

PERBANDINGAN AKURASI ALGORITMA C4.5 DAN NAÏVE BAYES UNTUK DETEKSI DINI GANGGUAN AUTISME PADA ANAK

Bayu Sugara¹, Dedi Adidarma², Sularso Budilaksono³
^{1,2,3}Program Studi Magister Ilmu Komputer, STMIK Nusa Mandiri Jakarta
Jalan Kramat Raya No.18, Jakarta Pusat 10420
Telp. (021) 31908575
email: bayusugaraa@gmail.com¹, dediaridarma8@gmail.com², sularso@upi-yai.co.id³

ABSTRAK

Deteksi dini gangguan autisme merupakan hal yang terpenting dalam proses tumbuh kembang anak. Namun, kebanyakan orang tua masih belum memahami tentang gangguan autisme yang dialami oleh anaknya dan bagaimana cara penanganannya. Para orang tua lebih memilih untuk mencari dan mendatangi seorang dokter atau pakar tumbuh kembang anak untuk berkonsultasi. Tetapi, ketersediaan dokter ataupun pakar tumbuh kembang anak sekarang tidaklah terlalu banyak sehingga para orang tua harus mengeluarkan biaya yang relatif banyak hanya untuk berkonsultasi kepada seorang dokter ataupun pakar. Oleh karena itu diperlukan cara atau metode untuk memudahkan para orang tua dalam mendeteksi dini gangguan autisme yang dialami oleh anaknya. Perbandingan kinerja Algoritma C4.5 dan Naïve Bayes bertujuan untuk mengukur tingkat akurasi terbaik dari masing-masing algoritma untuk diterapkan dalam deteksi dini gangguan autisme pada anak. Hasil dari komparasi Algoritma C4.5 dan Naïve Bayes menunjukkan bahwa dengan menggunakan Algoritma C4.5 menghasilkan nilai akurasi sebesar 72% sedangkan menggunakan Algoritma Naïve Bayes menunjukkan hasil 73.33%.

Kata Kunci: Gangguan Autisme, Data Mining, Algoritma C4.5, Algoritma Naive Bayes, Tingkat Akurasi

ABSTRACT

Early detection of autism disorders is the most important thing in the process of child development. However, most parents still do not understand about autism disorders experienced by their children and how to handle them. Parents prefer to look for and visit a doctor or child development expert to consult. However, the availability of doctors or child development experts now is not too much so that parents have to spend relatively a lot of money just to consult a doctor or expert. Therefore, a method or method is needed to make it easier for parents to detect early autism disorders experienced by their children. The performance comparison of Algorithms C4.5 and Naïve Bayes aims to measure the best level of accuracy of each algorithm to be applied in the early detection of autism disorders in children. The results of the comparison of the Algorithms C4.5 and Naïve Bayes show that using the C4.5 Algorithm produces an accuracy value of 72% while using the Naïve Bayes Algorithm shows 73.33%.

Keywords: Autism Disorder, Data Mining, C4.5 Algorithm, Naive Bayes Algorithm, Accuracy Level

1. PENDAHULUAN

1.1. Latar Belakang

Deteksi dini gangguan autisme merupakan hal yang terpenting dalam proses tumbuh kembang anak. Namun, kebanyakan orang tua masih belum memahami tentang gangguan autisme yang dialami

oleh anaknya dan bagaimana cara penanganannya. Para orang tua lebih memilih mencari seorang dokter ataupun datang langsung ke rumah sakit hanya untuk berkonsultasi. Tetapi, ketersediaan dokter dalam bidang tumbuh kembang anak sekarang tidaklah terlalu banyak sehingga para orang tua harus mengeluarkan biaya yang relatif

besar hanya untuk berkonsultasi kepada seorang dokter.

Seiring dengan berkembangnya teknologi saat ini, terutama dalam penggunaan internet. Internet sudah tidak bisa lagi dipisahkan dengan kehidupan masyarakat saat ini. Masyarakat dapat dengan mudahnya mengakses internet dimanapun dan kapanpun. Hal ini mendorong para seorang dokter untuk mengembangkan internet guna mendukung dan mempermudah pekerjaan mereka, khususnya dalam mendiagnosa/mendeteksi suatu penyakit. Berdasarkan penelitian sebelumnya yang menggunakan algoritma C4.5 dan Naïve Bayes untuk mendeteksi suatu penyakit menghasilkan nilai akurasi yang tinggi, maka teknik *data mining*

1.2. Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan diatas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

- a. Bagaimana memberi pengetahuan terhadap para orang tua mengenai gangguan autisme dan mempermudah para orang tua mendeteksi gangguan autisme berdasarkan gejala-gejala yang dialami oleh anaknya.
- b. Bagaimana memanfaatkan teknik *data mining* dengan menggunakan algoritma C4.5 dan Naïve Bayes untuk mendeteksi dini gangguan autisme.

