

DETEKSI GANGGUAN AUTIS PADA ANAK MENGGUNAKAN METODE *MODIFIED K-NEAREST NEIGHBOR*

Yuliasih Kripsandita¹, Deni Arifianto², Qurrota A'yun³

Program Studi Teknik informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

yuliasihkripsandita@gmail.com¹, deniarifianto@unmuhjember.ac.id²,
qurrota.ayun@unmuhjember.ac.id³

(Naskah masuk: 25 Februari 2021, diterima untuk diterbitkan: 26 April 2021)

ABSTRAK

Autism Spectrume Disorder merupakan gangguan perkembangan dimana seorang anak memperlihatkan suatu perilaku menjauhkan diri dari lingkungan sosialnya, seakan – akan hidup di dunianya sendiri. Semakin dini mengetahui anak menderita gangguan autis sangatlah penting, karena semakin dini autis diobati semakin besar peluang untuk sembuh. Dengan adanya perkembangan teknologi pendeteksian gejala autis dapat dilakukan lebih awal menggunakan sistem deteksi autis berbasis web. Metode *Modified K-Nearest Neighbor* merupakan perkembangan dari metode konvensional *K-Nearest Neighbor*. Deteksi autis pada anak menggunakan metode *Modified K-Nearest Neighbor* dengan data yang digunakan diambil dari website *UCI Machine Learning Repository*, dengan jumlah data 292 data dan 2 class output. Pengujian dilakukan dengan mencari kedekatan dari data *training* dan data *testing* untuk menghitung *Weight voting*, setelah mendapatkan hasil *weight voting* akan dicari mayoritas datanya. Berdasarkan pengujian yang telah dilakukan didapatkan hasil akurasi tertinggi sebesar 96,67%, hasil presisi tertinggi sebesar 97,33% dan hasil recall tertinggi sebesar 100% pada $K = 13$. Untuk K optimal dari pengujian ini ditunjukkan pada $K = 3$.

Kata Kunci : Deteksi Autisme, Klasifikasi, Metode *Modified K-Nearest Neighbor*

ABSTRACT

Autism Spectrume Disorder is a development disturbance in which a child shows a behavior of distancing himself from his social environment, as though living in his own world. The earlier it is known that a child has an autistic disorder is very important, because the earlier autism is treated, the greater the chance of recovery. With the use of technological developments, autism symptom detection can be done earlier using a web-based autism detection system. Modified K-Nearest Neighbor method is a development of the conventional K-Nearest Neighbor method. Autism detection in children uses the Modified K-Nearest Neighbor method with the data used taken from the UCI Machine Learning Repository website, with a total of 292 data and 2 output classes. Testing is done by looking for the closeness of each training data to determine the validity value, after that look for the closeness of training data and testing data to calculate weight voting, after getting the results of weight voting the majority of the data will be searched. Based on the results of the test which have been done, it was found out that the highest accuracy results were 96.67%,

highest precision results were 97,33% and highest recall results were 100% at $K = 13$. For the optimal k of this test was shown at $K = 3$.

Keywords: Autism Detection, Classification, Modified K-Nearest Neighbor Method.

1. PENDAHULUAN

Autisme merupakan salah satu gangguan tumbuh kembang yang menyebabkan fungsi otak tidak bekerja secara normal sehingga dapat berpengaruh pada kemampuan berkomunikasi dan berinteraksi sosial seseorang. Gejala pada autisme dapat diketahui dari adanya penyimpangan tumbuh kembang seorang anak (Sunu, 2012). Kurangnya pengetahuan pada gejala awal autisme menyebabkan peningkatan jumlah penderitanya.

Untuk mengurangi kesalahan dalam deteksi dan mengurangi keterlambatan dalam diagnosa maka dibutuhkan sebuah sistem yang mampu mendeteksi gejala autis sejak dini dengan menggunakan penerapan teknik data mining metode klasifikasi. Metode klasifikasi dapat dilakukan dengan memetakan suatu item data ke dalam kelas berdasarkan kelas data yang telah didefinisikan sebelumnya (Agarwal, 2014).

