

PERBANDINGAN ALGORITMA MACHINE LEARNING UNTUK ANALISIS DAN DETEKSI GANGGUAN SPEKTRUM AUTISME DISORDER

Taryadi¹⁾, Era Yuniyanto²⁾

Program Studi Sistem Informasi / STMIK Widya Pratama^{1,2)}
paktari.1218@gmail.com¹⁾, era.yuniyanto@gmail.com²⁾

Abstrak

Autism Spectrum Disorder (ASD) merupakan gangguan saraf dimana seseorang memiliki efek seumur hidup pada interaksi dan komunikasi dengan orang lain. Autisme dapat didiagnosis pada setiap tahap dalam satu kehidupan karena dalam dua tahun pertama kehidupan gejala autisme biasa muncul. Gejala ASD muncul dimulai pada masa kanak-kanak dan terus berlanjut hingga remaja dan dewasa. Didorong dengan meningkatnya penggunaan teknik pembelajaran mesin dalam dimensi penelitian diagnosis medis, dalam makalah ini ada upaya untuk mengeksplorasi kemungkinan penggunaan Naïve Bayes, Regresi Logistik, Support Vector Machine, Neural Network, KNN dan Convolutional Neural Network untuk memprediksi dan menganalisis masalah ASD pada anak, remaja, dan orang dewasa. Teknik yang diusulkan dievaluasi pada tiga set data ASD non-klinis yang tersedia untuk umum. Dataset pertama terkait skrining ASD pada anak memiliki 292 instance dan 21 atribut. Dataset kedua yang terkait dengan skrining ASD Subjek dewasa berisi total 704 kejadian dan 21 atribut. Dataset ketiga terkait skrining ASD pada subjek Remaja terdiri dari 104 kejadian dan 21 atribut. Setelah menerapkan berbagai teknik pembelajaran mesin dan menangani nilai yang hilang, hasil prediksi dengan model berbasis Convolutional Neural Network memiliki tingkat kinerja dan akurasi yang lebih baik yaitu sebesar 99,53% untuk deteksi ASD pada dewasa, 98,30% untuk deteksi ASD pada anak-anak dan 96,88% untuk deteksi pada Remaja.

Kata kunci: *Autism Spectrum Disorder, Machine Learning, Klasifikasi, Deteksi Dini.*

1. Pendahuluan

Masalah gangguan spektrum autisme atau Autism Spectrum Disorder (ASD) telah meningkat pesat saat ini di antara semua usia populasi manusia. Deteksi dini penyakit saraf ini dapat sangat membantu dalam pemeliharaan kesehatan mental dan fisik subjek. Dengan maraknya penerapan model berbasis pembelajaran mesin dalam prediksi berbagai penyakit manusia, deteksi dini berdasarkan berbagai parameter kesehatan dan fisiologis sekarang tampaknya mungkin. Faktor ini memotivasi kami untuk meningkatkan minat dalam deteksi dan analisis penyakit ASD untuk meningkatkan metodologi pengobatan yang lebih baik (Constantino, et al., 2017). Deteksi ASD menjadi tantangan karena ada beberapa gangguan mental lain yang gejalanya sedikit mirip dengan gejala ASD, sehingga membuat tugas ini sulit.

ASD merupakan masalah yang berkaitan dengan perkembangan otak manusia. Seseorang yang telah menderita ASD umumnya tidak dapat

melakukan interaksi sosial dan komunikasi dengan orang lain (Thabtah, 2018; Vaishali, et al., 2018). Dalam hal ini, kehidupan seseorang biasanya terpengaruh sepanjang hidupnya. Sangat menarik untuk mengetahui bahwa faktor lingkungan dan genetik dapat menjadi faktor penyebab penyakit ini. Gejala masalah ini dapat dimulai pada usia tiga tahun dan dapat berlanjut seumur hidup. Pasien yang menderita penyakit ini tidak dapat diobati secara tuntas, namun efeknya dapat dikurangi untuk beberapa waktu jika gejalanya terdeteksi lebih awal. Dengan asumsi bahwa gen manusia bertanggung jawab untuk itu, penyebab pasti ASD belum diketahui oleh para ilmuwan.

