

# PERMODELAN PREDIKTIF *AUTISTIC SPECTRUM DISORDER* DENGAN ALGORITMA C.45

Oktafian Farhan<sup>1</sup>; Agus Subekti<sup>2</sup>

<sup>1</sup>Program Pascasarjana Magister Ilmu Komputer  
STMIK Nusa Mandiri  
[www.nusamandiri.ac.id](http://www.nusamandiri.ac.id)  
[oktafian.ofn@bsi.ac.id](mailto:oktafian.ofn@bsi.ac.id)

<sup>2</sup>Pusat Penelitian Elektronika dan Telekomunikasi  
Lembaga Ilmu Pengetahuan Indonesia (LIPI)  
[www.lipi.go.id](http://www.lipi.go.id)  
[agus.subekti@lipi.go.id](mailto:agus.subekti@lipi.go.id)



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

**Abstract**—Autism is a developmental disability experienced throughout the life of a patient with Autistic Spectrum Disorder (ASD). The sooner it is handled, the more likely the child will return to normal. For this reason, a new method is needed that can help parents to quickly recognize the symptoms of autism in their children. In a previous study conducted by Fadi Fayez Tabhtah a data set was produced to detect whether a child has autism or not. But the research only produces data sets, he does not examine more in which algorithm is suitable for the data sets that have been produced. The data set attributes have some missing value, which invite a question about the accuracy of data. In this study researchers used the CRISP-DM method and test the accuracy of data sets of previous studies using the C.45 algorithm. Furthermore, the WEKA application using feature selection and influence of the missing value for each attribute and find the most significant. These attributes are then tested with the C.45 algorithm so that the predictive model of the data set is obtained. The A6 attribute of the decision tree calculation does not appear at all as a branch. A new model is obtained where the A6 attribute is omitted, so that when measured by the C.45 algorithm, a better accuracy value is obtained. The results of the new model were then tested on the new questionnaire data, which produced precise predictions.

**Keywords:** ASD, children, data set, C.45 algorithm

**Intisari**— Autisme merupakan disabilitas perkembangan yang dialami sepanjang hidup penderita Autistic Spectrum Disorder (ASD). Semakin cepat ditangani, semakin besar kemungkinan anak akan kembali normal. Untuk

alasan ini, diperlukan metode baru yang dapat membantu orang tua dengan cepat mengenali gejala autisme pada anak-anak mereka. Dalam studi sebelumnya yang dilakukan oleh Fadi Fayez Tabhtah, suatu data set dihasilkan untuk mendeteksi apakah seorang anak memiliki autisme atau tidak. Tetapi penelitiannya hanya menghasilkan data set, ia tidak memeriksa lebih lanjut dimana algoritma cocok untuk data set yang telah dihasilkan. Atribut data set ternyata memiliki nilai yang salah, yang mengundang pertanyaan tentang keakurasian data. Dalam penelitian ini peneliti menggunakan metode CRISP-DM dan menguji keakuratan data set penelitian sebelumnya menggunakan algoritma C.45. Selanjutnya, aplikasi WEKA menggunakan pemilihan fitur dan pengaruh dari nilai yang salah untuk setiap atribut dan menemukan atribut yang paling signifikan. Atribut-atribut ini kemudian diuji dengan algoritma C.45 sehingga model prediksi dari data set diperoleh. Atribut A6 dari perhitungan pohon keputusan tidak muncul sama sekali sebagai cabang. Sebuah model baru diperoleh di mana atribut A6 dihilangkan, sehingga ketika diukur oleh algoritma C.45, nilai akurasi yang lebih baik diperoleh. Hasil model baru kemudian diuji pada data kuesioner baru, yang menghasilkan prediksi yang tepat.

