

Optimasi Akurasi Identifikasi Gangguan Spektrum Autisme Pada Anak Menggunakan Support Vector Machine

Lika Alaika

Jurusan Ilmu Komputer, FMIPA, Universitas Negeri Semarang.
Email: likaalaika@students.unnes.ac.id.

Abstrak

Autisme atau Gangguan Spektrum Autisme (GSA) merupakan kelainan perkembangan sistem saraf otak yang mengganggu proses berpikir sehingga mempengaruhi kemampuan komunikasi, interaksi, perilaku serta minat penderitanya. Gejala gangguan perkembangan ini umumnya diwaspadai sejak berumur 6 bulan. Meskipun autisme tidak bisa disembuhkan tetapi terapi dapat dilakukan dengan tujuan agar penderita dapat menyesuaikan diri dalam kehidupan sehari-hari. Berdasarkan data dari WHO, rasio anak yang memiliki gangguan autisme yaitu 1:160, pada tahun 2013 diperkirakan lebih dari 112.000 anak menderita autisme di Indonesia. Laporan Pusat Pengendalian dan Pencegahan Penyakit Amerika Serikat (CDC) menyatakan 1:100 anak didunia memiliki gangguan perkembangan ini. Dengan tercatatnya kasus yang tinggi dan seiring dengan perkembangan teknologi informasi maka data tersebut membutuhkan metode pengolahan yang tepat agar dapat dimanfaatkan secara maksimal. Salah satu teknik untuk mengolah data menjadi informasi yaitu *data mining*. Support Vector Machine termasuk salah satu metode klasifikasi yang dapat digunakan untuk mendiagnosis autisme. Pada penelitian ini, dataset yang digunakan yaitu *Autistic Spectrum Disorder Screening for Children Dataset* dari UCI *machine learning repository* yang terdiri dari 292 *record* dan 21 atribut, *tools* yang digunakan yaitu Weka 3.8.5. Tujuan dari penelitian ini adalah mengetahui akurasi metode Support Vector Machine untuk diagnosis autisme pada anak. Penelitian ini menghasilkan sebesar 99.6575% dan 95.8904%.

Kata kunci: SVM, Data Mining, Autisme, GSA.

Abstract

Autism or autism spectrum disorder (ASD) is a developmental disorder of the brain's nervous system that interferes with the thought process so that it affects the ability of communication, interaction, behavior, and interests of the sufferer. Symptoms of this developmental disorder are generally watched out for since the age of 6 months. Although autism cannot be cured, therapy can be done with the aim that the sufferer can adjust to everyday life. Based on data from the WHO, the ratio of children with autism disorders is 1:160, in 2013 it is estimated that more than 112,000 children suffer from autism in Indonesia. The United States Centers for Disease Control and Prevention (CDC) report states that 1:100 children in the world have this developmental disorder. With the high number of cases recorded and in line with the development of information technology, the data requires appropriate processing methods so that it can be utilized optimally. One technique to process data into information is data mining. The Support Vector Machine is one of the classification methods that can be used to diagnose autism. In this study, the dataset used is the Autistic Spectrum Disorder Screening for Children Dataset from the UCI machine learning repository consisting of 292 records and 21 attributes, the tools used are Weka 3.8.5. The purpose of this study was to determine the accuracy of the Support Vector Machine method for the diagnosis of autism in children. This study yielded 99.6575% and 95.8904%.

Keywords: SVM, Data Mining, Autisme, GSA.