Gen manusia mempengaruhi perkembangan dengan mempengaruhi lingkungan. Ada beberapa faktor risiko yang mempengaruhi ASD seperti berat badan lahir rendah, saudara dengan ASD dan memiliki orang tua tua, dan lain-lain. Selain itu, ada beberapa masalah interaksi sosial dan komunikasi seperti : tidak ada kepekaan rasa

sakit, tidak mampu kontak mata, tidak merespon suara, tidak ingin dipeluk dan tidak mampu mengekspresikan gerak tubuh.

Orang dengan ASD juga mengalami kesulitan dengan minat yang terbatas dan pengulangan perilaku yang konsisten. Perilaku spesifik yang seriang dilakukan oleh penderita ASD seperti mengulangi perilaku tertentu seperti mengulang kata atau frasa berkali-kali, mudah marah ketika rutinitas akan berubah, memiliki sedikit minat pada topik tertentu seperti angka, fakta, dan rasa sensitifitas yang kurang dibandingkan orang lain dalam beberapa kasus seperti cahaya dan kebisingan (Thabtah, et al., 2018).

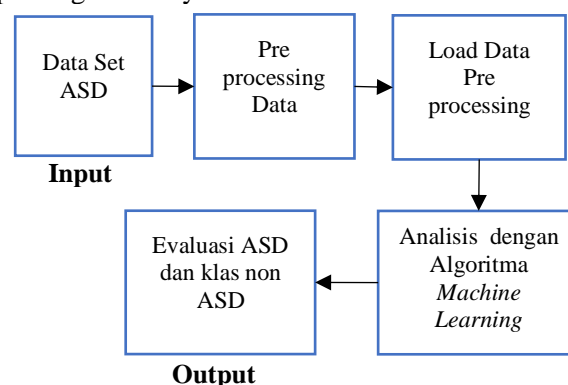
Deteksi dini dan pengobatan adalah langkah paling penting yang harus diambil untuk mengurangi gejala masalah gangguan spektrum autisme dan untuk meningkatkan kualitas hidup orang yang menderita ASD. Namun, tidak ada prosedur tes medis untuk mendeteksi autisme. Gejala ASD biasanya dikenali dengan observasi. Pada remaja yang lebih tua yang bersekolah, gejala ASD biasanya diidentifikasi oleh orang tua dan guru. Setelah itu gejala ASD dievaluasi oleh tim pendidikan khusus sekolah. Tim sekolah menyarankan anak-anak ini mengunjungi dokter kesehatan untuk tes yang diperlukan. Pada orang dewasa mengidentifikasi gejala ASD sangat sulit dibandingkan anak yang lebih tua dan remaja karena beberapa gejala ASD mungkin tumpang tindih dengan gangguan kesehatan mental lainnya. Sangat mudah untuk mengidentifikasi perubahan perilaku pada anak dengan mudah dengan observasi karena dapat dilihat lebih awal pada usia 6 bulan daripada pencitraan otak khusus autisme karena pencitraan otak dapat diidentifikasi setelah usia 2 tahun.

Deteksi dini dan pengobatan adalah langkah paling penting yang harus diambil untuk mengurangi gejala masalah gangguan spektrum autisme dan untuk meningkatkan kualitas hidup orang yang menderita ASD. Namun, tidak ada prosedur tes medis untuk mendeteksi autisme. Gejala ASD biasanya dikenali dengan observasi. Pada remaja yang lebih tua yang bersekolah, gejala ASD biasanya diidentifikasi oleh orang tua dan guru. Setelah itu gejala ASD dievaluasi oleh tim pendidikan khusus sekolah. Tim sekolah menyarankan anak-anak ini mengunjungi dokter kesehatan untuk tes yang diperlukan. Pada orang

dewasa mengidentifikasi gejala ASD sangat sulit dibandingkan anak yang lebih tua dan remaja karena beberapa gejala ASD mungkin tumpang tindih dengan gangguan kesehatan mental lainnya. Sangat mudah untuk mengidentifikasi perubahan perilaku pada anak dengan mudah dengan observasi karena dapat dilihat lebih awal pada usia 6 bulan daripada pencitraan otak khusus autisme karena pencitraan otak dapat diidentifikasi setelah usia 2 tahun.

