

Penerapan Principal Component Analysis pada Model Deteksi Dini Anak Autisme

Aries Saifudin¹, Endar Nirmala², Irpan Kusyadi³

Teknik Informatika, Universitas Pamulang, Jl. Raya Puspitek No. 46 Buaran, Serpong, Tangerang Selatan, Banten, Indonesia, 15417

e-mail: ¹aries.saifudin@unpam.ac.id, ²dosen00216@unpam.ac.id, ³dosen00673@unpam.ac.id

Submitted Date: January 19th, 2024

Reviewed Date: April 21st, 2024

Revised Date: April 26th, 2024

Accepted Date: April 30th, 2024

Abstract

ASD (Autism Spectrum Disorder) is a neurological disorder that causes lifelong disturbances in children resulting in mental illness. Treatment can help but cannot be cured. Currently ASD is detected by understanding a child's behavior and intellectual activity. This diagnosis can be subjective, time-consuming, inconclusive, does not provide precise insight into genetics and is unsuitable for early detection. In Autism, a major challenge faced in many healthcare conditions is timing of diagnosis. It can take up to 6 months to diagnose a child with autism with certainty because of the lengthy process, and a child must see many different specialists to diagnose autism, from a developmental pediatrician, neurologist, psychiatrist or psychologist. Machine Learning Methods can make relevant changes to speed up the process. In this study, it is proposed to apply PCA (Principal Component Analysis). PCA is basically the basis of multivariate data analysis that applies the projection method. This analysis technique is usually used to summarize multivariate data tables on a large scale so that they can be used as a collection of smaller variables or a summary index. From there, variables are then analyzed to find out certain trends, variable clusters, and outliers. In this study it is proposed to implement the PCA (Principal Component Analysis) algorithm, namely PCA (Principal Component Analysis), Kernel PCA, Sparse PCA, and Incremental PCA. In this study using the experimental method by making applications to implement the proposed algorithm. Then test the model using the secondary dataset and measure the performance of the model. The research results show that the model that applies Sparse PCA gives the best results, which means that the application of PCA can be used to reduce the number of features and increase model performance.

Keywords: Children; autism; Detection; Principal Component Analysis

Abstrak

ASD (*Autism Spectrum Disorder*) adalah gangguan saraf yang menyebabkan gangguan seumur hidup pada anak yang mengakibatkan penyakit jiwa. Perawatan dapat membantu tetapi tidak dapat disembuhkan. Saat ini ASD dideteksi dengan memahami perilaku dan aktivitas intelektual seorang anak. Diagnosis ini bisa subjektif, memakan waktu, tidak meyakinkan, tidak memberikan wawasan yang tepat tentang genetika dan tidak cocok untuk deteksi dini. Dalam Autisme, tantangan besar yang dihadapi dalam banyak kondisi perawatan kesehatan adalah waktu diagnosis. Diperlukan waktu hingga 6 bulan untuk mendiagnosis anak autis dengan pasti karena proses yang lama, dan seorang anak harus menemui banyak spesialis berbeda untuk mendiagnosis autisme, mulai dari dokter anak perkembangan, ahli saraf, psikiater atau psikolog. Metode Pembelajaran Mesin dapat membuat perubahan yang relevan untuk mempercepat proses. Pada penelitian ini diusulkan penerapan PCA (Principal Component Analysis). PCA pada dasarnya adalah dasar untuk analisis data multivariat yang menerapkan metode proyeksi. Teknik analisis ini biasanya digunakan untuk merangkum tabel data multivariat yang besar sehingga dapat digunakan sebagai kumpulan variabel yang lebih kecil atau sebagai indeks ringkasan. Dari sana, variabel



dianalisis untuk menemukan tren spesifik, cluster variabel, dan outlier. Pada penelitian ini diusulkan untuk mengimplementasikan algoritma PCA (Principal Component Analysis), yaitu PCA (Principal Component Analysis), Kernel PCA, Sparse PCA, dan Incremental PCA. Pada penelitian ini menggunakan metode eksperimen dengan membuat aplikasi untuk menerapkan algoritma yang diusulkan. Kemudian menguji model menggunakan dataset sekunder dan mengukur kinerja model. Hasil penelitian menunjukkan bahwa model yang menerapkan Sparse PCA memberikan hasil terbaik yang berarti bahwa penerapan PCA dapat digunakan untuk mereduksi jumlah fitur dan meningkatkan kinerja model.

Kata kunci: Anak; Autis; Deteksi; Principal Component Analysis

1 Pendahuluan

Masalah *Autism Spectrum Disorder* (ASD) saat ini telah meningkat pesat 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 mereka berdasarkan berbagai parameter kesehatan dan fisiologis sekarang tampaknya mungkin dilakukan. Faktor ini memotivasi kami untuk meningkatkan minat dalam deteksi dan analisis penyakit ASD untuk meningkatkan metodologi pengobatan yang lebih baik. Deteksi ASD menjadi tantangan karena ada beberapa gangguan mental lain yang sedikit gejalanya sangat mirip dengan gejala ASD, sehingga membuat tugas ini menjadi sulit (Thabtah, 2019).