**Kata kunci :** ASD, anak-anak, data set, algoritma C.45

## PENDAHULUAN

Salah satu gangguan kejiwaan primer adalah Autistik Spektrum Disorder (ASD). ASD adalah gangguan mental yang membatasi penggunaan linguistik, komunikatif, kognitif, keterampilan,

serta keterampilan sosial dan kemampuan. Baru-baru ini, ASD telah dipelajari dalam perilaku sains menggunakan metode cerdas berbasis pembelajaran mesin (*machine learning*) untuk mempercepat waktu *screening* atau meningkatkan kepekaan, spesifisitas atau akurasi proses diagnosis. Pembelajaran mesin menganggap masalah diagnosis ASD sebagai tugas klasifikasi dalam model prediktif yang dibangun berdasarkan kasus historis dan kontrol. (Thabtah, 2017)

Pengkajian dari hasil penelitian yang dilakukan oleh Fadi Fayed Thabtah mengenai Dalam beberapa penelitian, sebenarnya sudah ada yang membahas mengenai metode pendeteksian secara dini bagi anak-anak yang peningkatan kinerja diagnosis dengan alat diduga menderita autisme.

Salah satu masalah penting dalam penelitian autisme saat ini adalah diagnostik yang ada sehingga individu dapat memiliki layanan yang lebih spesifik, lebih baik, dan lebih cepat sedini mungkin. Ini dapat dilakukan dengan banyak cara, seperti mengurangi waktu diagnosis secara efisien atau meningkatkan akurasi prediktif diagnosis tanpa mengorbankan validitas atau sensitivitas tes. (Thabtah, 2018)

*Autistic Spectrum Disorder Screening Data for Children* dimana diharapkan dengan kajian ulang ini diperoleh model pendeteksian dini terhadap anak yang mengidap autisme melalui metode *machine learning* dengan akurat.

Apakah setiap variabel dalam *Data set* berpengaruh dalam mendeteksi gejala autisme pada anak ?. Apakah model pola yang dihasilkan algoritma C45 dapat mendeteksi dengan akurat autisme pada anak ?. Pernyataan utama masalah di atas dijelaskan ke pertanyaan penelitian berikut:

1. Bagaimana hasil pengukuran *data set* dengan menggunakan algoritma Naive Bayes?
2. Bagaimana hasil pengukuran *data set* dengan menggunakan algoritma *Decision tree* C.45 ?
3. Bagaimana korelasi antar variabel dengan metode Pearce ?
4. Bagaimana pengaruh saat *feature selection* pada *data set* ?
5. Bagaimana model pola yang diperoleh algoritma *decision tree* C.45 ?

## BAHAN DAN METODE

### Metode *Data mining*

*Data mining* adalah proses menemukan pola yang menarik, dan pengetahuan dari data yang berjumlah besar (Han, Pei, & Kamber, 2011). Metode *Data mining* yang digunakan

model CRISP-DM, yang terdiri dari 6 fase yaitu (Shearer, 2000):

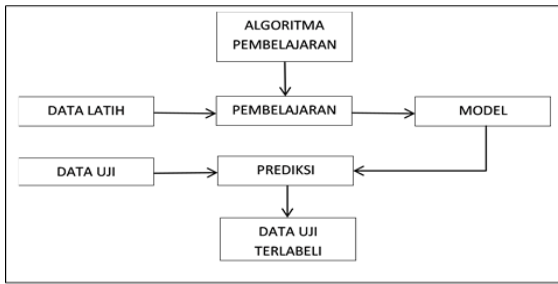
- a. *Business Understanding* atau pemahaman domain (penelitian).
- b. *Data Understanding* atau pemahaman data
- c. *Data preparation* atau persiapan data
- d. *Modeling* adalah fase menentukan teknik *data mining* yang digunakan, menentukan tools *data mining*, teknik *data mining*, algoritma *data mining*, menentukan parameter/atribut dengan nilai yang optimal.
- e. *Evaluation* adalah fase interpretasi terhadap hasil *data mining* yang ditunjukkan dalam proses pemodelan
- f. *Deployment* atau penyebaran adalah fase penyusunan laporan atau presentasi.

### Objek Penelitian

Objek dari penelitian ini adalah, anak-anak yang memiliki masalah dalam proses tumbuh kembang yang diperoleh melalui *data set*.

