

# Penerapan Algoritma Support Vector Machine Untuk Mendeteksi Autisme

Miftahul Khoiriah\*, Rakhmat Kurniawan

Fakultas Sains Dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: <sup>1,\*</sup>khoiriahmiftahul8@email.com, <sup>2</sup>rakhmat.kr@uinsu.ac.id

Email Penulis Korespondensi: khoiriahmiftahul8@gmail.com

Submitted: 26/07/2024; Accepted: 08/08/2024; Published: 09/08/2024

**Abstrak**—Autisme merupakan salah satu jenis gangguan perkembangan yang dapat menyebabkan suatu kondisi neurologis mengganggu fungsi otak dan berdampak pada proses pertumbuhan seseorang, keterampilan komunikasi, dan kemampuan interaksi sosial. Secara umum gangguan spektrum autisme dapat dideteksi pada bayi sejak dini sebagai 6 bulan. Hal yang mengganggu perkembangan anak tersebut terjadi karena susunan fungsi otak yang terganggu. Disabilitas yang tersebar luas ini digambarkan sebagai gangguan spektrum karena ukuran yang cukup besar dengan variasi bagaimana seseorang memanasifestasikan gejala dan Tingkat keparahannya. Dengan melakukan pendeteksian ini, dapat memudahkan orang tua mengetahui apakah anak mengidap penyakit autis atau tidak sehingga mengetahui tindakan apa yang akan dilakukan. Penelitian ini dilakukan menggunakan metodologi penelitian bersifat kuantitatif, dimana pendekatan penelitian tersebut menitikberatkan pada pengumpulan dan analisis data yang dapat diukur dalam bentuk numerik dengan teknik statistik sehingga mendapatkan angka dan penyamarataan. Pendekatan ini melibatkan hubungan fenomena dan sebab akibat menggunakan sample yang lebih besar. Setelah tahapan-tahapan sebelumnya selesai, selanjutnya melanjutkan pengujian dari hasil prediksi dengan menggunakan data testing dan akurasi hingga mendapatkan hasil klasifikasi. Dari hasil klasifikasi diatas, nilai klasifikasi yang dihasilkan mencapai 100% menggunakan data uji dan menggunakan nilai akurasi. Algoritma *Support Vector Machine* (SVM) dengan kernel linear telah diterapkan pada dataset autisme pada anak. Model ini berhasil memisahkan kelas dengan baik, menunjukkan bahwa SVM merupakan algoritma yang efektif untuk masalah klasifikasi ini.

**Kata Kunci:** *Support Vector Machine*; Autisme; *Machine Learning*; ASD; Python

**Abstract**—Autism is a type of developmental disorder that can cause a neurological condition to disrupt brain function and impact a person's growth process, communication skills and social interaction abilities. In general, autism spectrum disorders can be detected in babies as early as 6 months. Things that interfere with a child's development occur because the structure of brain function is disturbed. This widespread disability is described as a spectrum disorder due to the considerable variation in how an individual manifests symptoms and their severity. By carrying out this detection, it can make it easier for parents to know whether their child has autism or not so they know what action to take. This research was conducted using a quantitative research methodology, where the research approach focuses on collecting and analyzing data that can be measured in numerical form using statistical techniques to obtain numbers and generalize. This approach involves the relationship between phenomena and cause and effect using a larger sample. After the previous stages are completed, then continue testing the prediction results using testing and accuracy data to obtain classification results. From the classification results above, the resulting classification value reaches 100% using test data and using accuracy values. Support Vector Machine (SVM) algorithm ) with a linear kernel has been applied to a dataset of autism in children. This model succeeded in separating classes well, showing that SVM is an effective algorithm for this classification problem.

**Keywords:** *Support Vector Machine*; Autisme; *Machine Learning*; ASD; Python

## 1. PENDAHULUAN

Menurut WHO (2023), ada 1 dari 100 anak-anak seluruh dunia menderita *Autisme*. Anak yang menderita Autisme dapat dibedakan pada proses pertumbuhan dan perkembangan fisik dengan anak normal pada umumnya. Pada bulan Mei 2014, Majelis Kesehatan Dunia ke-67 mengadopsi resolusi berjudul Upaya Komprehensif dan terkoordinasi untuk mengatasi Gangguan Spektrum Autisme, yang didukung oleh lebih dari 60 negara. Resolusi tersebut mengajak WHO untuk bekerja sama dengan negara-negara anggota dan lembaga mitra untuk memperkuat kapasitas nasional dalam mengatasi ASD dan gangguan perkembangan lainnya[1].

ASD (*Autism Spektrum Disorder*) ialah gangguan yang terjadi sejak batita, yang menjadikan kurangnya perkembangan dan membuat anak sulit membangun lingkungan sosial serta hubungan komunikasi pada anak disekitarnya[2]. Hasil penelitian melibatkan 132 ibu dan anak-anak mereka yang memiliki ASD ketika mereka berusia 9 bulan, 3 tahun, dan 5 tahun, membahas tentang masalah perilaku pada anak-anak dengan ASD tidak bersifat dua arah [3].

Autisme merupakan gabungan dari sikap sosial yang terganggu, bermasalah pada komunikasi serta perilaku yang berulang yang bisa dialami oleh seseorang dari batita sampai dewasa. Autisme bisa terjadi karena memiliki kelainan otak yang terjadi karena kurangnya sel saraf, kemungkinan terjadi pada masa kehamilan atau disaat persalinan[4]. Diagnosis autisme merupakan tindakan yang pertama agar memahami kondisi anak untuk melakukan penanganan autisme sejak dini. Deteksi gangguan autisme sejak dini adalah proses yang penting pada pertumbuhan dan perkembangan pada anak[5].