Gangguan Spektrum Autisme merupakan masalah yang berkaitan dengan perkembangan otak manusia. Seseorang yang mengalami Gangguan Spektrum Autisme umumnya tidak dapat melakukan interaksi sosial dan komunikasi dengan orang lain (Thabtah et al., 2018). Dalam hal ini, kehidupan seseorang biasanya terpengaruh sepanjang hidupnya. Menarik untuk diketahui bahwa faktor lingkungan dan genetik dapat menjadi faktor penyebab penyakit ini. Gejala masalah ini mungkin dimulai pada usia tiga tahun dan dapat berlanjut seumur hidup. Tidak mungkin untuk menyelesaikan pengobatan pasien yang menderita penyakit ini (Constantino et al., 2007), namun efeknya dapat berkurang untuk beberapa waktu jika gejalanya terdeteksi secara dini. Dengan asumsi bahwa gen manusia bertanggung jawab untuk itu, penyebab pasti ASD belum dikenali oleh ilmuwan (Bone et al., 2014). Gen manusia mempengaruhi perkembangan dengan mempengaruhi lingkungan. Ada beberapa faktor risiko yang mempengaruhi GSA seperti berat

badan lahir rendah, saudara kandung dengan ASD dan memiliki orang tua, dan lain-lain. Selain itu, terdapat beberapa masalah interaksi dan komunikasi sosial seperti tertawa dan cekikikan yang tidak pantas, tidak ada kepekaan nyeri, tidak bisa melakukan kontak mata dengan baik, tidak ada respons yang tepat terhadap suara, mungkin tidak ingin berpelukan, tidak bisa mengekspresikan gerak tubuh mereka, tidak ada interaksi dengan orang lain, keterikatan objek yang tidak pantas, ingin hidup sendiri, menggunakan kata-kata gema dan lain-lain.

Orang dengan ASD juga mengalami kesulitan dengan minat yang terbatas dan pengulangan perilaku yang konsisten. Daftar berikut menyajikan contoh spesifik dari jenis-jenis perilaku. Mengulangi perilaku tertentu seperti mengulang kata atau frasa sepanjang waktu. Orang tersebut akan marah ketika rutinitas akan berubah. Memiliki sedikit minat pada hal-hal tertentu dari topik seperti angka, fakta, dan lain-lain. Kurang sensitif dibandingkan orang lain dalam beberapa kasus seperti cahaya, kebisingan, dan lain-lain.

Deteksi dan pengobatan dini merupakan langkah terpenting yang harus dilakukan untuk mengurangi gejala gangguan spektrum autisme dan meningkatkan kualitas hidup penderita GSA. Namun, tidak ada prosedur tes medis untuk mendeteksi autisme (Wall, Kosmicki, et al., 2012). Gejala ASD biasanya dikenali dengan observasi. Pada lansia dan remaja yang bersekolah, gejala ASD biasanya dikenali oleh orang tua dan gurunya. Setelah itu gejala ASD ada dan dievaluasi oleh tim pendidikan khusus sekolah. Tim sekolah ini menyarankan anak-anak ini mengunjungi dokter perawatan kesehatan mereka untuk tes yang diperlukan. Pada orang dewasa, mengidentifikasi gejala ASD sangat sulit daripada anak-anak dan remaja yang lebih tua karena beberapa gejala ASD mungkin tumpang tindih

dengan gangguan kesehatan mental lainnya (Wall, Dally, et al., 2012). Perubahan perilaku pada anak dapat dengan mudah diidentifikasi melalui observasi karena dapat dilihat pada usia 6 bulan lebih awal dibandingkan dengan pencitraan otak spesifik Autisme karena pencitraan otak dapat diidentifikasi setelah usia 2 tahun.

Vaishali R. dan Sasikala R. (R & R, 2018) telah mengusulkan metode untuk mengidentifikasi Autisme dengan kumpulan perilaku optimal. Dalam pekerjaan ini, dataset diagnosis ASD dengan 21 fitur yang diperoleh dari repositori pembelajaran mesin UCI bereksperimen dengan pembungkus pemilihan fitur biner firefly berbasis kecerdasan swarm. Hipotesis alternatif dari eksperimen mengklaim bahwa model pembelajaran mesin dimungkinkan untuk mencapai akurasi klasifikasi yang lebih baik dengan subset fitur minimum. Dengan menggunakan pembungkus seleksi fitur kunang-kunang biner tujuan tunggal berbasis kecerdasan Swarm, ditemukan bahwa 10 fitur di antara 21 fitur dari dataset ASD cukup untuk membedakan antara pasien ASD dan non-ASD. Hasil yang diperoleh dengan pendekatan ini membenarkan hipotesis dengan menghasilkan akurasi rata-rata pada kisaran 92,12% -97,95% dengan subset fitur optimal yang kira-kira sama dengan akurasi rata-rata yang dihasilkan oleh seluruh dataset diagnosis ASD.