2. Metode Penelitian

Gambar 1 menunjukkan langkah-langkah dalam alur kerja yang diusulkan yang melibatkan pra-pemrosesan data, pelatihan, dan pengujian dengan model tertentu, evaluasi hasil dan prediksi ASD. Pengolahan data menggunakan bahasa pemrograman Python 3.



Gambar 1. Tahapan deteksi ASD yang diusulkan

2.1 DataSet

Dataset untuk tujuan penelitian ini telah dikumpulkan dari UCI Repository yang tersedia untuk umum (Wall, et al. 2012; Thabtah, 2017). Dalam penelitian ini terutama tiga jenis dataset telah digunakan. Ringkasan rinci dari dataset diberikan di bawah ini: Kumpulan data ini memiliki 20 atribut umum yang digunakan untuk prediksi. Atribut ini tercantum di bawah ini:

Tabel 1. Daftar Dataset ASD yang digunakan

No	Nama Dataset	Sumber	Jumlah Atribut	Jumlah Instan
1	Data Screening ASD Dewasa	Repositori UCI <i>Machine Learning</i>	21	704
2	Data Screening untuk anak ASD	Repositori UCI <i>Machine Learning</i>	21	292

3	Data Screening untuk Remaja ASD	Repositori UCI Machine Learning	21	104
---	---------------------------------	---------------------------------	----	-----

Tabel 2. Daftar atribut dari dataset

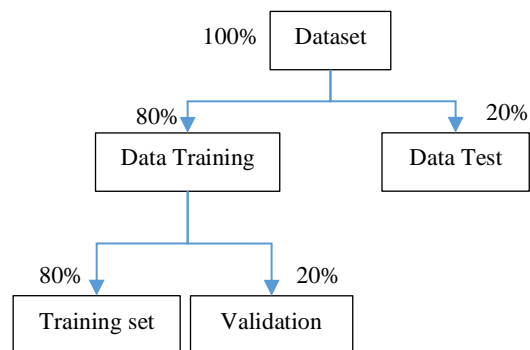
ID Atribut	Deskripsi Atribut
1	Umur
2	Jenis kelamin
3	Nasionaliti
4	Pasien menderita penyakit kuning sejak lahir
5	Setiap anggota keluarga menderita gangguan perkembangan pervasif
6	Siapa yang memenuhi eksperimen?
7	Negara tempat tinggal pengguna
8	Penyaringan Aplikasi yang digunakan oleh pengguna sebelumnya atau tidak?
9	Tipe test screening
10-19	Berdasarkan jawaban metode penyaringan dari 10 pertanyaan
20	Nilai screening

2.2 Data Pre-Processing

Pra-pemrosesan data adalah teknik mengubah data mentah menjadi format yang berarti dan dapat dimengerti (Bone, et al., 2015). Data dunia nyata umumnya tidak lengkap dan tidak konsisten karena mengandung banyak kesalahan dan nilai nol. Sebuah data pra-olahan yang baik selalu menghasilkan hasil yang baik. Berbagai metode pra-pemrosesan data digunakan untuk menangani data yang tidak lengkap dan tidak konsisten seperti menangani nilai yang hilang, deteksi outlier, diskritisasi data, dan reduksi data. Masalah nilai yang hilang dalam kumpulan data ini telah ditangani dengan metode imputasi.

2.3 Training and Testing Model

Seluruh dataset telah dibagi menjadi dua bagian yaitu satu bagian adalah pelatihan dataset dan yang lainnya adalah pengujian dataset dengan rasio 80:20 masing-masing. Untuk tujuan validasi silang lagi data pelatihan telah dibagi menjadi dua bagian. Satu bagian adalah dataset pelatihan dan bagian lainnya adalah dataset validasi dengan rasio masing-masing 80:20. Gambar 2 menunjukkan set pelatihan, pengujian, dan validasi akhir yang klasifikasinya telah dilakukan (Li, et al., 2015).



