara data uji dan data latih

$C_i$  = jenis atau kelas i

### 3. METODOLOGI

#### 3.1 Data Penelitian

Data yang digunakan pada penelitian merupakan data sekunder yang didapat dari skripsi sebelumnya yang . Data terdiri dari:

1. 125 data sekunder dari skripsi Nyoman Wisnu W yang merupakan data rekam medis pasien.
2. 15 daftar fitur yang berasal dari hasil rekam medis pasien
3. 4 kelas risiko hipertensi yaitu Normal, Pre Hipertensi, Stadium 1 dan Stadium 2.

#### 3.2 Alur NWKNN



Gambar 1 Alur NWKNN

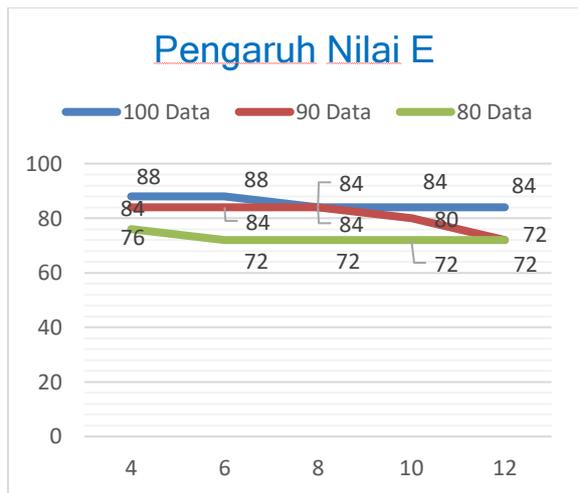
Ada empat proses utama dalam algoritma NWKNN, proses-proses tersebut yaitu proses menghitung kedekatan ketetanggaan menggunakan *Cosine Similarity*, mengurutkan nilai *Cosine Similarity*, pembobotan setiap jenis gejala dan menghitung skor di setiap jenis sesuai nilai K. Proses pertama yaitu dengan menghitung tetangga terdekat menggunakan *Cosine Similarity* untuk mengetahui data mana yang masuk klasifikasi kelas yang dicari. Kedua yaitu mengurutkan nilai hasil *Cosine Similarity* sesuai dengan terbesar ke terkecil. Ketiga yaitu proses pembobotan tiap kelas risiko hipertensi, proses ini berfungsi untuk memberikan bobot pada tiap kelas risiko dengan bobot terbesar

diberikan pada kelas dengan jumlah paling sedikit dan bobot kecil diberikan pada kelas paling banyak. Teraakhir yaitu proses perhitungan nilai skor hasil klasifikasi, proses ini berfungsi untuk mengetahui hasil klasifikasi dari risiko hipertensi itu sendiri.

**4. PENGUJIAN DAN ANALISIS**

Pada bab ini dilakukan proses pengujian dan analisis. Terdapat 4 proses pengujian yang dijelaskan didalam bab ini, yaitu pengujian pengaruh perubahan nilai K, pengaruh perubahan nilai E, pengaruh perubahan data latih terhadap data uji, perbandingan akurasi didalam NWKNN menggunakan *Cosine Similarity* dan *Euclidean Distance*. Fungsi dari pengujian berguna dalam mengetahui hasil akurasi dari penggunaan metode NWKNN yang diterapkan didalam sistem.

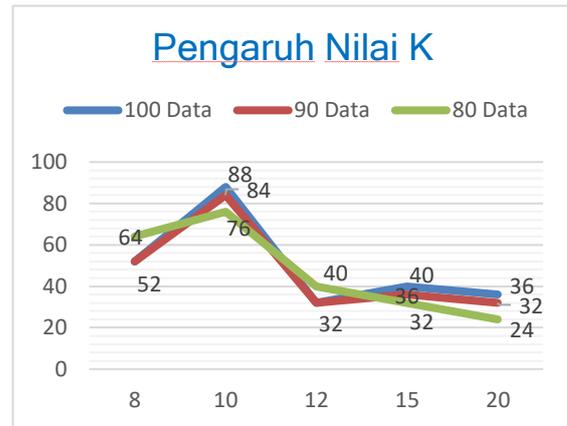
**4.1 Pengujian Dan Analisis Pengaruh Nilai E**



Gambar 2 Hasil Pengujian Nilai E

Grafik menunjukkan bahwa pengujian perubahan nilai E cenderung stabil nilai akurasinya pada nilai E yang berbeda beda walaupun pada rentang 100 dan 80 data. Hal ini menunjukkan bahwa perubahan nilai E di setiap rentang data latih tidak begitu memberikan perubahan akurasi yang besar.