Fadi Thabtah (Thabtah, 2017) telah mengusulkan model penyaringan ASD menggunakan Machine Learning Adaption dan DSM-5. Alat skrining telah digunakan untuk mewujudkan satu atau lebih tujuan dalam skrining ASD. Dalam makalah ini, peneliti membahas klasifikasi ASD Machine Learning dengan pro dan kontranya. Peneliti mencoba untuk menyoroti masalah yang menyertai alat skrining ASD yang ada dan konsistensi dari alat tersebut menggunakan DSM-IV daripada manual DSM-5.

Mythili dan Shanavas (Mythili & Shanavas, 2014) memiliki studi tentang ASD menggunakan Teknik Klasifikasi. Tujuan utama dari makalah ini adalah untuk mendeteksi masalah autisme dan tingkat autisme. Dalam Neural Network ini, teknik SVM dan Fuzzy dengan tools WEKA digunakan untuk menganalisis perilaku dan interaksi sosial siswa. Kosmicki dkk (Kosmicki et al., 2015) mengira metode pencarian untuk set paling sedikit sifat untuk deteksi autisme. Dalam hal ini, penulis

menggunakan pendekatan pembelajaran mesin untuk mengevaluasi penilaian klinis ASD. ADOS dilakukan pada subset perilaku anak-anak berdasarkan spektrum autisme. ADOS memiliki empat modul. Dalam pekerjaan ini, 8 algoritma pembelajaran mesin yang berbeda digunakan, yang melibatkan identifikasi fitur mundur bertahap pada lembar skor dari 4.540 individu. Ini menggunakan 9 dari 28 perilaku dari modul 2 dan 12 dari 28 perilaku dari modul 3 untuk mengidentifikasi risiko ASD dengan akurasi keseluruhan masing-masing 98,27% dan 97,66%.

Li B., A. Sharma, J Meng, S. Purushwalkam, dan E. Gowen (Li et al., 2017) telah menggunakan pengklasifikasi pembelajaran mesin untuk mendeteksi orang dewasa autis dengan metode imitasi. Tujuan dari penelitian ini adalah untuk mengetahui permasalahan mendasar terkait dengan kondisi uji diskriminatif dan parameter kinematik. Dataset berisi 16 peserta ASD yang memiliki rangkaian gerakan tangan. Dalam 40 kendala kinematik dari 08 kondisi imitasi telah diekstraksi dengan menggunakan metode pembelajaran mesin. Penelitian ini menunjukkan bahwa untuk sampel kecil, penerapan metode pembelajaran mesin dapat dilakukan untuk menganalisis data berdimensi tinggi dan klasifikasi diagnostik autisme. Tingkat sensitivitas yang dicapai oleh RIPPER yang memiliki fitur Va (87.30%), CHI (80.95%), IG (80.95%), Correlation (84.13%), CFS (84.13%), dan "no feature selection" (80.00%) pada kumpulan data AQ-Adolescent.

Beberapa penelitian di atas menunjukkan bahwa penelitian masih dilanjutkan karena belum ada model yang mendeteksi ASD secara tepat atau menghasilkan akurasi yang sempurna (Koehler et al., 2024) (Qureshi et al., 2023). Terbukti dari bagian yang dibahas di atas bahwa ada kebutuhan untuk mengeksplorasi kemungkinan penerapan model berbasis mesin pembelajaran (machine learning) yang dalam untuk mendeteksi ASD pada populasi manusia. Sebagian besar pekerjaan yang dibahas di atas menggunakan pendekatan pembelajaran mesin konvensional dan oleh karena itu kinerjanya terbatas. Dalam pekerjaan ini, performa beberapa model pembelajaran mesin telah dibandingkan dengan model pembelajaran dalam untuk tujuan ini. Model terpisah telah disiapkan untuk kumpulan populasi terpisah

(dibahas pada bagian di bawah) dan dibandingkan secara individual.

Pada penelitian ini diusulkan penerapan PCA (Principal Component Analysis). PCA pada dasarnya merupakan dasar dari analisis data multivariat yang menerapkan metode proyeksi. Teknik analisis ini biasanya digunakan untuk meringkas tabel data multivariat dalam skala besar hingga bisa dijadikan kumpulan variabel yang lebih kecil atau indeks ringkasan. Dari situ, kemudian variabel dianalisis untuk mengetahui tren tertentu, kluster variabel, hingga outlier.

2 Metodologi

Pada penelitian ini digunakan pendekatan kuantitatif. Pada pendekatan penelitian kuantitatif, dilakukan analisa kuantitatif secara teliti terhadap beberapa generasi informasi yang berbentuk kuantitatif. Umumnya pendekatan kuantitatif memiliki tiga bentuk yang berbeda, yaitu pendekatan inferensial, pendekatan eksperimental, dan pendekatan simulasi. Pada pendekatan inferensial, sampel yang diperoleh digunakan untuk membuat dugaan karakteristik populasi, relasinya, dan lain-lain. Pada pendekatan ini, peneliti tidak memiliki kontrol atas karakteristik, variabel, dan responden yang diteliti. Pendekatan eksperimental ditandai dengan adanya kontrol atas lingkungan penelitian oleh peneliti. Eksperimen adalah suatu proses yang sistematis di mana peneliti memiliki kontrol atas variabel berdasarkan pertimbangan agar sesuai dengan tujuan penelitian. Simulasi berarti operasi model numerik yang mewakili struktur proses dinamis. Pada pendekatan simulasi, lingkungan buatan dibuat di mana informasi yang diperlukan dapat dihasilkan.

