

Penerapan Teknik Vote Menggunakan C4.5 Naïve Bayes Dan K-nearest Neighbor Pada Data Gangguan Autisme

Application of Vote Technique Using C4.5 Naïve Bayes and K-nearest Neighbor on Autism Disorder Data

Arifandi Pratama¹, Agung Nilogiri^{2*}, Habibatul Azizah³

¹Mahasiswa Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Email : arifandy0805@gmail.com

²Dosen Fakultas Teknik, Universitas Muhammadiyah Jember* Koresponden Author
Email : agungnilogiri@unmuhjember.ac.id

³ Dosen Fakultas Teknik, Universitas Muhammadiyah Jember
Email : habibatulazizah@unmuhjember.ac.id

Abstrak

Autisme merupakan gangguan pada perkembangan otak yang memengaruhi kemampuan penderita dalam berkomunikasi dan berinteraksi dengan orang lain,. Pada penelitian yang dilakukan oleh (Zhang, dkk, 2014) pada data Breast-cancer dengan algoritma C4.5 didapatkan akurasi 75,5%. Penelitian pada dataset gangguan autisme pada anak pernah dilakukan oleh (Sugara, dkk., 2018) pada penelitiannya yang menggunakan algoritma C4.5 didapatkan akurasi sebesar 72%. Berdasarkan latar belakang tersebut penelitian dilakukan untuk meningkatkan akurasi pada data gangguan autisme. Untuk itu akan digunakan teknik *voting* pada algoritma C4.5, atribut yang digunakan yaitu GJ01 hingga GJ24. Metode yang akan digunakan pada *vote* yaitu C4.5, *K-nearest Neighbor* dan *naive bayes*. *Ensemble method* merupakan teknik untuk meningkatkan akurasi. Salah satu contoh dari *ensemble method* adalah *voting* atau bisa disebut *majority vote*. lalu digunakan ensemble method majority vote didapatkan akurasi 88,89%, dimana akurasi tersebut mendapatkan peningkatan akurasi sebesar 13,39% .

Kata Kunci : *Gangguan Autisme, Algoritma C4.5, vote, K-nearest neighbor, Naive bayes*

Abstract

Autism is a disorder in brain development that affects the patient's ability to communicate and interact with other people, by classifying the symptoms of people with autism, it will be faster to find out the disorders experienced. In a study conducted by (Zhang, et al, 2014) on breast-cancer data with the C4.5 algorithm, an accuracy of 75.5% was obtained. Research on autistic disorder datasets in children has been carried out by (Sugara, et al., 2018) in his research using the C4.5 algorithm, an accuracy of 72% was obtained. Based on this background, research was conducted to increase the accuracy of data on autism disorders. technique will be used *voting* on the C4.5 algorithm, the attributes used are GJ01 to GJ24. The methods that will be used in *voting* are C4.5, *K-nearest Neighbor* and *naive bayes*. *Ensemble method* is a technique to improve accuracy. One example of *the ensemble method* is *voting* or can be called *majority vote*. Then the majority vote ensemble method was used to get an accuracy of 88.89%, where the accuracy got an increase in accuracy of 13.39%.

Keywords: *Autism Disorder, C4.5 Algorithm, vote, K-nearest neighbor, Naive bayes*

1. Pendahuluan

A. Latar Belakang

Autisme merupakan gangguan yang sejak dulu menjadi salah satu misteri di kedokteran dengan perkembangan yang sangat kompleks. Gangguan autisme sudah ada sejak lama, namun banyak yang tidak mengetahui sebagai gangguan autis. Pada cerita zaman dulu sering dianggap hal yang tidak normal pada anak, gejala autis sudah menunjukkan perilaku yang tidak normal seperti pada umumnya. Menolak ketika digendong, sering menangis saat malam hari dan tidur pada siang hari. Seringkali membuat orang tuanya bingung dengan bahasa yang tidak mereka mengerti. Mereka bisa mencakar, mengigit ataupun menyerang saat dalam kondisi marah. Terkadang tertawa seperti ada yang mengajaknya bercanda dan orang tua ada yang menganggapnya tertukar dengan anak peri, sehingga tidak bisa menyesuaikan perilaku dengan kehidupan manusia normal pada umumnya.