Aplikasi berbasis web merupakan suatu aplikasi, yang dapat diakses melalui penjelajah web menggunakan jaringan seperti internet. Aplikasi web menjadi populer karena kemudahan dalam mengaksesnya. Jadi pada penelitian ini akan dibuat deteksi autis berbasis web, agar memudah orangtua untuk mengetahui lebih awal gejala autis pada anak.

Metode *Modified K-Nearest Neighbor* merupakan pengembangan dari metode konvensional *K-Nearest Neighbor*. Dimana dilakukan modifikasi dengan menambahkan proses validitas dan proses *weight voting* (Parvin et al., 2010).

Sebagai pendukung proses olah data penulis memutuskan menggunakan metode *Modified K-Nearest Neighbor*. Karena pada penelitian sebelumnya dengan judul "*Modified K-Nearest Neighbor* untuk Prediksi Curah Hujan" yang dilakukan oleh (Muhammad, 2015) didapatkan hasil nilai akurasi terbaik sebesar 82,46%. Pada penelitian selanjutnya dengan judul "Deteksi Penyakit Kucing dengan menggunakan *Modified K-Nearest Neighbor* Teroptimasi" yang dilakukan oleh (Fitri et al., 2017) didapatkan hasil untuk K optimal 1 sebesar 100% menggunakan algoritma genetika. Dan pada penelitian selanjutnya dengan judul "Klasifikasi Kelompok Varietas Unggul padi menggunakan *Modified K-Nearest Neighbor*" didapatkan hasil akurasi tertinggi sebesar 79,96% dan nilai akurasi terendah 51,2%. Melihat hasil dari penelitian - penelitian sebelumnya diharapkan pada penelitian ini mampu menghasilkan nilai akurasi yang lebih baik.

Berdasarkan paparan yang telah dikemukakan diatas, penulis tertarik untuk mengimplentasikan deteksi dini autis melalui penelitian dengan judul "Deteksi Gangguan Autis pada Anak menggunakan metode *Modified K-Nearest Neighbor*". Jawaban dari setiap pertanyaan pada *screening* kemudian dijadikan bahan masukan untuk menklasifikasi autis. Penggunaan metode *Modified K-Nearest Neighbor* diharapkan dapat memberikan hasil nilai akurasi, presisi dan recall yang tinggi.

2. METODOLOGI PENELITIAN

2.1 Data Penelitian

Pada data penelitian ini data yang digunakan merupakan data sekunder yang diperoleh dari media *internet* pada *website UCI Machne Learning* (<https://archive.ics.uci.edu/ml/index.php>) tentang *Autism Spectrume Disorder* dari *Auckland New Zeland* pada tahun 2017 dengan jumlah data 150 dan 10 atribute

2.2 Metode *K-Nearest Neighbor*

Metode *K-Nearest Neighbor* merupakan metode yang paling dasar dari contoh *based learning* dan juga termasuk kelompok *instance-based learning*. *K*-dalam penelitian. *Nearest Neighbor* dilakukan dengan mencari kelompok objek dalam data *training* yang paling dekat dengan dengan objek pada data *testing* (Saxena et al., 2014). Metode KNN dapat dihitung menggunakan persamaan *Euclidean Distance*.

Euclidean Distance merupakan perhitungan jarak dari dua titik dalam *Euclidean Space*, untuk mempelajari hubungan antar dua jarak (Nurliza, 2018). Berikut ini rumus dari *Euclidean Distance* yang digunakan untuk menghitung jarak antar titik pada data *training* (x) dan titik pada data *testing* (y):

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

Keterangan:

$d(x, y)$: jarak *Euclidean*

x_i : data *training* ke- i

y_i : data *testing* ke- i

i : 1,2,3,...,n

2.3 Metode Modified *K-Nearest Neighbor*

Metode *Modified K-Nearest Neighbor* merupakan suatu pengembangan dari metode *K-Nearest Neighbor* yang diusulkan oleh (Parvin et al., 2010) yang sebagian bertujuan untuk perubahan pada tingkat akurasi yang rendah pada metode