### Algoritma

Algoritma akan dibandingkan hasilnya untuk mendapatkan algoritma yang paling baik nilai akurasi. *Classification* adalah satu bentuk analisis data yang menghasilkan model untuk mendeskripsikan kelas data yang penting. *Classification* memprediksi kategori (*discrete, unordered*) ke dalam label *class*. (Han et al., 2011). Algoritma Klasifikasi *Decision tree* J-48 yang terdapat pada tools aplikasi WEKA. J48 merupakan implementasi dari algoritma C4.5 yang memproduksi *Decision tree* (Witten & Frank, 2005). Metode C4.5 merupakan salah satu metode yang dapat digunakan untuk membentuk sebuah pohon keputusan (*Decision tree*). Klasifikasi *Decision tree* terbukti dapat memberikan sebuah keputusan yang rumit atau kompleks menjadi lebih simple atau sederhana, sehingga keputusan yang diambil dapat lebih menggambarkan atau menginterpretasikan permasalahan yang ada. Algoritma Klasifikasi *Naive Bayes*, dimana kaidah Bayes merupakan kaidah yang memperbaiki atau merevisi suatu probabilitas dengan cara memanfaatkan informasi tambahan. Maksudnya, dari probabilitas awal (*prior probability*) yang belum diperbaiki dengan rumuskan berdasarkan informasi yang tersedia saat ini, kemudian dibentuklah probabilitas berikutnya (*posterior probability*).



Sumber : (Adinugroho. & Sari, 2018)

Gambar 1. Skema Dasar Proses Klasifikasi

Data set yang digunakan dalam penelitian ini adalah data set hasil dari penelitian yang dilakukan oleh Fadi Thabtah dengan judul penelitian “Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment”(Thabtah, 2017). Sehingga diperoleh data set yang terdapat di UCI Machine Learning Repository tentang Autistic Spectrum Disorder Screening Data for Children Data set (Dheeru & Karra Taniskidou, 2017).

Data set ini terdiri dari 292 data dengan 20 atribut, dimana 10 atribut merupakan pertanyaan informasi objek penelitian dan 10 pertanyaan diambil dari Autism Spectrum Quotient (AQ-10) yang memiliki nilai 1 atau 0 pada setiap pertanyaan.

**Penentuan Algoritma**

Pengujian dilakukan pada tools WEKA dengan uji percentage split dan cross validation baik Naive Bayes maupun J48. Selanjutnya dipilih yang paling tinggi akurasi. Hasil yang diperoleh dapat dilihat pada Tabel 1.

Tabel 1 Hasil pengujian Algoritma

Mode Pengujian	Algoritma Klasifikasi	Predeksi i Benar	Prediksi Salah	Akurasi
Use Training	Naive Bayes	151-0	0-141	100%
	Decision Tree J-48	151-0	0-141	100%
	Tree J-48	151-0	0-141	100%
Percentage Split 90%	Naive Bayes	12-0	1-16	96,55%
	Decision Tree J-48	12-0	0-17	100%
	Tree J-48	12-0	0-17	100%
Percentage Split 80%	Naive Bayes	28-1	1-28	96,55%
	Decision Tree J-48	29-0	0-29	100%
	Tree J-48	29-0	0-29	100%
Percentage Split 70%	Naive Bayes	47-2	1-38	96,55%
	Decision Tree J-48	49-0	0-39	100%
	Tree J-48	49-0	0-39	100%
Percentage Split 60%	Naive Bayes	63-2	0-52	98,52%
	Decision Tree J-48	65-0	0-52	100%
	Tree J-48	65-0	0-52	100%
Cross-validation Folds=10	Naive Bayes	149-2	1-140	98,97%
	Decision Tree J-48	151-0	0-141	100%
	Tree J-48	151-0	0-141	100%

Sumber: (Farhan & Subekti, 2018)

