

Algoritma Naïve Bayes Classifier Untuk Memprediksi Gejala Autism Spectrum Disorders Pada Anak-Anak

Sri Rahayu¹, Jajang Jaya Purnama², Hendri Mahmud Nawawi³, Fitria Septia Nugraha⁴
Ilmu Komputer, STMIK Nusa Mandiri Jakarta, Indonesia^{1,2,3,4}

srahayu110527@gmail.com¹, jajangja2412@bsi.ac.id², mahmudhend94@gmail.com³, fitraseptia7@gmail.com⁴

Abstrak--Manusia merupakan makhluk sosial yang tidak bisa terlepas dari hubungan dengan manusia lainnya demi keberlangsungan hidup bersama, salah satunya dengan berkomunikasi. Dengan berkomunikasi, pikiran, perasaan dan kebutuhan seseorang dengan dunia luarnya dapat tersampaikan. Pengaruh besar ini dibuktikan oleh Peter Drucker seorang analis manajemen Amerika yang menilai bahwa di negara-negara yang sudah maju, maka setiap pembelanjaan dalam bentuk dollar, selain untuk makanan dan pakaian dihabiskan untuk kepentingan berkomunikasi. Melihat begitu pentingnya komunikasi ini, terlihat sangat disayangkan, masih saja ada beberapa orang yang terlahir dengan kesulitan dalam hal berkomunikasi, berinteraksi bahkan merespon interaksi dari orang lain akibat dari berbagai gangguan yang terjadi dalam organisme tubuhnya sejak ia dilahirkan. Salah satu disorders (gangguan) itu adalah disebut dengan Autism Spectrum Disorders (ASD), yaitu kelainan perkembangan saraf yang mempengaruhi perilaku dan cara seseorang berkomunikasi, mereka cenderung menghabiskan waktunya sendiri seperti memiliki dunia sendiri. Para penderita Autism Spectrum Disorders (ASD) bisa ditolong dengan berbagai obat-obatan dan terapi yang dilakukan setelah diketahui gejala-gejala yang menunjukkan bahwa itu adalah gejala Autism Spectrum Disorders (ASD). Melihat kondisi seperti ini, penulis bermaksud untuk melakukan penelitian dengan melakukan klasifikasi terhadap gejala-gejala Autism Spectrum Disorders (ASD), yang terjadi pada anak-anak. Dengan adanya klasifikasi ini diharapkan para orang tua lebih sigap dalam mengetahui kondisi kesehatan anaknya dari gangguan (disorders) yang mungkin saja terjadi, sehingga dapat langsung melakukan tindakan sebagai usaha untuk meminimalisir gejala yang terjadi dan diharapkan tindakan sejak dini. Klasifikasi dilakukan dengan sebuah algoritma Naïve Bayes yang dikenal merupakan salah satu model klasifikasi yang baik dan sering digunakan. Hasil dari penelitian ini mendapatkan accuracy sebesar 60,48% menggunakan rapid miner 9.1 dengan 248 dataset.

Kata Kunci: Naïve Bayes, ASD, Klasifikasi.

1. PENDAHULUAN

Seperti yang telah diketahui bersama, manusia merupakan makhluk sosial yang tidak bisa terlepas dari hubungan dengan manusia lainnya demi keberlangsungan hidup bersama, salah satunya dengan berkomunikasi. Harold D. Laswell salah seorang peletak dasar ilmu komunikasi lewat ilmu politik menyebut tiga fungsi dasar pentingnya manusia dalam berkomunikasi.

Pertama, adalah hasrat manusia untuk mengontrol lingkungannya. Melalui komunikasi manusia dapat mengetahui peluang yang ada untuk dimanfaatkan, dipelihara, dan menghindari pada hal-hal yang mengancam alam sekitarnya. Melalui komunikasi manusia dapat mengetahui suatu kejadian atau peristiwa. Bahkan melalui komunikasi manusia dapat mengembangkan

pengetahuannya, yakni belajar dari pengalamannya, maupun melalui informasi yang mereka terima dari lingkungan sekitarnya.

Kedua, adalah upaya manusia untuk dapat beradaptasi dengan lingkungannya. Proses kelanjutan suatu masyarakat sesungguhnya tergantung bagaimana masyarakat itu bisa beradaptasi dengan lingkungannya. Penyesuaian disini bukan saja terletak pada kemampuan manusia memberi tanggapan terhadap gejala alam seperti banjir, gempa bumi dan musim yang dipengaruhi perilaku manusia, tetapi juga lingkungan masyarakat tempat manusia hidup dalam tantangan. Dalam lingkungan seperti ini diperlukan penyesuaian, agar manusia dapat hidup dalam suasana yang harmonis.

Ketiga, adalah upaya untuk melakukan transformasi warisan sosialisasi. Suatu masyarakat yang ingin mempertahankan keberadaannya, maka anggota masyarakatnya dituntut untuk melakukan pertukaran nilai, perilaku, dan peranan. Misalnya bagaimana orang tua mengajarkan tatakrama bermasyarakat yang baik kepada anak-anaknya. Bagaimana sekolah difungsikan untuk mendidik warga negara. Bagaimana media massa menyalurkan hati nurani khalayaknya, dan bagaimana pemerintah dengan kebijaksanaan yang dibuatnya untuk mengayomi kepentingan anggota masyarakat yang dilayaninya.

Ketiga fungsi tersebut menunjukkan bahwa kunci dari interaksi sosial di masyarakat adalah dengan berkomunikasi. Dengan berkomunikasi, pikiran, perasaan dan kebutuhan seseorang dengan dunia luarnya dapat tersampaikan. Pengaruh besar ini dibuktikan oleh Peter Drucker seorang analis manajemen Amerika yang menilai bahwa di negara-negara yang sudah maju, maka setiap pembelanjaan dalam bentuk dollar, selain untuk makanan dan pakaian dihabiskan untuk kepentingan berkomunikasi. Drucker membuktikannya dengan menunjuk pembayaran telepon, telex, perangko, pajak TV, radio, surat kabar, seminar, menonton, rekreasi, buku, majalah, komputer semuanya untuk memenuhi kebutuhan informasi.

Melihat begitu pentingnya komunikasi ini, terlihat sangat disayangkan, masih saja ada beberapa orang yang terlahir dengan kesulitan dalam hal berkomunikasi, berinteraksi bahkan merespon interaksi dari orang lain akibat dari berbagai gangguan yang terjadi dalam organisme tubuhnya sejak ia dilahirkan. Salah satu *disorders* (gangguan) itu adalah disebut dengan *Autism Spectrum Disorders* (ASD), yaitu kelainan perkembangan saraf yang mempengaruhi perilaku dan cara seseorang berkomunikasi, mereka cenderung menghabiskan waktunya sendiri seperti memiliki dunia sendiri. Pada umumnya, mereka memiliki masalah dalam komunikasi nonverbal (bahasa tubuh, ekspresi wajah, kontak mata dan nada suara), mereka juga mungkin mengulangi tindakan tertentu secara berulang-ulang.