1. PENDAHULUAN

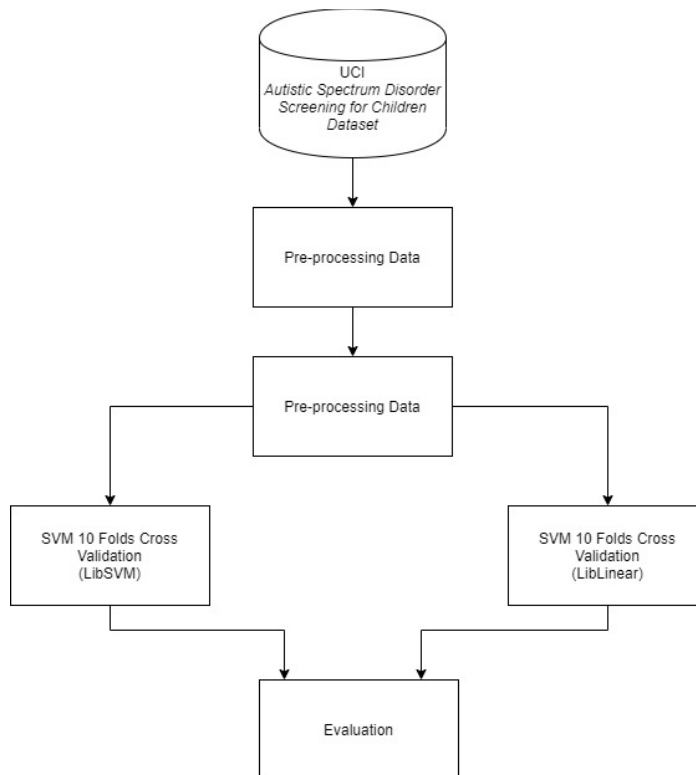
Autisme atau biasa disebut Gangguan Spektrum Autisme (GSA) adalah salah satu jenis gangguan perkembangan dengan gejala yang mengakibatkan adanya kelainan saraf sehingga menyebabkan fungsi otak tidak bekerja secara normal dan mengganggu proses berpikir manusia sehingga mempengaruhi kemampuan komunikasi dan interaksi seseorang, gangguan perilaku serta dapat membatasi minat penderitanya. Gejala autisme dapat diketahui dari adanya suatu penyimpangan dari ciri-ciri tumbuh kembang anak secara normal [1]. Autisme bukan merupakan penyakit kejiwaan karena merupakan suatu gangguan yang terjadi pada otak yang menyebabkan berkurangnya fungsi otak dan dapat mempengaruhi perilaku penderitanya. Penyebab autisme belum diketahui tetapi berdasarkan penelitian autisme disebabkan oleh interaksi antara faktor genetik dan berbagai paparan negatif dari lingkungan. Gejala gangguan perkembangan ini umumnya dapat diwaspadai sejak anak berumur 6 bulan. Meskipun autisme tidak bisa disembuhkan tetapi berbagai terapi dapat dilakukan untuk menangani autisme yang bertujuan agar penderita dapat menyesuaikan diri dalam kehidupan sehari-hari. Hasil dari penelitian membuktikan bahwa jika anak-anak dengan autisme diobati sedini mungkin, beban yang terjadi akibat gangguan perkembangan ini sangat berkurang [2]. Autisme merupakan salah satu gangguan perkembangan yang sering salah teridentifikasi atau tertukar dengan gangguan perkembangan lainnya seperti *attention deficit hyperactivity disorder* (ADHD), tunagrahita berat atau bahkan gangguan pendengaran [3]. Berdasarkan data yang dihimpun WHO, rasio anak yang memiliki gangguan autisme yaitu 1:160 [4], pada tahun 2013 diperkirakan lebih dari 112.000 anak penderita autisme di Indonesia. Laporan Pusat Pengendalian dan Pencegahan Penyakit Amerika Serikat (CDC) menyatakan 1% atau 1:100 anak di dunia memiliki gangguan perkembangan ini.

Dengan tercatatnya kasus yang tinggi dan seiring dengan perkembangan teknologi informasi maka data tersebut membutuhkan metode pengolahan yang tepat agar dapat dimanfaatkan secara maksimal. Salah satu teknik untuk mengolah data menjadi informasi yaitu data mining. Support Vector Machine adalah teknologi machine learning yang dikembangkan pada pertengahan tahun 1990 yang biasanya digunakan untuk suatu masalah khusus seperti pengenalan angka dan tulisan tangan, klasifikasi teks dan lain-lain dengan performa yang cukup baik [5][6]. Support Vector Machine merupakan sistem pembelajaran mesin yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan learning bias yang berasal dari teori pembelajaran. Beberapa kelebihan dari metode Support Vector Machine adalah metode ini bekerja baik untuk sekumpulan data yang tidak dapat dipisahkan secara linear [7].