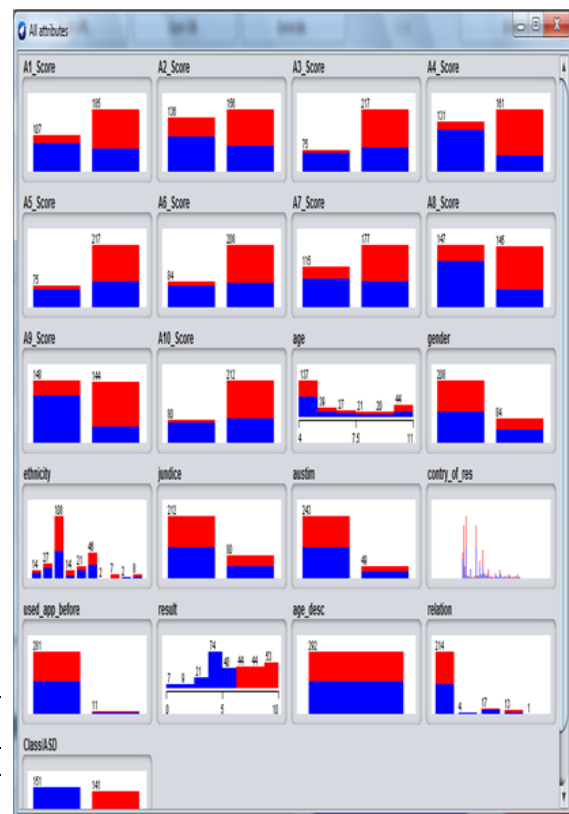
**Uji korelasi**

Untuk mengukur hubungan antar pertanyaan satu dengan pertanyaan lain maka

pada penelitian ini dilakukan pengukuran korelasi, perhitungan menggunakan rumus korelasi PEARSON. Hasilnya korelasi tertinggi ada pada atribut 4 dengan atribut 6 yaitu 0,354 (level moderat).

**Sifat Atribut**

Dari sifat-sifat masing-masing atribut, terlihat bahwa atribut terbanyak dalam bentuk binari 1/0 (A1-A10) karena merupakan skor nilai jawaban. Selebihnya merupakan atribut keterangan yang akan kita ukur apakah ada pengaruhnya terhadap ASD seperti yang terlihat pada gambar 1.



Sumber: (Farhan & Subekti, 2018)

Gambar 2. Sifat Atribut

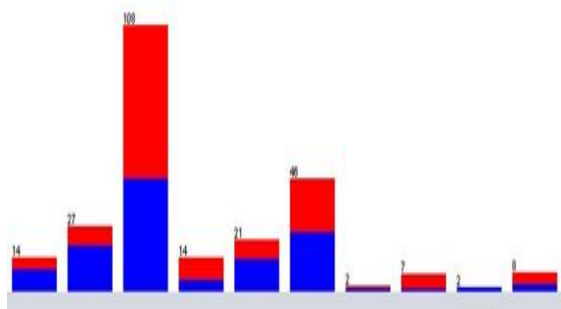
**Missing Value**

Dari sifat atribut, dapat dilihat ada 2 atribut yang mempunyai missing value yaitu A13 Ethnicity dan A20 Relation, yang masing-masing missing value sebesar 43 data (15%). Dalam WEKA, imputansi dapat dilakukan dengan mengisi data yang hilang dengan nilai mediannya. Dimana nilai tersebut tidak jauh dengan yang lain dengan cara ReplaceMissingValue. (Adinugroho. & Sari, 2018)

Selected attribute

Name: ethnicity	Missing: 43 (15%)	Distinct: 10	Type: Nominal	Unique: 0 (0%)
No.	Label	Count	Weight	
1	Others	14	14.0	
2	Middle Eastern	27	27.0	
3	White-European	108	108.0	
4	Black	14	14.0	
5	South Asian	21	21.0	
6	Asian	46	46.0	
7	Pasifika	2	2.0	
8	Hispanic	7	7.0	
9	Turkish	2	2.0	

Class: Class/ASD (Nom) Visualize All



Sumber: (Farhan & Subekti, 2018)  
 Gambar 3. Missing Value atribut Ethnicity Sebelum di Replace

Gambar 3 menunjukkan data set memiliki missing value sebanyak 43 data (15%), data tersebut merupakan data apa adanya.