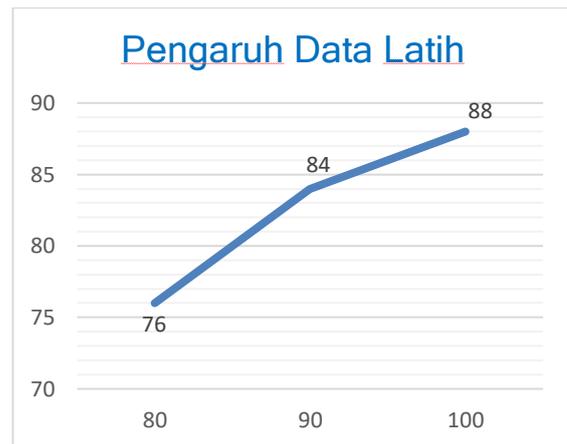
**4.2 Pengujian dan Analisis Pengaruh Nilai K**



Gambar 3 Hasil Pengujian Nilai K

Grafik menunjukkan bahwa perubahan nilai K cenderung mempengaruhi hasil akurasi. Hasil akurasi yang dihasilkan cenderung tidak stabil akan tetapi hasil akurasi terbaik dihasilkan saat K=10, setelah K=10 dan nilai K semakin besar hasil akurasi cenderung menurun. Penurunan hasil akurasi terjadi karena ketika semakin besarnya nilai K tetangga yang digunakan maka semakin banyak data yang mempunyai jumlah jenis yang mendominasi masuk kedalam ketetanggaan yang telah ditentukan. Sehingga ketika dilakukan klasifikasi, data uji akan sering masuk kedalam jenis yang salah..

**4.3 Pengujian Dan Analisis Pengaruh Perubahan Jumlah Data Latih**



Gambar 4 Hasil Pengujian Pengaruh Perubahan Jumlah Data Latih

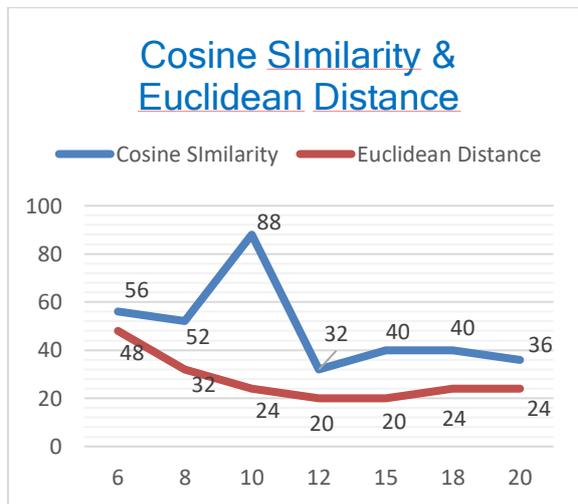
Grafik menunjukkan bahwa hasil akurasi yang diperoleh pada pengujian ini bervariasi. Semakin banyak jumlah data yang digunakan pada data latih, akurasi yang diperoleh juga semakin meningkat. Hasil akurasi terendah pada saat jumlah data latih sebanyak 80 dan 90 data, dan

tertinggi pada saat jumlah data latih sebanyak 100 data. Hal ini disebabkan ketika semakin banyak data latih yang digunakan maka akan semakin banyak data yang akan dibandingkan dengan data uji dan memungkinkan untuk memperoleh hasil *Cosine Similarity* yang mempunyai *similarity* yang mendekati data uji yang akan di klasifikasi cenderung tinggi untuk masuk kedalam ketetanggannya.

**4.4 Pengujian Dan Analisis Akurasi K Fold Cross Validation**

Grafik menunjukkan bahwa hasil akurasi yang diperoleh pada pengujian ini bervariasi. Hasil rata-rata akurasi terendah terdapat pada data uji B. Hal ini disebabkan Karena bervariasinya jumlah kelas risiko hipertensi yang tidak seimbang. Pemilihan data yang dilakukan secara acak juga mempengaruhi hasil akurasi pada sistem. Sementara rata-rata terbaik didapat pada data uji c dikarenakan kelas risiko yang memenuhi hasil klasifikasi mempunyai *similarity* yang tepat pada data latih dan uji.

**4.5 Pengujian Dan Analisis Perbandingan Akurasi Metode Nwknn Menggunakan Euclidean Distance dan Cosine Similarity**



Gambar 5 Hasil Pengujian Metode Nwknn menggunakan *Euclidean Distance* dan *Cosine Similarity*

Grafik menunjukkan bahwa hasil akurasi yang dihasilkan dengan menggunakan *Euclidean Distance* saat nilai *K* kecil akurasinya lebih rendah dibandingkan dengan menggunakan *Cosine Similarity*. Ketika *K* semakin besar dengan menggunakan *Cosine Similarity* hasil akurasi yang dihasilkan cenderung lebih baik walaupun hasil akurasi tersebut tidak stabil dibandingkan dengan *Euclidean Distance* yang bahkan menunjukkan akurasi sampai 20% ketika

*K* bernilai besar pada klasifikasi risiko hipertensi ini. Hal tersebut bisa saja terjadi karena jika menggunakan *Euclidean Distance* untuk mengetahui kedekatan ketetangaan antara data uji dan data latih harus menghitung jarak, sehingga ketika nilai *K* besar maka akan semakin banyak tetangga yang harus di proses untuk diklasifikasi bahkan data yang berjauhan dengan data uji akan tetap dihitung walaupun tidak mempunyai kedekatan jarak. Hal ini mengakibatkan data uji cenderung akan terklasifikasi kedalam jenis yang lebih mendominasi pada data latih.