Disorders (gangguan) memang berbeda dengan *Deases* (Penyakit) yang bisa disembuhkan secara total, namun demikian, para penderita *Autism Spectrum Disorders* (ASD) bisa ditolong dengan berbagai obat-obatan dan terapi yang dilakukan setelah diketahui gejala-gejala yang menunjukkan bahwa itu adalah gejala *Autism Spectrum Disorders* (ASD). Bisa dikatakan sembuh apabila beberapa gejala yang dialami menghilang seperti yang dikatakan oleh Prof. Rimland seorang ahli dalam bidang autisme di Amerika, sesuai pepatah “*if it looks like a duck, walks like a duck, quacks like a duck, it must be a*

duck”, maka analoginya pada “mantan” penyandang autisme ini yaitu “*if they look recovered, are thought to be recovered, (that means) they are recovered*”. jika memang ada gejala sisa, lalu kenapa? Karena tidak ada orang yang bisa mengaku bahwa dirinya 100% normal. Selain itu, banyak penyakit yang sembuh tapi meninggalkan gejala sisa, misalnya pada *Autism Spectrum Disorders* (ASD) ini, namun setidaknya dengan melakukan diagnosa gejalanya sejak dini, gangguan ini bisa diminimalisir setidaknya saat diketahui, orang-orang disekitarnya bisa merawat dengan perlakuan khusus. Maka dari itu, penulis ingin memprediksi gejala yang terjadi pada penderita *Autism Spectrum Disorders* (ASD) di kalangan anak-anak dengan sebuah Algoritma Klasifikasi yaitu *Naïve Bayes* dengan teknik Data Mining.

Dengan adanya klasifikasi ini diharapkan para orang tua lebih sigap dalam mengetahui kondisi kesehatan anaknya dari gangguan (*disorders*) yang mungkin saja terjadi, sehingga dapat langsung melakukan tindakan sebagai usaha untuk meminimalisir gejala yang terjadi dan diharapkan tindakan sejak dini ini membuat gejala lain yang akan terjadi justru tidak terjadi dan yang sudah terjadi akan terkikis meskipun tidak normal kembali setidaknya gejala yang ada berkurang.

2. LANDASAN TEORI

2.1. Pengertian Data Mining

Didalam bukunya Kusri dan Luthfi, Algoritma Data Mining, 2009, dalam (Hasibuan *et al.*, 2017) ada beberapa pengertian data mining menurut para ahli, yaitu :

1. Menurut Larose data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi didalam *database*. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang terkait dari berbagai *database* besar.
2. Menurut *Gartner Group* data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika.

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali,

atau menambang pengetahuan dari data atau informasi yang kita miliki. Data mining, sering juga disebut sebagai *Knowledge Discovery in Database (KDD)*. KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar.

1. Metode Pelatihan

Secara garis besar metode pelatihan yang digunakan dalam teknik-teknik data mining dibedakan ke dalam dua pendekatan, yaitu :

- *Unsupervised learning*, metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*). Guru disini adalah label dari data.
- *Supervised learning*, yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai output atau label selama proses *training*.

2. Pengelompokan Data Mining

Ada beberapa teknik yang dimiliki data mining berdasarkan tugas yang bisa dilakukan, yaitu :

1. Deskripsi

Para peneliti biasanya mencoba menemukan cara untuk mendeskripsikan pola dan *trend* yang tersembunyi dalam data.

2. Estimasi

Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numerik dari pada kategori.

3. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi di masa depan).

4. Klasifikasi

Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Clustering

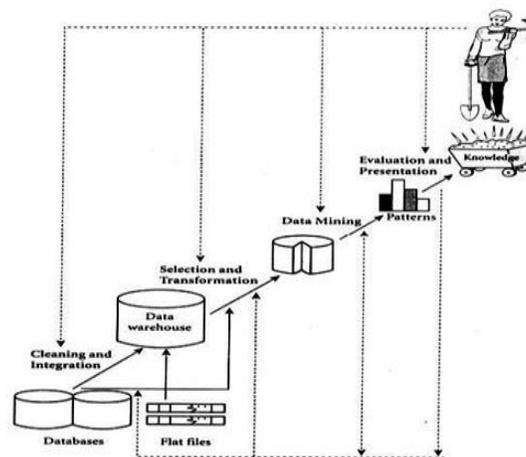
Clustering lebih kearah pengelompokkan *record*, pengamatan, atau kasus dalam kelas yang memiliki kemiripan.

6. Asosiasi

Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

3. Tahap-tahap Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada Gambar 1. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledgebase*.



Gambar 1 : Tahap-tahap Data Mining

Tahap-tahap data mining adalah sebagai berikut :

1. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.

2. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.

3. Seleksi data (*data selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

4. Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.

5. Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (*pattern evaluation*)

Untuk mengindikasi pola-pola menarik ke dalam *knowledge base* yang ditemukan.

7. Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. (Ridwan, Suyono, & Sarosa, 2013)

2.2. Pengertian Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”, berupa *decision tree*, formula matematis atau *neural network*. Proses *classification* biasanya dibagi menjadi dua fase: *learning* dan *test*. Pada *fase learning*, sebagian data yang telah diketahui kelas datanya diumpungkan untuk membentuk model perkiraan. Kemudian pada *fase test* model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui. (Andrea, dkk : 2011) dalam (Sumber & Telekomunikasi, 1978)

2.3. Metode Naïve Bayes

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan *Naïve Bayes* merupakan 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. Keuntungan penggunaan *Naïve Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naïve Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

Persamaan dari teorema Bayes adalah :

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)}$$

Dimana :

X = Data dengan *class* yang belum diketahui.

H = Hipotesis Data X merupakan suatu *class* spesifik

P(H|X) = probabilitas hipotesis H berdasarkan kondisi x (*posteriori prob.*)

P(H) = Probabilitas hipotesis H (*prior prob.*)

P(X|H) = probabilitas X berdasarkan kondisi tersebut

P(X) = probabilitas dari X

Untuk menjelaskan metode *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naïve Bayes* diatas disesuaikan sebagai berikut :

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Posterior = Peluang munculnya kelas

Prior = Kelas sebelum masuknya sampel

Likelihood = Kemunculan karakteristik sampel pada kelas

Evidence = Kemunculan karakteristik sampel secara global

(Ridwan et al., 2013).