Gejala ASD umumnya muncul pada saat bayi dan balita, sehingga menimbulkan tantangan bagi individu, keluarga, serta masyarakat secara umum. ASD mempunyai tingkat heritabilitas yang signifikan, dimana sering terjadi dalam keluarga[6]. Dengan meningkatnya jumlah kasus ASD pada anak-anak, penting untuk melakukan

deteksi dini dan memberikan perawatan yang tepat[7]. Untuk mengurangi kesalahan dalam proses diagnosis, diperlukan sistem yang memberikan dukungan melalui pengetahuan profesional, sehingga autisme dapat dideteksi dengan lebih mudah dan akurat[8].

Autisme merupakan salah satu jenis gangguan perkembangan yang dapat menyebabkan suatu kondisi neurologis mengganggu fungsi otak dan berdampak pada proses pertumbuhan seseorang, keterampilan komunikasi, dan kemampuan interaksi sosial. Secara umum gangguan spektrum autisme dapat dideteksi pada bayi sejak umur 6 bulan[9]. Penyebab perkembangan anak yang terganggu terjadi disebabkan dengan susunan fungsi otak yang terganggu. Anak yang menderita autisme disebabkan terjadinya kelainan pada otak karena jumlah sel syaraf yang tumbuh pada masa kehamilan maupun sesudah masa persalinan. Disabilitas yang tersebar luas ini digambarkan sebagai gangguan spektrum karena ukuran yang cukup besar dengan variasi bagaimana seseorang memanifestasikan gejala dan Tingkat keparahannya[10].

Namun masih banyak orang yang kurang mengerti tentang gejala-gejala yang terjadi jika seseorang terkena gangguan autisme, sehingga terlambat untuk ditangani serta tenaga medis yang fokus pada perkembangan otak penderita autisme sangat sedikit[11]. Ada beberapa yang menunjukkan mengenai orang yang memiliki kebutuhan khusus seperti autisme ini masih banyak yang diabaikan, itu yang menyebabkan terhambatnya penanganan dan pengobatan bagi penderita autisme[12].

Agar memudahkan untuk mendeteksi gejala yang terjadi dengan perkembangan ilmu komputer dapat membantu untuk mendeteksi gejala tersebut menggunakan *machine learning*. Pada bidang kesehatan, *machine learning* membantu saat mendeteksi penyakit lebih cepat menggunakan hasil prediksi yang telah diterima dokter, sehingga dapat dilakukan penanganan lebih cepat[13]. Penerapan *machine learning* pasti membutuhkan data dari penggunaannya. *Machine learning* mempunyai peran dalam pengembangan yang sangat luas, apalagi pada pengembangan data analitik[14]. Teknik *machine learning* ditawarkan agar memudahkan dalam pendeteksian secara otomatis, seperti yang diketahui jika mendeteksi secara manual akan memakan waktu yang lama sehingga penanganan akan lambat[15].

Perbedaan penelitian ini dengan penelitian terdahulu yang dilakukan oleh Riki Supriyadi., et al mendapatkan nilai akurasi menggunakan *support vector machine* sebesar 81,5603%, dan menggunakan *naïve bayes* mendapatkan nilai akurasi sebesar 96,4539%[16]. Penelitian terdahulu yang dilakukan oleh Erina Seviyanti Dewi menggunakan *Naïve Bayes* mendapatkan hasil penelitian yang menggunakan nilai *batchsize* 50,100 dan 200 dengan rasio data latih dan uji 9:1, 4:1, 7:3, 3:2, 1:1, 2:3, 3:7, 1:4 dan 1:9. Sehingga hasil akurasi yang didapat menggunakan *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* dengan rasio data 1:1 mendapatkan nilai 98.6301%. Jika berdasarkan algoritma *naïve bayes* pada rasio 7:3 dan 2:3 mendapatkan hasil bernilai 1. Dimana saat pelatihan dengan rasio 1:9 adalah yang terendah, dan saat pengujian dengan rasio 9:1 adalah yang terendah[15].

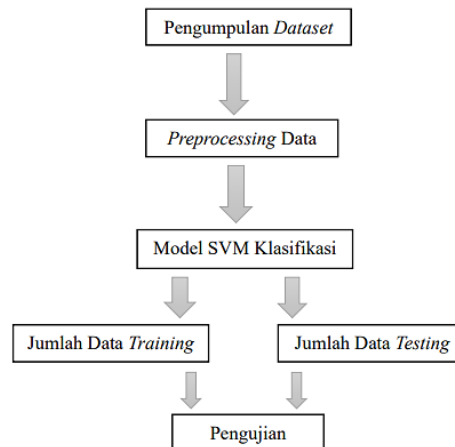
Penelitian terdahulu juga yang dilakukan oleh Bayu Sugara., et al menggunakan perbandingan metode antara Algoritma C4.5 Dan Nive Bayes mendapatkan hasil penelitian menggunakan metode cross validation dengan aplikasi *ripedminer*. Dimana algoritma C4.5 menampilkan hasil akurasi dengan nilai 72%, sedangkan algoritma *naïve bayes* menampilkan hasil akurasi dengan nilai 73,33%[9]. Dan penelitian yang juga telah dilakukan oleh Lika Alaika., et al menggunakan perbandingan antara metode *support vector machine* dan *correlation-base feature selection* mendapatkan hasil akurasi 94,91% dengan *support vector machine* dan 96,61% dengan menggunakan *support vector machine* dan *correlation-base feature selection*[14]. Sedangkan penelitian ini mendapatkan nilai akurasi mencapai kurang lebih 100%.