Tujuan dari metode kuantitatif adalah untuk dapat memahami bagaimana sesuatu dikonstruksi, bagaimana dibangun, dan bagaimana cara kerjanya. Penelitian kuantitatif umumnya didorong oleh hipotesis, kemudian dibuat rumusan dan pengujian secara ketat untuk menunjukkan bahwa hipotesisnya salah. Sehingga usaha yang dilakukan adalah membuktikan bahwa hipotesis yang dibuat adalah salah, jika hipotesisnya tahan uji, maka hipotesis tersebut dianggap benar. Tetapi jika tidak tahan uji, maka hipotesisnya dianggap salah.

Sudut pandang kuantitatif menekankan bahwa pengukuran merupakan dasar yang dapat digunakan untuk menunjukkan hubungan antara

observasi dan formalisasi model, teori, dan hipotesis. Penelitian dan metode kuantitatif akan menghasilkan pengembangan model, teori, dan hipotesis yang berkaitan dengan fenomena alam.

Pendekatan yang akan digunakan pada penelitian ini adalah pendekatan eksperimen. Pada penelitian eksperimen dilakukan dengan cara menginvestigasi hubungan sebab-akibat menggunakan pengujian yang dikontrol oleh peneliti. Pada penelitian semiekperimental sering mendapatkan kendala pada tidak cukupnya akses terhadap sampel, masalah etika, dan sebagainya. Untuk pengembangan, evaluasi, dan pemecahan masalah proyek biasanya dilakukan dengan eksperimen.

Tujuan dari penelitian ini adalah mengembangkan aplikasi Deteksi Dini Anak Autis. Suatu penelitian harus mengikuti aturan yang diakui agar dapat diakui/diterima, maka penelitian ini dilakukan sesuai tahapan yang ditunjukkan pada Gambar 3.



Gambar 1 Tahapan Penelitian

Tahapan pada Gambar 3 dijelaskan sebagai berikut:

a. Pengumpulan data

Pengumpulan data bukan hanya sekedar mengambil data yang ada, tetapi harus mampu mendeskripsikan data yang ada, dan memiliki kontribusi terhadap pengetahuan. Data tersebut harus dapat memberikan penjelasan, hubungan, perbandingan, prediksi, generalisasi, dan teori

(Dawson, 2009). Berdasarkan sumbernya, data dibedakan menjadi dua, yaitu:

- Data primer, yaitu data yang dikumpulkan langsung dari sumber data. Pengumpulan data ini memerlukan waktu, dan biaya yang lebih banyak dari data sekunder. Contoh sumber data primer adalah kuisisioner, observasi, wawancara, dan eksperimen yang dilakukan langsung oleh peneliti.
- Data sekunder, yaitu data yang diperoleh dari peneliti/pihak lain, walaupun data tersebut sebelumnya digunakan dengan tujuan yang berbeda. Data sekunder dapat diperoleh relatif lebih cepat, dan dengan biaya rendah. Contoh sumber data sekunder adalah kantor statistik baik pemerintah maupun swasta, perpustakaan, toko buku, maupun internet.

Beberapa alasan penggunaan data sekunder adalah:

- Karena pengumpulan data primer relatif lebih sulit, memakan waktu, dan mahal.
- Karena ada kemungkinan memiliki data yang tidak cukup.
- Karena lebih masuk akal untuk menggunakan data yang ada, walaupun datanya dalam bentuk yang berbeda.
- Karena lebih dapat menjelaskan, atau pelengkap dari data primer dikumpulkan.
- Karena dapat mengkonfirmasi, memodifikasi atau bertentangan dengan yang ditemukan.
- Karena memungkinkan untuk memusatkan perhatian pada analisis dan interpretasi.
- Karena tidak dapat melakukan penelitian pada sesuatu yang telah dilakukan/terjadi.
- Karena lebih banyak data yang dikumpulkan daripada yang pernah digunakan.

b. Pengolahan awal data

Data yang sudah dikumpulkan diolah menggunakan algoritma soft-computing untuk mengurangi data yang tidak relevan, atau data dengan atribut yang hilang. Pengolahan juga dapat berupa konversi nilai-nilai redundan (berlebihan), atau nilai yang terlalu beragam ke

dalam kelompok yang lebih kecil untuk mempermudah pembentukan model.

c. Model/metode yang diusulkan

Untuk menggambarkan alur model/metode yang diusulkan dan menjelaskan cara kerja model/metode yang diusulkan. Model/metode ini digambarkan secara skematik dan disertai dengan formula penghitungan. Model/metode yang diusulkan akan dibentuk dari data yang sudah diolah, dan hasil pengolahan model akan diukur dengan model yang ada saat ini.

d. Eksperimen dan pengujian model

Menjabarkan bagaimana eksperimen yang dilakukan hingga terbentuknya model, serta menjelaskan cara menguji model yang terbentuk.

e. Evaluasi dan validasi hasil

Evaluasi dilakukan dengan mengamati hasil prediksi menggunakan algoritma soft-computing. Validasi dilakukan dengan mengukur hasil prediksi dibandingkan dengan data asli/awal. Pengukuran kinerja dilakukan dengan membandingkan nilai kesalahan (error) hasil prediksi masing-masing algoritma sehingga dapat diketahui algoritma yang lebih akurat.