Gambar 2. Pelatihan Akhir, Pengujian dan Validasi Set

3. Hasil dan Pembahasan

Hasilnya diukur dalam hal spesifisitas, sensitivitas, dan akurasi dengan menggunakan matriks konfusi dan laporan klasifikasi. Hasilnya tergantung pada seberapa akurat model dilatih.

3.1 Metrik Evaluasi Kinerja

Mengukur kinerja adalah kunci untuk memeriksa seberapa baik model klasifikasi bekerja untuk mencapai target. Metrik evaluasi kinerja digunakan untuk mengevaluasi efektivitas dan kinerja model klasifikasi pada kumpulan data pengujian. Penting untuk memilih metrik yang tepat untuk mengevaluasi kinerja model seperti matriks konfusi, akurasi, spesifisitas, sensitivitas, dan lain-lain. Rumus berikut digunakan untuk menemukan metrik kinerja:

Tabel 3. Elemen Matriks Kinerja

Nilai ASD	Aktual	Nilai Prediksi ASD	
		Positif Benar (TP)	Positif Salah (FP)
		Negatif Salah (FN)	Negatif Benar (TN)

$$\text{Spesifisitas} = \frac{\text{NB}}{\text{NB} + \text{PB}} \quad (1)$$

$$\text{Sensitivitas positif sejati} = \frac{\text{TN}}{\text{NB} + \text{NS}} \quad (2)$$

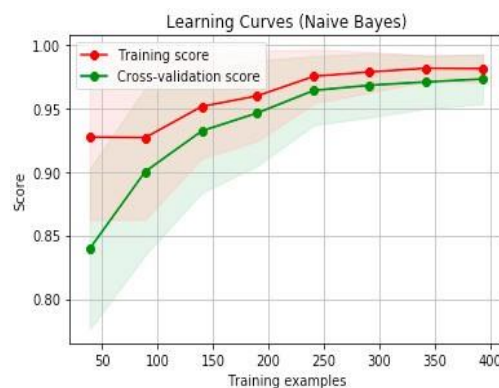
$$\text{Akurasi} = \frac{\text{PB} + \text{NB}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (3)$$

Hasil eksperimen dari berbagai pendekatan algoritma pembelajaran mesin dengan semua pemilihan fitur telah ditunjukkan untuk data skrining ASD untuk Dewasa, anak-anak, dan remaja. Dalam hal ini, 21 fitur dipilih untuk menemukan spesifisitas, sensitivitas, dan akurasi model yang diprediksi. Untuk implementasi algoritma Naïve menggunakan Gaussian Naïve Bayes. Untuk SVM menggunakan Kernel RBF dengan nilai gamma 0,1. Di KNN menggunakan N sebanyak 5 ($N=5$). Model ANN menggunakan Adam Optimizer dengan tingkat pembelajaran 0,01 dan 100 masa. Ukuran kinerja keseluruhan dari semua pengklasifikasi pembelajaran mesin dengan semua 3 set data telah ditunjukkan di bawah ini secara mendetail:

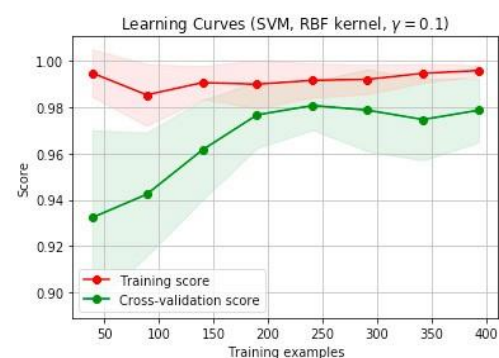
Tabel 4. Hasil Keseluruhan untuk Data Skrining ASD untuk Dewasa

Algoritma	Kekhususan	Sensitivitas	Akurasi
Regresi Logistik	95,75 %	96,96 %	96.69 %
SVM	95,74 %	88,89 %	98.11 %
Naive Bayes	93,61 %	96,96 %	96.22 %
KNN	91,48 %	96,96 %	95.75 %
ANN	97,87 %	97,57 %	97.64 %
CNN	99,89 %	99,39 %	99.53 %