Hasil Cross Validation fold 10 Sebelum di Replace

```

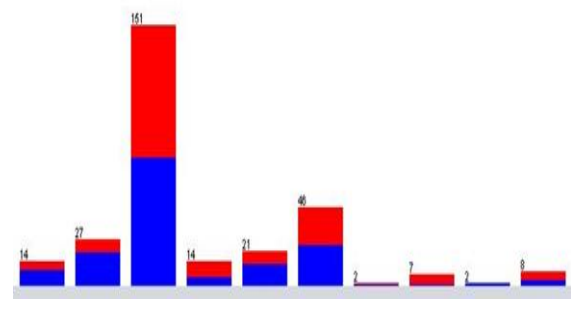
=== Run information ===
Correctly Classified Instances 292 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0
Root mean squared error 0
Relative absolute error 0 %
Root relative squared error 0 %
Total Number of Instances 292
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC
ROC Area PRC Area Class
1.000 1.000 NO
1.000 0.000 1.000 1.000 1.000 1.000
1.000 1.000 YES
Weighted Avg. 1.000 0.000 1.000 1.000 1.000
1.000 1.000 1.000
=== Confusion Matrix ===
a b <-- classified as
151 0 | a = NO
0 141 | b = YES
    
```

Dengan Cross Validation, data-data dengan missing value diuji, dan peroleh nilai akurasi 100%.

Selected attribute

Name: ethnicity	Missing: 0 (0%)	Distinct: 10	Type: Nominal	Unique: 0 (0%)
No.	Label	Count	Weight	
1	Others	14	14.0	
2	Middle Eastern	27	27.0	
3	White-European	151	151.0	
4	Black	14	14.0	
5	South Asian	21	21.0	
6	Asian	46	46.0	
7	Pasifika	2	2.0	
8	Hispanic	7	7.0	
9	Turkish	2	2.0	

Class: Class/ASD (Nom) Visualize All



Sumber: (Farhan & Subekti, 2018)  
 Gambar 4. Missing Value atribut Ethnicity Setelah di Replace

Gambar 4 menunjukkan data missing value sebanyak 43 data (15%), data tersebut diganti dengan data bernilai maksimum.

Hasil Cross Validation fold 10 Setelah di Replace

```

=== Summary ===
Correctly Classified Instances 292 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0
Root mean squared error 0
Relative absolute error 0 %
Root relative squared error 0 %
Total Number of Instances 292
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-
Measure MCC ROC Area PRC Area Class
1.000 1.000 NO
1.000 0.000 1.000 1.000 1.000
1.000 1.000 YES
Weighted Avg. 1.000 0.000 1.000 1.000
1.000 1.000 1.000
=== Confusion Matrix ===
a b <-- classified as
151 0 | a = NO
0 141 | b = YES
    
```

Dengan Cross Validation, data dengan missing value yang telah diganti diuji, dan peroleh nilai akurasi 100%.

**Feature Selection**

*Correlation based Feature Selection* (CFS) adalah algoritma filter sederhana yang memberi peringkat fitur *subset* berdasarkan korelasi fungsi evaluasi heuristik. Bias dari fungsi evaluasi adalah terhadap *subset* yang mengandung fitur yang sangat berkorelasi dengan kelas dan tidak berkorelasi satu sama lain. Pencarian *best first* dapat dimulai dengan tanpa fitur (*no attribute*) atau dengan semua fitur (*all attribute*). pencarian ke arah maju menuju ruang pencarian dengan menambahkan satu fitur ; di bagian terakhir pencarian bergerak mundur melalui ruang pencarian dengan menghapus satu fitur. Untuk mencegah pencarian *best first* dari menjelajahi seluruh ruang pencarian fitur *subset*, kriteria berhenti diberlakukan. Pencarian akan berakhir jika lima *subset* yang diperluas berturut-turut menunjukkan tidak ada perubahan atas *subset* terbaik yang diperoleh.(Hall, 1999).

Penyortiran atribut dilakukan dengan cara menggunakan *feature selected* atribut pada aplikasi Weka dengan tujuan atribut yang difilter atau tidak dipilih tersebut akan diabaikan dalam uji keakurasian *data set*.

```

=== Attribute Selection on all input data ===
Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 110
  Merit of best subset found: 1
Attribute Subset Evaluator (supervised, Class
(nominal): 21 Class/ASD):
  CFS Subset Evaluator
  Including locally predictive attributes
Selected attributes: 1,2,3,4,5,6,7,8,9,10,18 : 11
  A1_Score
  A2_Score
  A3_Score
  A4_Score
  A5_Score
  A6_Score
  A7_Score
  A8_Score
  A9_Score
  A10_Score
result
    
```

Sumber: Hasil pengujian WEKA

Hasil seleksi fitur diatas, menunjukkan hanya ada 11 atribut yang saling mempengaruhi.