Penelitian ini dilakukan dengan tujuan untuk mendeteksi suatu penyakit gangguan kesehatan seperti autisme menjadi lebih mudah dan cepat. Pendeteksian penderita autisme dilakukan menggunakan karakteristik gejala yang dialami oleh penderita. Hasil dari penelitian akan menampilkan nilai akurasi kurang lebih sebesar 100%, sehingga ketika nilai yang ditampilkan sebesar 97% maka hasil deteksi sudah dapat dikatakan benar. Dengan nilai tersebut diharapkan dapat memudahkan para medis dan membantu untuk mendeteksi autisme, Sistem ini juga diharapkan dapat membantu dalam memberikan rekomendasi dan merancang model diagnostik yang mampu mengevaluasi semua hambatan dan tantangan yang dihadapi oleh anak-anak.

## **2. METODOLOGI PENELITIAN**

### **2.1 Tahapan Penelitian**

Penelitian ini dilakukan menggunakan metodologi penelitian bersifat kuantitatif, dimana pendekatan penelitian tersebut menitikberatkan pada pengumpulan dan analisis data yang dapat diukur dalam bentuk numerik dengan teknik statistik sehingga mendapatkan angka dan penyamarataan. Pendekatan ini melibatkan hubungan fenomena dan sebab akibat menggunakan sample yang lebih besar. Penelitian ini memiliki tahapan yang terdapat pada gambar 1 berikut.



Gambar 1. Tahapan Penelitian

a. Pengumpulan Dataset

Pada penelitian ini data klasifikasi yang digunakan adalah dataset public dari *autism dataset dor toddlers*, jenis data yang diambil dari Kaggle yang dapat diakses menggunakan link [https://www.kaggle.com/datasets/vaishnavi\\_sirigiri/autism-dataset-for-toddlers](https://www.kaggle.com/datasets/vaishnavi_sirigiri/autism-dataset-for-toddlers) yang memiliki data sebanyak 1054 data dengan periode tahun 2024. Dataset tersebut memberi informasi yang terdiri dari atribut, tipe dan deskripsi.

b. Preprocessing Data

Pada preprocessing data memiliki tahapan yang dilakukan yaitu pembersihan data, seleksi data, transformasi data, integrasi data sampai mendapatkan informasi dengan mengevaluasi pola dari data tersebut. Tujuan preprocessing data adalah meningkatkan kelengkapan akurasi dari data berbasis *database* yang telah disaring, dan memperbaiki kesalahan termasuk nilai yang hilang, sehingga data dapat divalidasi menjadi sebuah format.

c. Model SVM Klasifikasi

Terdapat algoritma yang diterapkan pada data ini yaitu *Support Vector Machine*. Untuk melakukan prediksi pada data yang digunakan terbagi menjadi dua, yaitu data *training* dan data *testing*. Pada pembagian data *training* 70% dan data *testing* 30%, sehingga untuk data *training* sebanyak 738 data dan data *testing* sebanyak 316 data. Dengan menggunakan algoritma tersebut memiliki tujuan agar dapat menghasilkan model terbaik yang mampu melakukan prediksi dengan nilai tinggi dan hasil yang optimal terhadap *Autism Dataset For Toddlers*.

d. Pengujian

Setelah selesai seluruh tahapan, selanjutnya adalah melakukan pengujian yang bertujuan untuk mendapatkan hasil dan menunjukkan apakah dengan menggunakan model algoritma tersebut bisa menghasilkan akurasi yang tepat. Ketika pengujian berhasil, maka saat melakukan prediksi autism bisa mengetahui apa faktor paling utama yang mempengaruhi seseorang mengidap ASD sehingga dapat dilakukan penanganan yang tepat.

## 2.2 Machine Learning

*Machine learning* merupakan teknik yang mampu menghasilkan prediksi tentang hasil suatu sistem berdasarkan model yang telah dibuat sebelumnya[17]. *Machine learning* dapat memudahkan dalam menyelesaikan permasalahan yang terjadi salah satunya adalah klasifikasi pada sebuah penyakit. Untuk mengoptimalkan pekerjaan pengklasifikasi perlu dilakukan dengan memilih fitur yang relevan menggunakan pemilihan fitur. Dengan menghilangkan fitur yang tidak perlu yang akan berdampak pada keakuratan hasil, pemilihan fitur mencoba untuk menurunkan dimensi atribut[18].

Pada *machine learning* terdapat pendekatan menggunakan data dan hasil yang dapat memprediksi hasil dengan benar yaitu *supervised learning*. Pada pendekatan ini biasanya digunakan untuk mengerjakan dua macam permasalahan yaitu klasifikasi dan regression. Pada *supervised learning* biasanya memberi informasi mengenai hasil dari perbandingan klasifikasi yang dikerjakan oleh sistem dengan klasifikasi yang sebenarnya sering disebut *confusion matrix*. Untuk mengukur *performance* dari *confusion matrix*, digunakan persamaan (1) agar menemukan nilai prediksi.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Penjumlahan antara jumlah prediksi benar (TP) dan jumlah prediksi negatif yang benar(TN), dibagi dengan penjumlahan antara jumlah prediksi benar (TP) jumlah prediksi negatif benar(TN) jumlah prediksi positif salah(FP) dan jumlah prediksi negatif salah (FN).