2.4. Pengertian Autism Spectrum Disorders (ASD)

Kementerian Pemberdayaan Perempuan dan Perlindungan Anak Republik Indonesia pada tahun 2013 menuliskan pengertian Anak dengan gangguan spectrum autisma atau *autism spectrum disorders* (ASD) adalah anak yang mengalami gangguan dalam tiga area dengan tingkatan berbeda-beda, yaitu kemampuan komunikasi dan interaksi sosial, serta pola-pola perilaku yang repetitif dan stereotipi dalam buku Panduan Penanganan Anak Berkebutuhan Khusus Bagi Pendamping (Orang Tua, Keluarga, dan Masyarakat).

Autism Spectrum Disorder (ASD) adalah gangguan perkembangan terutama dalam berperilaku, yang secara umum disebabkan oleh kelainan struktur otak atau fungsi otak. ASD ini terlihat sebelum anak berusia tiga tahun dan dapat diketahui dari interaksi sosial dan komunikasi yang terbatas dan berulang-ulang.

Penyandang ASD semakin meningkat di seluruh dunia. Pada tahun 2006 diperkirakan prevalensi jumlah

penyandang ASD 1:100 kelahiran (Kelana & Elmy 2007). Di Indonesia hingga saat ini penyandang ASD belum diketahui secara pasti jumlahnya, akan tetapi diperkirakan lebih dari 400.000 orang (Kelana & Elmy 2007) dalam (Ridwan et al., 2013).

2.5. Penelitian Terkait

Penelitian terkait ini digunakan sebagai acuan peneliti, kemana arah penelitian menggunakan *dataset* yang didapat dari *UCI Repository* bernama *Autism_Child_Data.arff* ini. Berikut ini beberapa penelitian terkait yang sebelumnya sudah dilakukan oleh para peneliti terdahulu :

1. Fadi Thabtah seorang mahasiswa di *Nelson Marlborough Institute of Technology, New Zealand* dalam jurnalnya berjudul *Autism Spectrum Disorder Screening : Machine Learning Adaptation and DSM-5 Fulfillment* dalam *Association for Computing Machinery*. ACM ISBN 978-1-4503-5224-6/17/05 pada tahun 2017 membahas tentang fokus pada mesin terbaru belajar studi yang ditangani ASD sebagai masalah klasifikasi dan menganalisis secara kritis kelebihan dan kekurangan mereka. Bahkan, mereka menunjukkan langkah-langkah yang diperlukan untuk mengklaim pembangunan alat diagnostik cerdas berdasarkan pembelajaran mesin oleh mengganti aturan buatan tangan di dalam alat skrining ASD dengan model prediksi. Terakhir, kami menyoroti urgensi memperbaiki alat skrining klinis ASD untuk mencerminkan perubahan yang diusulkan dalam manual DSM-5. Penyebaran DSM-5 menuntut perubahan dalam cara kode *algoritma diagnostic* dalam alat skrining ASD berperilaku dalam proses mengklasifikasikan kasus. Ada kebutuhan untuk memeriksa kembali pertanyaan atau fitur dalam alat diagnostik ASD untuk memenuhi kriteria baru dari DSM-5. Ini membutuhkan pemetaan kriteria ASD baru ke fitur atau atribut yang digunakan dalam alat diagnosis klinis juga mengevaluasi cara kerja algoritma diagnostik. Itu Penyesuaian akan mengarahkan peneliti ke bagaimana Pervasif berbeda Development Disorders (PDD) tumpang tindih dengan kriteria DSM-5 yang baru dan ini akan membantu meningkatkan alat diagnostik saat ini. (Thabtah, 2017)
2. Dalam jurnal penelitian *Informatics for Health and Social Care* yang juga ditulis oleh Fadi Thabtah pada

tahun 2018 berjudul *Machine Learning in Autistic Spectrum Disorder Behavioral Research : A Review and Ways Forward* menjelaskan tentang memeriksa dan menganalisis secara kritis studi pembelajaran mesin ini pada ASD dan aspek konseptualisasi, implementasi, dan data terkait yang disebutkan di atas. Pada penelitian ini, algoritma SVM dipercaya dapat memodelkan klasifikasi ASD karena prediktifnya yang tinggi. Ketika diintegrasikan ke dalam alat diagnostik yang ada, model prediksi ini khususnya SVM dapat sangat bermanfaat bagi para ahli dan peneliti sebagai alat efektif dalam pengambilan keputusan. (Thabtah, 2018)

3. Ayse Demirhan berasal dari *Electrical and Electronics Engineering Department* di *Faculty of Technology, Gazi University, Turkey* dalam *Mugla Journal of Science and Technology* yang berjudul *Performance of Machine Learning Methods in Determining the Autism Spectrum Disorder Cases* pada tahun 2018 menerangkan tentang evaluasi kinerja metode mesin pembelajaran untuk klasifikasi kasus ASD menggunakan 104 data pemindaian remaja. Semua proses pemilihan parameter dilakukan menggunakan CV 10 kali lipat untuk mengurangi bias yang akan timbul dari parameter. Metode pembelajaran mesin SVM, KNN dan RF digunakan untuk menganalisis jawaban yang diberikan untuk kuesioner yang mencakup pertanyaan terbukti efektif dalam membedakan kasus ASD, akurasi, sensitifitas dan spesifisitas *metric* kinerja digunakan untuk evaluasi model yang dilatih. Tingkat akurasi menggunakan metode SVM adalah 95%, KNN 89%, dan RF adalah 100%. (Demirhan, 2018)

Dari penelitian terkait yang dikaji oleh penulis, sebagai originalitas penelitian, dalam penelitian ini prediksi dilakukan pada penderita gangguan *Autism Spectrum Disorders* di kalangan anak-anak dengan harapan dapat mendeteksi lebih dini maka lebih dini pula dalam menyikapinya entah itu menjaga, mengobati ataupun menyiasati gejala lain yang mungkin timbul. Parameter yang digunakan juga yang berkaitan dengan kepribadiannya.

3. PEMBAHASAN

3.1. Penerapan Metode Naive Bayes

Data yang digunakan untuk penelitian ini diambil dari *UCI Repository* dengan nama file

Autism_Child_Data.arff pada tahun 2018 yang pada awalnya ada 292 data dengan 20 atribut dan 1 class. Namun, dengan proses *cleaning* data, yang digunakan dalam penelitian ini sebanyak 248 data dengan 8 atribut dan 1 class. Berikut merupakan atribut yang digunakan dalam penelitian ini :

1. usia, adalah usia anak-anak yang akan dianalisa apakah menderita gangguan ASD atau tidak.
2. jenis kelamin, apakah anak tersebut perempuan atau laki-laki.
3. keturunan, anak tersebut berasal dari keturunan atau etnis mana.
4. riwayat penyakit kuning, misalnya sebelumnya atau pada saat itu si anak menderita penyakit kuning atau tidak.
5. riwayat keluarga autis, memiliki riwayat keluarga yang menderita ASD juga atau tidak.
6. kota asal, merupakan tempat tinggal anak saat ini.
7. pernah melakukan skrining, sebelumnya pernah melakukan pemeriksaan atau belum.
8. kedekatan, maksudnya adalah si anak ini berhubungan paling dekat atau lebih dekat dengan siapanya.