K-Nearest Neighbor. Pengembangan dilakukan dengan memodifikasi *metode K-Nearest Neighbor* yang bertujuan untuk meningkatkan kinerja dan akurasi pada metode *K-Nearest Neighbor*. Pada metode *Modified K-Nearest Neighbor* terdapat penambahan proses yakni proses validitas data pada setiap dataset yang telah dilakukan perhitungan jarak euclidean antar data *training*. Dan juga penambahan proses *weight voting* yang didapatkan dari hasil nilai validitas data dikalikan dengan hasil perhitungan jarak *euclidean* data *training* dan data *testing* (Parvin et al., 2010).

a. Perhitungan Nilai Validitas

Dalam metode *Modified K-Nearest Neighbor* suatu data akan melalui tahap validitas setelah mendapatkan hasil dari perhitungan jarak *euclidean* antar data *training* (Parvin et al., 2010). Untuk perhitungan nilai validitas digunakan persamaan sebagai berikut:

$$Validity(i) = \frac{1}{H} \sum_{i=1}^H S(lbl(x), lbl(Ni(x))) \quad (2)$$

Keterangan :

H : nilai tetangga yang digunakan

$lbl(x)$: label dari dataset

$Ni(x)$: label dari tetangga terdekat dengan x

i : 1,2,3.... H .

S : Fungsi yang memperhitungkan kesamaan antara satu dataset dengan dataset yang lain.

Berikut definisi dari fungsi S :

$$S(a, b) = \begin{cases} 1 & \text{jika } a = b \\ 0 & \text{jika } a \neq b \end{cases} \quad (3)$$

Keterangan :

a : Kelas a pada data latih

b : kelas lain selain a pada data latih

b. Perhitungan *Weight Voting*

Weight Voting dataset dalam metode *Modified K-Nearest Neighbor* merupakan

salah satu dari modifikasi yang dilakukan pada metode konvensional *K-Nearest Neighbor*. Dimana dilakukan penentuan kelas dari data objek yang baru dengan cara melakukan pemilihan berdasarkan bobot pada *dataset*. Setiap *dataset* akan diberikan bobot dengan melakukan perhitungan dengan persamaan seperti dibawah ini (Parvin et al., 2010) :

$$W(i) = Validity(i) \times \frac{1}{de+a} \quad (4)$$

Keterangan :

$W(i)$: Bobot

$Validity(i)$: Validitas dari i tetangga terdekat antara dataset dan data uji

de : Jarak Euclidean antar data latih dengan data uji

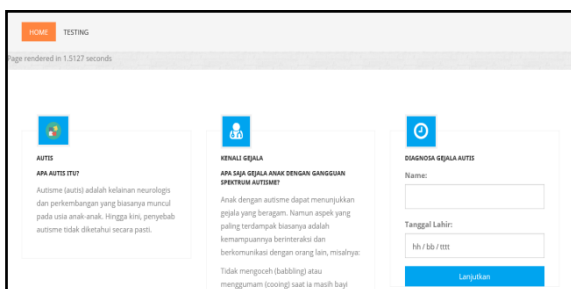
a : Konstanta

Teknik *weight voting* akan memberikan nilai yang tinggi pada *dataset* yang memiliki nilai validitas yang besar serta kesamaan yang tinggi dengan data testing (Parvin et al., 2010)

3. HASIL DAN PEMBAHASAN

3.1 Implementasi

Bagian ini menjelaskan mengenai implementasi mengenai antarmuka yang terdiri dari implementasi halaman beranda, halaman deteksi autisme dan halaman *testing* (proses pengujian metode MKNN).