1.3. Tujuan

Adapun tujuan dicapai dalam penelitian pendeteksian dini gangguan autisme pada anak ini adalah sebagai berikut:

- a. Memberikan pengetahuan kepada orang tua mengenai gangguan autisme.
- b. Mengimplementasikan Algoritma C4.5 dan Naïve Bayes untuk mendeteksi dini gangguan autisme dengan berdasarkan gejala-gejala yang dialami.
- c. Membandingkan nilai akurasi dari Algoritma C4.5 dan Naïve Bayes.

2. METODOLOGI

Metode penelitian yang digunakan dalam penelitian ini adalah metode pengumpulan data. Metode pengumpulan data dapat dijelaskan sebagai berikut:

- a. Observasi/Pengamatan
Observasi adalah suatu cara pengumpulan data dengan pengamatan langsung dan pencatatan secara sistematis terhadap objek yang akan diteliti. Observasi dilakukan dalam penelitian ini dengan cara mendatangi langsung lokasi lembaga yang menangani anak berkebutuhan khusus atau autisme yang terletak di Bekasi.
- b. Wawancara

dengan metode algoritma C4.5 dan Naïve Bayes ini digunakan dalam penelitian pendeteksian dini gangguan autisme.

Hasil dari metode algoritma C4.5 dan Naïve Bayes serta klasifikasi yang digunakan dalam penelitian ini menunjukkan dengan menggunakan Algoritma C4.5 menghasilkan nilai akurasi sebesar 72% sedangkan menggunakan Algoritma Naïve Bayes menunjukkan nilai akurasi sebesar 73.33%. Penelitian ini membuat para orang tua dapat dengan mudahnya melakukan pendeteksian gangguan autisme pada anaknya dengan pengetahuan yang diberikan berdasarkan gejala-gejala yang dialaminya dan dengan cepat bagaimana menangani gangguan autisme yang dialami oleh anaknya.

Wawancara dilakukan kepada para pakar/dokter yang ahli dalam bidang autisme dan/atau perkembangan anak guna mendapatkan informasi mengenai gangguan dan gejala-gejala autisme.

- c. Studi Pustaka
Studi Pustaka ini dilakukan cara pengumpulan data atau informasi yang berkaitan dengan penelitian ini maupun penelitian sebelumnya. Metode ini bersumber dari jurnal-jurnal penelitian sebelumnya, paper, artikel, buku serta mencari referensi materi pendukung lainnya menggunakan *internet*.
- d. Kuesioner
Penelitian ini juga dilakukan dengan cara membuat kuesioner yang dibagikan kepada 50 Orang tua. Kuesioner ini bertujuan untuk mengetahui gangguan autisme yang dialami oleh anaknya dengan mengisi kuesioner yang telah disediakan.

3. LANDASAN TEORI

Dalam penelitian ini, menggunakan tinjauan pustaka didasarkan pada beberapa jurnal dan paper. Diantaranya sebagai berikut:

3.1.1. Penelitian Terdahulu

Dalam penelitian yang dilakukan oleh Rahmawati yang berjudul Analisa Komparasi Algoritma Naïve Bayes dan C4.5 Untuk Prediksi Penyakit *Liver* mengemukakan bahwa penelitian yang dilakukan pembuatan model menggunakan algoritma Naïve bayes dan C4.5 menggunakan data pasien penderita *Liver*. Model yang dihasilkan, dikomparasi untuk mengetahui algoritma yang paling baik dalam penentuan identifikasifikasi penyakit *Liver*. Untuk mengukur kinerja kedua algoritma tersebut digunakan metode pengujian *Cross Validation* dan *Split Percentace*, dan pengukurannya dengan menggunakan *confusion matrix*. Dengan demikian algoritma C4.5 dapat

memberikan pemecahan untuk permasalahan dalam mengidentifikasi penyakit *Liver*. (Rahmawati, 2015).

Sedangkan pada penelitian dikemukakan bahwa Penggunaan Metode *Business Intelligence* diterapkan dalam melakukan pendeteksian dini untuk mengenali anak berkebutuhan khusus. Dengan metode klasifikasi, atribut yang paling berpengaruh berhasil diperoleh, yaitu kemampuan belajar. Dan hasil perbandingan untuk kedua algoritma C4.5 dan Naïve Bayes menunjukkan kedua algoritma sama-sama memberikan hasil yang sangat baik/ tidak ada perbedaan yang signifikan. (Grand dan Richardus, 2017).

Jayawardanu dan Hansun mengemukakan pendapat bahwa hal yang dilakukan untuk meminimalisir ketidaktahuan tersebut adalah dengan membangun sebuah sistem yang dapat mendeteksi katarak secara dini tanpa harus mengunjungi klinik mata. Sistem ini dibangun berdasarkan *website*, dengan mengimplementasikan metode *learning decision tree* dengan algoritma C4.5. Algoritma C4.5 dipilih karena dinyatakan sebagai algoritma *learning* yang memiliki tingkat keakuratan paling tinggi.