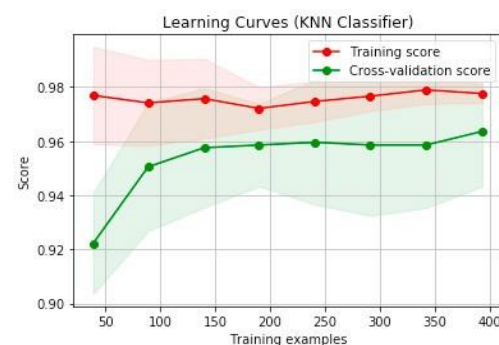
Evaluasi berbagai model pembelajaran mesin pada dataset diagnosis dewasa ASD memperoleh tingkat akurasi padakisanan 95,75% - 99,53 % pada dataset asli. Pengklasifikasi K-NN dengan $K=5$ menghasilkan akurasi paling rendah sebesar 95,75%. CNN menghasilkan akurasi prediksi 99,53 % pada dataset asli. Kurva pembelajaran dari semua algoritma *machine learning* juga menggambarkan hasil dari model prediksi seperti terlihat pada tabel 4. Sedangkan gambar 3, 4, 5, 6, dan 7 menunjukkan kurva pembelajaran dari tiap metode *machine learning* yang digunakan.



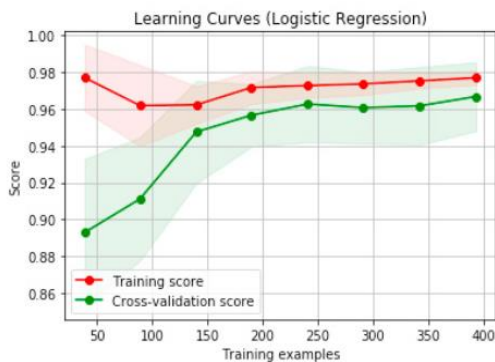
Gambar 3. Kurva Pembelajaran metode Naive Bayes untuk dataset dewasa



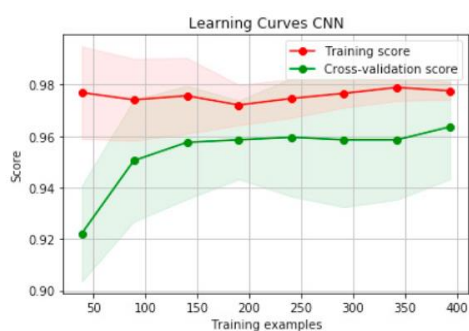
Gambar 4. Kurva Pembelajaran metode SVM untuk dataset dewasa



Gambar 5. Kurva Pembelajaran metode KNN untuk dataset dewasa



Gambar 6. Kurva Pembelajaran metode Regresi Logistik untuk dataset dewasa



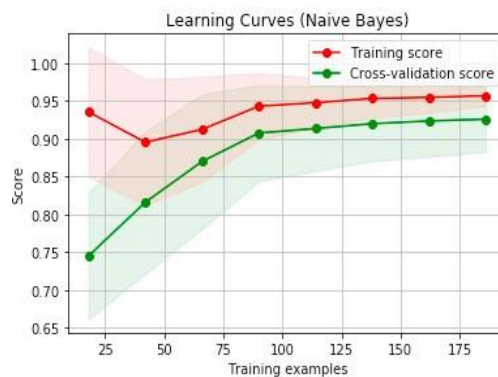
Gambar 7: Kurva Pembelajaran metode CNN untuk dataset dewasa

Tabel 5. Hasil Keseluruhan untuk Data Skringing Gangguan Spektrum Autistik untuk anak-anak