Kedelapan atribut tersebut berbentuk kategorikal sehingga, tahapan Proses Naïve Bayes :

1. Baca data training/latih

Data training atau data latih yang digunakan adalah sebanyak 248 data yang merupakan kategorikal sebagai berikut :

Tabel 1. Data Training

| no | usia | jk | keturunan | penyakit_kuning | riwayat_keluaraga_autis | kota | pernah_skrining | kedekatan | diagnosa |
|----|------|----|------------------|-----------------|-------------------------|------------------|-----------------|-----------|----------|
| 1 | 6 | l | Other s | tidak | tidak | Jordan | tidak | orang tua | normal |
| 2 | 6 | l | 'Middle Eastern' | tidak | tidak | Jordan | tidak | orang tua | normal |
| 3 | 5 | l | Other s | ya | tidak | 'United States' | tidak | orang tua | ASD |
| 4 | 5 | l | White - European | tidak | tidak | 'United Kingdom' | tidak | orang tua | ASD |
| 5 | 5 | p | 'Middle Eastern' | tidak | tidak | Bahrain | tidak | ora | AS |

| | | | | | | | | | | | | |
|-----|----|---|--------------------|-------|-------|------------------------|-------|-------|------------------------|-------|-----------|--------|
| | | | | | | le Eastern' | k | k | in | k | ng tua | D |
| 6 | 11 | p | 'Middle Eastern' | tidak | tidak | Bahrain | tidak | tidak | Bahrain | tidak | orang tua | ASD |
| 7 | 10 | l | 'White - European' | ya | tidak | 'United Kingdom' | | tidak | 'United Kingdom' | tidak | sendiri | ASD |
| 8 | 4 | l | 'White - European' | ya | tidak | 'United States' | | tidak | 'United States' | tidak | orang tua | ASD |
| 9 | 4 | p | Black | tidak | tidak | 'United Arab Emirates' | | tidak | 'United Arab Emirates' | tidak | orang tua | normal |
| 10 | 6 | l | 'White - European' | tidak | tidak | Europe | | tidak | Europe | tidak | orang tua | ASD |
| ... | | | | | | | | | | | | |
| 248 | 4 | p | 'South Asian' | tidak | tidak | India | | tidak | India | tidak | orang tua | normal |

2. Menghitung jumlah kelas/label "P(H)"

$$P(\text{ASD}) = \frac{\text{jumlah data "ASD"}}{\text{jumlah data}} = \frac{126:248}{} = 0,50806 = 50,81\%$$

$$P(\text{normal}) = \frac{\text{jumlah data "normal"}}{\text{jumlah data}} = \frac{122:248}{} = 0,49193 = 49,19\%$$

Sehingga ketika dijumlahkan hasilnya adalah 100 %.

3. Menghitung jumlah kasus per kelas "P(X|H)"

- a. Atribut Usia

Probabilitas usia dengan keterangan "ASD"

$P(\text{usia4})|\text{ASD}/P(\text{ASD}) = \frac{\text{jumlah data "usia4" dengan keterangan "ASD"}}{\text{jumlah data "ASD"}}$

$$P(\text{usia4})|\text{ASD}/P(\text{ASD}) = \frac{41}{126} = 0,32539 = 32,54\%$$

$$P(\text{usia5})|\text{ASD}/P(\text{ASD}) = \frac{15}{126} = 0,11904 = 11,90\%$$

$$P(\text{usia6})|\text{ASD}/P(\text{ASD}) = \frac{16}{126} = 0,12698 = 12,70\%$$

$$P(\text{usia7})|\text{ASD}/P(\text{ASD}) = \frac{11}{126} = 0,08730 = 8,73\%$$

$$P(\text{usia8})|\text{ASD}/P(\text{ASD}) = \frac{11}{126} = 0,08730$$

$$= 8,73\%$$

$$P(\text{usia9})|ASD/P(ASD) = 9/126 = 0,07142$$

$$= 7,14\%$$

$$P(\text{usia10})|ASD/P(ASD) = 9/126 = 0,07142$$

$$= 7,14\%$$

$$P(\text{usia11})|ASD/P(ASD) = 14/126 = 0,11111$$

$$= 11,11\%$$

Sehingga ketika dijumlahkan hasilnya adalah 100 %.

Probabilitas usia dengan keterangan “normal”

$P(\text{usia4})|normal/P(normal)$ = jumlah data “usia4” dengan keterangan “normal” dibagi jumlah data “normal”

$$P(\text{usia4})|normal/P(normal) = 37/122 = 0,303279$$

$$= 30,33\%$$

$$P(\text{usia5})|normal/P(normal) = 21/122 = 0,17213$$

$$= 17,21\%$$

$$P(\text{usia6})|normal/P(normal) = 17/122 = 0,139344$$

$$= 13,93\%$$

$$P(\text{usia7})|normal/P(normal) = 13/122 = 0,10655$$

$$= 10,656\%$$

$$P(\text{usia8})|normal/P(normal) = 13/122 = 0,10655$$

$$= 10,656\%$$

$$P(\text{usia9})|normal/P(normal) = 8/122 = 0,065574$$

$$= 6,56\%$$

$$P(\text{usia10})|normal/P(normal) = 8/122 = 0,065574$$

$$= 6,56\%$$

$$P(\text{usia11})|normal/P(normal) = 9/122 = 0,073770$$

$$= 7,38\%$$

Sehingga ketika dijumlahkan hasilnya adalah 100 %.

b. Atribut Jenis Kelamin

Probabilitas jenis kelamin dengan keterangan “ASD”

$P(\text{jkL})|ASD/P(ASD)$ = jumlah data “jkL” dengan keterangan “ASD” dibagi jumlah data “ASD”

$$P(\text{jkL})|ASD/P(ASD) = 90/126 = 0,71428$$

$$= 71\%$$

$$P(\text{jkP})|ASD/P(ASD) = 36/126 = 0,285714$$

$$= 29\%$$

Sehingga ketika dijumlahkan hasilnya adalah 100 %.