Machine learning adalah ilmu pengetahuan yang berperan besar dalam bidangnya. Tanpa disadari seluruh pengguna *machine learning* menggunakan produk yang dihasilkan oleh penerapan *machine learning*. Pada era perkembangan teknologi *machine learning* sudah sering digunakan untuk membantu dalam mendiagnosa suatu penyakit pada gangguan autisme dengan metode klasifikasi. *Decision tree* merupakan metode yang sering digunakan dalam klasifikasi. Salah satunya algoritma *decision tree C4.5*. *Ensemble method* adalah menggabungkan beberapa klasifikasi tree untuk menghasilkan kinerja prediksi yang lebih baik daripada klasifikasi tree tunggal. sehingga meningkatkan akurasi model. Salah satunya adalah *Majority Voting* yang merupakan peningkatan dari algoritma C4.5 oleh karena itu pemilihan metode yang digunakan membuktikan kelebihan pada *Majority Voting* mendapatkan akurasi yang lebih baik dibandingkan dengan metode klasifikasi biasa, dengan demikian pada penelitian ini akan dilakukan klasifikasi algoritma C4.5 dengan metode *Majority Voting*.

B. Rumusan Masalah

Dari uraian latar belakang dapat dirumuskan permasalahan sebagai berikut:

1. Berapa tingkat presisi *Majority Voting* algoritma C4.5 pada deteksi dini gangguan autisme.
2. Berapa tingkat akurasi algoritma C4.5 sebelum dan sesudah menggunakan metode teknik *Majority Voting*.

C. Batasan Masalah

Permasalahan pada tulisan ini dibatasi dengan batasan-batasan sebagai berikut:

1. Data yang akan digunakan pada penelitian ini adalah data dari penelitian oleh (Sugara, Widyatmoko, Prakoso & Saputro, 2017).
2. Priode yang digunakan pada penelitian ini pada tahun 2017.
3. Data yang dihitung sebagai sample adalah data gejala autisme dengan jumlah record sebanyak 50 data.
4. Penentuan gangguan autisme berdasarkan 24 parameter.

D. Tujuan Penelitian

Tujuan penelitian adalah untuk mengetahui:

1. Mengetahui tingkat presisi *Majority Voting* algoritma C4.5 pada deteksi dini gangguan autisme.
2. Mengetahui tingkat akurasi algoritma C4.5 sebelum dan sesudah menggunakan metode teknik *Majority Voting*.

E. Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memperoleh manfaat dan memberikan kegunaan sebagai berikut:

1. Dapat digunakan sebagai alternatif perhitungan klasifikasi C4.5.
2. Peningkatan akurasi algoritma C4.5 sebelum dan sesudah menggunakan *Majority Voting*.

2. Tinjauan Pustaka

A. Autisme

Autisme berasal dari kata “auto” yang berarti sendiri. Karena istilah ini penderita autis umumnya lebih memilih untuk menghindari dari segala bentuk interaksi

secara langsung yang membuat mereka seolah hidup sendirian. Gangguan autisme bukanlah gangguan motorik biasa. Pengaruh terhadap pola pikir dan tindakan anak sangat besar, pada masa depan anak sangat berpengaruh dan tidak menutup kemungkinan menjadi abnormal selamanya jika dibiarkan tanpa mendapatkan terapi khusus (Sunu, 2012).

B. Machine Learning

Machine learning menyelesaikan masalah dengan menggunakan data dari pembelajaran yang dilakukan manusia. Penerapan *machine learning* yaitu klasifikasi dan prediksi, klasifikasi adalah metode yang mengurutkan data dengan mengklasifikasikan atribut satu sama lain. Sedangkan prediksi digunakan untuk *output* dari *input* data yang telah diperoleh melalui data *training*.

Kemampuan yang belajar dengan mesin yang telah diprogram dari awal dengan rangkaian perintah tertentu. Mesin *learning* memiliki tugas yang dinamis dengan kapabilitas yang tinggi.

C. Klasifikasi

Klasifikasi adalah suatu pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target. Sehingga algoritma-algoritma untuk menyelesaikan masalah klasifikasi dikategorisasikan ke dalam *supervised learning* atau pembelajaran yang diawasi. Teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini menggunakan *supervised induction*, yang memanfaatkan kumpulan pengujian dari record yang terklasifikasi untuk menentukan kelas-kelas tambahan. Salah satu contoh yang mudah dan populer adalah dengan *Decision tree* yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi.

D. Algoritma C4.5

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- Pilih atribut sebagai akar.
- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.

- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus seperti yang tertera berikut (David Hartanto, 2014):

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{s_i}{s} * Entropy(S_i)$$

Dimana:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai *Gain* adalah dengan mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar dari Entropi adalah sebagai berikut: $Entropy(S) = -\sum_{i=1}^n p_i \log_2 p_i$

Dimana :

S : Himpunan kasus

n : Jumlah partisi S

p_i : Proporsi dari S_i terhadap S

E. Naïve Bayes

Naïve Bayes merupakan sebuah pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema bayes dan mengansumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. *Naïve Bayes* juga didefinisikan sebagai pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. *Naïve Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai

output. Dengan kata lain, diberikan nilai *output*, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Saleh, 2015).

$$P(H|X) = P(X|H).p(H)/p(H)$$

Dimana :

X : data dengan class yang belum diketahui

H : hipotesis data menggunakan suatu class spesifik

P(H|X) : probabilitas hipotesis H berdasar kondisi X (parteriori probabilitas)

P(H) : probabilitas hipotesis H (prior probabilitas)

P(X|H) : probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : probabilitas

F. K-nearest Neighbor

Algoritma *K-nearest neighbor* (KNN) merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data terdekat dari objek tersebut, KNN juga merupakan algoritma supervised learning dimana hasil klasifikasi data baru berdasarkan pada kategori sebagian besar jarak terdekat menuju K. Algoritma ini menghitung berdasarkan jarak minimum dari data baru ke K terdekat yang sudah ditetapkan. Jarak antara data baru dengan data *learning* dihitung dengan cara mengukur jarak antara titik yang mewakili data baru dengan titik yang mewakili data *learning* dengan rumus *Euclidean Distance* (Fitroh, 2015).

$$(x,y)=\sqrt{\sum(xik-yik)^2mk=1}$$

Dimana,

xik = nilai *x* pada *training* data

yik = nilai *x* pada *testing* data

m = batas jumlah banyaknya data

G. Majority Voting

Aturan keputusan *Majority Voting* yang memilih salah satu dari banyak alternatif, berdasarkan kelas prediksi dengan suara terbanyak. Pemungutan suara mayoritas tidak memerlukan parameter apa pun untuk mengklasifikasi klasifikasi individual yang telah dilatih. Dalam hal pemungutan suara

terbobot, bobot pemungutan suara harus bervariasi di antara kelas *output* yang berbeda di setiap *classifikasi*. Masalah pembobotan dapat dilihat sebagai masalah optimisasi.

H. Cross Validation

Cross-validation merupakan metode menggeneralisasi pendekatan ini dengan mensegmentasi data kedalam *k* partisi berukuran sama. Selama proses, salah satu dari partisi dipilih untuk *training*, sedangkan sisanya untuk *testing*. Prosedur ini diulangi *k* kali sedemikian sehingga setiap partisi digunakan untuk *testing* tepat satu kali. Total error ditentukan dengan menjumlahkan error untuk semua *k* proses tersebut (Tan, dkk. 2005).

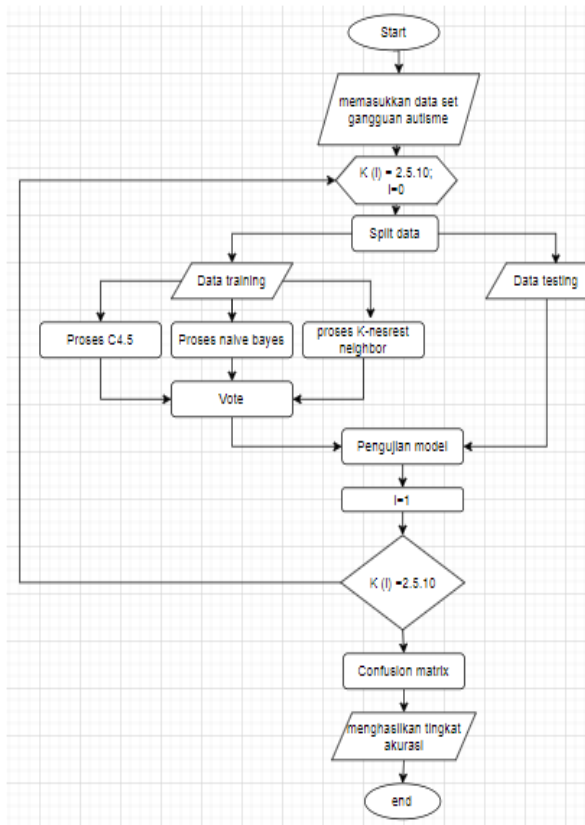
I. Confusion Matrix

Confusion Matrix adalah pengukuran terhadap kinerja suatu sistem klasifikasi merupakan hal yang penting. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya (Rosandy, 2016).