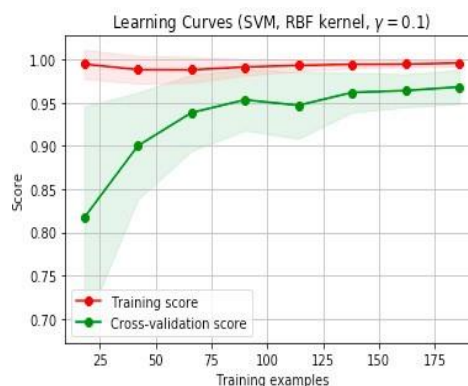
Algoritma	Kekhususan	Sensitivitas	Akurasi
Regresi Logistik	100 %	96,77 %	98,30 %
SVM	100 %	96,79 %	98,30 %
Naive Bayes	96,42 %	93,54 %	94,91 %
KNN	96,42 %	80,64 %	88,13 %
ANN	96,42 %	100 %	98,30 %
CNN	100 %	96,78 %	98,30 %

Evaluasi berbagai model pembelajaran mesin pada dataset diagnosis anak ASD tingkat akurasi pengamatan pada kisaran 88,13% - 98,30 % pada dataset asli seperti ditunjukkan pada tabel 5. Pengklasifikasi K-NN dengan K=5 menghasilkan akurasi paling rendah sebesar 88,13%. CNN, SVM, ANN, dan RL menghasilkan akurasi prediksi 98,30% pada dataset asli. Kurva pembelajaran dari semua algoritma *machine learning* juga

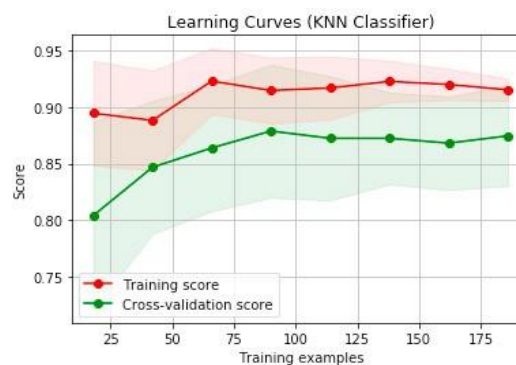
menggambarkan hasil dari model prediksi. Bentuk penggambaran kurva pembelajaran dari masing-masing metode *machine learning* yang digunakan ditunjukkan pada gambar 8 – 12.



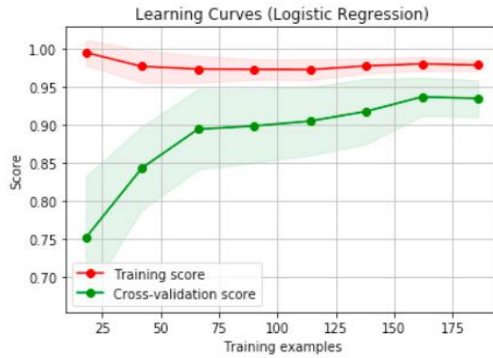
Gambar 8: Kurva Pembelajaran metode Naive Bayes untuk dataset anak-anak.



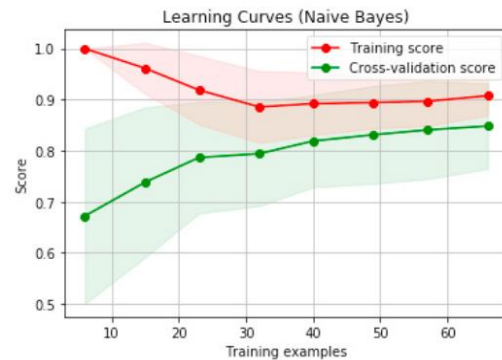
Gambar 9: Kurva Pembelajaran metode SVM untuk dataset anak-anak.



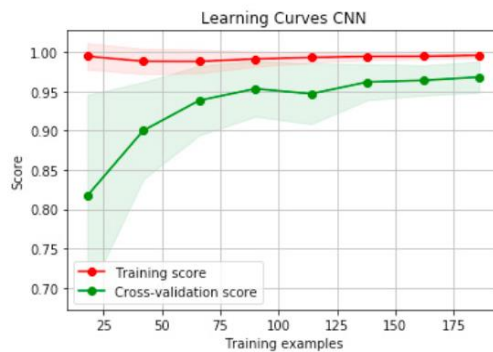
Gambar 10: Kurva Pembelajaran metode KNN untuk dataset anak-anak.