Support Vector Machine termasuk salah satu metode klasifikasi yang dapat digunakan untuk mendiagnosis autisme. Pada penelitian ini, dataset yang digunakan yaitu *Autistic Spectrum Disorder Screening for Children Dataset* yang diperoleh dari UCI *machine learning repository* yang terdiri dari 292 record dan 21 atribut, sedangkan tools yang digunakan untuk menerapkan model yaitu menggunakan aplikasi Weka 3.8.5 dengan tujuan dari penelitian ini adalah untuk mengetahui tingkat akurasi metode Support Vector Machine untuk diagnosis autisme pada anak.

2. METODE

Metode penelitian yang akan digunakan dalam penelitian ini yaitu menggunakan algoritma klasifikasi Support Vector Machine menggunakan pilihan tes *Cross-validation* dengan 10 *Folds*. Adapula fungsi yang digunakan yaitu LibSVM dan LibLinear yang telah tersedia di aplikasi Weka 3.8.5. Hasil akhir yang didapat yaitu diketahuinya tingkat akurasi diagnosis gangguan spektrum autisme pada anak menggunakan algoritma klasifikasi Support Vector Machine. **Gambar 1.** merupakan gambaran kerangka penelitian yang akan dilakukan.



Gambar 1. Kerangka penelitian

Berdasarkan kerangka penelitian yang telah disajikan dalam **Gambar 1.** maka metode penelitian tersebut dapat dijabarkan menjadi seperti berikut: a) Dataset yang digunakan dalam penelitian ini yaitu *Autistic Spectrum Disorder for Children Dataset* yang berasal dari *UCI machine learning repository* yang memiliki 292 *record* dan 21 atribut. b) Pre-processing data merupakan teknik pengolahan data awal yang digunakan dalam penelitian ini dengan cara melakukan *Replace Missing Value* yang dilakukan di tahap *Preprocess* di Weka yang berguna untuk mengisi nilai pada atribut yang hilang dengan menggunakan nilai rata-rata dan modus yang didapatkan dari dataset yang telah diunggah. c) Proses Support Vector Machine 10 *Folds Cross Validation* ini terbagi menjadi dua proses yang digunakan yaitu LibSVM dan LibLinear, kedua fungsi tersebut merupakan fungsi yang menjalankan algoritma Support Vector Machine, adapun perbedaannya yaitu LibSVM dapat digunakan untuk

tipe data yang lebih fleksibel dan LibLinear dapat meminimalkan kesalahan hitung pada tipe data linear. Pada proses ini validasi akan dilakukan dengan cara membagi dataset menjadi 10 bagian dengan rincian satu bagian sebagai data uji dan bagian lain sebagai data latih. Proses ini akan dilakukan secara berulang dimulai dari data uji hingga data latih yang terakhir, hal ini menjadikan semua data dalam dataset telah teruji. Proses ini bertujuan agar menghasilkan prediksi diagnosis gangguan spektrum autisme pada anak ditemukan akurasi. d) Evaluation berfungsi untuk melihat kinerja diagnosis gangguan spektrum autisme pada anak yang terkalkulasi dalam bentuk presentase akurasi serta nilai *Area Under ROC Curve* (AUC) yang digunakan untuk memberikan matriks numerik single sehingga dapat digunakan untuk membandingkan hasil kinerja. Nilai yang dihasilkan oleh AUC dapat dijadikan ukuran untuk melihat prediksi yang berbentuk angka dengan rentang 0 sampai dengan 1, dengan perkiraan angka yang mendekati satu menjadikan nilai dengan tingkat akurasi yang lebih baik karena meningkatkan efektivitas klasifikasi.

Aplikasi yang digunakan untuk menerapkan penelitian ini yaitu aplikasi Weka versi 3.8.5 menggunakan algoritma klasifikasi Support Vector Machine yang merupakan algoritma yang bertujuan untuk menemukan batas pemisah optimal yang disebut *hyperplane* untuk membedakan kelas satu dan kelas yang lainnya [8]. Sampel yang paling dekat dengan *hyperplanes* ini disebut vektor dukungan dan perbedaan tersebut dinyatakan sebagai jumlah bobot dari subset sampel yang membatasi masalah [9]. Algoritma ini bekerja berdasarkan prinsip dasar dengan menyisipkan *hyperplane* yang ada diantara kelas dan berperan sebagai penjaga jarak maksimum dari titik data terdekat [8].