Untuk mengukur *precision* yang merupakan prediksi benar positif menggunakan persamaan (2) dengan membandingkan hasil keseluruhan yang diprediksi positif.

$$precision = \frac{TP}{TP+FP} \tag{2}$$

Jumlah prediksi positif benar (TP) dibagi dengan penjumlahan antara jumlah prediksi positif benar (TP) dan jumlah prediksi positif salah (FP).

Untuk menemukan kembali hasil informasi dari prediksi benar positif dengan seluruh data yang benar positif dengan persamaan (3).

$$recall = \frac{TP}{TP+FN} \tag{3}$$

Jumlah prediksi positif benar (TP) dibagi dengan penjumlahan antara jumlah prediksi positif benar (TP) dan jumlah prediksi negatif salah (FN).

### 2.3 Support Vector Machine

*Support Vector Machine* (SVM) adalah suatu sistem yang menggunakan pembelajaran teori fungsi linear dengan sebuah fitur yang dilatih menggunakan Algoritma berdasarkan pada teori optimal[19]. SVM memiliki tujuan untuk memisahkan titik-titik data input dengan cara menemukan *hyperplane* terbaik pada ruang berdimensi-N. *Hyperplane* adalah sebuah garis yang memisahkan kedua kelas untuk menemukan menggunakan perhitungan *margin* dan mencari titik maksimumnya[20].

Untuk mendapatkan *hyperplane* yang tepat, langkah awal yang dilakukan adalah menotasikan data menjadi  $\vec{x}_i \in R^2$  dan menotasikan masing-masing label menjadi  $y_i \in \{1,0\}$  untuk  $i = 1,2,\dots,l$  dimana  $l$  merupakan jumlah dari data. Untuk memisahkan *hyperplane* dari kedua kelas yaitu 1 dan 0 dengan baik menggunakan asumsi yang dilakukan dengan persamaan (4)

$$\vec{w} \cdot \vec{x} + b = 0 \tag{4}$$

Vektor bobot  $\vec{w}$  yang dikalikan dengan vektor fitur  $\vec{x}$  lalu dijumlahkan dengan bias atau *intercept*  $b$ . Untuk merumuskan kelas 1 menggunakan pola  $\vec{x}_1$  merupakan pertidaksamaan (5)

$$\vec{w} \cdot \vec{x}_1 + b \leq 1 \tag{5}$$

Vektor bobot untuk menentukan orientasi serta kemiringan *hyperlane*  $\vec{w}$  yang menggunakan data latih ke- $l$  dengan vektor fitur  $\vec{x}$  yang mengevaluasi kondisi margin dan menentukan jarak *hyperlane* dari asal koordinat dengan  $b$ . Dengan kondisi  $\leq 1$  menunjukkan titik data  $\vec{x}$  berada pada sisi negatif dari *hyperlane* atau di dalam dan pada batas margin negatif. Dan untuk merumuskan kelas 0 menggunakan pertidaksamaan (6) dengan pola  $\vec{x}_1$

$$\vec{w} \cdot \vec{x}_1 + b \leq 0 \tag{6}$$

Vektor bobot untuk menentukan orientasi serta kemiringan *hyperlane*  $\vec{w}$  yang menggunakan data latih ke- $l$  dengan vektor fitur  $\vec{x}$  yang mengevaluasi kondisi margin dan menentukan jarak *hyperlane* dari asal koordinat dengan  $b$ . Dengan kondisi  $\leq 0$  menunjukkan titik data  $\vec{x}$  berada pada sisi negatif dari *hyperlane* atau pada *hyperlane* itu sendiri. Dengan memaksimalkan nilai pada *hyperplane* dan melakukan persamaan (7) pada titik dapat menemukan margin.

$$\frac{1}{\|\vec{w}\|} \tag{7}$$

Digunakan untuk menghitung margin maksimum antara dua kelas yang dipisahkan dengan *hyperlane*. Semakin kecil nilai  $\|\vec{w}\|$ , semakin besar margin antara dua kelas. Untuk mencari titik minimum dengan merumuskan menjadi *quadratic programming problem* menggunakan persamaan (8) dan mengamati *constraint* pada persamaan (9)

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \tag{8}$$

Digunakan untuk memaksimalkan margin antara dua kelas yang dipisahkan dengan *hyperlane*. Dengan meminimalkan  $\|\vec{w}\|^2$  akan langsung memaksimalkan margin.  $\frac{1}{2}$  digunakan untuk memudahkan proses optimasi dalam fungsi objektif yang akan membantu perhitungan menjadi lebih sederhana.

$$y_i(\vec{w} \cdot \vec{x}_i) - 1 \geq 0, \forall i \tag{9}$$

$y_i$  merupakan label kelas dari data latih ke- $i$ , dengan nilai 1 dan -1.  $(\vec{w} \cdot \vec{x}_i)$  merupakan nilai proyeksi dari titik data pada *hyperlane*.  $-1 \geq 0$  ialah kondisi yang memastikan bahwa proyeksi pada titik data, setelah dikalikan dengan label lalu dikurangi 1 adalah positif atau nol, berarti titik data berada di luar atau pada margin