3.1.2. Autisme

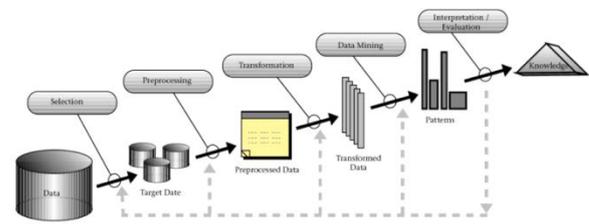
Menurut pendapat Tutik dkk (2009), Autisme merupakan gangguan perkembangan mental pada anak yang menyebabkan seorang anak sulit untuk berinteraksi sosial. Diagnosa autisme biasanya dilakukan oleh seorang pakar/ahli dibidang tumbuh kembang anak. Namun sebenarnya, orang tua juga dapat melakukan pengamatan perilaku anak dalam kesehariannya terutama dari cara berkomunikasi, berinteraksi sosial dengan anak sebayanya dan kemampuan berimajinasi pada anak.

3.1.3. Data Mining

Data mining adalah proses menemukan pengetahuan yang menarik, seperti asosiasi, pola, perubahan, struktur yang signifikan dan anomali, dari sejumlah besar data yang disimpan dalam *database* atau gudang data atau repositori informasi lainnya. Telah banyak digunakan dalam beberapa tahun terakhir karena ketersediaan data dalam jumlah besar dalam bentuk elektronik, dan ada kebutuhan untuk mengubah data tersebut menjadi informasi yang berguna dan pengetahuan untuk aplikasi yang besar. Aplikasi ini ditemukan di bidang-bidang seperti *Artificial Intelligence*, *Machine Learning*, Analisis Pasar, Statistik dan Sistem *Database*, Manajemen Bisnis dan Pendukung Keputusan.

Data mining adalah proses mempekerjakan satu atau lebih teknik *machine learning* untuk menganalisis dan mengekstraksi *knowledge* secara otomatis. Penggunaan *data mining* memiliki tujuan

untuk mengetahui pola universal dari data-data yang ada. Dalam menghasilkan suatu *knowledge* dari pola yang ada, diperlukan penerapan metode *scientific* yang disebut dengan *Knowledge Discovery in*



Database (KDD). (Jayawardanu dan Hansun, 2015:2).

Gambar 1. Tahapan proses KDD (Jayawardanu dan Hansun, 2015)

Terdapat 5 (lima) Tahapan pada proses *Knowledge Discovery in Database* (KDD) dapat dijelaskan sebagai berikut:

1. *Selection*
Pertama, lakukan seleksi data dari *multiple data source* dengan membersihkan data yang memiliki *noise* ataupun *missing value*.
2. *Preprocessing*
Kedua, lakukan penggabungan terhadap seluruh *data source* yang telah terkumpul sebelum tahapan dari proses selanjutnya dilanjutkan.
3. *Transformation*
Ketiga, lakukan transformasi data ke dalam bentuk yang lebih sesuai untuk dilakukan *data mining*.
4. *Data Mining*
Keempat, terapkan metode *data mining* pada sistem, sehingga dapat menghasilkan pola dari data yang terkumpul.
5. *Interpretation/Evaluation*
Kelima, lakukan interpretasi dan evaluasi dari pola yang didapatkan, sehingga dapat diidentifikasi apakah pola tersebut sudah dapat mewakili *knowledge* yang ingin dicapai.

Dalam proses KDD, metode *data mining* tersebut untuk mengekstraksi pola dari data. Pola itu bisa jadi ditemukan tergantung pada tugas *data mining* yang diterapkan. Umumnya ada dua jenis tugas penambangan data yaitu Tugas penambangan data deskriptif yang menggambarkan umum sifat data yang ada, dan data mining yang prediktif tugas yang berusaha melakukan prediksi berdasarkan data yang ada. *Data mining* dapat dilakukan pada data yang bersifat kuantitatif, tekstual, atau multimedia.

Aplikasi *data mining* bisa menggunakan jenis yang berbeda parameter untuk memeriksa data. Mereka termasuk asosiasi (pola dimana satu acara terhubung ke acara lain), analisis urutan atau jalur (pola di mana satu peristiwa mengarah ke acara

lain), klasifikasi (identifikasi pola baru dengan target yang telah ditetapkan) dan pengelompokan (pengelompokan benda identik atau sejenis).

Penambangan data dalam *data mining* dapat melibatkan beberapa langkah-langkah kunci yang akan dijelaskan sebagai berikut:

1. *Problem definition*

Langkah pertama adalah mengidentifikasi tujuan. Berdasarkan tujuan yang telah ditetapkan, rangkaian alat yang benar bisa jadi diterapkan pada data untuk membangun perilaku yang sesuai model.

2. *Data exploration*

Jika kualitas data tidak sesuai untuk model yang akurat maka rekomendasi pada data masa depan. Strategi pengumpulan dan penyimpanan bisa dilakukan pada saat ini. Untuk analisis, semua data perlu dikonsolidasikan sehingga bisa jadi diperlakukan secara konsisten.