Berbeda ketika metode Nwknn menggunakan *Cosine Similarity* untuk mengetahui kedekatan ketetangaan antara data uji dan data latih digunakan *similarity*/kemiripan antara data uji dan data latih untuk mengetahui hasil klasifikasinya. Dengan *Cosine Similarity* akan ditentukan kemiripan nilai atau pola yang ada pada data uji dan data latih. Walaupun nilai *K* semakin besar data uji akan tetap terklasifikasi walaupun akurasi yang dihasilkan tidak terlalu baik atau maksimum karena yang di proses untuk klasifikasi adalah kemiripannya, meskipun data berjauhan dengan data uji data tersebut masih sedikit memiliki kemiripan dengan data uji yang ada.

**5. KESIMPULAN**

Cara mengklasifikasi risiko hipertensi dengan menggunakan metode Nwknn yaitu dengan cara memasukan data rekam medis pasien yang meliputi 15 fitur kemudian ditentukan nilai *similarity* antara rekam medis tersebut terhadap data latih dan diproses hingga didapatkan nilai skor, dimana jenis dengan hasil skor tertinggillah yang masuk kedalam klasifikasi. Berdasarkan hasil pengujian, metode Nwknn dapat melakukan klasifikasi risiko hipertensi dengan hasil akurasi terbaik pada *K*=10, *E*=4 dan data latih sebanyak 100 data dengan akurasi 88%. Untuk menentukan kedekatan ketetangaan menunjukkan bahwa metode Nwknn dengan menggunakan *Cosine Similarity* memberikan hasil rata-rata akurasi jauh lebih baik dari pada metode Nwknn dengan menggunakan *Euclidean Distance* pada kasus ini.

**6. DAFTAR PUSTAKA**

Adeniya, D., Wei, Z. & Yongquan, Y., 2014. *Automated Web Usage Data Mining and Recommendation System Using K-*

- Nearest Neighbor (KNN) classification method*. Elsevier, pp. 90-108.
- Andi, S. N., 2015. *Penerapan Algoritma Modified K-Nearest Neighbor (MKNN) untuk Pengklasifikasian Penyakit Attention Deficit Hiperactive Disorder (ADHD) pada Anak*. Informasi dan Teknologi Ilmiah (INTI), pp. 7-10.
- Ari, M. & Rika, A., 2016. *Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree*. Scientific Journal of Informatics, pp. 19-26.
- Faldy, H., Indrianti & Novanto, Y., 2015. *Analisis Sentimen Opini Film Berbahasa Indonesia Berbasis Kamus Menggunakan Metode Neighbor-Weighted K-Nearest Neighbor (NWKNN)*. Malang: Universitas Brawijaya.
- Herwati & Wiwi, S., 2011. *Terkontrolnya Tekanan Darah Penderita Hipertensi Berdasarkan Pola Diet Kebiasaan Olah Raga di Padang Tahun 2011*. Jurnal Kesehatan Masyarakat, pp. 8-14.
- Indriati & Ridok, A., 2016. *Sentiment Analysis for Review Mobile Application Using Neighbor Weighted K-Nearest Neighbor (NWKNN)*. Journal of Environmental Engineering & Sustainable Technology, pp. 23-32.
- Jiawei, H., Micheline, K. & Jian, P., 2012. *Data Mining Concepts and Techniques*. United States of America: Elsevier.
- Nindy, A., 2015. *Klasifikasi Risiko Hipertensi Menggunakan Fuzzy Tsukamoto dan Algoritma Genetika*. Malang: Universitas Brawijaya.
- Nyoman, W., 2016. *Klasifikasi Penyakit Hipertensi Menggunakan Algoritma Naïve Bayes*. Yogyakarta: Universitas Sanata Dharma.
- Putri, N. F., Dian, E. K. & Indriati, 2016. *Identifikasi Jenis Attention Deficit Hyperactivity Disorder (ADHD) Pada Anak Usia Dini menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NWKNN)*. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), pp. 194-200.
- Ridok, A. & Latifah, R., 2015. *Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN*. Konferensi Nasional Sistem & Informatika, pp. 222-227.
- Valerian, A., 2015. *Klasifikasi Dokumen Tanaman Obat menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NWKNN)*. Malang: Universitas Brawijaya.
- Yogiantoro, M., 2009. *Hipertensi Esensial Ilmu Penyakit Dalam Jilid II Edisi V*. Jakarta: Interna