Gambar 3. 1 Tampilan Halaman Beranda



Gambar 3. 2 Tampilan halaman Deteksi Autis

Pilih Data Uji: 2-Fold | 1-75
 Pilih Nilai K: Nilai K=3
 Hitung Data

DATA TRAINING DAN TESTING												
No	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈	A ₉	A ₁₀	Jenis Data	Class/ASD
1	1	1	0	0	1	1	0	1	0	0	Training	2
2	1	1	0	0	1	1	0	1	0	0	Training	2
3	1	1	0	0	0	1	1	1	0	0	Training	2
4	0	1	0	0	1	1	0	0	0	1	Training	2
5	1	1	1	1	1	1	1	1	1	1	Training	1
6	0	0	1	0	1	1	0	1	0	1	Training	2
7	1	0	1	1	1	1	0	1	0	1	Training	1
8	1	1	1	1	1	1	1	1	0	0	Training	1
9	1	1	1	1	1	1	1	0	0	0	Training	1
10	0	0	1	1	1	0	1	1	0	0	Training	2

Gambar 3. 3 Tampilan Halaman Testing

Setelah dilakukan pengujian pada halaman testing didapatkan hasil akurasi, presisi dan recall seperti pada tabel dibawah ini :

3.2 Pengujian K-fold

Pengujian K-fold dilakukan menggunakan *2-fold*, *3 fold* dan *5-fold* dengan jumlah data 150 data. Setelah dilakukan pengujian didapatkan hasil akurasi, presisi dan recall yang ditunjukkan pada tabel dibawah ini .

Tabel 3. 1 Hasil Akurasi, Presisi dan Recall Pada 2-fold Skenario 1

Nilai K	2-Fold Skenario 1		
	Akurasi	Presisi	Recall
3	86,67%	86,67%	76,92%
5	81,33%	81,33%	68,89%
7	77,33%	77,33%	64,58%
9	72%	72%	59,62%
11	60%	60%	50,82%
13	60%	49,33%	44,93%

Tabel 3.1 menunjukkan hasil dari pengujian dengan 2-fold skenario 1, didapatkan hasil akurasi tertinggi sebesar 86,67%, presisi tertinggi 86,67% dan recall tertinggi 76,92% yang terdapat pada K = 3.

Tabel 3. 2 Hasil Akurasi, Presisi dan Recall pada 2-fold Skenario 2

Nilai K	2-Fold Skenario 2		
	Akurasi	Presisi	Recall
3	91,89%	86,96%	100%
5	90,54%	90,24%	92,50%
7	87,84%	82,98%	97,50%
9	75,68%	68,97%	100%
11	74,32%	67,80%	100%
13	75,68%	68,97%	100%

Tabel 3.2 menunjukkan hasil dari pengujian dengan 2-fold skenario 2, didapatkan hasil akurasi tertinggi sebesar 91,89%, presisi tertinggi 90,24% dan recall tertinggi 100%.

Tabel 3. 3 Hasil Akurasi, Presisi dan Recall Pada 3-fold Skenario 1

Nilai K	3-Fold Skenario 1		
	Akurasi	Presisi	Recall
3	90%	80,77%	100%
5	88%	77,88%	100%
7	88%	77,78%	100%
9	86%	75%	100%
11	90%	83,33%	95,24%
13	88%	77,78%	100%

Tabel 3.3 menunjukkan hasil dari pengujian dengan 3-fold skenario 1, didapatkan hasil akurasi tertinggi sebesar

90%, presisi tertinggi 83,33% dan recall tertinggi 100%.

Tabel 3. 4 Hasil Akurasi, Presisi dan Recall Pada 3-fold Skenario 2

Nilai K	3-Fold Skenario 2		
	Akurasi	Presisi	Recall
3	92%	86,67%	100%
5	90%	86,21%	96,15%
7	90%	83,87%	100%
9	86%	78,79%	100%
11	80%	72,22%	100%
13	78%	70,27%	100%

Tabel 3.4 menunjukkan hasil dari pengujian dengan 3-fold skenario 2, didapatkan hasil akurasi tertinggi sebesar 90%, presisi tertinggi 86,67% dan recall tertinggi 100%.