Permasalahan bisa dipecahkan menggunakan bermacam komputasi salah satunya *Langrange Multiplier*, dengan persamaan (10)

$$L(\vec{w}, b, a) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i((\vec{w} \cdot \vec{x}_i + b) - 1)), i = 1, 2, \dots, l \tag{10}$$

$\frac{1}{2} \|\vec{w}\|^2$  digunakan untuk memaksimalkan margin antara dua kelas data. *Langrange Multiplier* akan menangani *constraint* dengan  $\alpha_i$ , jika *constraint* tidak aktif maka akan menjadi 0 sebaliknya jika aktif maka positif.  $y_i((\vec{w} \cdot \vec{x}_i + b) - 1)$  digunakan untuk setiap titik data dengan *constraint margin* yang memastikan setiap titik berada disisi pada margin yang benar.

Dari persamaan diatas yang termasuk *langrange multiplier* adalah  $\alpha_i$ , dengan memiliki nilai nol atau  $\alpha_i \geq 0$ . Untuk meminimumkan nilai  $L$  atas  $\vec{w}$  dan  $b$  serta memaksimumkan nilai  $L$  kepada  $\alpha_i$  menggunakan persamaan (11) yang bersifat mengoptimalkan

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j \tag{11}$$

$\sum_{i=1}^l \alpha_i$  digunakan untuk melakukan penjumlahan semua *langrange multiplier*  $\alpha_i$ .  $\alpha_i, \alpha_j$  merupakan *langrange multiplier* pada titik data ke- $i$  dan ke- $j$ .  $y_i y_j$  digunakan untuk label kelas dari titik data ke- $i$  dan ke- $j$  dengan nilai 1 atau -1. Dan untuk hasil kali skalar antara vektor fitur dari titik data ke- $i$  dan ke- $j$  menggunakan  $\vec{x}_i \cdot \vec{x}_j$ .

Menurut sifatnya titik optimal  $L = 0$ , untuk memodifikasi menjadi maksimum masalah yang hanya ada  $\alpha_i$  dapat dilakukan dengan persamaan (12) dimana hitungan  $\alpha_i$  mendapatkan hasil yang bersifat positif. Data positif itu lah yang disebut *Support Vector*.

$$\alpha_i \geq 0 (i = 1, 2, \dots, l) \sum_{i=1}^l \alpha_i y_i = 0 \tag{12}$$

*Constraint*  $\alpha_i \geq 0$  memastikan setiap  $\alpha_i$  tidak negatif. *Constraint*  $\sum_{i=1}^l \alpha_i y_i = 0$  memastikan keseimbangan pada ruang solusi dan bagain penting dari formula dual *support vector machine*.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Penerapan Algoritma Support Vector Machine

Pada penelitian ini menggunakan dataset yang diambil dari Kaggle, memiliki keseluruhan jumlah sebesar 1054 data dengan 80% *training data* dan 20% *testing data*. Dataset yang digunakan pada penelitian ini memiliki 18 atribut dengan 10 pertanyaan yang diperoleh dari instrument dan akan digunakan untuk mendeteksi autisme. Dapat dilihat pada Tabel 1 berikut atribut dan keterangan yang ada pada dataset penelitian ini.

**Tabel 1.** Atribut Pada Dataset Penelitian

Atribut	Tipe	Keterangan
Jawaban pertanyaan 1 (A1)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan
Jawaban pertanyaan 2 (A2)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan
Jawaban pertanyaan 3 (A3)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan
Jawaban pertanyaan 4 (A4)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan
Jawaban pertanyaan 5 (A5)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan
Jawaban pertanyaan 6 (A6)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan
Jawaban pertanyaan 7 (A7)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan
Jawaban pertanyaan 8 (A8)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan
Jawaban pertanyaan 9 (A9)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan
Jawaban pertanyaan 10 (A10)	Binary (0,1)	Jawaban pertanyaan berdasarkan kode skrining yang digunakan

Atribut	Tipe	Keterangan
Usia	Number	Usia anak (bulan)
Skor Hasil	Integer	1-10 (kurang dari atau sama dengan 3 tidak Autis. Lebih dari 3 Autis)
Jenis Kelamin	Character	Laki-laki atau Perempuan
Etnis	String	Etnis
Penyakit Kuning	Boolean (Yes Atau No)	Apakah memiliki penyakit kuning
Keluarga Pengidap Autis	Boolean (Yes Atau No)	Apakah anggota keluarga mengidap Autis
Yang Menyelesaikan Tes	String	Jawaban yang menyelesaikan oleh medis, orang tua, dan lain-lain
Tipe Tes	Boolean (Yes Atau No)	Termasuk ASD atau Tidak ASD (ditetakan otomatis oleh aplikasi ASDTests).(Yes atau No)

Pada penelitian ini, dataset dilakukan pelabelan terlebih dahulu untuk menentukan apakah data tersebut dikatakan negatif atau positif. Pada pelabelan, atribut yang digunakan adalah QChat-10-Score, jika nilai yang ditampilkan QChat-10-Score adalah lebih kecil dari 3 atau sama dengan 3 maka label menyatakan negatif. Pada tabel 2 menampilkan beberapa dataset yang sudah diberi label.