3. *Data preparation*

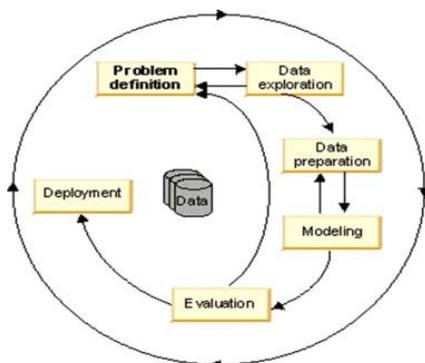
Tujuan dari langkah ini adalah untuk membersihkan dan ubah data sehingga hilang dan nilai yang tidak valid diperlakukan dan semua nilai valid yang diketahui dibuat konsisten untuk analisis yang lebih kuat.

4. *Modeling*

Berdasarkan data dan yang diinginkan hasil, algoritma data mining atau kombinasi Algoritma dipilih untuk analisis. Algoritma ini termasuk teknik klasik seperti statistik, lingkungan dan *clustering* tapi juga generasi penerus teknik seperti pohon keputusan, jaringan dan algoritma berbasis aturan. Algoritma spesifik dipilih berdasarkan tujuan tertentu yang ingin dicapai dan kualitas data yang akan dianalisis.

5. *Evaluation and Deployment*

Berdasarkan hasil algoritma data mining, analisis dilakukan terhadap tentukan kesimpulan utama dari analisis dan buat serangkaian rekomendasi untuk dipertimbangkan.



Gambar 2. Representasi proses data mining (Krishnaiah, Narmisha dan Chandra, 2013)

3.1.4. **Algoritma C4.5**

Algoritma C4.5 adalah program yang memberi kontribusi satu set data berlabel dan menghasilkan pohon keputusan sebagai keluaran. Pohon keputusan tindak lanjut ini kemudian diverifikasi terhadap data uji berlabel yang tidak terlihat untuk menghitung generalisasinya. C4.5 adalah program yang digunakan untuk menghasilkan peraturan taksonomi dengan menggunakan pohon keputusan dari sekumpulan data yang diberikan.

Algoritma C4.5 merupakan perpanjangan dari algoritma ID3 dasar dan dirancang oleh Quinlan. C4.5 adalah salah satu algoritma pembelajaran yang banyak digunakan. Algoritma C4.5 membangun pohon keputusan dari serangkaian data pelatihan yang mirip dengan Algoritma ID3, dengan menggunakan konsep entropi informasi. C4.5 juga dikenal sebagai klasifikasi statistik. (Manikantan dan Latha, 2013:2).

Algoritma ini dapat menyelesaikan masalah secara sistematis dengan membentuk suatu *decision tree* dengan langkah-langkah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk masing-masing record dari atribut.
3. Membagi kasus ke dalam cabang.
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang menghasilkan suatu keputusan yang sesuai.

Dalam memilih sebuah atribut menjadi akar, dilakukan perhitungan nilai dari atribut yang ada. Nilai gain yang paling tinggi dijadikan root pada pohon keputusan. Untuk menghitung nilai *gain* digunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (1)$$

Keterangan:

- S : Himpunan kasus,
- A : Atribut,
- n : Jumlah partisi dalam atribut,
- |S_i| : Jumlah kasus pada partisi ke-i,
- |S| : Jumlah kasus

Sedangkan untuk perhitungan nilai *entropy* dapat dilakukan dengan rumus:

$$Entropy(S) = \sum_{i=0}^n - P_i \cdot \log_2 P_i \quad (2)$$

Keterangan:

- S : Himpunan kasus,
- n : Jumlah partisi dalam atribut,

π_i : Proposi dari S_i terhadap S

Setelah perhitungan selesai, maka dihasilkan pohon keputusan yang dapat diubah menjadi *rules* menggunakan kaidah *IF-THEN-ELSE*. Penjelasan *rules* tersebut adalah sebagai berikut:

a. *IF-THEN Rule*

Peraturan induksi: adalah proses penggalian berguna 'jika kemudian' aturan dari data berdasarkan signifikansi statistik. Sebuah sistem *Rule* berdasarkan membangun satu *set if-then*-aturan. Pengetahuan merupakan telah pengumpulan bentuk dan strategi penyimpanan dapat dibuat di ini. Untuk analisis, semua data perlu dikonsolidasikan sehingga dapat diobati secara konsisten.

b. *IF* kondisi *THEN* Kesimpulan

Aturan ini terdiri dari dua bagian. Aturan yang (bagian *IF*) mengandung satu atau lebih kondisi sekitar nilai prediktor atribut mana sebagai akibat aturan (*THEN* bagian) berisi prediksi tentang nilai atribut tujuan. Prediksi yang akurat dari nilai atribut tujuan akan meningkatkan proses pengambilan keputusan. *IF-THEN* aturan prediksi sangat populer di data mining. Mereka mewakili menemukan pengetahuan pada tingkat tinggi abstraksi. Aturan Induksi Metode memiliki potensi untuk menggunakan kasus diambil untuk prediksi.