Probabilitas jenis kelamin dengan keterangan “normal”

$P(\text{jkL})|normal/P(normal)$ = jumlah data “jkL” dengan keterangan “normal” dibagi jumlah data “normal”

$$P(\text{jkL})|normal/P(normal) = 84/122 = 0,68852$$

$$= 69\%$$

$$P(\text{jkP})|normal/P(normal) = 38/122 = 0,31147$$

$$= 31\%$$

Sehingga ketika dijumlahkan hasilnya adalah 100 %.

c. Atribut Keturunan

Probabilitas keturunan dengan keterangan “ASD”

$P(\text{others})|ASD/P(ASD)$ = jumlah data “others” dengan keterangan “ASD” dibagi jumlah data “ASD”

$$P(\text{others})|ASD/P(ASD) = 50/126 = 0,39682$$

$$= 3,97\%$$

$$P(\text{middle eastern})|ASD/P(ASD) = 8/126 = 0,06349$$

$$= 6,35\%$$

$$P(\text{white european})|ASD/P(ASD) = 62/126 = 0,49206$$

$$= 49,21\%$$

$$P(\text{black})|ASD/P(ASD) = 9/126 = 0,07142$$

$$= 7,14\%$$

$$P(\text{south asian})|ASD/P(ASD) = 8/126 = 0,06349$$

$$= 6,35\%$$

$$P(\text{asian})|ASD/P(ASD) = 22/126 = 0,17460$$

$$= 17,46\%$$

$$P(\text{pasifika})|ASD/P(ASD) = 1/126 = 0,07936$$

$$= 0,79\%$$

$$P(\text{hispanic})|ASD/P(ASD) = 6/126 = 0,04761$$

$$= 4,76\%$$

$$P(\text{turkish})|ASD/P(ASD) = 0/126 = 0$$

$$= 0\%$$

$$P(\text{latino})|ASD/P(ASD) = 5/126 = 0,03968$$

$$= 3,97\%$$

$$= 3,97\%$$

Sehingga ketika dijumlahkan hasilnya adalah 100 %.

Probabilitas keturunan dengan keterangan “normal”

$P(\text{others})|normal/P(normal)$ = jumlah data “others” dengan keterangan “normal” dibagi jumlah data “normal”

$$P(\text{others})|normal/P(normal) = 9/122 = 0,07377$$

$$= 7,38\%$$

$$P(\text{middle eastern})|normal/P(normal) = 18/122 = 0,14754$$

$$= 14,75\%$$

$$P(\text{white european})|normal/P(normal) = 46/122 = 0,37704$$

$$= 37,70\%$$

$$P(\text{black})|normal/P(normal) = 5/122 = 0,49193$$

$$= 4,10\%$$

$$P(\text{south asian})|normal/P(normal) = 13/122 = 0,10655$$

$$= 10,66\%$$

$$P(\text{asian})|normal/P(normal) = 24/122 = 0,19672$$

$$= 19,67\%$$

$$= 19,67\%$$

$$\begin{aligned}
 P(\text{fasifika})|\text{normal}/P(\text{normal}) &= 1/122 = \\
 0,00819 &= 0,82\% \\
 P(\text{hispanic})|\text{normal}/P(\text{normal}) &= 1/122 = \\
 0,00819 &= 0,82\% \\
 P(\text{turkish})|\text{normal}/P(\text{normal}) &= 2/122 = 0,01639 \\
 &= 1,64\% \\
 P(\text{latino})|\text{normal}/P(\text{normal}) &= 3/122 = 0,02459 \\
 &= 2,46\%
 \end{aligned}$$

Sehingga ketika dijumlahkan hasilnya adalah 100 %.

d. Atribut Penyakit Kuning

Probabilitas penyakit kuning dengan keterangan “ASD”
 $P(\text{ya})|\text{ASD}/P(\text{ASD}) = \text{jumlah data “ya” dengan keterangan “ASD” dibagi jumlah data “ASD”}$

$$\begin{aligned}
 P(\text{ya})|\text{ASD}/P(\text{ASD}) &= 31/126 = 0,24603 \\
 &= 24,60\% \\
 P(\text{tidak})|\text{ASD}/P(\text{ASD}) &= 95/126 = \\
 0,75396 &= 75,40\%
 \end{aligned}$$

Sehingga ketika dijumlahkan hasilnya adalah 100 %.

Probabilitas keturunan dengan keterangan “normal”
 $P(\text{ya})|\text{normal}/P(\text{ya}) = \text{jumlah data “ya” dengan keterangan “normal” dibagi jumlah data “normal”}$

$$\begin{aligned}
 P(\text{ya})|\text{normal}/P(\text{normal}) &= 30/122 = 0,24590 \\
 &= 24,59\% \\
 P(\text{tidak})|\text{normal}/P(\text{normal}) &= 92/122 = 0,75409 \\
 &= 75,41\%
 \end{aligned}$$

Sehingga ketika dijumlahkan hasilnya adalah 100 %.

e. Atribut Keluarga Autis

Probabilitas keluarga autis dengan keterangan “ASD”
 $P(\text{ya})|\text{ASD}/P(\text{ASD}) = \text{jumlah data “ya” dengan keterangan “ASD” dibagi jumlah data “ASD”}$

$$\begin{aligned}
 P(\text{ya})|\text{ASD}/P(\text{ASD}) &= 21/126 = 0,16667 \\
 &= 16,67\% \\
 P(\text{tidak})|\text{ASD}/P(\text{ASD}) &= 105/126 = \\
 0,83333 &= 83,33\%
 \end{aligned}$$

Sehingga ketika dijumlahkan hasilnya adalah 100%.

Probabilitas keluarga autis dengan keterangan “normal”
 $P(\text{ya})|\text{normal}/P(\text{ya}) = \text{jumlah data “ya” dengan keterangan “normal” dibagi jumlah data “normal”}$

$$\begin{aligned}
 P(\text{ya})|\text{normal}/P(\text{normal}) &= 24/122 = 0,24590 \\
 &= 24,59\% \\
 P(\text{tidak})|\text{normal}/P(\text{normal}) &= 98/122 = 0,80327 \\
 &= 80,33\%
 \end{aligned}$$

Sehingga ketika dijumlahkan hasilnya adalah 100%.