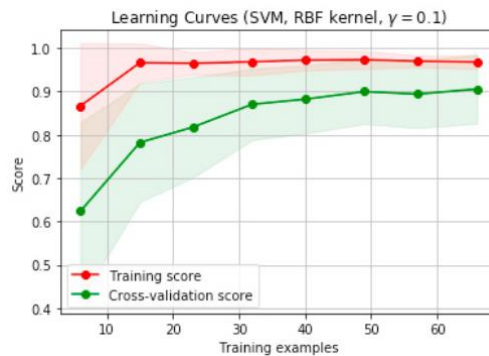
Gambar 11: Kurva Pembelajaran metode Regresi Logistik untuk dataset anak-anak.



Gambar 13: Kurva Pembelajaran metode Naive Bayes untuk dataset remaja.



Gambar 12: Kurva Pembelajaran metode CNN untuk dataset anak-anak.

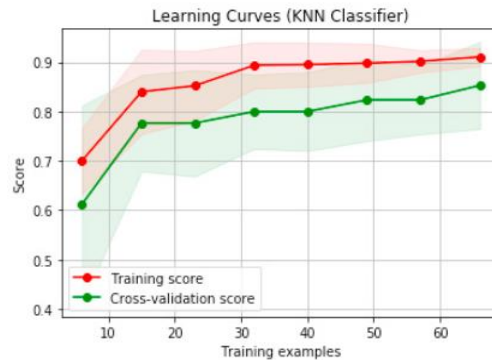


Gambar 14: Kurva Pembelajaran metode SVM untuk dataset remaja.

Tabel 6: Hasil Keseluruhan untuk Data Skringing Gangguan Spektrum Autistik untuk Remaja

Classifier	Kekhususan	Sensitivitas	Akurasi
Logistic Regression	100 %	66,66 %	85,71 %
SVM	100 %	88,88 %	95,23 %
Naive Bayes	91,66 %	88,88 %	90,47 %
KNN	100 %	55,55 %	80,95 %
ANN	100 %	77,77 %	90,47 %
CNN	100 %	93,35 %	96,88 %

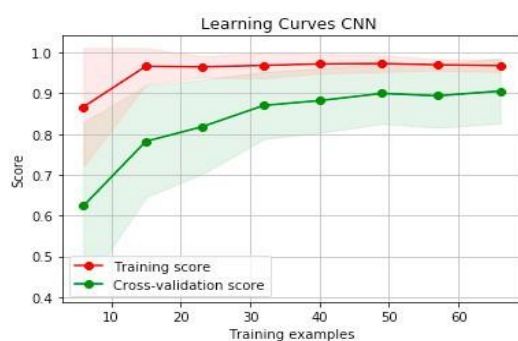
Evaluasi berbagai model pembelajaran mesin pada dataset diagnosis ASD remaja tingkat akurasi pada kisaran (80,95% - 96,88%) pada dataset asli. Pengklasifikasi K-NN dengan K=5 menghasilkan akurasi paling rendah sebesar 80,95%. Pengklasifikasi CNN menghasilkan akurasi prediksi tertinggi 96,88% pada dataset asli seperti ditampilkan pada tabel 6. Visualisasi kurva pembelajaran dengan berbagai model *machine learning* yang digunakan ditunjukkan pada gambar 13 – 17.



Gambar 15: Kurva Pembelajaran metode KNN untuk dataset remaja.



Gambar 16: Kurva Pembelajaran metode Regresi Logistik untuk dataset remaja.



Gambar 17: Kurva Pembelajaran metode CNN untuk dataset remaja.

Hasil perbandingan metode data skrining ASD pada anak dengan model sebelumnya untuk menangani data yang hilang dan penanganan setelah nilai yang hilang didapatkan hasil adanya peningkatan tingkat akurasi pada kisaran 98,30% seperti terlihat pada tabel 7.