3.1.5. Pohon Keputusan (*Decision Tree*)

Pohon keputusan ini adalah struktur representasi pengetahuan yang terdiri dari simpul dan cabang yang diorganisasikan dalam bentuk pohon semacam itu bahwa setiap simpul non-daun internal diberi label dengan nilai atributnya. Cabang-cabangnya keluar dari internal node diberi *label* dengan nilai atribut pada *node* tersebut. Setiap simpul diberi *label* dengan kelas (nilai tujuan atribut). Model berbasis pohon yang meliputi klasifikasi dan regresi pohon, adalah implementasi umum dari pemodelan induksi. Model pohon keputusan yang terbaik cocok untuk data mining. Mereka murah untuk membangun, mudah ditafsirkan, mudah diintegrasikan dengan sistem *database* dan mereka memiliki akurasi yang sebanding atau lebih baik aplikasi. Ada banyak algoritma pohon keputusan seperti itu sebagai algoritma HUNTS (ini adalah salah satu algoritma paling awal), CART, ID3, C4.5 (algoritma ID3 versi selanjutnya), SLIQ, SPRINT. (Krishnaiah, 2013:4).

Menurut pendapat Adhataro (2013), Dalam pembelajaran *decision tree*, ID3 (*Iterative Dichotomiser 3*) adalah algoritma yang diciptakan oleh Ross Quinlan biasa membuat pohon keputusan dari *dataset*. ID3 biasanya digunakan di mesin belajar dan pengolahan bahasa alami. Teknik pengambilan keputusan melibatkan membangun pohon untuk memodelkan proses klasifikasi. Begitu

pohon dibangun, itu diterapkan pada masing-masing tupel di database dan hasilkan klasifikasi untuk tuple tersebut. Permasalahan yang dihadapi kebanyakan algoritma pohon keputusan adalah sebagai berikut:

- Memilih atribut pemisahan
- Memesan atribut pemisahan
- Jumlah perpecahan yang harus diambil
- Saldo struktur pohon dan pemangkasan
- Menghentikan kriteria

3.1.6. Naïve Bayes

Naïve Bayes merupakan metode yang tidak memiliki aturan, Naïve Bayes menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data training. Naïve bayes merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam data mining, algoritma ini juga dikenal dengan nama *Idiot's Bayes*, *Simple Bayes* dan *Independence Bayes* (Bramer, Max, 2007).

Klasifikasi Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Klasifikasi Bayesian didasarkan pada teorema Bayes, diambil dari nama seorang ahli matematika yang juga menteri *Prebysterian Inggris*, Thomas Bayes (1702-1761) (Larose, 2015). Klasifikasi Bayesian memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* (Maimon, Oded and Rokach, Lior, 2010).

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (3)$$

Keterangan:

y = data dengan kelas yang belum diketahui

x = hipotesis data y merupakan suatu kelas spesifik

$P(x|y)$ = probabilitas hipotesis x berdasar kondisi y (*posteriori probability*)

$P(x)$ = probabilitas hipotesis x (*prior probability*)

$P(y|x)$ = probabilitas y berdasarkan kondisi pada hipotesis x

$P(y)$ = probabilitas dari y

4. HASIL DAN PEMBAHASAN

4.1. Basis Pengetahuan

Pada basis pengetahuan, akan dijelaskan tabel gangguan autisme, tabel gejala autisme, tabel keputusan autisme, *rule-rule* dan pohon keputusan. Yang akan dijelaskan secara rinci sebagai berikut:

4.1.1. Tabel Gangguan Autisme

Tabel Gangguan Autisme akan menampilkan gangguan-gangguan autisme/perkembangan yang dialami pada anak.

Tabel 1. Tabel gangguan autisme

Kode Gangguan	Nama Gangguan
GG01	Gangguan Perilaku
GG02	Gangguan Komunikasi
GG03	Gangguan Interaksi Sosial

4.1.2. Tabel Gejala Autisme

Tabel Gejala Autisme akan menampilkan data gejala-gejala autisme/perkembangan yang dialami pada anak.