f. Atribut Kota Asal

Probabilitas kota asal dengan keterangan “ASD”
 $P(\text{jordan})|\text{ASD}/P(\text{ASD}) = \text{jumlah data “ya” dengan keterangan “ASD” dibagi jumlah data “ASD”}$

$$\begin{aligned}
 P(\text{jordan})|\text{ASD}/P(\text{ASD}) &= 1/126 = 0,00793 \\
 &= 0,79\% \\
 P(\text{united state})|\text{ASD}/P(\text{ASD}) &= 31/126 = 0,24603 \\
 &= 24,60\% \\
 P(\text{united kingdom})|\text{ASD}/P(\text{ASD}) &= 24/126 = 0,19047 \\
 &= 19,05\% \\
 P(\text{bahrain})|\text{ASD}/P(\text{ASD}) &= 2/126 = 0,01587 \\
 &= 1,59\% \\
 P(\text{united arab emirates})|\text{ASD}/P(\text{ASD}) &= 0/126 = 0 \\
 &= 0\% \\
 P(\text{europe})|\text{ASD}/P(\text{ASD}) &= 1/126 = 0,00793 \\
 &= 0,79\% \\
 P(\text{malta})|\text{ASD}/P(\text{ASD}) &= 1/126 = 0,00793 \\
 &= 0,79\% \\
 P(\text{bulgaria})|\text{ASD}/P(\text{ASD}) &= 1/126 = 0,00793 \\
 &= 0,79\% \\
 P(\text{south africa})|\text{ASD}/P(\text{ASD}) &= 2/126 = 0,01587 \\
 &= 1,59\% \\
 P(\text{india})|\text{ASD}/P(\text{ASD}) &= 17/126 = \\
 0,13492 &= 13,49\% \\
 P(\text{iraq})|\text{ASD}/P(\text{ASD}) &= 0/126 = 0 \\
 &= 0\% \\
 P(\text{new zealand})|\text{ASD}/P(\text{ASD}) &= 5/126 = 0,03968 \\
 &= 3,97\% \\
 P(\text{australia})|\text{ASD}/P(\text{ASD}) &= 12/126 = 0,09523 \\
 &= 9,52\% \\
 P(\text{georgia})|\text{ASD}/P(\text{ASD}) &= 1/126 = 0,007937 \\
 &= 0,79\% \\
 P(\text{armenia})|\text{ASD}/P(\text{ASD}) &= 3/126 = 0,02381 \\
 &= 2,38\% \\
 P(\text{turkey})|\text{ASD}/P(\text{ASD}) &= 0/126 = 0 \\
 &= 0\% \\
 P(\text{pakistan})|\text{ASD}/P(\text{ASD}) &= 1/126 = 0,00793
 \end{aligned}$$

| | | | |
|--------------------------------|-------------------|--------------------|---------|
| P(canada) ASD/P(ASD) | = 0,79% | = 5/126 = 0,039683 | = 3,97% |
| P(oman) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(brazil) ASD/P(ASD) | = 2/126 = 0,01587 | | = 1,59% |
| P(south korea) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(costa rica) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(sweden) ASD/P(ASD) | = 1/126 = 0,00794 | | = 0,79% |
| P(philippines) ASD/P(ASD) | = 3/126 = 0,02381 | | = 2,38% |
| P(argentina) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(japan) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(bangladesh) ASD/P(ASD) | = 2/126 = 0,01587 | | = 1,59% |
| P(bhutan) ASD/P(ASD) | = 1/126 = 0,00793 | | = 0,79% |
| P(ghana) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(italy) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(egypt) ASD/P(ASD) | = 2/126 = 0,01587 | | = 1,59% |
| P(libya) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(isle of man) ASD/P(ASD) | = 1/126 = 0,00793 | | = 0,79% |
| P(outlying islands) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(afghanistan) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(Austria) ASD/P(ASD) | = 1/126 = 0,00793 | | = 0,79% |
| P(germany) ASD/P(ASD) | = 1/126 = 0,00793 | | = 0,79% |
| P(ireland) ASD/P(ASD) | = 1/126 = 0,00793 | | = 0,79% |
| P(malaysia) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(nepal) ASD/P(ASD) | = 1/126 = 0,00793 | | = 0,79% |
| P(netherlands) ASD/P(ASD) | = 0/126 = 0 | | = 0% |
| P(nigeria) ASD/P(ASD) | = 1/126 = 0,00793 | | |

| | | |
|----------------------------|-------------------|---------|
| P(Saudi arabia) ASD/P(ASD) | = 1/126 = 0,00793 | = 0,79% |
| P(romania) ASD/P(ASD) | = 0/126 = 0 | = 0% |
| P(syria) ASD/P(ASD) | = 0/126 = 0 | = 0% |

Sehingga ketika dijumlahkan hasilnya adalah 100%.

Probabilitas kota asal dengan keterangan “normal”
 $P(\text{jordan})|\text{normal}/P(\text{normal}) = \text{jumlah data "ya" dengan keterangan "normal"} / \text{jumlah data "normal"}$

| | | |
|------------------------------------|--------------------|----------|
| P(jordan) normal/P(normal) | = 8/122 = 0,06557 | = 6,56% |
| P(United States) normal/P(normal) | = 11/122 = 0,09016 | = 9,02% |
| P(United Kingdom) normal/P(normal) | = 25/122 = 0,20491 | = 20,49% |
| P(bahrain) normal/P(normal) | = 0/122 = 0 | = 0% |
| P(United Arab) normal/P(normal) | = 5/122 = 0,04098 | = 4,10% |
| P(europe) normal/P(normal) | = 0/122 = 0 | = 0% |
| P(malta) normal/P(normal) | = 0/122 = 0 | = 0% |
| P(bulgaria) normal/P(normal) | = 0/122 = 0 | = 0% |
| P(south africa) normal/P(normal) | = 0/122 = 0 | = 0% |
| P(india) normal/P(normal) | = 25/122 = 0,20491 | = 20,49% |
| P(iraq) normal/P(normal) | = 2/122 = 1,63934 | = 1,64% |
| P(new zealand) normal/P(normal) | = 8/122 = 0,06557 | = 6,56% |
| P(australia) normal/P(normal) | = 11/122 = 0,09016 | = 9,02% |
| P(georgia) normal/P(normal) | = 2/122 = 0,00819 | = 0,82% |
| P(armenia) normal/P(normal) | = 0/122 = 0 | = 0% |

| | |
|--------------------------------------|------------------------------|
| P(turkey) normal/P(normal) | = 2/122 = 1,63934 = 1,64% |
| P(pakistan) normal/P(normal) | = 0/122 = 0 |
| P(canada) normal/P(normal) | = 2/122 = 1,63934 = 1,64% |
| P(oman) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(brazil) normal/P(normal) | = 0/122 = 0 |
| P(south korea) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(costa rica) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(sweden) normal/P(normal) | = 0/122 = 0 |
| P(philippines) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(argentina) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(japan) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(bangladesh) normal/P(normal) | = 4/122 = 0,03278 = 3,28% |
| P(bhutan) normal/P(normal) | = 0/122 = 0 |
| P(ghana) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(italy) normal/P(normal) | = 2/122 = 1,63934 = 1,64% |
| P(egypt) normal/P(normal) | = 0/122 = 0 |
| P(libya) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(isle of man) normal/P(normal) | = 0/122 = 0 |
| P(mexico) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(outlying islands) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(afghanistan) normal/P(normal) | = 2/122 = 1,63934 = 1,64% |
| P(austria) normal/P(normal) | = 0/122 = 0 |
| P(germany) normal/P(normal) | = 0/122 = 0 |

| | |
|----------------------------------|------------------------------|
| P(ireland) normal/P(normal) | = 0/122 = 0 |
| P(malaysia) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(nepal) normal/P(normal) | = 0/122 = 0 |
| P(netherlands) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(nigeria) normal/P(normal) | = 0/122 = 0 |
| P(Saudi arabia) normal/P(normal) | = 0/122 = 0 |
| P(romania) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |
| P(syria) normal/P(normal) | = 2/122 = 0,00819 = 0,82% |