Dalam Autisme, tantangan besar yang dihadapi dalam banyak kondisi perawatan kesehatan adalah waktu diagnosis. Diperlukan waktu hingga 6 bulan untuk mendiagnosis anak autis dengan pasti karena proses yang lama, dan seorang anak harus menemui banyak spesialis berbeda untuk mendiagnosis autisme, mulai dari dokter anak perkembangan, ahli saraf, psikiater atau psikolog. Dengan cara tradisional saat ini, waktu yang dibutuhkan untuk menyelesaikan diagnosis Autisme relatif lama.

Oleh karena itu, metode Pembelajaran Mesin dapat membuat perubahan yang relevan untuk mempercepat proses. Diketahui bahwa intervensi dini merupakan kunci untuk memperbaiki anak autis. Jelas mempercepat waktu diagnosis bahkan lebih penting dalam kasus Autisme. Teknologi big data dan pembelajaran mesin dapat membuat kemajuan besar untuk memprediksi dan mempercepat proses diagnosis dan pengobatan yang kompleks dan memakan waktu. Sistem pembelajaran mesin dapat dikembangkan untuk memanfaatkan sejumlah

besar data kesehatan dan medis yang tersedia untuk pemodelan prediktif dan analisis prediktif. Dalam makalah ini, perbandingan dari beberapa teknik dan model pembelajaran mesin akan diuji dan dianalisis.

Data diproses sebelumnya untuk membuat prediksi berdasarkan kategori yang berbeda di mana pengujian diklasifikasikan sebagai Autistik. Ada banyak algoritma klasifikasi yang dapat diterapkan. Setiap pengklasifikasi berbeda dalam cara pengumpulan data, pemfilteran data, ekstraksi fitur, dan menggunakan proses ini untuk memasukkan model untuk dipelajari. Dalam pekerjaan ini, efektivitas beberapa algoritma pembelajaran mesin untuk mengevaluasi efektivitas pengobatan atau prediksi hasil pengobatan ASD dinilai.

Metode ensemble biasanya digunakan untuk meningkatkan akurasi prediksi dengan menggabungkan prediksi beberapa model pembelajaran mesin. Kearifan tradisional telah menggabungkan apa yang disebut pelajar "lemah". Namun, pendekatan yang lebih modern adalah membuat ansambel dari koleksi model yang kuat namun beragam yang dipilih dengan baik.

Membangun model ansambel yang kuat memiliki banyak kesamaan dengan membangun tim manusia yang sukses dalam bisnis, sains, politik, dan olahraga. Setiap anggota tim memberikan kontribusi yang signifikan dan kelemahan serta bias individu diimbangi oleh kekuatan anggota lainnya.

Jenis ansambel yang paling sederhana adalah prediksi rata-rata tak berbobot dari model yang membentuk perpustakaan model. Misalnya, jika pustaka model menyertakan tiga model untuk target interval (seperti yang ditunjukkan pada gambar berikut), rata-rata tidak tertimbang akan memerlukan pembagian jumlah nilai yang diprediksi dari tiga model kandidat dengan tiga. Dalam rata-rata tanpa bobot, setiap model memiliki bobot yang sama saat model ansambel dibuat.

2.1 Prediksi rata-rata untuk membentuk model

Secara lebih umum, Anda bisa memikirkan tentang menggunakan rata-rata tertimbang. Misalnya, Anda mungkin yakin bahwa beberapa model lebih baik atau lebih akurat dan Anda ingin menetapkan bobot yang lebih tinggi secara manual.

Penumpukan (terkadang disebut Generalisasi Bertumpuk) adalah paradigma yang berbeda. Tujuan penumpukan adalah untuk mengeksplorasi ruang model yang berbeda untuk masalah yang sama. Identy adalah bahwa Anda dapat menyerang masalah pembelajaran dengan berbagai jenis model yang mampu mempelajari beberapa bagian dari masalah, tetapi tidak seluruh ruang masalah. Jadi, Anda dapat membangun beberapa pelajar yang berbeda dan Anda menggunakannya untuk membuat prediksi menengah, satu prediksi untuk setiap model yang dipelajari. Kemudian Anda menambahkan model baru yang belajar dari prediksi perantara target yang sama.

Model terakhir ini dikatakan ditumpuk di atas yang lain, oleh karena itu dinamai model. Dengan demikian, Anda dapat meningkatkan kinerja Anda secara keseluruhan, dan sering kali Anda mendapatkan model yang lebih baik daripada model perantara individual. Namun perhatikan, itu tidak memberi Anda jaminan apa pun, seperti yang sering terjadi pada teknik pembelajaran mesin apa pun.