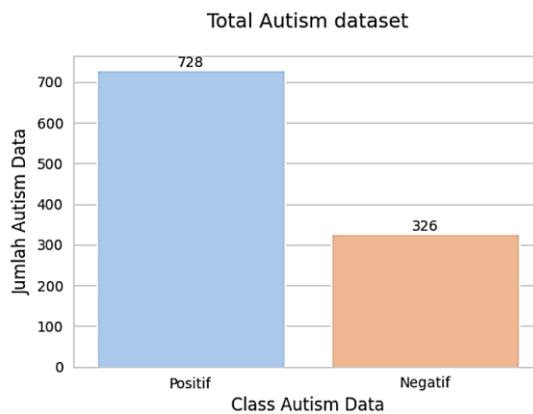
$$Qchat - 10score = A1 + A2 + A3 + A4 + A5 + A6 + A7 + A8 + A9 + A10$$

$$Qchat - 10score = 1 + 1 + 0 + 0 + 0 + 1 + 1 + 0 + 0 + 0 = 4$$

Tabel 2. Pelabelan Dataset Penelitian

No.	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Qchat-10-score	Label
1	0	0	0	0	0	0	1	1	0	1	3	Negatif
2	1	1	0	0	0	1	1	0	0	0	4	Positif
3	1	0	0	0	0	0	1	1	0	1	4	Positif
4	1	1	1	1	1	1	1	1	1	1	10	Positif
5	1	1	0	1	1	1	1	1	1	1	9	Positif
6	1	1	0	0	1	1	1	1	1	1	8	Positif

Pada perhitungan diatas dengan menjumlahkan keseluruhan A1-A10, sehingga mendapatkan hasil yang akan menentukan label pada setiap data seperti pada tabel 2. Jika nilai yang dihasilkan sama dengan 3 atau lebih kecil dari 3 maka data tersebut negatif, sedangkan jika nilai yang dihasilkan lebih dari 3 maka data tersebut positif. Pada gambar 2 terlihat grafik dari pelabelan yang telah dilakukan, dengan jumlah positif yang lebih besar yaitu 728 data dan jumlah negatif yang kecil yaitu 326 data.



Gambar 2. Grafik Pembagian Pelabelan Dataset

Untuk perhitungan menggunakan persamaan *support vector machine* sehingga mendapatkan *hyperlane* dengan menggunakan 2 *subset* yang ada pada data yang telah diberi label dengan mengkonversi label positif menjadi 1 dan negatif menjadi -1. Berikut pada tabel 3 dataset kecil yang akan digunakan untuk perhitungan.

Tabel 2. Dataset Kecil

Qchat-10-Score	Label
3	Negatif
6	Positif
4	positif

Qchat-10-Score	Label
7	Positif

Untuk memaksimalkan margin antara dua kelas, sehingga fungsi yang akan diminimalkan menggunakan persamaan

$$\min \frac{1}{2} \|\omega\|^2$$

Dengan kendala

$$y_i(\omega \cdot x_i + b) \geq 1$$

Untuk menyelesaikan sistem persamaan dari kendala yang diberikan, dengan mencari parameter  $\omega$  dan  $b$  pada 2 titik dari 2 kelas yang berbeda

Titik (3, -1) dan (6,1)

Sehingga membentuk sistem persamaan berikut lalu menghitung parameter  $\omega$  dan  $b$ :

Pada titik (3,-1)

$$-1(\omega \cdot 3 + b) = 1$$

$$-3\omega - b = 1$$

Pada titik (6,1)

$$1(\omega \cdot 6 + b) = 1$$

$$6\omega + b = 1 \quad (3.3)$$

Sehingga menemukan nilai  $\omega$  dan  $b$ , selanjutnya lakukan penyelesaian sistem persamaan tersebut dengan menjumlahkan kedua persamaan:

$$-3\omega - b + 6\omega + b = 1 + 1$$

$$3\omega = 2$$

$$\omega = \frac{2}{3}$$

Substitusikan  $\omega$  ke dalam salah satu persamaan:

$$6\left(\frac{2}{3} + b = 1\right)$$

$$4 + b = 1$$

$$b = 1 - 4$$

$$b = -3$$

Jadi, parameter optimal adalah:

$$\omega = \frac{2}{3}$$

$$b = -3$$

Dengan parameter tersebut, dapat ditemukan persamaan hyperlane adalah:

$$\frac{2}{3}x - 3 = 0$$

$$x = \frac{9}{2} = 4.5$$

Sehingga *hyperlane* yang ditemukan untuk memisahkan dua kelas dengan garis vertikal pada  $x = 4.5$ . Karena penelitian ini menggunakan kernel linier, setelah proses pelabelan lakukan perhitungan kernel menggunakan data yang telah diberi label menggunakan rumus (13), sehingga menghasilkan kernel matriks

$$K(x_i, x_j) = x_i \cdot x_j$$

$$K([0,0,0,0,0,0,1,1,0,1,3], [1,1,0,0,0,1,1,0,0,0,4])$$

$$K = (0 \times 1) + (0 \times 1) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 1) + (1 + 1) + (1 \times 0) + (0 \times 0) + (1 \times 0) +$$

$$(3 \times 4) = 0 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 12 = 13$$

Menggunakan kernel linier antara data pertama dan kedua menghasilkan nilai yaitu 13, pada tabel 2 menampilkan hasil beberapa data yang telah dilakukan perhitungan menggunakan kernel linier.