Adapun hasil pengujian dihitung menggunakan *Confusion Matrix* untuk mencari akurasi dan AUC. *Confusion matrix* adalah metode yang digunakan untuk mengukur kinerja model klasifikasi berdasarkan perhitungan objek, dengan data hasil prediksi diantara dua kelas yang menghasilkan kelas positif dan negatif [10]. *Confusion matrix* yang disajikan berisi informasi aktual dan prediksi klasifikasi [8]. Proses evaluasi *confusion matrix* dapat memperoleh nilai *precision*, *recall*, dan *accuracy* yang didapat dari rumus berikut:

$$Precision = \frac{TP}{(TP+FP)} \quad (1)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (2)$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (3)$$

Keterangan:

TP: Jumlah kasus positif yang diklasifikasikan sebagai positif.

FP: Jumlah kasus negatif yang diklasifikasikan sebagai positif.

TN: Jumlah kasus negatif yang diklasifikasikan sebagai negatif.

FN: Jumlah kasus positif yang diklasifikasikan sebagai negatif.

Adapun **Tabel 1.** menjelaskan contoh tabel *confusion matrix* yang menunjukkan klasifikasi dua kelas.

Tabel 1. Model *confusion matrix*.

<i>Classification</i>	<i>Predicted Class</i>	
	<i>Class = Yes</i>	<i>Class = No</i>
<i>Class = Yes</i>	<i>True Positives (TP)</i>	<i>False Negatives (FN)</i>
<i>Class = No</i>	<i>False Positive (FP)</i>	<i>True Negatives (TN)</i>

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini menggunakan dataset yang diambil dari UCI *machine learning repository* dengan judul *Autistic Spectrum Disorder for Children Dataset* dengan 292 *record* dan 21 atribut yang disajikan seperti terlampir di **Tabel 2.**

Tabel 2. Dataset

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	age	gender	...	class
1	1	0	0	1	1	0	1	0	0	6	m	...	NO
1	1	0	0	1	1	0	1	0	0	6	m	...	NO
1	1	0	0	0	1	1	1	0	0	6	m	...	NO
0	1	0	0	1	1	0	0	0	1	5	f	...	NO
1	1	1	1	1	1	1	1	1	1	5	m	...	YES
0	0	1	0	1	1	0	1	0	1	4	m	...	NO
1	0	1	1	1	1	0	1	0	1	5	m	...	YES
1	1	1	1	1	1	1	1	0	0	5	f	...	YES
1	1	1	1	1	1	1	0	0	0	11	f	...	YES
0	0	1	1	1	0	1	1	0	0	11	f	...	NO
...
0	0	1	0	1	1	1	1	0	1	11	m	...	NO

Adapula deskripsi dari atribut dataset tersebut akan dijelaskan seperti yang tertera di **Tabel 3.**

Tabel 3. Atribut dataset

Attribute	Type	Description
Age	Number	years
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (yes or no)	Whether the case was born with jaundice
Family member with PDD	Boolean (yes or no)	Whether any immediate family member has a PDD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician ,etc.
Country of residence	String	List of countries in text format
Used the screening app before	Boolean (yes or no)	Whether the user has used a screening app
Screening Method Type	Integer (0,1,2,3)	The type of screening methods chosen based on age category (0=toddler, 1=child, 2= adolescent, 3= adult)
Question 1 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 2 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 3 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 4 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 5 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 6 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 7 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 8 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 9 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 10 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Screening Score	Integer	The final score obtained based on the scoring algorithm of the screening method used. This was computed in an automated manner

Tahap awal penggunaan metode ini yaitu dimulai dari tahap *pre-processing* data yang ada dengan menggunakan aplikasi Weka. *Pre-processing* data yang dilakukan yaitu dengan menggunakan fungsi *Replace missing value* yang berfungsi untuk mengisi data yang hilang dan digantikan dengan mean dan modus yang didapatkan dari data sebelum proses *pre-processing*. Setelah itu maka akan dilakukan proses klasifikasi menggunakan algoritma Support Vector machine dengan pilihan test *10 Folds Cross Validation* dimana dataset akan dibagi menjadi 10 bagian dengan rincian satu bagian sebagai data uji dan bagian lain sebagai data latih. Proses ini akan dilakukan secara berulang dimulai dari data uji hingga data latih yang terakhir, hal ini menjadikan semua data dalam dataset telah teruji. Setelah semua proses selesai maka akan didapatkan nilai akurasi bersama dengan waktu prosesnya yang akan dijabarkan pada **Tabel 4**.