2.2 Data Pendukung

Dataset yang digunakan merupakan data sekunder yang diambil dari UCI Dataset (*Autistic Spectrum Disorder Screening Data for Children Data Set*) dengan link <https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++>.

Tabel 1 Spesifikasi 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

Attribute	Type	Description
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

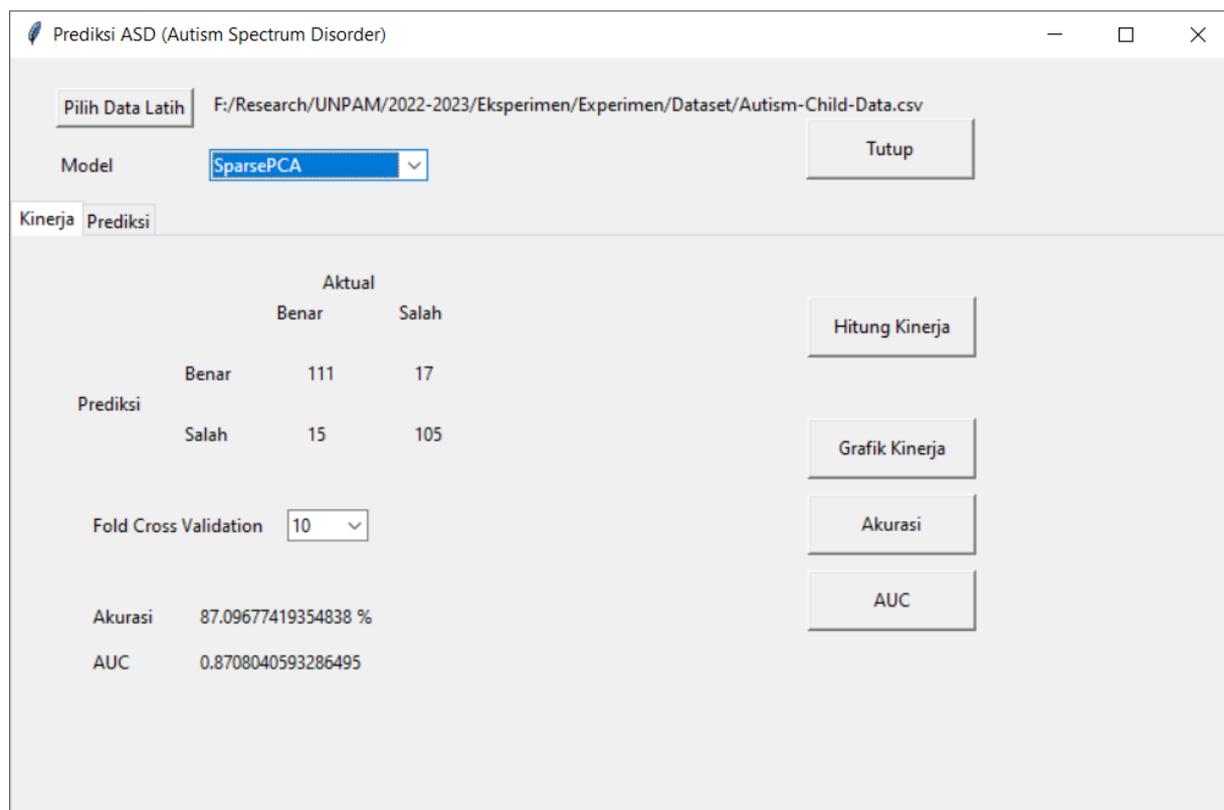
2.3 Model yang Diusulkan

Pada model yang diusulkan digunakan dataset ASD yang telah dikumpulkan. Dataset ASD yang terkumpul dilakukan preprocessing yang mencakup, cleaning, transformasi, dan standarisasi. Dataset akan dibagi menjadi dua bagian menggunakan 10-fold cross validation, yaitu data latih (K-1 fold) dan data uji (1 fold). Data latih digunakan untuk melatih model yang menerapkan KNN, Naïve Bayes, Decision Tree, SVM, Logistic Regression, dan PCA dari kelima algoritma tersebut. Model yang telah dilatih

kemudian divalidasi menggunakan data uji. Hasil validasi digunakan untuk mengukur kinerja model. Kemudian semua model yang terbentuk dibandingkan untuk mengetahui model terbaiknya.

3 Hasil dan Pembahasan

Pada penelitian ini digunakan bahasa pemrograman Python untuk menerapkan model prediksi/diagnosa *Autism Spectrum Disorder* (ASD) yang diusulkan. Tampilan aplikasi ditunjukkan pada Gambar 2.