Sehingga ketika dijumlahkan hasilnya adalah 100%.

g. Atribut Pernah Skinning

Probabilitas pernah skinning dengan keterangan “ASD”
 $P(ya)|ASD/P(ASD)$ = jumlah data “ya” dengan keterangan “ASD” dibagi jumlah data “ASD”

$$P(ya)|ASD/P(ASD) = 3/126 = 0,02381 = 2,38\%$$

$$P(tidak)|ASD/P(ASD) = 123/126 = 0,97619 = 97,62\%$$

Sehingga ketika dijumlahkan hasilnya adalah 100%.

Probabilitas pernah skinning dengan keterangan “normal”

$P(ya)|normal/P(normal)$ = jumlah data “ya” dengan keterangan “normal” dibagi jumlah data “normal”

$$P(ya)|normal/P(normal) = 3/122 = 0,02459 = 2,46\%$$

$$P(tidak)|normal/P(normal) = 119/122 = 0,97541 = 97,54\%$$

Sehingga ketika dijumlahkan hasilnya adalah 100%

h. Atribut Kedekatan

Probabilitas kedekatan dengan keterangan “ASD”

$P(\text{orang tua})|ASD/P(ASD)$ = jumlah data “orang tua” dengan keterangan “ASD” dibagi jumlah data “ASD”

$$P(\text{orang tua})|ASD/P(ASD) = 112/126 = 0,88889 = 88,89\%$$

$$P(\text{relatif})|ASD/P(ASD) = 5/126 = 0,88889 = 3,97\%$$

$$P(\text{ahli medis})|ASD/P(ASD) = 7/126 = 0,05556 = 5,56\%$$

$$P(\text{sendiri})|ASD/P(ASD) = 2/126 = 0,01587 = 1,59\%$$

Sehingga ketika dijumlahkan hasilnya adalah 100%

Probabilitas kedekatan dengan keterangan “normal”
 $P(\text{orang tua})|normal/P(normal)$ = jumlah data “orang tua” dengan keterangan “normal” dibagi jumlah data “normal”

$$P(\text{orang tua})|normal/P(normal) = 101/122 = 0,82786 = 82,79\%$$

$$P(\text{relatif})|normal/P(normal) = 12/122 = 0,09836 = 9,84\%$$

$$P(\text{ahli medis})|normal/P(normal) = 6/122 = 0,04918 = 4,92\%$$

$$P(\text{sendiri})|normal/P(normal) = 3/122 = 0,02459 = 2,46\%$$

4. Kalikan semua variabel kelas “ $P(X|H) \times P(H)$ ” pada data testing.

Tabel 2. Data Testing

| usi a | jk | ket uru nan | pen yak it_k uni ng | riw aya t_k elu arg a_a utis | kota | per nah _sk rini ng | ke de ka ta n | dia gno sa |
|----------|----|--------------------------------|---------------------------------|--|-----------------------------|---------------------------------|---------------------------|------------------|
| 10 | 1 | Wh ite- Eur ope an | ya | tida k | 'Unit ed King dom' | tida k | se n di ri | ? |

- Probabilitas ASD dan normal :
 $P(\text{diagnosis}="ASD") = 126/248$
 $P(\text{diagnosis}="normal") = 122/248$

- Probabilitas usia10
 $P(\text{usia10}="ASD") = 9/126$
 $P(\text{usia10}="normal") = 8/122$

- Probabilitas jenis kelamin
 $P(l="ASD") = 90/126$
 $P(l="normal") = 38/122$

- Probabilitas keturunan
 $P(\text{white-european}="ASD") = 62/126$
 $P(\text{white-european}="normal") = 46/122$

- Probabilitas penyakit kuning
 $P(\text{ya}="ASD") = 31/126$
 $P(\text{ya}="normal") = 92/122$

- Probabilitas riwayat keluarga autisme
 $P(\text{tidak}="ASD") = 105/126$
 $P(\text{tidak}="normal") = 98/122$

- Probabilitas kota asal
 $P(\text{united kingdom}="ASD") = 24/126$
 $P(\text{united kingdom}="normal") = 25/122$

- Probabilitas pernah melakukan skrining
 $P(\text{tidak}="ASD") = 123/126$
 $P(\text{tidak}="normal") = 119/122$

- Probabilitas kedekatan
 $P(\text{sendiri}="ASD") = 2/126$
 $P(\text{sendiri}="normal") = 3/122$

Kemudian, kalikan semua hasil variabel ASD dan normal pada setiap atribut yang digunakan.

$$\text{Variabel ASD} = 9/126 \times 90/126 \times 62/126 \times 31/126 \times 105/126 \times 24/126 \times 123/126 \times 2/126 = 0.001519182$$

$$\text{Variabel normal} = 8/122 \times 38/122 \times 46/122 \times 92/122 \times 98/122 \times 25/122 \times 119/122 \times 3/122 = 0.002292846$$

5. Bandingkan hasil per kelas.

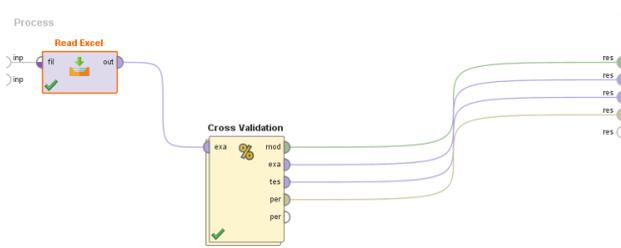
Setelah dikalikan berdasarkan variabel ASD dan normal, nilai terbesar yang dihasilkan adalah variabel normal, maka data testing diatas pada tabel 2, diagnosanya adalah normal.