Tabel 7: Perbandingan hasil dengan metode yang ada pada Data Skrining Gangguan Spektrum Autistik pada Anak

Model	Akurasi sebelum menangani nilai yang hilang	Akurasi setelah menangani nilai yang hilang
SVM	97,95	98,30
ANN	97,60	98,30
CNN	97,50	98,30

4. Kesimpulan

Dalam penelitian ini deteksi Gangguan Spektrum Autisme dicoba menggunakan berbagai pembelajaran mesin dan teknik pembelajaran mendalam. Berbagai metrik

evaluasi kinerja digunakan untuk menganalisis kinerja model yang diterapkan untuk deteksi ASD pada dataset non-klinis dari tiga kelompok usia yaitu Anak, Remaja, dan Dewasa. Ketika membandingkan hasil dengan penelitian terbaru lainnya pada masalah ini mendapatkan hasil yang lebih baik dari pengklasifikasi CNN daripada SVM dengan memasukkan semua atribut fiturnya setelah menangani nilai yang hilang. Dalam pekerjaan ini setelah menangani nilai yang hilang, baik model berbasis SVM dan CNN menunjukkan akurasi prediksi yang sama sekitar 98,30 % untuk dataset ASD anak (Mythili et al., 2014; Kosmicki, et al., 2015). Namun untuk dua set data lainnya, model berbasis CNN mampu mencapai hasil akurasi tertinggi daripada semua teknik pembuatan model lainnya yang dipertimbangkan, hasil ini menyarankan model berbasis CNN dapat diimplementasikan untuk mendeteksi Spektrum Autisme Gangguan alih-alih pengklasifikasi pembelajaran mesin konvensional lainnya yang disarankan dalam penelitian sebelumnya (Masood, et al., 2018).

DAFTAR PUSTAKA

Daftar pustaka menggunakan style Chicago, dan harus menggunakan referensi (baik dengan aplikasi pengelola seperti mendeley atau zotero atau yang lainnya) Daftar pustaka yang ditulis manual tidak diperkenankan. Contoh sebagaimana berikut.

F. Thabtah, Machine learning Autistic spectrum disorder reseac. 2018, Informatics for Health and Social Care.

Vaishali, R., dan R. Sasikala. A machine learning based approach to classify Autism with optimum behaviour sets. 2018. International Journal of Engineering & Technology.

F. Thabtah, F. Kamalov, dan K. Rajab. 2018. Computational intelligence approach to detect autistic features for autism screening. 2018. International journal of medical informatics.

J.N. Constantino, P. D. Lavesser, Y. I. Zhang, A. M. Abbacchi, T. Gray, dan R.D. Todd.

- Rapid quantitative assessment of autistic social impairment by classroom teachers. 2017. *Journal of the American Academy of Child & Adolescent Psychiatry*.
- D. Bone, M.S. Goodwin, M.P. Black, Chi-Chun Lee, K.Audhkhasi, dan S.Narayanan. Applying machine learning to facilitate autism diagnostics: pitfalls and promises. 2015. *Journal of autism and developmental disorders* 45(5).
- B. Li, A. Sharma, J. Meng, S. Purushwalkam, dan E. Gowen. Applying machine learning to identify autistic adults using imitation: An exploratory study. 2017 *PloS one*, 12(8).
- D.P. Wall, J. Kosmicki, T. F. Deluca, E. Harstad, dan V.A. Fusaro. Use of machine learning to shorten observation- based screening and diagnosis of autism. 2012. *Translational psychiatry*.
- D. Bone, C.C. Lee, M.P. Black, M.E. Williams, S. Lee, P. Levitt, dan S. Narayanan. The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody. 2014. *Journal of Speech, Language, and Hearing Research*.
- F. Thabtah. Autistic Spectrum Disorder Screening Data for Adult. 2017. <https://archive.ics.uci.edu/ml/machine-learning-databases/00426/>.
- J. A. Kosmicki, V. Sochat, M. Duda, dan D. P. Wall. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. 2015. *Translational psychiatry*, 5(2).
- S. Masood, A. Rai, A. Aggarwal, M.N. Doja, dan M.Ahmad. Detecting distraction of drivers using convolutional neural network. 2018. *Pattern Recognition Letters*.