Gambar 2 Tampilan Aplikasi Prediksi ASD (*Autism Spectrum Disorder*)

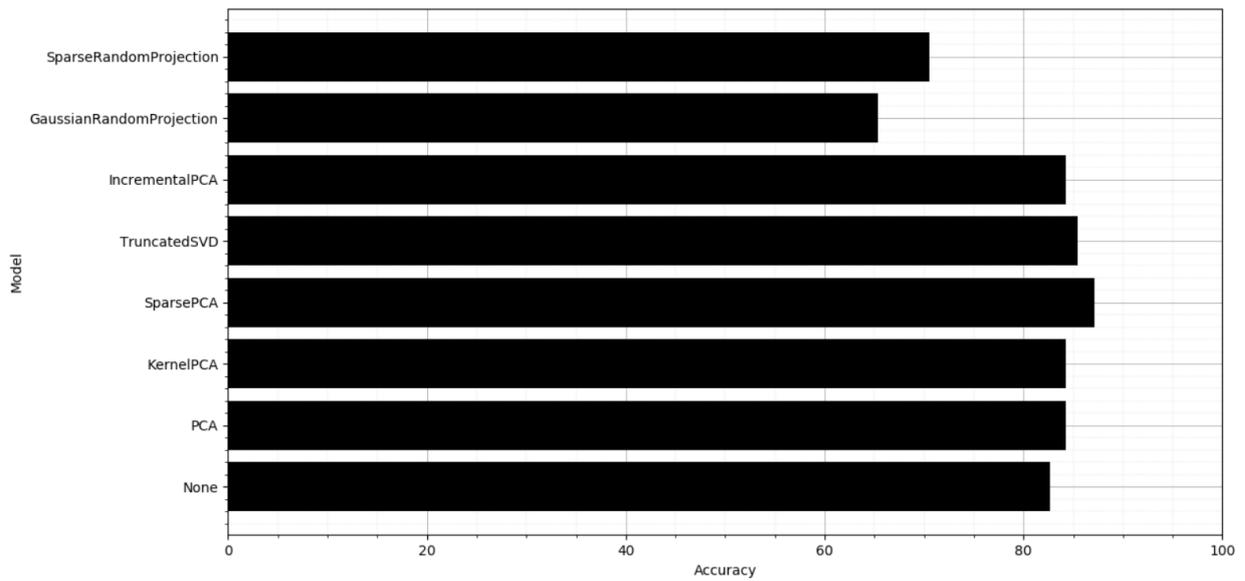
Hasil penerapan model memberikan hasil kinerja seperti pada Tabel 2.

Tabel 2 Kinerja Model

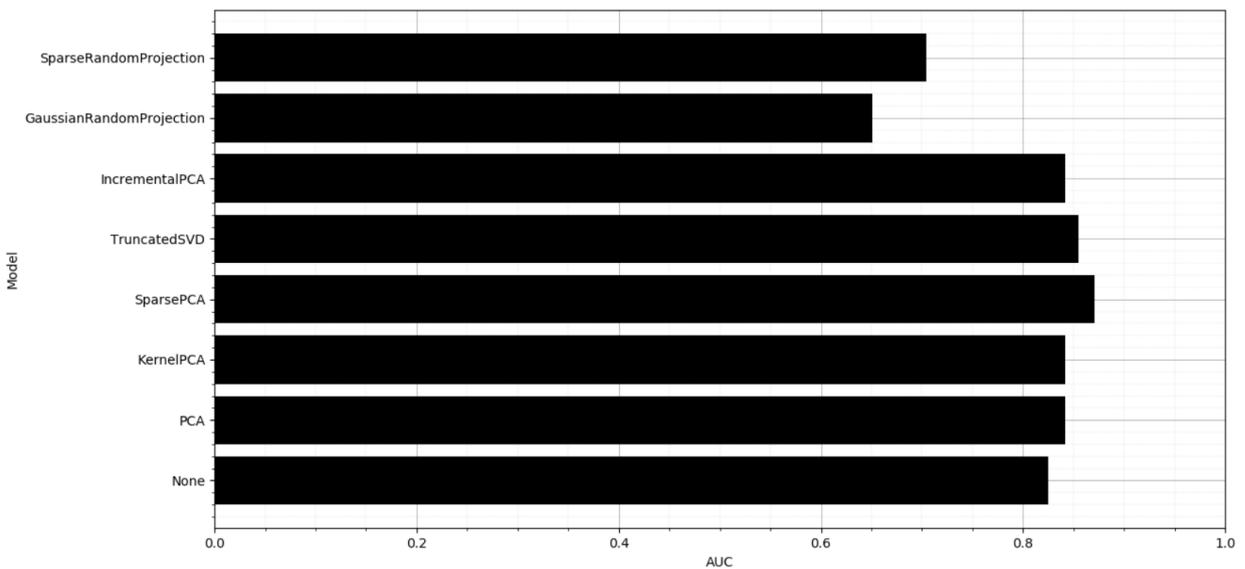
Model	Akurasi	AUC
None	82,66%	0,8252
PCA	84,27%	0,8422
Kernel PCA	84,27%	0,8422
Sparse PCA	87,10%	0,8708
Truncated SVD	85,48%	0,8545
Incremental PCA	84,27%	0,8422
Gaussian Random Projection	65,32%	0,6513
Sparse Random Projection	70,56%	0,7045

Kemudian nilai kinerja model divisualisasi menggunakan grafik untuk perbandingannya.