Dari proses seleksi atribut yang telah diuraikan di atas, maka dapat disimpulkan hasil dari seleksi atribut tersebut dalam sebuah tabel 2, di bawah ini:

Tabel 2. Rekapitulasi Hasil Seleksi Fitur

Atribut Data Awal	Atribut Data Akhir
A1_Score	Selected
A2_Score	Selected
A3_Score	Selected
A4_Score	Selected
A5_Score	Selected
A6_Score	Selected
A7_Score	Selected
A8_Score	Selected
A9_Score	Selected
A10_Score	Selected
Age	Not Selected
Gender	Not Selected
Ethnicity	Not Selected
Jundice	Not Selected
Autism	Not Selected
Country_of_Res	Not Selected
Used_app_before	Not Selected
Result	Selected
Age_Desc	Not Selected
Relation	Not Selected

Sumber: (Farhan & Subekti, 2018)

Hanya ada 11 atribut yang paling saling mempengaruhi dalam mendeteksi ASD atau tidaknya seorang anak.

**HASIL DAN PEMBAHASAN**

Dari hasil penentuan parameter, maka atribut yang memiliki pengaruh korelasi yang dipakai adalah: A1 – A10 dan A18 (*Result*). Atribut selebihnya tidak menentukan ASD seorang anak sama sekali.

Dari hasil uji keakurasian *data set* hasil *select attributes* menghasilkan *tree* dengan memiliki hanya dua cabang, hal ini dikarenakan parameter yang dijadikan acuannya adalah atribut *result*, dimana pada atribut ini akan menjadikan atribut *Class/ASD* bernilai “NO” jika akumulasi atribut *A1\_Score* sampai dengan atribut *A10\_Score* berjumlah <= 6, dan akan memberikan nilai “YES” jika akumulasi atribut *A1\_Score* sampai dengan atribut *A10\_Score* berjumlah > 6. Dengan kata lain penentuan atribut *Class/ASD* bernilai “NO” atau “YES” tergantung dari nilai yang ada pada atribut *result*, dan bukan dari pengamatan atribut *A1\_Score* sampai *A10\_Score*.

Dikarenakan nilai “NO” atau “YES” yang diberikan dari *data set* hasil *select attributes* di atas bukan merupakan bukan hasil pengamatan dari atribut *A1\_Score* sampai dengan atribut *A10\_Score*, maka peneliti memutuskan untuk tidak menggunakan atribut *result*. Hal ini dimaksudkan agar hasil keputusan yang diberikan oleh atribut *Class/ASD* yaitu nilai “NO” atau “YES” merupakan hasil dari pengamatan *A1\_Score* sampai dengan atribut *A10\_Score*.

Dari hasil pengujian tersebut dihasilkan bahwa keakurasian dari model baru saat tidak

menggunakan atribut *result* adalah seperti yang dihasilkan di bawah ini:

*Cross validation Fold 10* atribut A1\_Score sampai dengan A10\_Score

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances          268
91.7808 %
Incorrectly Classified Instances      24      8.2192
%
Kappa statistic                       0.8353
Mean absolute error                    0.1189
Root mean squared error                0.2844
Relative absolute error                23.8005 %
Root relative squared error            56.9078 %
Total Number of Instances             292
=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-
Measure MCC  ROC Area  PRC Area  Class
0.836  0.905  0.851  NO      0.934  0.922
      0.901  0.066  0.927  0.901  0.914
0.836  0.905  0.897  YES
      Weighted Avg.  0.918  0.083  0.918  0.918
0.918  0.836  0.905  0.873

=== Confusion Matrix ===
  a  b  <-- classified as
141 10 | a = NO
 14 127 | b = YES
    