**Tabel 4.** Perhitungan Dengan Kernel Linier

	Data 1	Data 2	Data 3	Data 4	...	Data 1051	Data 1052	Data 1053	Data 1054
Data 1	11	13	15	33	...	16	30	11	19
Data 2	13	20	18	44	...	21	39	13	27
Data 3	15	18	20	44	...	21	40	15	26
Data 4	33	44	44	110	...	55	99	33	66
...	...	...	...	...	...	...	...	...	...
Data 1051	16	21	21	55	...	30	50	15	32
Data 1052	30	39	40	99	...	50	90	30	59
Data 1053	11	13	15	33	...	15	30	12	20
Data 1054	19	27	26	66	...	32	59	20	42

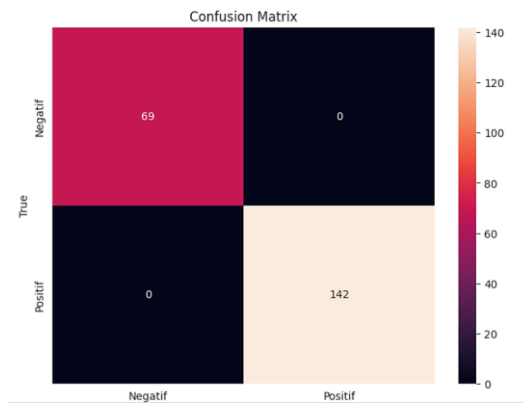
Pada tabel 4 dapat dilihat hasil perhitungan kernel linier pada data yang telah diberi label. Perhitungan antara data 1 dan data 1 sampai data 1054 dan data 1054 sehingga menghasilkan angka yang ada pada tabel.

### 3.1.1 Evaluasi Model

Sebelum melakukan pengujian, lakukan pembagian pada data uji menggunakan confusion matriks yang dapat dilihat pada tabel 5 serta pada gambar 3 dapat dilihat terdapat grafik yang dihasilkan dari pembagian data uji berikut.

**Tabel 5.** Pembagian Data Uji Dengan Confusion Matriks

	Prediksi Negatif	Prediksi Positif
Label Negatif (0)	69	0
Label Positif (1)	0	142



**Gambar 3.** Pembagian True Pada Data Uji

Pada gambar 3 yang menampilkan grafik untuk pembagian true pada data uji, untuk jumlah true negatif terlihat pada warna merah dan untuk jumlah true positif terlihat pada warna coklat. Setelah model dilatih, performa model dievaluasi menggunakan data uji. Matriks evaluasi yang digunakan meliputi akurasi, precision, recall, dan F1-score. Evaluasi ini bertujuan untuk memastikan bahwa model mampu mengklasifikasikan data dengan baik.

a. Akurasi

Pada proses perhitungan nilai akurasi menggunakan persamaan manual sebagai berikut:

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Total\ Data\ Uji} \times 100\%$$

$$Akurasi = \frac{2}{3} \times 100\% = 66.67\%$$

b. Precision

Pada perhitungan nilai precision menggunakan persamaan manual sebagai berikut:

$$Precision = \frac{TP}{TP+FP} = \frac{2}{2+1} = \frac{2}{3} = 67\%$$

c. Recall

Pada perhitungan nilai Recall menggunakan persamaan manual sebagai berikut:

$$Recall = \frac{TP}{TP+FN} = \frac{2}{2+0} = 1.00\text{ atau }100\%$$

d. F1-Score

Pada perhitungan nilai precision menggunakan persamaan manual sebagai berikut:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0.67 \times 1.00}{0.67 + 1.00} = 2 \times \frac{0.67}{1.67} = 2 \times 0.40 = 0.80 \text{ atau } 80\%$$

### 3.2 Pengujian

Setelah tahapan-tahapan sebelumnya selesai, selanjutnya melanjutkan pengujian dari hasil prediksi dengan menggunakan data *testing* dan akurasi hingga mendapatkan hasil klasifikasi pada gambar 5 berikut.

Hasil Klasifikasi Autism Dataset				
	precision	recall	f1-score	support
Negatif	1.00	1.00	1.00	69
Positif	1.00	1.00	1.00	142
accuracy			1.00	211
macro avg	1.00	1.00	1.00	211
weighted avg	1.00	1.00	1.00	211

**Gambar 4.** Hasil Klasifikasi Akurasi

Dari hasil klasifikasi diatas, nilai klasifikasi yang dihasilkan mencapai 100% menggunakan data uji dan menggunakan nilai akurasi dengan algoritma *Support Vector Machine*.

## 4. KESIMPULAN

Algoritma Support Vector Machine (SVM) dengan kernel linear telah diterapkan pada dataset autisme pada anak. Model ini berhasil memisahkan kelas dengan baik, menunjukkan bahwa SVM merupakan algoritma yang efektif untuk masalah klasifikasi ini. Hasil evaluasi model menunjukkan metrik performa yang baik. Akurasi, precision, recall, dan F1-score masing-masing memberikan nilai yang memuaskan, menunjukkan bahwa model mampu mendeteksi autisme pada anak dengan tingkat keakuratan yang tinggi. Beberapa fitur dalam dataset, seperti riwayat keluarga dengan ASD, jenis kelamin, dan jawaban dari beberapa pertanyaan terkait gejala autisme, terbukti memiliki pengaruh signifikan terhadap hasil prediksi model. Standarisasi data menggunakan StandardScaler terbukti penting untuk meningkatkan performa model. Hal ini menunjukkan bahwa preprocessing data adalah langkah krusial dalam membangun model yang akurat dan andal.