Tabel 2. Tabel gejala autisme

Kode Gejala	Nama Gejala
GJ01	Tidak memiliki kontak mata
GJ02	Suka diam/menyendiri
GJ03	Tidak suka dipeluk
GJ04	Tidak dapat merespon jika dipanggil orang
GJ05	Suka melakukan kegiatan/gerakan secara berulang-ulang
GJ06	Suka terpaku terhadap benda-benda tertentu
GJ07	Suka menyukai hal yang aneh seperti mencium-cium benda
GJ08	Suka mengungkapkan emosi (sedih, senang, marah dll) dengan sendirinya tanpa sebab
GJ09	Tidak bisa diam
GJ10	Tidak dapat berbicara
GJ11	Bisa berbicara namun tidak jelas
GJ12	Sering berbicara berlebihan
GJ13	Suka mengucapkan bahasa/kata-kata yang aneh secara berulang-ulang
GJ14	Tidak dapat menunjuk sesuatu dengan jari sendiri
GJ15	Tidak dapat menunjukkan keinginan dengan kata-kata
GJ16	Suka menarik-narik orang lain jika menginginkan sesuatu
GJ17	Tidak ada usaha dalam berkomunikasi
GJ18	Menghindar jika didekati
GJ19	Tidak dapat berinteraksi dengan lingkungan sekitar
GJ20	Tidak tertarik dengan orang lain
GJ21	Tidak peduli dengan sekitarnya
GJ22	Tidak suka dengan keramaian
GJ23	Tidak suka bermain dengan teman sebayanya
GJ24	Tidak dapat bersosialisasi dengan orang lain

4.1.3. Tabel Keputusan Gangguan Autisme

Tabel keputusan gangguan ini menampilkan relasi antara gangguan autisme dengan gejala autismenya.

Tabel 3. Tabel keputusan gangguan autisme

Gejala	Gangguan Autisme		
	GG01	GG02	GG03
GJ01	Ya	Ya	Ya
GJ02	Ya	Tidak	Ya
GJ03	Ya	Tidak	Ya
GJ04	Tidak	Ya	Tidak
GJ05	Ya	Tidak	Tidak
GJ06	Ya	Tidak	Tidak
GJ07	Ya	Tidak	Tidak
GJ08	Ya	Tidak	Tidak
GJ09	Ya	Tidak	Tidak
GJ10	Tidak	Ya	Tidak
GJ11	Tidak	Ya	Tidak
GJ12	Tidak	Ya	Tidak
GJ13	Tidak	Ya	Tidak
GJ14	Tidak	Ya	Tidak
GJ15	Tidak	Ya	Tidak
GJ16	Tidak	Ya	Tidak
GJ17	Tidak	Ya	Tidak
GJ18	Tidak	Tidak	Ya
GJ19	Tidak	Tidak	Ya
GJ20	Tidak	Tidak	Ya
GJ21	Tidak	Tidak	Ya
GJ22	Tidak	Tidak	Ya
GJ23	Tidak	Tidak	Ya
GJ24	Tidak	Tidak	Ya

4.2. Algoritma C4.5

Pada tahapan ini dilakukan analisa perhitungan terhadap kasus permasalahan, dalam penelitian ini menggunakan Algoritma C4.5. Akan dijelaskan secara rinci detail Algoritma C4.5 menjadi analisa perhitungan, rule-rule yang dihasilkan dan gambaran pohon keputusan (*decision tree*).

4.2.1. Analisa Perhitungan

Dalam penelitian ini akan dilakukan eksperimen dengan menggunakan metode klasifikasi *data mining* Algoritma C4.5 dan *Decision Tree* terhadap data anak dengan gejala-gejala yang dialaminya. Data akan diolah dengan menggunakan algoritma C4.5 dan menghasilkan model, maka terhadap model yang dihasilkan tersebut dilakukan pengujian menggunakan *k-fold cross validation*.

	Ya	14	0	14	0	0	
	Tidak	36	18	6	12	1.45914792	
GJ05							0.35357947
	Ya	16	13	3	0	0.69621226	
	Tidak	34	5	17	12	1.43699052	

Dari hasil perhitungan nilai *entropy* dan nilai *gain* yang terdapat pada tabel 5 terlihat bahwa atribut GJ04 (Tidak dapat merespon jika dipanggil orang) mempunyai nilai *gain* paling tinggi yaitu 0.5029344. Oleh karena itu GJ04 (Tidak dapat merespon jika dipanggil orang) akan menjadi akar (node pertama) dari pohon keputusan yang terbentuk. Setelah didapatkan nilai *gain* yang tertinggi maka selanjutnya dihitung dengan cara yang sama dengan menggunakan persamaan *entropy* untuk mendapatkan nilai *entropy* dan persamaan *gain* untuk mendapatkan nilai *gain*.

Setelah dilakukan hasil perhitungan nilai *entropy* dan nilai *gain*, maka akan terbentuk pohon keputusan seperti gambar 4

Gambar 4. Pohon keputusan (*Decision Tree*)

Setelah pohon keputusan (*decision tree*) terbentuk, selanjutnya dilakukan perubahan menjadi *rule*, berikut ini merupakan pembentukan *rule tree*.