```

Sumber: Hasil pengujian WEKA?

**Perhitungan algoritma C.45**

Selanjutnya dengan perhitungan algoritma C.45 pada atribut A1\_Score sampai dengan A10\_Score

Langkah-langkah untuk memperoleh atribut sebagai akar adalah dengan menghitung jumlah kasus dan jumlah target atribut. Setelah itu menghitung nilai entropy yang digunakan untuk menentukan seberapa informatif sebuah *input* atribut untuk menghasilkan *output* atribut. (Kusrini & Luthfi, 2009).

Dengan rumus entropi:

$$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots\dots(1)$$

$$Gain (S,A) = Entropy (S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy (S_i) \dots\dots (2)$$

Memuat Cabang.

Setelah proses pencarian nilai atribut yang memiliki nilai *gain* tertinggi diperoleh, maka selanjutnya atribut tersebut digunakan sebagai *node*. *Node* ini memiliki *instance* sehingga *instance* dijadikan sebagai cabang dari *node*.

Membagi kasus dalam cabang.

Setiap nilai pada *instance* memiliki nilai yang berbeda. Nilai *instance* ini diklasifikasikan

berdasarkan makna dari nilai *instance* tersebut agar menjadi lebih sederhana.

```

=== Classifier model (full training set) ===
J48 pruned tree
-----
A4_Score = 0
| A9_Score = 0: NO (91.0/4.0)
| A9_Score = 1
| | A5_Score = 0: NO (12.0/1.0)
| | A5_Score = 1
| | | A2_Score = 0
| | | | A8_Score = 0: NO (11.0/1.0)
| | | | A8_Score = 1: YES (8.0/1.0)
| | | A2_Score = 1: YES (9.0)
A4_Score = 1
| A10_Score = 0
| | A5_Score = 0: NO (15.0)
| | A5_Score = 1
| | | A1_Score = 0: NO (6.0)
| | | A1_Score = 1
| | | | A9_Score = 0
| | | | | A2_Score = 0: NO (4.0)
| | | | | A2_Score = 1: YES (4.0/1.0)
| | | | A9_Score = 1: YES (7.0)
| | | A10_Score = 1
| | | A7_Score = 0
| | | | A3_Score = 0: NO (6.0)
| | | | A3_Score = 1
| | | | | A9_Score = 0
| | | | | | A1_Score = 0: NO (7.0/1.0)
| | | | | | A1_Score = 1
| | | | | | | A5_Score = 0: NO (4.0/1.0)
| | | | | | | A5_Score = 1: YES (11.0)
| | | | | | A9_Score = 1: YES (18.0)
| | | A7_Score = 1: YES (79.0/1.0)
Number of Leaves : 16
Size of the tree : 31
    
```

Sumber: Hasil pengujian WEKA?

Dari model *Decision tree pruned* yang dihasilkan di atas terlihat bahwa atribut A6\_Score tidak muncul, sehingga atribut ini dapat diabaikan. Adapun hasil keakurasian model baru tanpa menyertakan atribut A6\_Score adalah seperti hasil perhitungan di bawah ini

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances          269          92.1233 %
Incorrectly Classified Instances       23          7.8767 %
Kappa statistic                       0.8421
Mean absolute error                    0.1166
Root mean squared error                0.279
Relative absolute error                23.3434 %
Root relative squared error            55.8265 %
Total Number of Instances             292

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC
ROC Area  PRC Area  Class
0.909  0.853  NO      0.940  0.940  0.925  0.843
      0.901  0.060  0.934  0.901  0.917  0.843
0.909  0.908  YES
      Weighted Avg.  0.921  0.080  0.922  0.921  0.921
0.843  0.909  0.880
=== Confusion Matrix ===
  a  b  <-- classified as
142  9 | a = NO
    
```

14 127 | b = YES

Sumber: Hasil pengujian WEKA?

**Pseudocode**

Pada awalnya model Program Prediksi *Autisme Syndrome Disorder*:

Kamus

pertanyaan1,pertanyaan2,pertanyaan3, pertanyaan4,pertanyaan5,pertanyaan6,pertanyaan7,pertanyaan8,pertanyaan9,pertanyaan10:

Integer

result : integer

ASD : String

Adalah menjadi :

Kamus

pertanyaan1,pertanyaan2,pertanyaan3, pertanyaan4,pertanyaan5,pertanyaan7,pertanyaan8,pertanyaan9,pertanyaan10 : Integer

result : integer

ASD : String

**Evaluasi akhir**

Data training yang sudah dipersiapkan pada fase-fase sebelumnya (model baru) akan menjadi acuan data testing yang peneliti ambil dari lapangan. Dimana pengujian dilakukan pada 10 anak dengan data testing hasil dari interview langsung kepada responden melalui orang tuanya, dimana responden 1, 3, 6,8 dan 9 adalah 5 anak dengan ASD.