Tabel 4. Hasil akurasi

Algoritma Klasifikasi	Test Options	Akurasi		Waktu	
		LibSVM	LibLinear	LibSVM	LibLinear
Support Vector Machine	10 Folds Cross Validation	99.6575 %	95.8904 %	0.01 s	0.03 s

Adapun beberapa perbedaan hasil perhitungan menggunakan dua fungsi yang telah diuji akan dipaparkan pada **Tabel 5**.

Tabel 5. Perbedaan dua fungsi

	LibSVM	LibLinear
Correctly Classified Instances	291	280
Incorrectly Classified Instances	1	12
Kappa statistic	0.9931	0.9178
Mean absolute error	0.0034	0.0411
Root mean squared error	0.0585	0.2027
Relative absolute error	0.6857 %	8.2285 %
Root relative squared error	11.7107 %	40.5669 %

Correctly classified instances menunjukkan jumlah data yang benar sedangkan *incorrectly classified instances* menunjukkan jumlah data yang diklasifikasikan salah. *Kappa statistic* menunjukkan ukuran yang menyatakan konsistensi pengukuran dan ada rata-rata selisih mutlak dengan nilai prediksi serta hasil dari pengakarannya dan nilai error absolut relatif dan hasil pengakarannya. Dari hasil tersebut juga didapatkan

confusion matrix yang berfungsi untuk mencari nilai akurasi AUC yang dijelaskan dalam **Tabel 6**.

Tabel 6. Hasil *confusion matrix*

LibSVM	LibLinear
a b <-- classified as	a b <-- classified as
150 1 a = NO	142 9 a = NO
0 141 b = YES	3 138 b = YES

Berdasarkan tabel diatas maka fungsi LibSVM menunjukkan ada 150 data yang dikelompokkan benar dan 1 data salah untuk *class no* dan 141 data benar dan 0 data salah untuk *class yes*, sedangkan untuk fungsi LibLinear menunjukkan hasil 142 data benar dan 9 data salah untuk *class no* dan 138 data benar dan 3 data salah untuk *class yes*. Adapun perhitungan nilai AUC yang telah didapatkan dari Support Vector Machine LibSVM dan LibLinear akan dijelaskan pada **Tabel 7** dan **Tabel 8**.

Tabel 7. Nilai AUC LibSVM

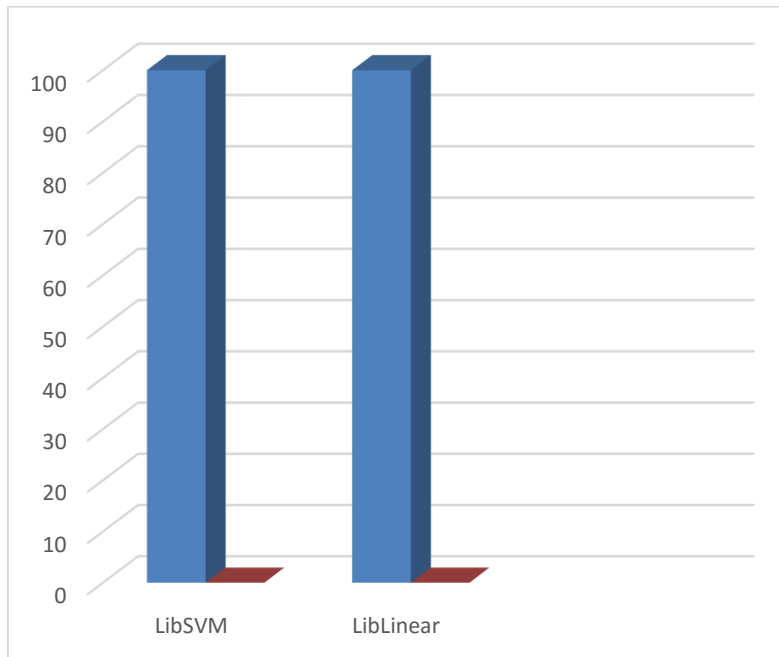
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,993	0,000	1,000	0,993	0,997	0,993	0,997	0,997	NO
	1,000	0,007	0,993	1,000	0,996	0,993	0,997	0,993	YES
Weighted Avg.	0,997	0,003	0,997	0,997	0,997	0,993	0,997	0,995	