Tabel 3. 5 Hasil Akurasi, Presisi dan Recall Pada 3-fold Skenario 3

Nilai K	3-Fold Skenario 3		
	Akurasi	Presisi	Recall
3	87,76%	80%	100%
5	85,71%	77,42%	100%
7	75,51%	66,67%	100%
9	71,43%	63,16%	100%
11	71,43%	63,16%	100%
13	69,39%	61,54%	100%

Tabel 3.5 menunjukkan hasil dari pengujian dengan 3-fold skenario 3, didapatkan hasil akurasi tertinggi sebesar 87,76%, presisi tertinggi 80% dan recall tertinggi 100%.

Tabel 3. 6 Hasil Akurasi, Presisi dan Recall Pada 5-fold Skenario 1

Nilai K	5-Fold Skenario 1		
	Akurasi	Rresisi	Recall
3	86,67%	78,95%	100%
5	90%	83,33%	100%
7	90%	83,33%	100%
9	90%	83,33%	100%
11	93,33%	88,24%	100%
13	96,67%	97,33%	100%

Tabel 3.6 menunjukkan hasil dari pengujian dengan 5-fold skenario 1,

didapatkan hasil akurasi tertinggi sebesar 96,67%, presisi tertinggi 97,33% dan recall tertinggi 100%.

Tabel 3. 7 Hasil Akurasi, Presisi dan Recall Pada 5-fold Skenario 2

Nilai K	5-Fold Skenario 2		
	Akurasi	Rresisi	Recall
3	83,33%	66,67%	100%
5	76,67%	60%	90%
7	76,67%	60%	90%
9	76,67%	60%	90%
11	76,67%	58,82%	90%
13	80%	62,50%	100%

Tabel 3.7 menunjukkan hasil dari pengujian dengan 5-fold skenario 2, didapatkan hasil akurasi tertinggi sebesar 83,33%, presisi tertinggi 66,67% dan recall tertinggi 100%.

Tabel 3. 8 Hasil Akurasi, Presisi dan Recall Pada 5-fold Skenario 3

Nilai K	5-Fold Skenario 3		
	Akurasi	Rresisi	Recall
3	90%	83,33%	100%
5	86,67%	78,95%	100%
7	80%	71,43%	100%
9	80%	71,43%	100%
11	76,67%	68,18%	100%
13	80%	71,43%	100%

Tabel 3.8 menunjukkan hasil dari pengujian dengan 5-fold skenario 3, didapatkan hasil akurasi tertinggi sebesar 86,67%, presisi tertinggi 83,33% dan recall tertinggi 100%.

Tabel 3. 9 Hasil Akurasi, Presisi dan Recall Pada 5-fold Skenario 4

Nilai K	5-Fold Skenario 4		
	Akurasi	Rresisi	Recall
3	90%	93,33%	87,50%
5	96,67%	94,12%	100%
7	93,33%	88,89%	100%
9	80% ^o	72,73%	100%
11	80%	72,73%	100%
13	80%	72,73%	100%

Tabel 3.9 menunjukkan hasil dari pengujian dengan 5-fold skenario 4, didapatkan hasil akurasi tertinggi sebesar 96,67%, presisi tertinggi 94,12% dan recall tertinggi 100%.

Tabel 3. 10 Hasil Akurasi, Presisi dan Recall Pada 5-fold Skenario 5

Nilai K	5-Fold Skenario 5		
	Akurasi	Rresisi	Recall
3	86,21%	86,21%	100%
5	89,66%	83,33%	100%
7	79,31%	71,43%	100%
9	72,41%	65,22%	100%
11	72,41%	65,22%	100%
13	68,97%	62,5%	100%

Tabel 3.10 menunjukkan hasil dari pengujian dengan 5-fold skenario 5, didapatkan hasil akurasi tertinggi sebesar 89,66%, presisi tertinggi 86,21% dan recall tertinggi 100%.