Hasil prediksi yang diperoleh menggunakan WEKA dengan model baru dimana Atribut-atribut yang di Uji:

=== Evaluation on test set ===

=== Predictions on test set ===

inst#	actual	predicted	error	prediction
1	1:?	2:YES	0.75	
2	1:?	1:NO	0.917	
3	1:?	2:YES	0.987	
4	1:?	1:NO	0.956	
5	1:?	1:NO	0.917	
6	1:?	2:YES	0.987	
7	1:?	1:NO	0.917	
8	1:?	2:YES	0.987	
9	1:?	2:YES	0.987	
10	1:?	1:NO	0.917	

=== Summary ===

Total Number of Instances 0

Ignored Class Unknown Instances 10

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure
MCC	ROC Area	PRC Area	Class		
?	NO	?	?	?	?
?	YES	?	?	?	?
?	Weighted Avg.	?	?	?	?
?					

=== Confusion Matrix ===

a b <-- classified as

0 0 | a = NO

0 0 | b = YES

Sumber: Hasil pengujian WEKA?

Hasil pengujian menunjukkan 5 responden yaitu responden 1,3, 6, 8 dan 9 memiliki indikasi (prediksi) gejala ASD. Sedangkan responden lainnya dapat dikatakan normal (tidak ASD).

**KESIMPULAN**

Uji komparasi *data set* yang dilakukan dengan aplikasi Weka menunjukkan algoritma *Decision tree* C.45 lebih baik dibandingkan menggunakan algoritma *Naive Bayes*. Korelasi tertinggi pada hubungan A4 dengan A6 dan a tidak ada pengaruh *missing value dataset* terhadap *result*. Penelitian ini juga menunjukkan bahwa prediksi yang dihasilkan sangat bergantung kepada atribut *result* sehingga pola *tree* yang dihasilkan hanya memiliki dua pola dan tidak mempertimbangkan pola kemunculan dari atribut A1\_Score sampai dengan A10\_Score, sehingga tidak mempergunakan lagi atribut *result* agar prediksi yang dihasilkan memperhitungkan atribut A1\_Score sampai dengan A10\_Score. Atribut A6\_Score tidak berpengaruh, dan ketika dilakukan uji akurasi yang baru tanpa menyertakan atribut A6\_Score menghasilkan sebuah keakurasian yang nilainya lebih baik. Sehingga dalam penelitian ini dihasilkan sebuah *data set* baru yang hasil akurasinya lebih baik dari *data set* sebelumnya.

**REFERENSI**

Adinugroho., S., & Sari, Y. A. (2018). *Implementasi Data mining Menggunakan WEKA* (1st ed.). Malang: Brawijaya Press.

Dheeru, D., & Karra Taniskidou, E. (2017). {UCI} Machine Learning Repository.

Farhan, O., & Subekti, A. (2018). *Laporan Penelitian: Permodelan Prediktif Autistic Spectrum Disorder Dengan Algoritma C.45*. Jakarta.

Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. <https://doi.org/10.1.1.149.3848>

Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). USA: Morgan Kaufmann.

Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining* (1st ed.). Yogyakarta: Andi

- Publisher. Retrieved from <https://doi.org/10.1145/3107514.3107515>  
<http://andipublisher.com/produk-0907003050-algoritma-data-mining.html>
- Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 5(1). [https://doi.org/10.1016/0022-4073\(81\)90019-4](https://doi.org/10.1016/0022-4073(81)90019-4)
- Thabtah, F. (2017). Autism Spectrum Disorder screening: Machine learning adaptation and DSM-5 fulfillment. *Proceedings of the 1st International Conference on Medical and Health Informatics 2017 - ICMHI '17*, 1–6.
- Thabtah, F. (2018). Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health and Social Care*, 8157, 1–20. <https://doi.org/10.1080/17538157.2017.1399132>
- Witten, I., & Frank, E. (2005). *Data Mining. Practical Machine Learning Tools and Technicals*. <https://doi.org/10.1002/9780120884070>