3.2. Implementasi Dengan Rapidminer

3.2.1. Proses pemodelan

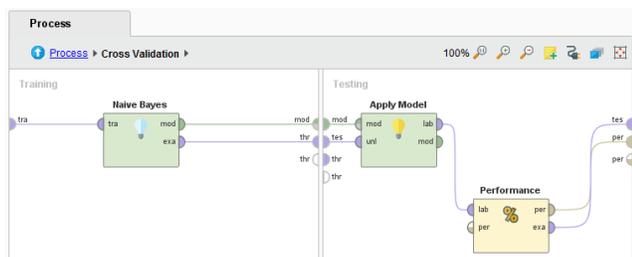
Berikut adalah pengolahan data Autisme pada Anak dengan menggunakan *naive bayes* pada *RapidMiner*:

1. *Import dataset* dan uji dengan *cross validation* untuk melakukan pengujian model.



Gambar 2. Cross Validation

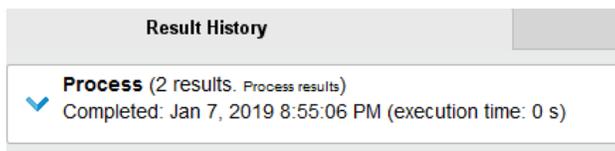
2. Didalam *cross validation*, masukkan model analisa datanya, yaitu menggunakan *naive bayes*, gunakan komponen *apply model* dan *performance* untuk melihat *performance* dari model yang digunakan.



Gambar 3. Model Naive Bayes

3.2.2. Hasil Pengujian Metode Naive Bayes

1. Result Overview



Gambar 4. Result Overview

Gambar 4 menunjukkan bahwa waktu eksekusi dari *dataset* yang diujikan berjalan selama 0 *second*, artinya tidak lama inilah salah satu kelebihan dari algoritma *naive bayes*.

2. Confusion Matrix

Berdasarkan data yang diujikan ke dalam rapid miner dengan algoritma *naive bayes*, berjumlah 248 data dengan 8 atribut dan 1 label yang telah dijelaskan sebelumnya, mendapatkan akurasi sebesar 60,48% dengan uraian seperti pada gambar :

accuracy: 60.53% +/- 8.89% (micro average: 60.48%)

| | true normal | true ASD | class precision |
|--------------|-------------|----------|-----------------|
| pred. normal | 66 | 42 | 61.11% |
| pred. ASD | 56 | 84 | 60.00% |
| class recall | 54.10% | 66.67% | |

Gambar 5. Performance Vector

Dari gambar diatas, kita dapat mengetahui jumlah *true positive* (TP) adalah 66 jumlah anak yang diklasifikasikan normal dan *false negative* (FN) sebanyak 42 orang anak yang masuk ke dalam klasifikasi normal tetapi masuk ke kelas ASD. jumlah *true negative* (TN) adalah 84 jumlah anak yang diklasifikasikan ASD dan *false positive* (FP) sebanyak 56 orang anak yang masuk ke dalam klasifikasi ASD tetapi masuk ke kelas normal. Berdasarkan *confusion matrix*, dihasilkan nilai akurasi sebesar 60,48% dengan perhitungan sebagai berikut :

$$\begin{aligned}
 Accuracy &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{(84 + 66)}{(84 + 42 + 66 + 56)} \times 100\% \\
 &= \frac{150}{248} \times 100\% \\
 &= 60,48\%
 \end{aligned}$$

3. Simple Distribution

Dari hasil pemodelan didapatkan *simple distribution* untuk label atribut dari masing-masing kelas ASD dan normal seperti gambar berikut :

SimpleDistribution

Distribution model for label attribute diagnosa

Class normal (0.492)
 8 distributions

Class ASD (0.508)
 8 distributions

Gambar 6. *Simple Distribution*

Artinya, dari 248 *dataset Autism* pada Anak, menghasilkan 0,492 = 49,2 % untuk kelas normal, dan 0,508 = 50,8% untuk kelas ASD.

4. KESIMPULAN & SARAN

4.1. Kesimpulan

Dari paparan yang telah dijelaskan diatas, dapat ditarik kesimpulan bahwa :

1. Teknik data mining berupa klasifikasi yang dilakukan pada 248 *dataset autism child* data atau data *autism* pada anak-anak yang diambil dari UCI *repository* yang dianalisis menggunakan algoritma *naive bayes* menghasilkan *accuracy* sebesar 60,48%.
2. Dengan *accuracy* sebesar 60,48%, kelas diagnosis ASD lebih besar dibandingkan dengan kelas diagnosis normal, yaitu 50,8 % dibanding 49,2%
3. Dengan menggunakan teknik data mining pengklasifikasian gejala-gejala *Autism Spectrum Disorders* atau ASD pada anak-anak, dapat memprediksi apakah gejala tersebut masih dalam kategori normal atau tidak (terkena gangguan), sehingga para orang tua bisa mendeteksi sejak dini mengenai gejala ini dan diharapkan segera melakukan penanganan dengan cepat dan tepat.

4.2. Saran

1. Dalam penelitian ini, klasifikasi menggunakan algoritma *Naive Bayes* saja, diharapkan peneliti selanjutnya dapat mengkomparasi beberapa algoritma sebagai perbandingan algoritma mana yang cocok untuk *dataset* seperti ini.
2. Akan lebih kompleks apabila seluruh atribut yang tersedia dalam *dataset* tanpa dilakukan *cleaning* data terlebih dahulu kemudian dianalisis dan atau bisa ditambahkan jumlah *dataset*nya itu sendiri dengan melakukan riset dari data lokal.
3. Pengimplementasian penelitian ini juga baik dan praktis apabila dibuatkan sebuah aplikasi *mobile* yang mendukung supaya mudah digunakan, cukup dengan memasukkan data testingnya saja, kemudian mesin berjalan dan memberikan diagnosa sesuai gejala yang diinputkan dan diolah oleh perhitungan (algoritma) yang digunakan dalam aplikasi tersebut.

DAFTAR PUSTAKA :

- [1] Demirhan, A. (2018). PERFORMANCE OF MACHINE LEARNING METHODS IN DETERMINING THE, 79–84. <https://doi.org/10.22531/muglajsci.422546>
- [2] Hasibuan, N. A., Silalahi, N., Nasution, S. D., Sutiksno, D. U., Nurdianto, H., Buulolo, E., ... Mining, A. D. (2017). IMPLEMENTASI DATA MINING UNTUK PENGATURAN LAYOUT, 4(4), 6–11.
- [3] Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, 7(1), 59–64.
- [4] Sumber, C. V., & Telekomunikasi, U. (1978). Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di, 81–92.
- [5] Thabtah, F. (2017). Autism Spectrum Disorder Screening : Machine Learning Adaptation and DSM-5 Fulfillment, 1–6.
- [6] Thabtah, F. (2018). Machine learning in autistic spectrum disorder behavioral research : A review and ways forward. *Informatics for Health and Social Care*, 0(0), 1–20. <https://doi.org/10.1080/17538157.2017.1399132>