Rule 1

IF GJ04 == Ya
THEN Gangguan Komunikasi

Rule 2

IF GJ04 == Tidak
AND GJ05 == Ya
THEN Gangguan Perilaku

Rule 3

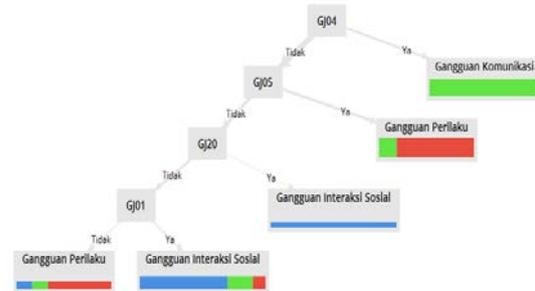
IF GJ04 == Tidak
AND GJ05 == Tidak
AND GJ20 == Ya
THEN Gangguan Interaksi Sosial

Rule 4

IF GJ04 == Tidak
AND GJ05 == Tidak
AND GJ20 == Tidak
AND GJ01 == Ya
THEN Gangguan Interaksi Sosial

Rule 5

IF GJ04 == Tidak
AND GJ05 == Tidak
AND GJ20 == Tidak
AND GJ01 == Tidak
THEN Gangguan Perilaku



Model yang telah dibentuk diuji tingkat akurasi dengan memasukan atau uji yang berasal dari data training dengan menggunakan *cross validation* pada aplikasi *ripedminer* untuk menguji tingkat akurasi. Dengan aplikasi *ripedminer* dihasilkan nilai akurasi, nilai *class recall* dan nilai *class precision* yang akan dijelaskan pada gambar dibawah ini:

Gambar 5. Nilai akurasi perhitungan Algoritma C4.5

Dari gambar 5 diatas dijelaskan bahwa gangguan interaksi sosial menghasilkan *class recall* sebesar

accuracy: 72.00% + 20.40% (mikro: 72.00%)

	true Gangguan Interaksi So...	true Gangguan Komunikasi	true Gangguan Perilaku	class precision
pred. Gangguan Interaksi S...	8	3	2	61.54%
pred. Gangguan Komunikasi	0	13	1	92.86%
pred. Gangguan Perilaku	4	4	15	65.22%
class recall	66.67%	65.00%	83.33%	

66,67% dan *class precision* sebesar 61.54%, sedangkan gangguan komunikasi menghasilkan *class recall* sebesar 65.00% dan *class precision* sebesar 92.86% serta gangguan perilaku menghasilkan *class recall* sebesar 83.33% dan *class precision* sebesar 65.22% dan nilai akurasi dari perhitungan algoritma C4.5 tersebut diatas adalah sebesar 72%.

4.3. Naïve Bayes

Pada tahapan ini dilakukan analisa perhitungan terhadap kasus permasalahan, dalam penelitian ini menggunakan Naïve Bayes. Akan dijelaskan secara rinci detail Naïve Bayes sebagai berikut:

P TOTAL	Gangguan Perilaku	Gangguan Komunikasi	Gangguan Interaksi Sosial
	36%	40%	24%
	18	20	12

P(GJ01=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	50%	75%	75%
Tidak	50%	25%	25%

P(GJ02=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	100%	40%	100%
Tidak	0%	60%	0%

P(GJ03=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	100%	35%	100%
Tidak	0%	65%	0%

P(GJ04=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	0%	70%	0%
Tidak	100%	30%	100%

P(GJ05=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	72%	15%	0%
Tidak	28%	85%	100%

P(GJ06=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	67%	15%	0%
Tidak	33%	85%	100%

P(GJ07=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	72%	20%	8%
Tidak	28%	80%	92%

P(GJ08=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	22%	0%	8%
Tidak	78%	100%	92%

P(GJ09=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	72%	20%	33%
Tidak	28%	80%	67%

P(GJ10=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	0%	55%	8%
Tidak	100%	45%	92%

P(GJ11=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	33%	70%	0%
Tidak	67%	30%	100%

P(GJ12=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	44%	65%	8%
Tidak	56%	35%	92%

P(GJ13=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	0%	60%	0%
Tidak	100%	40%	100%

P(GJ14=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	22%	65%	33%
Tidak	78%	35%	67%

P(GJ15=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	6%	55%	8%
Tidak	94%	45%	92%

P(GJ16=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	17%	70%	0%
Tidak	83%	30%	100%

P(GJ17=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	0%	45%	0%
Tidak	100%	55%	100%

P(GJ18=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	39%	20%	92%
Tidak	61%	80%	8%

P(GJ19=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	33%	20%	83%
Tidak	67%	80%	17%

P(GJ20=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	22%	5%	33%
Tidak	78%	95%	67%

P(GJ21=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	50%	25%	83%
Tidak	50%	75%	17%

P(GJ22=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	56%	25%	67%
Tidak	44%	75%	33%

P(GJ23=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	33%	15%	75%
Tidak	67%	85%	25%

P(GJ24=↓ ... C=Perilaku) ... C=Komunikasi) ... C=Sosial)
Ya	28%	20%	75%
Tidak	72%	80%	25%