Nilai akurasi ditampilkan pada Gambar 4 dan nilai AUC ditunjukkan pada Gambar 5.



Gambar 4 Akurasi model



Gambar 5 AUC model

Berdasarkan nilai yang didapat, menunjukkan bahwa model yang menerapkan Sparse PCA memberikan nilai kinerja yang terbaik. Rata-rata penggunaan algoritma PCA dapat meningkatkan kinerja model deteksi dini anak autisme, kecuali *Sparse Random Projection* dan *Gaussian Random Projection*.

4 Kesimpulan

Berdasarkan hasil eksperimen yang telah dilakukan dapat disimpulkan sebagai berikut:

- Model prediksi/diagnosa *Autism Spectrum Disorder* (ASD) dapat ditingkatkan menggunakan PCA.
- Model yang menerapkan Sparse PCA dapat meningkatkan kinerja model sebesar 5,37% untuk nilai akurasi dan 5,53% untuk AUC.

5 Saran

Hasil penelitian menunjukkan bahwa model yang diusulkan dengan menerapkan PCA belum mencapai nilai kinerja yang sempurna, maka

disarankan untuk penelitian selanjutnya dapat menerapkan teknik optimasi yang lain. Teknik optimasi yang dapat digunakan misalnya bagging, boosting, atau yang lainnya.

6 References

- Bone, D., Goodwin, M. S., & Lee, M. P. B. C. (2014). *Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises*. 101. <https://doi.org/10.1007/s10803-014-2268-6>
- Constantino, J. N., Lavesser, P. D., Zhang, Y., Abbacchi, A. M., Gray, T., & Todd, R. D. (2007). Rapid Quantitative Assessment of Autistic Social Impairment by Classroom Teachers. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46(12), 1668–1676. <https://doi.org/10.1097/chi.0b013e318157cb23>
- Dawson, C. W. (2009). Projects in Computing and Information Systems. In *Information Systems Journal* (Vol. 2). http://www.sentimentaltoday.net/National_Academy_Press/0321263553.Addison.Wesley.Publishing.Company.Projects.in.Computing.and.Information.Systems.A.Students.Guide.Jun.2005.pdf
- Koehler, J. C., Dong, M. Sen, Bierlich, A. M., Fischer, S., Späth, J., Plank, I. S., Koutsouleris, N., & Falter-Wagner, C. M. (2024). Machine learning classification of autism spectrum disorder based on reciprocity in naturalistic social interactions. *Translational Psychiatry*, 14(1), 76. <https://doi.org/10.1038/s41398-024-02802-5>
- Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a Minimal Set of Behaviors for Autism Detection Through Feature Selection-based Machine Learning. *Translational Psychiatry*, 5(2), e514-7. <https://doi.org/10.1038/tp.2015.7>
- Li, B., Sharma, A., Meng, J., Purushwalkam, S., & Gowen, E. (2017). Applying Machine Learning to Identify Autistic Adults Using Imitation: An Exploratory Study. *PLoS ONE*, 12(8), 1–19. <https://doi.org/10.1371/journal.pone.0182652>
- Mythili, M. S., & Shanavas, A. R. M. (2014). A Study on Autism Spectrum Disorders using Classification Techniques. *International Journal of Soft Computing and Engineering (IJSCE)*, 5, 2231–2307. <http://www.ijscce.org/wp-content/uploads/papers/v4i5/E2433114514.pdf>
- Qureshi, M. S., Qureshi, M. B., Asghar, J., Alam, F., & Aljarboub, A. (2023). Prediction and Analysis of Autism Spectrum Disorder Using Machine Learning Techniques. *Journal of Healthcare Engineering*, 2023, 1–10. <https://doi.org/10.1155/2023/9815989>
- R, V., & R, S. (2018). A Machine Learning based Approach to Classify Autism with Optimum Behavior Sets. *International Journal of Engineering & Technology*, 7(4), 4216–4219.
- Thabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. *ACM International Conference Proceeding Series, Part F1293*, 1–6. <https://doi.org/10.1145/3107514.3107515>
- Thabtah, F. (2019). Machine Learning in Autistic Spectrum Disorder Behavioral Research: A Review and Ways Forward. *Informatics for Health and Social Care*, 44(3), 278–297. <https://doi.org/10.1080/17538157.2017.1399132>
- Thabtah, F., Kamalov, F., & Rajab, K. (2018). A New Computational Intelligence Approach to Detect Autistic Features for Autism Screening. *International Journal of Medical Informatics*, 117, 112–124. <https://doi.org/10.1016/j.ijmedinf.2018.06.009>
- Wall, D. P., Dally, R., Luyster, R., Jung, J. Y., & DeLuca, T. F. (2012). Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism. *PLoS ONE*, 7(8). <https://doi.org/10.1371/journal.pone.0043855>
- Wall, D. P., Kosmicki, J., Deluca, T. F., Harstad, E., & Fusaro, V. A. (2012). Use of Machine Learning to Shorten Observation-based Screening and Diagnosis of Autism. *Translational Psychiatry*, 2(December 2011). <https://doi.org/10.1038/tp.2012.10>