3. Metodologi Penelitian

A. Jenis Penelitian

Penjelasan mengenai tahapan-tahapan dalam penerapan teknik *Majority Voting* untuk mengetahui akurasi terbaik dari hasil klasifikasi menggunakan teknik pada dataset gangguan autisme pada anak. Metodologi yang digunakan pada penelitian ini terdiri dari beberapa tahap yaitu pendahuluan, pengumpulan data, analisis data, perancangan sistem dan penarikan kesimpulan. Digram metodologi penelitian berikut tahapannya pada gambar dibawah ini.



Gambar 1 Kerangka Penelitian
 Sumber: Hasil *Drawio*

B. Metode Analisis Data

Dalam penelitian ini, analisis yang digunakan adalah deskriptif untuk data gejala pada gangguan autisme pada anak menggunakan software Microsoft excel 2010. Kemudian digunakan teknik *Majority Voting* pada algoritma C4.5, Naive Bayes, *K-nearest Neighbor* (KNN).

C. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data gejala gangguan autisme pada anak dari penelitian oleh (Sugara, Widyatmoko, Prakoso & Saputro, 2017). Data yang digunakan terdiri dari 24 parameter dan juga 3 output yang dihasilkan. Data yang akan digunakan berjumlah 50 *record*. Data diambil pada suatu lembaga autisme di bekasi pada periode 2018 oleh (Sugara et al., 2018). Data yang nantinya akan digunakan sebagai contoh perhitungan adalah 30 *record* 20 data untuk *training* dan 10 data untuk *testing*.

D. Proses Algoritma C4.5

Data training yang akan digunakan pada contoh perhitungan C4.5 ini menggunakan data gejala gangguan autisme. Yang mana pemilihan data dilakukan secara acak. Dengan jumlah data 15 dan atribut 24.

Tabel 1 Hasil Perhitungan node 1.3

GJ04	GJ05	GJ19	GJ24	Gangguan
Tidak	Tidak	Ya	Ya	Gangguan Interaksi Sosial
Tidak	Tidak	Ya	Ya	Gangguan Perilaku
Tidak	Tidak	Ya	Ya	Gangguan Interaksi Sosial
Tidak	Tidak	Ya	Ya	Gangguan Interaksi Sosial

Sumber: Hasil Perhitungan

E. Proses K-nearest Neighbor

Menghitung jarak Euclid data baru terhadap data yang ada. Menunjukkan perhitungan jarak Euclid. Rumus Euclidean distance. Dari hasil prediksi pada perhitungan *K-nearest Neighbor* dengan $K=3$ didapatkan prediksi Gangguan Komunikasi.

Tabel 2 Hasil K-nearest Neighbor

No	Class Aktual	Class Prediksi
1	Gangguan Perilaku	Ya
5	Gangguan Komunikasi	Ya
6	Gangguan Komunikasi	Ya

Sumber: Hasil Perhitungan

F. Proses Naïve Bayes

1. Menyiapkan data *training* yang akan dilakukan klasifikasi .
2. Melakukan perhitungan nilai jumlah sub atribut pada masing-masing *output* yang ada pada masing-masing atribut .
3. Setelah dihitung lalu siapkan data *testing* yang akan digunakan untuk menentukan

- klasifikasi yang akan digunakan untuk penilaian pada data *testing*
4. lakukan perhitungan masing-masing pada masing-masing output dimana ada 3 yang akan dirubah menjadi c1,c2,c3. Dimana c1 adalah gangguan komunikasi, c2 adalah gangguan perilaku, dan c3.

Tabel 3 Hasil Perhitungan Naïve Bayes

C1	C2	C3
0,00106168	0	0
0	0	0,002646107
0	0	0,000211689
0	0	0,001323054
0	0	0

Sumber: Hasil Perhitungan

G. Evaluasi

Evaluasi akan dilakukan dengan membandingkan hasil klasifikasi pada 3 pohon keputusan yang telah ada. Suara terbanyak akan dijadikan hasil klasifikasi akhir.

4. Implementasi Sistem

A. Gambaran Dataset

Data penelitian yang akan digunakan yaitu data dari penelitian yang dilakukan oleh (Sugara, dkk, 2018). Data ini berupa data gangguan autisme pada anak yang terdiri dari 24 atribut yaitu GJ01 hingga GJ24 dan 3 output yaitu gangguan interaksi sosial, gangguan komunikasi, dan gangguan perilaku. Data ini diambil pada periode 2018 dengan jumlah data awal 50 *record*. Lalu karena dataset tidak *balance* maka data awal berubah menjadi 36 data dengan 12 data masing-masing *output*.