Tabel 8. Nilai AUC LibLinear

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,940	0,021	0,979	0,940	0,959	0,919	0,960	0,952	NO
	0,979	0,060	0,939	0,979	0,958	0,919	0,960	0,929	YES
Weighted Avg.	0,959	0,040	0,960	0,959	0,959	0,919	0,960	0,941	

Dengan TP menunjukkan jumlah kasus positif yang diklasifikasikan sebagai positif, FP menunjukkan jumlah kasus negatif yang diklasifikasikan sebagai positif dilengkapi dengan proses evaluasi *confusion matrix* masing-masing kelas serta berat rata-rata atribut yang dihasilkan.

Berdasarkan tabel hasil pengujian yang ada maka didapatkan perbandingan hasil dalam bentuk grafik yang dipaparkan pada **Gambar 2**.



Gambar 2. Grafik Perbandingan

4. SIMPULAN

Berdasarkan penelitian yang telah dilakukan, teknik data mining berupa klasifikasi menggunakan metode Support Vector Machine dengan dataset yang digunakan yaitu *Autistic Spectrum Disorder Screening for Children Dataset* yang diperoleh dari *UCI machine learning repository* yang terdiri dari 292 record dan 21 atribut yang diimplementasikan mampu mendeteksi gangguan spektrum autisme pada anak dengan akurasi sebesar 99.6575% dan 95.8904%. Dua fungsi yang digunakan yaitu LibSVM dan LibLinear tidak memiliki perbedaan terlalu jauh terhadap data yang digunakan. Dengan nilai akurasi tersebut maka disimpulkan bahwa algoritma klasifikasi Support Vector Machine memiliki tingkat akurasi yang cukup tinggi dengan diaplikasikan pada dataset yang telah didapatkan. Algoritma ini dapat memprediksi diagnosis terjadinya gangguan spektrum autisme pada anak dengan akurasi yang cukup tinggi sehingga dapat dijadikan acuan bagi orang tua ataupun praktisi yang menangani bidang ini untuk mengetahui sejak dini dan melakukan penanganan yang tepat ataupun sebagai dasar algoritma untuk melakukan klasifikasi dibidang yang lainnya.

5. REFERENSI

- [1] Sunu, et al. 2012. *Panduan Memecahkan Masalah Autisme: Unlocking Autisme*. Lintang Terbing. Yogyakarta.
- [2] S. Qiu, et al. 2020. *Prevalance of autism sprectrum disorder in Asia: A systematic review and meta-analysis*. *Psychiatry Res*. Vol 284.
- [3] YPAC. 2010. *Buku Pedoman Penanganan dan Pendidikan Autisme*. YPAC Press.

- [4] Mayada et al. 2012. *Gloval pravelence of autism and other pervasive developmental disorders. Autism Research:160-179.*
- [5] Y. Chen. 2020. *Mining of instant messaging data in the Internet of Things based on support vector machine. Computer commun Vol. 154.*
- [6] K. Shankar. 2020. *Optimal feature-based multi-kernel SVM approach for thyroid disease classification. Supercomputer Vol. 76 No. 2.*
- [7] Nelly Indriani, et al. 2017. *Peringkasan dan support vector machine pada klasifikasi dokumen. Jurnal INFOTEL*
- [8] Kurniawan I. 2020. *Prediksi Gejala Autism Spectrum Disorders pada Remaja Menggunakan Optimasi Particle Swarm Optimization dan Algoritma Support Vector Machine. Informatics for Educators and Professionals. 113-122.*
- [9] S. Salcedo-Sanz, et al. 2014. *Support vector machines in engineering: An overview. Wiley Interdisciplin Review Data Mining Knowledge Discover Vol. 4 No. 3.*
- [10] M. Y. Kurniawan dan M. E. Rosadi. 2017. *Optimasi Decision Tree Menggunakan Particle Swarm Optimization Pada Data Siswa Putus Sekolah. Jtiulm Vol. 2 No. 1.*