Gambar 6. Perhitungan Algoritma Naïve Bayes

Model yang telah dibentuk diuji tingkat akurasi dengan memasukan atau uji yang berasal dari data training dengan menggunakan *cross validation* pada aplikasi *ripedminer* untuk menguji tingkat akurasi. Dengan aplikasi *ripedminer* dihasilkan nilai akurasi, nilai *class recall* dan nilai *class precision* yang akan dijelaskan pada gambar dibawah ini:

accuracy: 73.33%

	true Gangguan Interak...	true Gangguan Komuni...	true Gangguan Perilaku	class precision
pred. Gangguan Intera...	4	1	0	80.00%
pred. Gangguan Komu...	0	4	0	100.00%
pred. Gangguan Perilaku	0	3	3	50.00%
class recall	100.00%	50.00%	100.00%	

Gambar 7. Nilai akurasi perhitungan Naïve Bayes

Dari gambar 7 diatas dijelaskan bahwa gangguan interaksi sosial menghasilkan *class recall* sebesar 100% dan *class precision* sebesar 80%, sedangkan gangguan komunikasi menghasilkan *class recall* sebesar 50.00% dan *class precision* sebesar 100% serta gangguan perilaku menghasilkan *class recall* sebesar 100% dan *class precision* sebesar 50% dan nilai akurasi dari perhitungan algoritma Naive Bayes tersebut diatas adalah sebesar 73.33%.

5. KESIMPULAN

5.1. Kesimpulan

Dari hasil penelitian yang sudah dilakukan, dapat disimpulkan bahwa deteksi dini gangguan autisme pada anak dapat diprediksi dengan memanfaatkan teknik *data mining* dengan menggunakan metode algoritma C4.5 dan Naive Bayes untuk memprediksi gangguan autisme pada anak berdasarkan gejala yang dialami anak. Setelah

dilakukan pengujian dengan metode *cross validation* pada aplikasi *ripedminer*, Algoritma C4.5 menunjukkan hasil nilai akurasi sebesar 72% sedangkan Algoritma Naïve Bayes menunjukkan nilai akurasi yang lebih besar dan lebih baik yaitu sebesar 73.33%.

DAFTAR PUSTAKA

- Aanchal Oswal, Vachana Shetty, Mustafa Badshah, Rohit Pitre, Manali Vashi. 2014. *A Survey On Disease Diagnosis Algorithms. International Journal of Advanced Re-search in Computer Engineering & Technology (IJAR-CET)*. Volume 3 Issue 11. 3757-3761.
- Akila and S. Chandramathi. 2015. *A Hybrid Method for Coro-nary Heart Disease Risk Prediction using Decision Tree and Multi Layer Perceptron. Indian Journal of Science and Technology*. Vol 8. 1-7.
- Badr Hssina, Abdelkarim Merbouha, Hanane Ez-Zikouri, Mohammed Erritali. *A comparative study of decision tree ID3 and C4.5. International Journal of Ad-vanced Computer Science and Applications*. 13-19.
- Grand dan Richardus Eko Indrajit. 2017. Aplikasi Deteksi Dini Untuk Mengenali Anak Berkebutuhan Khusus Menggunakan Metode *Business Intelligence*. Seminar Nasional Sains dan Teknologi 2017, Fakultas Teknik Universitas Muhammadiyah Jakarta. ISSN: 2407-1846. 1-11.
- Jayawardanu, Ivana Herliana and Seng Hansun. 2015. Rancang Bangun Sistem Pakar Untuk Deteksi Dini Katarak Menggunakan Algoritma C4.5. *Ultima Computing*. Vol. VII, No.2, ISSN: 2355-3286. 48-58.
- Krishnaiah, V. Dr.G.Narsimha, Dr.N.Subhash Chandra. 2013. *Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques. International Journal of Computer Science and Information Technologies*. Vol 1. 39-45.
- Kalpesh Adhataro, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao. 2013. *Predicting Students Performance Using ID3 and C4.5 Classification Algorithms. International Journal of Data Mining & Knowledge Management Process (IJDMP)*, Vol.3, No.5. 39-52.
- Manikantan, V and S. Latha. 2013. *Predicting The Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods. International Journal on Advanced Computer Theory and Engineering (IJACTE)*. Vol.2, Issue-2, ISSN: 2319-2526. 5-10.
- Mirajkar, Pallavi and Dr. G. Prasanna Laksmi. 2017. *Prediction of Cancer Risk in Perspective of Symptoms using Naïve Bayes Classifier. International Journal of Engineering Re-search in Computer Science and Engineering (IJERCSE)*. Vol 4, Issue 9. ISSN 2394-2320. 145-149.
- Rahmawati, Eva. 2015. Analisa Komparasi Algoritma Naïve Bayes dan C4.5 Untuk Prediksi Penyakit Liver. *Jurnal Techno Nusa Mandiri*. Vol X11 No.12, September 2015. 27-37.
- Tutik A, Gusti Ayu Kadek, Rosa Delima dan umi Proboyekti. 2009. Penerapan *Forward Chaining* Pada Program Diagnosa Anak Penderita Autisme. Yogyakarta. *Jurnal Informatika* Volume 5 Nomor 2 November 2009: 46-59. Diambil dari: <http://ti.ukdw.ac.id/ojs/index.php/informatika/article/viewFile/73/35>. (20 Oktober 2014)