B. Skenario Uji

Pengujian yang akan dilakukan menggunakan *tool Rapid miner studio* versi 9.5 sebagai pengaplikasiannya. Sebelum digunakan pada *tool Rapid miner* data terlebih dahulu dibagi menjadi data *training* dan data *testing*. Dimana dengan jumlah data yang berbeda sesuai dengan *k-fold cross validation*.

Data akan dibagi menjadi k sama dengan 2, 3, 4, dan 6.

C. Hasil Pengujian Metode C4.5

Pada hasil pengujian ini yang akan perlihatkan adalah hasil terbaik dari *cross validation* sebelum menggunakan teknik *Ensemble*, didapatkan akurasi terbaik pada *k-fold* 6 percobaan ke 3 sebesar 83,33%, dengan presisi dimana nilai positif gangguan interaksi sosial memiliki hasil 100%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 66,67%.

accuracy: 83.33%

	true Gangguan Interaksi Sosial	true Gangguan Komunikasi	true Gangguan Perilaku	class precision
pred. Gangguan Interaksi Sos.	1	0	0	100.00%
pred. Gangguan Komunikasi	0	2	0	100.00%
pred. Gangguan Perilaku	1	0	2	66.67%
class recall	50.00%	100.00%	100.00%	

Gambar 2 Hasil Akurasi Dan Presisi
 Sumber: Hasil *Rapid Miner*

D. Hasil Pengujian Naïve Bayes

Pada hasil pengujian ini yang akan perlihatkan adalah hasil terbaik dari *cross validation* sebelum menggunakan teknik *Ensemble*, didapatkan akurasi terbaik pada *k-fold* 6 percobaan ke 4 sebesar 83,33%, dengan presisi dimana nilai positif gangguan interaksi sosial memiliki hasil 100%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 50%.

accuracy: 83.33%

	true Gangguan Interaksi Sosial	true Gangguan Komunikasi	true Gangguan Perilaku	class precision
pred. Gangguan Interaksi Sos.	2	0	0	100.00%
pred. Gangguan Komunikasi	0	2	0	100.00%
pred. Gangguan Perilaku	1	0	1	50.00%
class recall	66.67%	100.00%	100.00%	

Gambar 3 Hasil Akurasi Dan Presisi
 Sumber: Hasil *Rapid Miner*

E. Hasil Pengujian K-nearest Neighbor

Pada hasil pengujian ini yang akan perlihatkan adalah hasil terbaik dari *cross validation* sebelum menggunakan teknik *Ensemble*, didapatkan akurasi terbaik pada *k-fold* 6 percobaan ke 6 sebesar 83,33%, dengan presisi dimana nilai gangguan interaksi sosial

memiliki hasil 100%, nilai positif pada gangguan komunikasi sebesar 66,67%, dan pada nilai positif gangguan memiliki perilaku presisi 100%.

accuracy: 83,33%

	true Gangguan Perilaku	true Gangguan Interaksi Sosial	true Gangguan Komunikasi	class precision
pred. Gangguan Perilaku	2	0	0	100.00%
pred. Gangguan Interaksi Sos...	0	1	0	100.00%
pred. Gangguan Komunikasi	1	0	2	66.67%
class recall	66.67%	100.00%	100.00%	

Gambar 4 Hasil Akurasi Dan Presisi

Sumber: Hasil *RapidMiner*

F. Pembahasan

Hasil dari beberapa pengujian k-fold diperoleh presentase akurasi dan presisi dimana P adalah presisi dan A adalah akurasi, berikut merupakan gambaran dari tabel daftar pengujian hasil akurasi dan presisi.