## REFERENCES

- [1] WHO, "autisme," WHO. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>, diakses 20 Mei 2024
- [2] N. Ratama and Munawaroh, "Implementasi Metode Fuzzy Tsukamoto Untuk Deteksi Dini Autisme Pada Balita Berbasis Android," *J. Inform. Rekayasa Elektron.*, vol. 3, no. 2, 2020.
- [3] Y. N. Lin, L. S. Iao, Y. H. Lee, and C. C. Wu, "Parenting Stress and Child Behavior Problems in Young Children with Autism Spectrum Disorder: Transactional Relations Across Time," *J. Autism Dev. Disord.*, vol. 51, no. 7, pp. 2381–2391, 2021, doi: 10.1007/s10803-020-04720-z.
- [4] H. Sulistyowati, D. Mayasari, and S. D. Hastining, "Pemerolehan Kosa Kata Anak Autism Spectrum Disorder (ASD)," *J. Obs. J. Pendidik. Anak Usia Dini*, vol. 6, no. 4, pp. 3091–3099, 2022, doi: 10.31004/obsesi.v6i4.2374.
- [5] T. Ghazi Pratama, A. Ridwan, and A. Prihandono, "Deteksi Dini Asd (Autism Spectrum Disorder) Menggunakan Machine Learning," *J. Ilmu Komput. dan Matematika*, vol. 4, no. 2, pp. 44–51, 2023.
- [6] J. Jennings Dunlap, "Autism Spectrum Disorder Screening and Early Action," *J. Nurse Pract.*, vol. 15, no. 7, pp. 496–501, 2019, doi: 10.1016/j.nurpra.2019.04.001.
- [7] A. Parmeggiani, A. Corinaldesi, and A. Posar, "Early features of autism spectrum disorder: A cross-sectional study," *Ital. J. Pediatr.*, vol. 45, no. 1, pp. 1–8, 2019, doi: 10.1186/s13052-019-0733-8.
- [8] Fadlan Isa Damanik and Said Iskandar Al-Idrus, "Diagnosa Autisme Pada Anak Dengan Sistem Pakar Menggunakan Metode Forward Chaining," *J. Student Res.*, vol. 1, no. 2, pp. 448–460, 2023, doi: 10.55606/jsr.v1i2.1063.
- [9] S. B. Bayu Sugara, Dedi Adidarma, "Perbandingan Akurasi Algoritma C4.5 dan Naïve Bayes untuk Deteksi Dini Gangguan Autisme pada Anak," *J. IKRA-ITH Inform.*, vol. 3, no. 1, pp. 119–128, 2019.
- [10] G. L. Kandouw, A. Dundu, and C. Elim, "Deteksi Dini Anak Gangguan Spektrum Autisme dan Interaksinya dengan Orang Tua dan Saudara Kandung," *e-CliniC*, vol. 6, no. 1, 2018, doi: 10.35790/ecl.6.1.2018.19504.
- [11] A. N. Katilik and J. A. Djie, "Penerapan Pendekatan Orff-Schulwerk untuk Meningkatkan Hasil Belajar Siswa dengan Autism Spectrum Disorder ( ASD ) dalam Pembelajaran Instrumen Ritmis Sederhana," *Seni Musik*, vol. 12, no. 1, pp. 91–109, 2022.
- [12] E. D. Isnannisa and L. M. Boediman, "Dir/floortime to increase communication between a child with autism and a mother with different sensory profile," *J. Psikol. Sains dan Profesi*, vol. 3, no. 3, pp. 177–187, 2019.
- [13] B. Sugara and A. Subekti, "Penerapan Support Vector Machine (Svm) Pada Small Dataset Untuk Deteksi Dini Gangguan Autisme," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 177–182, 2019, doi: 10.33480/pilar.v15i2.649.
- [14] L. Alaika, "Optimization of Accuracy to Autism Spectrum Disorder Identification for Children Using Support Vector

- Machine and Correlation-based Feature Selection,” *J. Adv. Inf. Syst. Technol.*, vol. 4, no. 1, pp. 1–12, 2022
- [15] E. S. Dewi, “Klasifikasi Autism Spectrum Disorder Menggunakan Algoritma Naive Bayes,” *MATHunesa J. Ilm. Mat.*, vol. 9, no. 1, pp. 27–35, 2021, doi: 10.26740/mathunesa.v9n1.p27-35.
- [16] R. Supriyadi, N. Maulidah, A. Fauzi, H. Nalatissifa, and S. Diantika, “Penerapan Algoritma Naive Bayes Dan Support Vector Machine Dalam Memprediksi Autisme,” *J. Swabumi*, vol. 10, no. 1, p. 2022, 2022
- [17] D. Agustriawan, “Penerapan Pendekatan Machine Learning Pada Pengembangan Basis Data Herbal Sebagai Sumber Informasi Kandidat Obat Kanker,” *J. Teknol. Ind. Pertan.*, vol. 29, no. 2, pp. 175–182, 2019, doi: 10.24961/j.tek.ind.pert.2019.29.2.175.
- [18] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [19] R. N. Yusra, O. S. Sitompul, and Sawaluddin, “Kombinasi K-Nearest Neighbor (KNN) dan Relief-F Untuk Meningkatkan Akurasi Pada Klasifikasi Data,” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 1, pp. 0–5, 2021.
- [20] N. Arifin, U. Enri, and N. Sulistiyowati, “Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 6, no. 2, p. 129, 2021, doi: 10.30998/string.v6i2.10133.