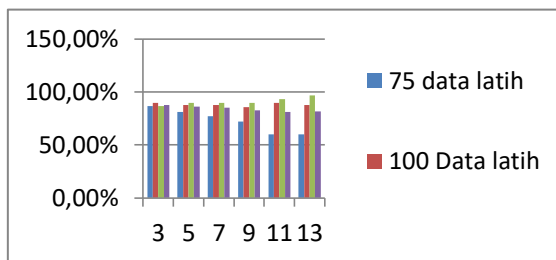
3.3 Pengujian Pengaruh Nilai K

Pengujian terhadap nilai K dilakukan menggunakan nilai K yang bervariasi, yakni nilai K = 3, K = 5, K = 7, K = 9, K = 11 dan K = 13 dengan jumlah data latih yang berbeda yaitu 75, 100, dan 120 data latih. Pengujian nilai K dilakukan dengan masing – masing data latih lalu dilakukan analisa terhadap nilai rata – rata yang didapatkan dari pengujian. Hasil pengujian pengaruh nilai K terhadap akurasi ditunjukkan pada tabel dibawah ini.

Tabel 3. 11 Hasil Pengujian Pengaruh Nilai K

NILAI K	75 data latih	100 Data latih	120 Data Latih	Rata – rata Akurasi
3	86,67%	90%	86,67%	87.78%
5	81,33%	88%	90%	86.44%
7	77,33%	88%	90%	85.11%
9	72%	86%	90%	82.67%
11	60%	90%	93,33%	81,11%
13	60%	88%	96,67%	81,56%

Dari pengujian pengaruh nilai K yang telah dilakukan didapatkan hasil dari nilai K = 3 menunjukkan rata – rata akurasi dengan presentase 87,78%. Untuk nilai K = 5 menunjukkan rata – rata akurasi dengan nilai presentase 86,44%. Untuk nilai K = 7 menunjukkan rata – rata akurasi dengan presentase 85,11%. Untuk nilai K = 9 menunjukkan rata – rata akurasi dengan presentase 82,67%. Untuk nilai K = 11 menunjukkan rata – rata akurasi dengan presentase 81,11%. Dan untuk nilai K = 13 menunjukkan rata – rata akurasi dengan presentase 81,56%. Grafik hasil pengujian pengaruh nilai K terhadap akurasi ditunjukkan pada Gambar 3.4



Gambar 3. 4 Grafik Pengujian Pengaruh Nilai K

4. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini yaitu nilai akurasi tertinggi sebesar 96,67%, nilai presisi tertinggi sebesar 97,33% dan nilai recall tertinggi sebesar 100% pada K = 11 dengan jumlah data latih 120 data. Untuk rata – rata akurasi tertinggi sebesar 87,78% terdapat pada K = 3.

Saran yang dapat diberikan untuk penelitian selanjutnya dapat menggunakan nilai K dan K-fold yang lebih beragam dan bisa mengembangkan sistem deteksi autis pada anak dengan penelitian ini sebagai acuan.

DAFTAR PUSTAKA

Agarwal, S. (2014). Data mining: Data mining concepts and techniques. *Proceedings -*

2013 *International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013.*
<https://doi.org/10.1109/ICMIRA.2013.45>

Fitri, A. D., Eka, R. D., & Wahyu, W. A. (2017). Deteksi Penyakit Kucing dengan Menggunakan Modified K-Nearest Neighbor Teroptimasi (Studi Kasus : Puskesmas Klinik Hewan dan Satwa Sehat Kota Kediri). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 1(11), 1295–1301.

Muhammad, B. L. (2015). Modified Nearest Neighbor Untuk Prediksi Curah Hujan. *Konferensi Nasional Sistem dan Informatika 2015*, 1, 272–277.
<http://ejournal.stikom-bali.ac.id/index.php/knsi/article/view/463/15>

Nurliza, N. N. (2018). *Penerapan Euclidean Distance Pada Pengenalan Pola Citra Sidik Jari*. 1–67.

Parvin, H., Alizadeh, H., & Minati, B. (2010). A Modification on K-Nearest Neighbor Classifier. *Global Journal of Computer Science and Technology*, 10(14), 37–41.

Saxena, K., Khan, Z., & Singh, S. (2014). Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm. *International Journal of Computer Science Trends and Technology (IJCSST)*.

Sunu, C. (2012). *Panduan memecahkan masalah autisme unlocking autism*. Lintangterbit.