Gambar 5 Hasil Akurasi Dan Presisi

Pengujian Ke-	K=2				K=3				K=4				K=6			
	A		P		A		P		A		P		A		P	
	GIS	GK	GP	GIS	GK	GP	GIS	GK	GP	GIS	GK	GP	GIS	GK	GP	
1	73,6 88%	75%	100 57,1 4%	75,0 7%	66,6 7%	100 7%	66,6 7%	66,6 7%	66,6 7%	100 3%	33,3 3%	66,6 60%	100 60%	100 50%	33,3 3%	
2	83,3 3%	85,71 %	100 75%	76,9 2%	100 %	100 %	50% %	66,6 7%	75% %	0% %	60% %	60% %	50% %	100 %	50% %	
3				66,6 7%	75% 7%	66,6 7%	100 %	88,8 9%	66,6 7%	100 %	100 %	66,6 7%	100 %	0% %	50% %	
4								66,6 7%	66,6 7%	60% %	100 %	66,6 7%	50% %	100 %	100 %	
5												83,3 3%	100 %	0,00 %	66,6 7%	
6												50% 50%	50% 50%	50% 50%	0% 0%	

Sumber: Hasil Perhitungan

5. Kesimpulan dan Saran

A. Kesimpulan

Dari hasil penelitian diatas maka dapat disimpulkan sebagai berikut:

1. Akurasi yang didapatkan sebesar 88,89% dan presisi yang didapatkan sebanyak 3 dimana nilai positif gangguan interaksi sosial memiliki hasil 100%, lalu nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan perilaku memiliki presisi 66,67%.
2. Perhitungan sesudah menggunakan teknik *Majority Vote* didapatkan akurasi yang sama pada k-fold 4 percobaan ke-3 dengan akurasi sebesar 88,89%. Dimana nilai positif gangguan interaksi sosial

memiliki hasil 100%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan perilaku memiliki presisi 66,67%. Dengan demikian pada penelitian yang telah dilakukan mendapatkan peningkatan akurasi sebesar 5,56%.

3. Ensemble method lebih baik dalam meningkatkan akurasi dan presisi pada teknik *Majority Vote* dibandingkan dengan menggunakan metode algoritma *C4.5*, *Naive Bayes* dan *K-nearest Neighbor* yang telah dibuktikan pada penelitian yang telah dilakukan.

B. Saran

Saran untuk penelitian berikutnya adalah:

1. Pada penelitian berikutnya dapat dikembangkan agar menggunakan platform web atau menggunakan tools *Rstudio*, *Weka* agar dapat berkembang dengan cara berbeda dibandingkan penelitian yang telah dilakukan.
2. Dikarenakan atribut yang terlalu banyak pada penelitian berikutnya dapat digunakan metode seleksi fitur seperti *genetic algorithm* atau *particle swarm optimization*. Lalu juga dapat melakukan penelitian berikutnya dengan metode *ensemble* lainnya. Contohnya *Bagging* dan *Random Forest*.

Daftar Pustaka

- Budhiman, M. 2002. *Autistic spectrum disorder*. Jakarta: Yayasan Autisma Indonesia.
- David Hartanto, Seng Hansun, 2014. Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa, *ULTIMATICS, Vol. VI, No.1 Juni 2014, ISSN: 2085-4552*.
- Fitroh, I. 2015. *Metode K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk peramalan kepadatan arus lalu lintas*. Semarang: Pasca Sarjana Universitas Dian Nuswantoro.
- Gorunescu, F. 2011. *Data Mining Concept Model and Techniques*. Berlin: Springer. ISBN 978-3-642-19720-8.

- Jojor Jennifer Sianipar, M. F. 2012. *Identifikasi Diagnosis Gangguan Autisme Pada Anak Menggunakan Metode Modified K-Nearest Neighbor MKNN*.
- Kusnawi, 2007. *Pengantar Solusi Data Mining*. Yogyakarta : STMIK AMIKOM.
- Munawaroh, Munjiati. 2013. *Manajemen Operasi*. Yogyakarta. LP3M UMY.
- Rosandy, T. 2016. *Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree Untuk Menganalisa Kelancaran Pembiayaan*. *Jurnal TIM Darmajaya*, 02(01), 52–62.
- Saleh, Ahmad. 2015. *Klasifikasi Gejala Depresi Pada Manusia dengan Metode Naive Bayes Menggunakan Java*, Yogyakarta.
- Sunu, Christoper. 2012. *Panduan memecahkan masalah autism unlocking autism*. Yogyakarta, Lintangterbit.
- Sugara, B., Widyatmoko, D., Prakoso, B. S., & Saputro, D. M. 2018. *Penerapan algoritma c4.5 untuk deteksi dini gangguan autisme pada anak*. 2018(Sentika).
- Tan, Et Al. 2007. *Intellectual capital and financial returns of companies*. *Journal of Intellectual Capital* Vol. 8 No. 1, 2007 pp. 76-95.
- YongZhang, H. 2014. *A Weighted Voting Classifier Based on Differential Evolutio*